MOX-Report No. 49/2015

Statistical inference for stochastic processes: two sample hypothesis tests

Ghiglietti, A.; Ieva, F.; Paganoni, A.M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Statistical inference for stochastic processes: two sample hypothesis tests

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

[a] Department of Mathematics, Università degli Studi di Milano, Milan, Italy
[b] Department of Mathematics, Politecnico di Milano, Milan, Italy

Abstract

In this paper, we present inferential procedures to compare the means of two samples of functional data. The proposed tests are based on a suitable generalization of Mahalanobis distance to the Hibert space of square integrable function defined on a compact interval. We do not require any specific distributional assumption on the processes generating the data. Test procedures are proposed for both the cases of known and unknown variance-covariance structures, and asymptotic properties of test statistics are deeply studied. A simulation study together with a real case data analysis are also presented.
Keywords: Functional Data, Distances in L^2, Hypothesis tests, Two sample problem.
AMS Subject Classification: 62M07

1 Introduction

In the last two decades statistical research has grown an increasing interest in the study of high dimensional data, having a wide range of applications in biology, chemometrics, medicine, meteorology and finance, among others. In all these cases, observed data may be points that belong to functions generated by a continuous time stochastic process with values in a suitable infinite dimensional Hilbert space, typically $L^2(T)$ with T compact interval of $\mathbb{R}$. Functional Data Analysis (FDA) gathers all the statistical models and tools fit for the study of this kind of data characterized by a number p of features observed for each statistical unit much larger than the sample size n (see Ramsey and Silverman (2002), Ramsey and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and references therein). Classical methodologies in FDA are concerned with the mean function and the covariance kernel of the process generating the data. The estimation and the inference on the mean function is typically realized by computing confidence bands that take into account the covariance structure of the process, 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.

Despite the massive methodological and theoretical development of statistical inference for functional data, testing hypothesis on generating distributions of two samples of curves has received only a little attention (Pomann et al, 2014). In particular concerning the study of differences in the mean functions of two independent samples of curves, Ramsay and Silverman (2005) proposed a pointwise t-test, Zhang et al. (2010) introduced a test based on L^2 norm for the Behrens-Fisher problem, Pini and Vantini (2013) proposed an interval testing procedure based on permutation tests, Staicu et al. (2014) developed a pseudo likelihood ratio test. In many cases, these procedures consist of a suitable dimensional reduction of the data which allows classical multivariate procedures to be applied. Although this approach may be satisfactory in some situations, the functional nature of the data is not fully exploited and some information may be lost due to the dimensional reduction.

Ghiglietti and Paganoni (2014) proposed a procedure to test the difference between the means of Gaussian processes based on a generalization of Mahalanobis distance (say d_p) that achieves two main goals: (i) to consider all the infinite components of data basis expansion and (ii) to share the same ideas which the Mahalanobis distance is based on. In this paper, we extend those inferential procedures to a wide range of situations, by relaxing some strong assumptions made in Ghiglietti and Paganoni (2014). Specifically, the proposed tests do not require any specific distribution assumption on the processes generating the data and allow comparison between samples with different covariance functions. In this wider context, we prove theoretical results on the convergence of d_p distance between sample mean and a fixed function $m(t) \in L^2(T)$ and between means of two independent samples. Additionally, we establish the rate of convergence and the exact asymptotic distributions of the distance d_p between functional sample means. The rate of convergence and the limiting distributions are sharply different when the true means of the processes are equal or different. Indeed, in the first case, we show that the exact rate of convergence is n^{-1} and the limiting distribution is a strictly positive random variable; in the second case, we prove a central limit theorem (CLT) for the distance d_p between functional sample means with Gaussian asymptotic distribution and the rate of convergence of $n^{-\frac{1}{2}}$.

The almost sure convergence established for the distance d_p of the sample means guaranteed the consistency of this estimator, while the second-order asymptotic results provide the probabilistic basis to construct test procedures for the comparison of the means of two functional populations. Indeed, the proposed critical regions are based on the limiting distribution established in the case of equality of the means, while the CLT is applied to compute the asymptotic power of the test given any difference between the means of the processes. Test procedures are proposed for both the cases of known and unknown variance-covariance structures. It is worth noting that all the results hold also for multivariate functional data case.

The rest of the paper is structured as follows: firstly, the distance d_p is introduced and its main properties are discussed in Section 2. Then, the asymptotic results on the behavior of this metric applied to random processes are presented in Section 3. Section 4 is concerned with inferential procedures for the comparison of the means in functional data analysis. Specifically, critical regions based on the distance d_p applied to functional sample means are proposed. Finally, a simulation study together with a real case data analysis are presented in Section 5. Appendixes A and B gather the proofs of the theorems stated in Section 3. All the analyzes are carried out with `R`.

2 Properties of a generalized Mahalanobis distance for functional objects

In this section, we present the generalized Mahalanobis distance for infinite dimensional spaces introduced in Ghiglietti and Paganoni (2014). This metric is characterized by features similar to the Mahalanobis distance commonly used in the multivariate context. Hence, inferential tools for infinite dimensional objects can be constructed in an analogous way to the procedures typically used for multivariate elements (see Section 4). We start by describing the motivation problem that required the introduction of the new distance.

Let y and w be realizations of a stochastic process $X \in L^2(T)$, where T is a compact interval of $\mathbb{R}$. 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 $v(s,t) = \mathbb{E}[(X(t)-m(t))(X(s)-m(s))]$. Then, denote by $\{\lambda_k; k \geq 1\}$ and $\{\varphi_k; k \geq 1\}$ the sequences of eigenvalues and eigenfunctions, respectively, associated to v.

Letting $\langle a,b\rangle = \int_T a(t)b(t)dt$ be the usual inner product in $L^2(T)$, the natural generalization of the Mahalanobis distance in the functional framework would be the following

$$d_M(y,w) = \sqrt{\sum_{k=1}^{\infty} \frac{(\langle y-w,\varphi_k\rangle)^2}{\lambda_k}}. \tag{2.1}$$

This distance takes into account the correlations and the variability described by the covariance structure of X. However, it is well known that in an infinite dimensional space d_M is not a proper distance in $L^2(T)$, since the series in (2.1) can diverge for some $y,w \in L^2(T)$. For this reason, a typical methodology is to fix an integer $K \in \mathbb{N}$ and consider the truncated version of the Mahalanobis distance, summing up in (2.1) only the first K components. Nevertheless, when this approach is used to measure the entire space $L^2(T)$, we can point out two main drawbacks can be highlighted: (i) the contribution given by the projections in the space orthogonal to $\varphi_1,..,\varphi_K$ 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. (ii) 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.

These issues are addressed by using the following equivalent representation for d_M in (2.1), that has been presented in Ghiglietti and Paganoni (2014)

$$d_M(y,w) := \sqrt{\int_0^{\infty} f(c;y,w)dc}, \tag{2.2}$$

where the function $f(\cdot;y,w) : \mathbb{R}^+ \to \mathbb{R}^+$ is defined as follows

$$f(c;y,w) := \sum_{k=1}^{\infty} (\langle y-w,\varphi_k\rangle)^2 \cdot \exp(-\lambda_k c). \tag{2.3}$$

The idea is to use (2.2) to compute a well-defined distance in $L^2(T)$ that is in a certain sense similar to d_M in (2.1). In fact, although $f(c; y, w)$ is finite for any $c \geq 0$, the function $f(\cdot; y, w)$ may not be integrable in $\mathbb{R}^+$ for some $y, w \in L^2(T)$, that occurs when the Mahalanobis distance is not defined. To deal with this case, we introduce a function $g(\cdot; p)$, tuned by a parameter $p > 0$, such that $\int_0^\infty g(c;p)dc < \infty$. In particular, without loss of generality, we define g such that $\int_0^\infty g(c;p)dc = p$. Since this ensures that $f(c; y, w) \cdot g(c; p)$ is integrable for any $p > 0$, the generalized Mahalanobis distance is defined as follows

$$\begin{aligned} d_p(y,w) \quad &:= \sqrt{\int_0^\infty f(c;y,w) \cdot g(c;p)dc} \\ &= \sqrt{\int_0^\infty \sum_{k=1}^\infty \left(\langle y - w, \varphi_k \rangle\right)^2 \cdot \exp(-\lambda_k c) \cdot g(c;p)dc}, \end{aligned} \tag{2.4}$$

that is finite for any pair of functions y and w.

Some conditions should be required to make the distance d_p sharing analogous features of the Mahalanobis distance. Specifically, g should be defined such that the functions $f \cdot g$ and f are similar concerning some essential properties, such as penalizing more the distances along components with smaller variability. To this aim, we assume that $g(0;p) = 1$ and $g(c;p)$ is non increasing and non negative, for any $p > 0$. Moreover, for any fixed $c > 0$, we assume that $g(c;p)$ is non decreasing in p and $\lim_{p\to\infty} g(c;p) = 1$, so that d_p tends to the Mahalanobis distance as p goes to infinity.

The distance $d_p(y, w)$ can be expressed in a more suitable way

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

where $d_{M,k}(y,w) = \sqrt{\left(\langle y - w, \varphi_k \rangle\right)^2 / \lambda_k}$ is the term representing the contribution of the Mahalanobis distance along the k^{th} component and

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

It follows that $h_k(p)/\lambda_k \to_k p$ and $h_k(p) \to_p 1$. Moreover, the distance d_p and the usual distance in L^2 (d_{L^2}) are equivalent since, for any $x, y \in L^2(T)$

$$\left(\frac{h_1(p)}{\lambda_1}\right) \cdot d_{L^2}(x,y) \;\leq\; d_p(x,y) \;\leq\; p \cdot d_{L^2}(x,y). \tag{2.5}$$

Possible choices for g are the following:

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

(2) $g(c;p) = \exp(-c/p)$. In this case $h_k(p) = \lambda_k/(\lambda_k + 1/p)$.

3 Asymptotic results

In this section, we present the main results concerning the first and second-order asymptotic properties of the functional sample mean. These results are essential to construct inferential procedures for testing the equality among the mean of one functional sample and a fixed function or for testing the equality of the means of two samples. The inferential framework is presented in Section 4.

Here, we consider a random function X and we establish the Large Law of Number (LLN) and the Central Limit Theorem (CLT) for the generalized Mahalanobis distance of the sample mean of i.i.d. realizations of X. Then, in Subsection 3.1 analogous results for the distance among the means of two independent random functions are obtained. In Subsection 3.2 we describe how these asymptotic results must be modified when the covariance structure is unknown and it must be estimated from the data. All the proofs are gathered in Appendix B.

Let $X_1, .., X_n$ be $n \geq 1$ i.i.d. realizations of a random function $X \in L^2(T)$, with $T \subset \mathbb{R}$ compact interval, and assume $\mathbb{E}[\|X\|^4] < \infty$; let m be the mean function of X and let $\{\lambda_k; k \geq 1\}$ and $\{\varphi_k; k \geq 1\}$ be the eigenvalues and the eigenfunctions, respectively, of the covariance operator of X. First, we report a functional strong law of large numbers concerning the distance d_p. The result is the following:

Theorem 3.1 *Let $\bar{X}_n$ be the pointwise sample mean $\bar{X}_n = (X_1 + ... + X_n)/n$. Then, for any $m_0 \in L^2(T)$, we have that*

$$d_p\left(\bar{X}_n, m_0\right) \xrightarrow{a.s.} d_p\left(m, m_0\right). \tag{3.1}$$

The proof of Theorem 3.1 is presented in Appendix B. Theorem 3.1 ensures that $d_p\left(\bar{X}_n, m_0\right)$ is a consistent estimator for the quantity $d_p\left(m, m_0\right)$. Hence, it seems reasonable to consider the statistics $d_p\left(\bar{X}_n, m_0\right)$ as statistics for testing the distance among m_0 and m, once we get its asymptotic probability distribution. This is provided by the following result, which shows the rate of convergence and the limiting distribution of $d_p\left(\bar{X}_n, m_0\right)$, for any function $m_0 \in L^2(T)$.

Theorem 3.2 *Let ψ_p be a positive random variable defined as follows*

$$\psi_p^2 := \sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p), \tag{3.2}$$

where $\{\chi^2_{1,k}; k \geq 1\}$ is a sequence of i.i.d chi-squared variables with 1 degree of freedom. Then, we have that

$$\sqrt{n} \cdot d_p\left(\bar{X}_n, m\right) \xrightarrow{\mathcal{D}} \psi_p, \tag{3.3}$$

For any $m_0 \neq m$, we have that

$$\sqrt{n} \cdot \left(d_p\left(\bar{X}_n, m_0\right) - d_p\left(m, m_0\right)\right) \xrightarrow{\mathcal{D}} r_p(m, m_0) \cdot Z, \tag{3.4}$$

where $r_p(m, m_0) := \left[d_p\left(m, m_0\right)\right]^{-1} \widetilde{d}_p\left(m, m_0\right)$ and

$$\widetilde{d}_p\left(m, m_0\right) := \sum_{k=1}^{\infty} \left(\frac{\langle m - m_0, \varphi_k\rangle}{\sqrt{\lambda_k}}\right)^2 \cdot \left(h_k(p)\right)^2. \tag{3.5}$$

The proof of Theorem 3.2 is presented in Appendix B, and it requires the auxiliary Theorem A.1, that is reported in Appendix A.

It is worth noting that the limiting distribution is not unique for any function m_0: when m_0 is the mean of the process X, the asymptotic distribution is a series of weighted chi-squared, while when $m_0 \neq m$, the asymptotic distribution is Gaussian. Moreover, note that, since $0 \leq h_k(p) \leq 1$ for any $p > 0$ and $k \geq 1$, we have that $\widetilde{d}_p(x, y) \leq d_p(x, y)$ for any $x, y \in L^2(T)$. Hence, the asymptotic variance $[r_p(m, m_0)]^2$ in (3.4) is always less than or equal to one.

Remark 3.3 *It is worth noting that analogous asymptotic results can be obtained for multivariate random functions $\boldsymbol{X} = (X_1, .., X_h)^T$, with $h \geq 2$, where $X_i \in L^2(t)$ for any $i \in \{1, .., h\}$. In that case, the mean $\boldsymbol{m} = \boldsymbol{E}[\boldsymbol{X}]$ is defined as a vector of functions in $L^2(T; \mathbb{R})$ such that $m_l = \boldsymbol{E}[X_l]$ for any $l = \{1, .., h\}$, and the covariance kernel $v(s, t) = \boldsymbol{Cov}[\boldsymbol{X(s)} \otimes \boldsymbol{X(t)}]$ is defined as a $h \times h$ matrix of functions such that $v_{l_1 l_2}(s, t) := \boldsymbol{Cov}[X_{l_1}(s), X_{l_2}(t)]$ for any $l_1, l_2 = \{1, .., h\}$. The scalar product between two elements y and w of $L^2(T)$ with values in $\mathbb{R}^h$ is defined as follows*

$$\langle y, w \rangle = \sum_{l=1}^{h} \int_T y_l(t) w_l(t) dt.$$

The eigenvalues $\{\lambda_k; k \geq 1\}$ and the eigenfunctions $\{\boldsymbol{\varphi_k}; k \geq 1\}$ of v are the elements solving $\sum_{l_2}^{h} \langle v_{l_1 l_2}, \varphi_{k l_2} \rangle = \lambda_k \varphi_{k l_1}$ for any $l_1 = \{1, .., h\}$. The generalized Mahalanobis distance d_p can be defined as in (2.4) using these quantities and the asymptotic results presented in this paper hold as well.

3.1 Asymptotic results for two populations of random functions

Here, we consider the results related to the distance d_p between the means of two independent random processes X_1 and X_2 of $L^2(T; \mathbb{R})$, with $\mathbb{E}[\|X_1\|^4], \mathbb{E}[\|X_2\|^4] < \infty$. Let $n_1, n_2 \geq 1$ and denote by $X_{11}, .., X_{1n_1}$ and $X_{21}, .., X_{2n_2}$ two i.i.d. samples of functions from X_1 and X_2, respectively. Consider the process X_j, $j = \{1, 2\}$, and call m_j its mean function and v_j its covariance function. Let $N := n_1 + n_2$ and assume that both n_1 and n_2 increase to infinity as N increases. To obtain asymptotic results for two populations, we need to define the right covariance function that generates the distance d_p defined in (2.4). To this end, since

$$\begin{aligned} &\boldsymbol{E}\left[(\bar{X}_{1n_1}(s) - \bar{X}_{2n_2}(s))(\bar{X}_{1n_1}(t) - \bar{X}_{2n_2}(t))\right] = \\ &(m_1(s) - m_2(s))(m_1(t) - m_2(t)) + \frac{v_1(s,t)}{n_1} + \frac{v_2(s,t)}{n_2}, \end{aligned}$$

the covariance function of $(1/n_1 + 1/n_2)^{-1/2} \cdot (\bar{X}_{1n_1} - \bar{X}_{2n_2})$ can be written as $(1 - c_N)v_1 + c_N v_2$, where $c_N = n_1/N$. Hence, the asymptotic covariance function can be identified whenever the following assumption holds

Assumption 1 *At least one of the following conditions must be satisfied:*

(c1) $v_1 = v_2$, *i.e.* X_1 *and* X_2 *have the same covariance function;*

(c2) $\exists c \in (0, 1)$, *such that* $n_1/N \to_N c$, *i.e. the sample sizes* n_1 *and* n_2 *increases to infinity at the same rate.*

Now, define $v := v_1 = v_2$ when (c1) holds, or $v := (1-c)v_1 + cv_2$ when (c2) holds, and denote by $\{\lambda_k; k \geq 1\}$ and $\{\varphi_k; k \geq 1\}$ the sequences of the eigenvalues and the eigenfunctions of v, respectively. This new covariance structure is considered to generate the distance d_p defined in (2.4). Thus, we can prove the following asymptotic results:

Theorem 3.4 *Consider Assumption 1. Then,*

$$d_p\left(\bar{X}_{1n_1}, \bar{X}_{2n_2}\right) \xrightarrow{a.s.}_N d_p\left(m_1, m_2\right). \tag{3.6}$$

If $m_1 = m_2$, we have

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot d_p\left(\bar{X}_{1n_1}, \bar{X}_{2n_2}\right) \xrightarrow{\mathcal{D}}_N \psi_p, \tag{3.7}$$

while if $m_1 \neq m_2$, we obtain

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot \left(d_p\left(\bar{X}_{1n_1}, \bar{X}_{2n_2}\right) - d_p\left(m_1, m_2\right)\right) \xrightarrow{\mathcal{D}}_N r_p(m_1, m_2) \cdot Z, \tag{3.8}$$

where we recall that $\psi_p^2 = \sum_{k=1}^{\infty} \chi_{1,k}^2 h_k(p)$ and $r_p(m_1, m_2) = \left[d_p\left(m_1, m_2\right)\right]^{-1} \widetilde{d}_p\left(m_1, m_2\right)$.

The proof of Theorem 3.4 is presented in Appendix B, and it requires the auxiliary Theorem A.1, that is reported Appendix A.

3.2 Asymptotic results in the case of unknown covariance function

All the results presented in Section 3 can be extended for the case of unknown covariance functions. In this situation, the eigenvalues and eigenfunctions used to compute the distance d_p in (2.4) are estimated from the data.

Consider an i.i.d. functional sample $X_1, .., X_n$ from the random process X. The covariance function v of X can be estimated by the following quantity

$$\widehat{v}_n := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i(s) - \bar{X}_n(s)\right)\left(X_i(t) - \bar{X}_n(t)\right).$$

Let us denote by $\{\widehat{\lambda}_k; k \geq 1\}$ the ordered eigenvalues of $\widehat{v}_n$ and $\{\widehat{\varphi}_k; k \geq 1\}$ the corresponding eigenfunctions. Naturally, since $\widehat{v}_n$ is computed using n functions, we have that $\widehat{\lambda}_k = 0$ for all $k \geq n$, and so the eigenfunctions $\widehat{\varphi}_n, \widehat{\varphi}_{n+1}, ..$ can be arbitrary chosen such that $\{\widehat{\varphi}_k; k \geq 1\}$ represents an orthonormal basis of $L^2(T)$.

The estimator of d_p based on the covariance estimator $\widehat{v}_n$ is defined as follows

$$\widehat{d}_p^2(y, w) := \sum_{k=1}^{n-1} \widehat{d}_{M,k}^2(y, w) \cdot \widehat{h}_k(p) + p \sum_{k=n}^{\infty} \left(\langle y - w, \widehat{\varphi}_k \rangle\right)^2, \tag{3.9}$$

with $y, w \in L^2(T)$, where $\widehat{d}_{M,k}^2(\cdot, \cdot)$ and $\widehat{h}_k(p)$ indicate the quantities $d_{M,k}^2(\cdot, \cdot)$ and $h_k(p)$, with λ_k replaced by $\widehat{\lambda}_k$ and φ_k replaced by $\widehat{\varphi}_k$. The definition of $\widehat{d}_p$ in (3.9) requires a specific explanation. Comparing the definitions of $\widehat{d}_p$ in (3.9) and d_p in (2.4), we note how the first $n-1$ components are similar, while the terms $k \geq n$ seem very different. The reason is that, since the span of the data $\{X_1 - \bar{X}_n, .., X_n - \bar{X}_n\}$ generates a space of dimension $(n-1)$, we have $\widehat{\lambda}_k = 0$ for all $k \geq n$ and so $\widehat{d}_{M,k}$ is undefined for $k \geq n$. Hence, we need to change the definition of $\widehat{d}_p$ in order to make all the components of the estimate $\widehat{d}_p$ as close as possible to the components of the distance d_p. Specifically, since $\lambda_k^{-1} h_k(p) \to p$ as $k \to \infty$, in (3.9) we have redefined $\widehat{d}_{M,k}\widehat{h}_k(p) := p \cdot (\langle y - w, \widehat{\varphi}_k \rangle)^2$ for any $k \geq n$.

Below, we present a theorem that extends the asymptotic results for one population of random functions to the case of estimated covariance function.

Theorem 3.5 *We have that*

$$\widehat{d}_p \left(\bar{X}_n, m_0\right) \;\xrightarrow{\;p\;}\; d_p \left(m, m_0\right). \tag{3.10}$$

Moreover,

$$\sqrt{n} \cdot \widehat{d}_p \left(\bar{X}_n, m\right) \;\xrightarrow{\;\mathcal{D}\;}\; \psi_p. \tag{3.11}$$

and, for any $m_0 \neq m$, we have

$$\sqrt{n} \cdot \left(\widehat{d}_p \left(\bar{X}_n, m_0\right) - \widehat{d}_p \left(m, m_0\right)\right) \;\xrightarrow{\;\mathcal{D}\;}\; r_p(m, m_0) \cdot Z, \tag{3.12}$$

where we recall that $\psi_p^2 = \sum_{k=1}^{\infty} \chi_{1,k}^2 h_k(p)$ and $r_p(m, m_0) = \left[d_p\left(m, m_0\right)\right]^{-1} \widetilde{d}_p\left(m, m_0\right)$. The proof of Theorem 3.5 is presented in Appendix B, and it requires the auxiliary Lemma A.2 that is reported in Appendix A.

Analogously, we can deal with the case of two populations presented in Subsection 3.1. In this case, for each $j = 1, 2$, consider the i.i.d. functional samples $X_{j1}, .., X_{jn_j}$ from the random process X_j, denote by $\bar{X}_{jn_j}$ the pointwise sample mean, i.e. $\bar{X}_{jn_j} = (X_{j1}, .., X_{jn_j})/n_j$, and let $\widehat{v}_{j,n_j}$ be the estimated covariance function, i.e.

$$\widehat{v}_{j,n_j} \;:=\; \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left(X_{i,j}(s) - \bar{X}_{jn_j}(s)\right) \left(X_{i,j}(t) - \bar{X}_{jn_j}(t)\right).$$

Now, let Assumption 1 be satisfied and define $v := v_1 = v_2$ when (c1) holds, or $v := (1-c)v_1 + cv_2$ when (c2) holds. Thus, assuming (c1), we define $\widehat{v}_N$ as the pooled estimator of v, i.e.

$$\widehat{v}_N \;:=\; \frac{(n_1 - 1)\widehat{v}_{1,n_1} + (n_2 - 1)\widehat{v}_{2,n_2}}{N - 2}. \tag{3.13}$$

On the other hand, assuming (c2), we define $\widehat{v}_N$ as the linear combination of the estimated covariance functions in each group, i.e.

$$\widehat{v}_N \;:=\; (1 - c_N)\widehat{v}_{1,n_1} + c_N \widehat{v}_{2,n_2},$$

where we remind that $c_N = n_1/N$ and $N = n_1 + n_2$. Then, by using the same arguments contained in the proof of Theorem 3.5, it is straightforward to extend the results concerning two populations presented in Theorem 3.4 to the case of estimated covariance function. For this reason, we omit it in the paper.

4 Inference on the mean

In this section we construct inferential procedures for testing the means of random processes and we discuss their main statistical properties. Specifically, we propose critical regions based on the generalized Mahalanobis distance d_p and we compute the asymptotic analytic expression of the corresponding power. Both the cases of known and unknown covariance function are analyzed. For each case, we present tests on the mean of a single population of random functions, and tests to compare the means of two independent populations.

4.1 Inference with known covariance function

First, we propose critical regions for testing the mean m of a process X with respect to an arbitrary function m_0 in $L^2(T)$, i.e.

$$H_0: \ m = m_0 \qquad vs \qquad H_1: \ m \neq m_0. \tag{4.1}$$

Consider an i.i.d. sample of random functions $X_1, .., X_n$ distributed as X. From (3.3) of Theorem 3.2, we have that the following critical region is asymptotically of level α

$$R^1_\alpha = \left\{\sqrt{n} \cdot d_p(\bar{X}_n, m_0) > \xi_{\alpha,p}\right\}, \tag{4.2}$$

where $\xi_{\alpha,p}$ indicates a quantile of the distribution of ψ_p, which can be numerically computed. Note that this probability distribution depends on the entire sequence $\{\lambda_k; k \geq 1\}$, on the choice of g and on the parameter p.

The power of test (4.1) based on the critical region (4.2) can be obtained using (3.4) of Theorem 3.2 as follows

$$\begin{aligned} P_{m\neq m_0}\left(R^1_\alpha\right) &= P_{m\neq m_0}\left(\sqrt{n} \cdot d_p(\bar{X}_n, m_0) > \xi_{\alpha,p}\right) \\ &\sim P_{m\neq m_0}\left(Z > \left[r_p(m, m_0)\right]^{-1} \cdot \left[\xi_{\alpha,p} - \sqrt{n} \cdot d_p(m, m_0)\right]\right), \end{aligned}$$

where we recall that $r_p(m, m_0) = \left[d_p\left(m, m_0\right)\right]^{-1} \widetilde{d}_p\left(m, m_0\right)$.

Now, consider the framework of Subsection 3.1, where we have two i.i.d. samples $X_{j1}, .., X_{jn_j}$ from the processes X_j, $j = 1, 2$, with the corresponding mean m_j and covariance function v_j. Let Assumption 1 hold and consider the following hypothesis test that compares the two mean functions

$$H_0: \ m_1 = m_2 \qquad vs \qquad H_1: \ m_1 \neq m_2. \tag{4.3}$$

Thus, a critical region asymptotically of level α can be obtain using (3.7) of Theorem 3.4 as follows

$$R^2_\alpha = \left\{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot d_p(\bar{X}_{1n_1}, \bar{X}_{2n_2}) > \xi_{\alpha,p}\right\}. \tag{4.4}$$

The power of (4.4) can be obtained using (3.8) of Theorem 3.4 as follows

$$\begin{aligned} P_{m_1\neq m_2}\left(R^2_\alpha\right) &= P_{m_1\neq m_2}\left(\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot d_p(\bar{X}_{1n_1}, \bar{X}_{2n_2}) > \xi_{\alpha,p}\right) \\ &\sim P_{m_1\neq m_2}\left(Z > \left[r_p(m_1, m_2)\right]^{-1} \cdot \left[\xi_{\alpha,p} - \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot d_p(m_1, m_2)\right]\right), \end{aligned}$$

where we recall that $r_p(m_1, m_2) = [d_p\,(m_1, m_2)]^{-1}\,\widetilde{d}_p\,(m_1, m_2)$.

Remark 4.1 *From Remark 3.3 we have that the analogous asymptotic results presented in Section 3 hold as well for multivariate random functions $\boldsymbol{X} = (X_1, .., X_h)^T \in L^2(T; \mathbb{R}^h)$, with $h \geq 2$. Then, the critical regions introduced in Section 4 based on those asymptotic results can be analogously used for testing the means of multivariate random functions.*

4.2 Inference with unknown covariance function

Here, we extend the inferential procedures presented in Section 4 to the case of unknown covariance structure. We propose tests similar to (4.2) and (4.4), where the covariance operator and the related eigenvalues and eigenfunctions are estimated from the data.

Theorem 3.5 ensures that, concerning the hypothesis tests (4.1) and (4.3), the critical regions obtained using covariance estimator show the same structure of those in (4.2) and (4.4) where the covariance function is assumed to be known. However, the asymptotic distribution of ψ_p depends on the eigenvalues of v, which are unknown here. Thus, to compute the critical regions we need a further asymptotic result which corresponds to Theorem 4.3 of Ghiglietti and Paganoni (2014)

Theorem 4.2 *Let $\{\chi^2_{1,k}; k \geq 1\}$ be a sequence of i.i.d. chi-squared with 1 d.f. independent of $\widehat{v}_n$. Let $\widehat{\xi}_{\alpha,p}$ be the $1-\alpha$ quantile of the conditional distribution of ψ_p given $\{\widehat{\lambda}_k; k \geq 1\}$. Then, we have that*

$$\widehat{\xi}_{\alpha,p} \xrightarrow{p}_n \xi_{\alpha,p} \tag{4.5}$$

The proof of Theorem 4.2 is reported in Theorem 4.3 of Ghiglietti and Paganoni (2014), and so we have omitted it here.

Consider the hypothesis tests (4.1) for the mean of a process X. By using Slustsky's Theorem, (3.11) and (4.5), we have that the following critical region is asymptotically of level α:

$$R^3_\alpha \;=\; \left\{\; \sqrt{n}\cdot \widehat{d}_p(\bar{X}_n, m) > \widehat{\xi}_{\alpha,p} \;\right\}. \tag{4.6}$$

Moreover, by following standard arguments and using (3.12), we can compute the power of test (4.6) as follows

$$\begin{aligned} P_{m\neq m_0}\left(R^3_\alpha\right) \quad &= P_{m\neq m_0}\left(\sqrt{n}\cdot d_p(\bar{X}_n, m_0) > \widehat{\xi}_{\alpha,p}\right) \\ &\sim P_{m\neq m_0}\left(Z > [r_p(m, m_0)]^{-1}\cdot\left[\xi_{\alpha,p} - \sqrt{n}\cdot \widehat{d}_p(m, m_0)\right]\right), \end{aligned}$$

where we recall that $r_p(m, m_0) = [d_p\,(m, m_0)]^{-1}\,\widetilde{d}_p\,(m, m_0)$.

Now, consider the framework of Subsection 3.1, where two i.i.d. samples $X_{j1}, .., X_{jn_j}$ from independent processes X_j, $j = 1, 2$, are used for the hypothesis tests (4.3). Let Assumption 1 hold. Hence, by following analogous arguments, it is possible to show that the following critical region is asymptotically of level α:

$$R^4_\alpha \;=\; \left\{\; \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot \widehat{d}_p(\bar{X}_{1n_1}, \bar{X}_{2n_2}) > \widehat{\xi}_{\alpha,p} \;\right\} \tag{4.7}$$

Moreover, the power can be derived as follows

$$\begin{aligned} P_{m_1 \neq m_2}\left(R^4_\alpha\right) &= P_{m_1 \neq m_2}\left(\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1/2} \cdot \widehat{d}_p(\bar{X}_{1n_1}, \bar{X}_{2n_2}) > \xi_{\alpha,p}\right) \\ &\sim P_{m_1 \neq m_2}\left([r_p(m_1,m_2)]^{-1} \cdot \left[\xi_{\alpha,p} - \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1/2} \cdot \widehat{d}_p(m_1,m_2)\right]\right), \end{aligned}$$

where we recall that $r_p(m_1, m_2) = [d_p\,(m_1, m_2)]^{-1}\, \widetilde{d}_p\,(m_1, m_2)$.

5 Simulations and Case Study

In this section, we illustrate the inferential properties of the tests proposed in the paper in a simulation setting, and we then present a real case study where the aim is to distinguish two populations of electrocardiograms (ECGs).

5.1 Simulation Study

In this simulation study, we consider two samples of i.i.d. curves generated by independent stochastic processes with different means (say $m_1,\ m_2 \in L^2(T)$ and the same covariance function. We wish to test the null hypothesis H_0: $m_1 = m_2$ in L^2 against the alternative H_1. The sample curves are generated as follows:

$$\begin{aligned} X_{1i}(t) &= m_1(t) \ + \ \sum_{k=1}^{\infty} U_{1i,k}\sqrt{\lambda_k}\varphi_k(t) \quad \text{for } t \in [0,1],\ i \in \{1,..,n_1\}, \\ X_{2i}(t) &= m_2(t) \ + \ \sum_{k=1}^{\infty} U_{2i,k}\sqrt{\lambda_k}\varphi_k(t) \quad \text{for } t \in [0,1],\ i \in \{1,..,n_2\}. \end{aligned}$$

We set

(i) the sample sizes $n_1 = n_2 = 300$ for ease of notation;

(ii) the random variables $\{U_{ji,k}; k \geq 1, j \in \{1,2\}, 1 \leq i \leq n_j\}$ as independent random variables uniformly distributed in $(-\sqrt{3}, \sqrt{3})$, so that $\boldsymbol{E}[U_k] = 0$ and $\boldsymbol{Var}[U_k] = 1$;

(iii) $\{\lambda_k; k \geq 1\}$ is the sequence of eigenvalues of the covariance function v defined as follows:

$$\lambda_k \ = \begin{cases} \frac{1}{k+1} & \text{if } k \in \{1,2,3\}, \\ \frac{1}{(k+1)^4} & \text{if } k \geq 4; \end{cases} \tag{5.1}$$

(iv) $\{\varphi_k; k \geq 1\}$ is the sequence of eigenfunctions of the covariance function v defined as follows:

$$\varphi_k \ = \begin{cases} 1 & \text{if } k = 1, \\ \sqrt{2}\sin(k\pi t) & \text{if } k \geq 2,\ k \textit{ even}, \\ \sqrt{2}\cos((k-1)\pi t) & \text{if } k \geq 3,\ k \textit{ odd}; \end{cases}$$

(v) the mean functions $m_1(t) = 4t(1-t)$ and $m_2(t) = m_1(t) + 3\sqrt{\lambda_4}\varphi_4(t)$ (see Figure 1, left panel).

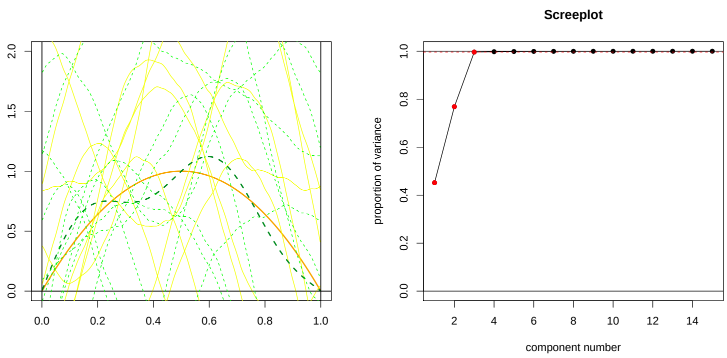

Fig. 1: Left panel: 10 curves from sample 1 (solid yellow lines) with the mean function m_1 (solid red line), and 10 curves from sample 2 (dashed light green lines) with the mean function m_2 (dashed green line). Right panel: screeplot of the estimated eigenvalues $\hat{\lambda}_k$.

First, consider the test defined by the critical region (4.7) presented in Subsection 4.2, i.e.

$$R^4_\alpha = \left\{ \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \cdot \widehat{d}_p(\bar{X}_{1n_1}, \bar{X}_{2n_2}) > \widehat{\xi}_{\alpha,p} \right\},$$

where we set $p = 1$ and $\alpha = 0.05$. We repeated the above test on 10^3 samples generated by simulation, and we obtained an empirical power of 0.865.

Now, let us consider other test statistics typically used in the functional principle component analysis (FPCA) approach (see Horvath and Kokoszka (2012)), where the inference is based on the projection of the difference between the sample means on the space generated by the components with higher variability. In the sequence of eigenvalues defined in (iv), the first three components explain most of the variance of the data. To see this in practice, we analyze the estimated eigenvalues $\hat{\lambda}_k$ obtained by the usual covariance estimator $\widehat{v}_N$ (see (3.13)). The screeplot of the $\hat{\lambda}_k$s, reported in Figure 1 (right panel), shows that the first three components contain a large percentage of the total variance of the processes. First, we consider the test statistics $T^{(1)}_{n_1n_2}$ and $T^{(2)}_{n_1n_2}$ presented in Horvath and Kokoszka (2012) as follows:

(1) $T^{(1)}_{n_1n_2} := \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \sum_{k=1}^3 \left(\frac{\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \hat{\varphi}_k \rangle}{\sqrt{\hat{\lambda}_k}}\right)^2$;

(2) $T^{(2)}_{n_1n_2} := \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \sum_{k=1}^3 \left(\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \hat{\varphi}_k \rangle\right)^2$.

As explained in Horvath and Kokoszka (2012), when the mean difference $(m_2(t) - m_1(t))$ is orthogonal to the linear span of $\{\varphi_1, \varphi_2, \varphi_3\}$, as in this simulation study, tests based on the statistics $T^{(1)}_{n_1n_2}$ and $T^{(2)}_{n_1n_2}$ will not reject H_0. Specifically, when m_1 and m_2 have the same projection on the subspace generated by $\{\varphi_1, \varphi_2, \varphi_3\}$, the power of such tests coincides with the nominal level α. This issue is addressed by the test based on R^4_α since, unlike $T^{(1)}_{n_1n_2}$ and $T^{(2)}_{n_1n_2}$, all the components provide a contribute to the distance $\widehat{d}_p$.

Another statistics that is proposed in Horvath and Kokoszka (2012) is based on the L^2 distance between the sample means:

$$(3)\quad U_{n_1n_2} := \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \int_0^1 \left(\bar{X}_{1n_1}(t) - \bar{X}_{2n_2}(t)\right)^2 dt.$$

Unlike $T^{(1)}_{n_1n_2}$ and $T^{(2)}_{n_1n_2}$, the statistics $U_{n_1n_2}$ is able to detect the mean difference along any component; however, the covariance structure of the process does not play any role in $U_{n_1n_2}$ and hence, since there is no adjustment for correlations and variability along the different components, this may lead to a severe loss of power. We repeated the test based on $U_{n_1n_2}$ at level $\alpha = 0.05$, with 10^3 samples generated by simulation, obtaining an empirical power equal to 0.563. Therefore, the test based on R^4_α is characterized by an higher power (0.865) since, unlike $U^{(1)}_{n_1n_2}$, the distance $\widehat{d}_{p,N}$ takes into account the covariance structure of the process.

Finally, the last competitor we consider is the one proposed in Pini at al (2013). This method implements a two populations interval testing procedure, based on permutation tests, for testing the difference between two functional populations evaluated on a uniform grid. Data are represented by means of the B-spline basis and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate (FWER). This procedure has been implemented by applying the R package *fdatest: Interval Testing Procedure for Functional Data* developed in Pini et al.(2015). to two functional samples generated by simulation and distributed as X_1 and X_2. Figure 2 reports the results. Since the curve of the adjusted p-values is above the significance level $\alpha = 0.05$ for any $t \in T$ (see Figure 2, left panel), the nonparametric testing procedure proposed in Pini et al. (2013) does not reject the null hypothesis, whereas the p-value of the test (4.7) applied to this dataset is 0.03.

Figure 2 reports the results obtained from the procedure.

5.2 Case Study

In this section we apply the inferential procedures proposed in Subsection 4.2 to a dataset composed by Electrocardiographic signals (ECGs). The basic statistical unit is the eight-variate function (the ECG) which describes the heart dynamics of a patient on the eight leads I, II, V1, V2, V3, V4, V5 and V6. We will focus on lead I (the most representative for the pathology we are going to consider) in order to have one functional data associated to each patient. The generalization to the multivariate functional case is also possible (see Remark 3.3).

The sample of curves we consider are extracted from a wider database (PROMETEO - PROgetto Milano Ecg Teletrasmessi Extra Ospedaliero, for further information, see Ieva and Paganoni (2013) and references therein) containing 6758 ECG signals both from healty and not healthy subjects.

Figure 3 shows the I lead of the ECG signal of 100 healthy patients (red curves, left panel) and of 100 patients affected by LBBB (blue curves, central panel), i.e., Left Bundle Brunch Block, a kind of Acute Myocardial Infarction. Their sample means, depicted in Figure 3 (right panel), appear pretty different in morphology.
By applying the test proposed in Subsection 4.2 (two balanced samples with different unknown variances), and by setting the parameters as reported in Table 5.2,

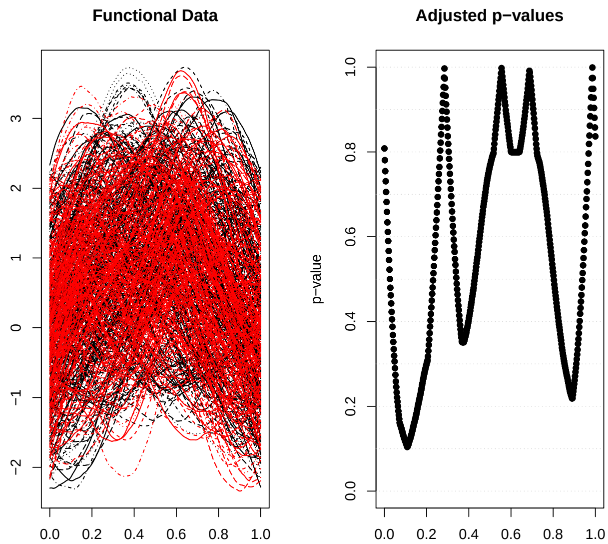

Fig. 2: Results of the nonparametric testing procedure proposed in Pini et al (2013) applied to the simulated dataset described in Subsection 5.1.

we get evidence for rejecting the null hypothesis of equal means (for all values of the tuning parameters p, the test statistics fell in the critical region R^4_α). We denote by $T_{0,p}$ the test statistic corresponding to the critical region in equation (4.7), computed for given values of the tuning parameters p. Since different choices for p do not change the result of the test and both $T_{0,p}$ and the quantiles are monotonic with respect to p, we decided to report only the extreme values corresponding to $p = 10^{-5}$ and $p = 1$.
Also the method proposed in Pini et al (2013) provides evidence for rejecting the null hypothesis in this case. Figure 4 shows the portion of the time domain the rejection is due to. In this case the output is helpful for identifying the time domain areas where the differences between the two functional populations are significant, but computational costs are considerable (43 hours).

6 Conclusions

In this paper, we propose inferential procedures to test difference between means of functional data. The proposed tests do not require any specific distribution assumption on the processes generating the data and allow comparison between samples with different covariance functions. The inference is based on a suitable generalization d_p of the Mahalanobis distance to the Hilbert space L^2. Theoretical results on the asymptotic behaviour of d_p distance between sample mean and a fixed function $m(t) \in L^2(T)$ and between means of two independent samples are proved. Additionally, the rate of convergence and the exact asymptotic distributions of the distance d_p between functional sample means has been studied. In fact, despite

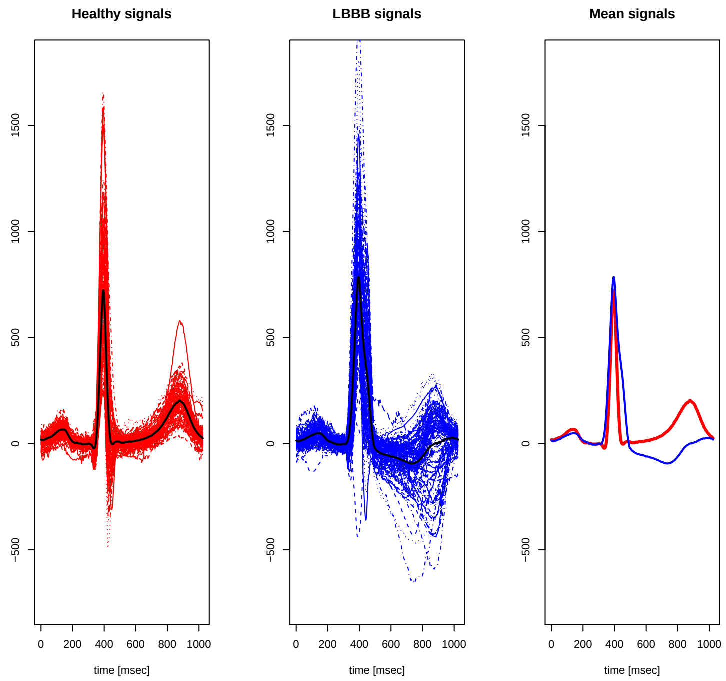

Fig. 3: Healthy (red lines, left pnel) and Left Bundle Brunch Block (blue lines, second panel) ECGs, with corresponding means (third panel).

the massive methodological and theoretical development of statistical inference for functional data, testing hypothesis on the distribution of stochastic processes generating two samples of curves has received little attention. The generalization to the inference on processes with cladag trajectories is the object of future work together with the study of properties of Skorohod distance in this wider context.

Acknowledgments

The authors wish to thank Mortara Rangoni Europe s.r.l. and the 118 Dispatch Centre of Milan for providing data within the project PROMETEO (PROgetto sull'area Milanese Elettrocardiogrammi Teletrasferiti dall'Extra Ospedaliero).

A Auxiliary results

In this section, we present some auxiliary results that are required to prove the main theorems of the paper, whose proofs are gathered in Appendix B. Specifically, Theorem A.1 represents the crucial result to prove the asymptotic distributions of $d_p(\bar{X}, m)$ and $d_p(\bar{X}, m_0)$ presented in (3.3) and (3.4) in Theorem 3.2. Naturally, the analogous results for two populations and unknown covariance function presented in Theorem 3.4 and 3.5, respectively, require Theorem A.1 as well. Theorem A.1 is presented below in a very general framework.

Test on the means of two populations	
Level of the test	$\alpha = 0.05$
Size of the samples	$n_1 = n_2 = 100$
Test Statistics	$T_{0,p=\{0.00001,1\}} = \{196.40, 1979.73\}$
Quantiles	$\hat{\xi}_{0.05,p==\{10^{-5},1\}} = \left\{2.22 * e^{-3}, 16.39\right\}$

Further parameters	
Time grid	$T = [0; 1024]$
Truncation parameter	$k = 200$
Tuning parameter	$p = \left\{10^{-5}, 1\right\}$

Tab. 1: Parameters settings for the test of the means of two populations of stochastic processes with different unknown covariances.

Theorem A.1 *Let $\{W_n; n \geq 1\}$ be a real-valued random sequence. Consider two sequences of continuous random variables $\{X_{j,n}; j, n \geq 1\}$ and $\{Y_{j,n}; j, n \geq 1\}$ such that $W_n = X_{j,n} + Y_{j,n}$ for any $j, n \geq 1$. Assume the following results hold*

(a) $\sup_{n\geq 1}\{\boldsymbol{E}[|Y_{j,n}|]\} \longrightarrow_j 0$;

(b) for any $j \geq 1$, there exists a r.v. X_j such that $X_{j,n} \xrightarrow{\mathcal{D}}_n X_j$;

(c) there exists a continuous r.v. X such that $X_j \xrightarrow{\mathcal{D}}_j X$;

(d) there exists a constant $M > 0$ such that $\boldsymbol{P}(X \in I) \leq M|I|$ for any interval $I \subset \mathbb{R}$, where $|I|$ is the length of I.

Then,

$$W_n \xrightarrow{\mathcal{D}} X. \tag{A.1}$$

Proof. To prove (A.1), we fix $t \in \mathbb{R}$ and we show that, for any $\epsilon > 0$,

$$\lim_{n\to\infty} |\boldsymbol{P}(W_n \leq t) - \boldsymbol{P}(X \leq t)| \leq \epsilon.$$

To this end, we establish the following

(1) $\lim_{n\to\infty} \boldsymbol{P}(W_n \leq t) \geq \boldsymbol{P}(X \leq t) - \epsilon$;

(2) $\lim_{n\to\infty} \boldsymbol{P}(W_n \leq t) \leq \boldsymbol{P}(X \leq t) + \epsilon$.

Part (1)

For any $j \in \mathbb{N}$ and $\nu > 0$, since $W_n = X_{j,n} + Y_{j,n}$ we have

$$\{W_n \leq t\} \supset \{X_{j,n} \leq t - \nu\} \bigcap \{|Y_{j,n}| < \nu\},$$

which implies the following

$$\boldsymbol{P}(W_n \leq t) \geq \boldsymbol{P}(X_{j,n} \leq t - \nu) - \boldsymbol{P}(|Y_{j,n}| > \nu). \tag{A.2}$$

We will show that for large values of j and n, the term $\boldsymbol{P}(|Y_{j,n}| > \nu)$ is arbitrary small and $\boldsymbol{P}(X_{j,n} \leq t - \nu)$ is arbitrary close to $\boldsymbol{P}(X \leq t)$.

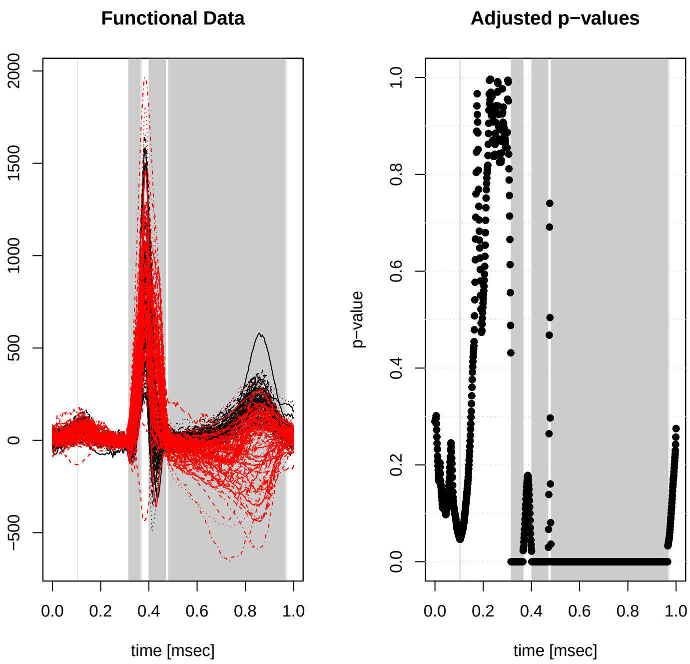

Fig. 4: Results of the nonparametric testing procedure proposed in Pini et al (2013).

Now, we present a list of preliminary results to make the proof easier to read. Choose $\nu := \epsilon(3M)^{-1}$ and, by applying Markov's inequality, note that

$$\boldsymbol{P}\left(|Y_{j,n}| > \nu\right) \;\leq\; \nu^{-1} \cdot \boldsymbol{E}[|Y_{j,n}|] \;\leq\; \nu^{-1} \cdot \sup_{n \geq n_0} \{\boldsymbol{E}[|Y_{j,n}|]\};$$

now, from (a) there exists $j_1 \geq 1$ such that $\sup_{n \geq n_0}\{\boldsymbol{E}[|Y_{j,n}|]\} \leq \epsilon^2/(9M)$ for any $j \geq j_1$; hence

$$\boldsymbol{P}\left(|Y_{j,n}| > \nu\right) \leq \frac{3M}{\epsilon} \cdot \frac{\epsilon^2}{9M} = \epsilon/3, \tag{A.3}$$

for any $j \geq j_1$. Moreover, from (c), there exists $j_2 \geq 1$ such that

$$|\boldsymbol{P}\left(X_j \leq t - \nu\right) - \boldsymbol{P}\left(X \leq t - \nu\right)| \leq \epsilon/3, \tag{A.4}$$

for any $j \geq j_2$; then, from (d), we have that

$$|\boldsymbol{P}\left(X \leq t - \nu\right) - \boldsymbol{P}\left(X \leq t\right)| \leq M\nu \;=\; \epsilon/3. \tag{A.5}$$

Now, consider (A.2) and fix $j \geq \max\{j_1; j_2\}$. From (b) and (A.3) we have that

$$\lim_{n \to \infty} \boldsymbol{P}\left(W_n \leq t\right) \geq \boldsymbol{P}\left(X_j \leq t - \nu\right) - \epsilon/3.$$

Moreover, from (A.4), we can obtain

$$\lim_{n \to \infty} \boldsymbol{P}\left(W_n \leq t\right) \geq \boldsymbol{P}\left(X \leq t - \nu\right) - 2\epsilon/3.$$

Finally, by using (A.5), we get

$$\lim_{n\to\infty} \boldsymbol{P}\left(W_n \leq t\right) \geq \boldsymbol{P}\left(X \leq t\right) - \epsilon.$$

Part (2)
The proof of this part is analogous to the previous one, once we note that

$$\left\{W_n \leq t\right\} \subset \left\{X_{j,n} \leq t+\nu\right\} \bigcup \left\{|Y_{j,n}| > \nu\right\},$$

which implies

$$\boldsymbol{P}\left(W_n \leq t\right) \leq \boldsymbol{P}\left(X_{j,n} \leq t+\nu\right) + \boldsymbol{P}\left(|Y_{j,n}| > \nu\right).$$

Then, from (A.3) we get $\boldsymbol{P}\left(|Y_{j,n}| > \nu\right) \leq \epsilon/3$ and by introducing results analogous to (A.4) and (A.5) we can obtain $\boldsymbol{P}\left(X_{j,n} \leq t+\nu\right) \leq \boldsymbol{P}\left(X \leq t\right) + 2\epsilon/3$, which concludes the proof. ■

Here, we present a Lemma required to prove the asymptotic results with estimated covariance function that are presented in Theorem 3.5. This result has been proved in Ghiglietti and Paganoni (2014), as Theorem 4.1.

Lemma A.2 *Let $\{(Y_n, W_n); n \geq 1\}$ be a pair of real-valued stochastic processes such that*

$$\sup_{n\geq 1}\left\{ \boldsymbol{E}\left[\|Y_n - W_n\|^4\right] \right\} < \infty, \tag{A.6}$$

and

$$\sum_{k=1} S_k < \infty, \qquad S_k := \sup_{n\geq 1} \mathbb{E}\left[\left(\langle Y_n - W_n, \varphi_k\rangle\right)^2\right], \tag{A.7}$$

Then, we have that

$$\mathbb{E}\left[|\widehat{d_p^2}(Y_n, W_n) - d_p^2(Y_n, W_n)|\right] \to_n 0. \tag{A.8}$$

B Proof of main results

Proof. [Proof of Theorem 3.1] Since $d_p\left(\bar{X}_n, m_0\right)$ is a non negative quantity, then (3.1) is equivalent to prove that

$$d_p^2\left(\bar{X}_n, m_0\right) \overset{a.s.}{\longrightarrow} d_p^2\left(m, m_0\right). \tag{B.1}$$

By noting that

$$d_p^2\left(\bar{X}_n, m_0\right) = d_p^2\left(\bar{X}_n - m + m, m_0\right) = \sum_{k=1}^{\infty}\left(\langle \bar{X}_n - m, \varphi_k\rangle + \langle m - m_0, \varphi_k\rangle\right)^2 \frac{h_k(p)}{\lambda_k},$$

we obtain the following decomposition

$$\begin{aligned} d_p^2\left(\bar{X}_n, m_0\right) - d_p^2\left(m, m_0\right) &= d_p^2\left(\bar{X}_n, m\right) \\ &+ 2\sum_{k=1}^{\infty}\left(\langle \bar{X}_n - m, \varphi_k\rangle\langle m - m_0, \varphi_k\rangle\right) \frac{h_k(p)}{\lambda_k}. \end{aligned} \tag{B.2}$$

From (B.2), we have that (B.1) can be obtained by establishing these two results

$$d_p\left(\bar{X}_n, m\right) \xrightarrow{a.s.} 0. \tag{B.3}$$

$$\sum_{k=1}^{\infty}\left(\langle \bar{X}_n - m, \varphi_k\rangle\langle m - m_0, \varphi_k\rangle\right)\frac{h_k(p)}{\lambda_k} \xrightarrow{a.s.} 0, \tag{B.4}$$

First, consider (B.3). Using the equivalence among the distance d_p and the usual-L^2 metric shown in (2.5), we have that

$$d_p\left(\bar{X}_n, m\right) \;\leq\; p \cdot \|\bar{X}_n - m\| \;\xrightarrow{a.s.}\; 0.$$

Then, consider (B.4). By using Holder's inequality and $h_k(p)/\lambda_k \leq p$, we have that the left hand side of (B.4) is less or equal to

$$p \cdot \sqrt{\sum_{k=1}^{\infty}\left(\langle \bar{X}_n - m, \varphi_k\rangle\right)^2} \cdot \sqrt{\sum_{k=1}^{\infty}\left(\langle m - m_0, \varphi_k\rangle\right)^2} = p \cdot \|\bar{X}_n - m\| \cdot \|m - m_0\|,$$

which tends to zero almost surely since $\|\bar{X}_n - m\| \xrightarrow{a.s.} 0$. This concludes the proof. ■

Proof. [Proof of Theorem 3.2] The proofs of both (3.3) and (3.4) are realized by applying Theorem A.1 to the corresponding random sequences.

We start by proving (3.3). Since $\sqrt{n} d_p\left(\bar{X}_n, m\right)$ is a non negative quantity, then (3.3) is equivalent to prove that

$$n \cdot d_p^2\left(\bar{X}_n, m\right) \xrightarrow{\mathcal{D}} \psi_p^2, \tag{B.5}$$

where we recall that $\psi_p^2 = \sum_{k=1}^{\infty} \chi_{1,k}^2 h_k(p)$. Using the notation of Theorem A.1, we define

$$\begin{aligned}
W_n &:= n \cdot d_p^2\left(\bar{X}_n, m\right), \\
X_{j,n} &:= n \cdot \sum_{k=1}^{j}\left(\langle \bar{X}_n - m, \varphi_k\rangle\right)^2 h_k(p)/\lambda_k, \\
Y_{j,n} &:= n \cdot \sum_{k=j+1}^{\infty}\left(\langle \bar{X}_n - m, \varphi_k\rangle\right)^2 h_k(p)/\lambda_k, \\
X_j &:= \sum_{k=1}^{j} \chi_{1,k}^2 h_k(p), \\
X &:= \sum_{k=1}^{\infty} \chi_{1,k}^2 h_k(p).
\end{aligned}$$

Then, once we prove that the assumptions of Theorem A.1 are satisfied, from Theorem A.1 we have that $W_n \xrightarrow{\mathcal{D}} X$, that is (B.5), and the proof of (3.3) is so concluded. We now show that the conditions (a)-(b)-(c)-(d) of Theorem A.1 hold:

(a) since $\boldsymbol{E}\left[|Y_{j,n}|\right] = \sum_{k=j+1}^{\infty} h_k(p)$ for any $n \geq 1$, we have that

$$\sup_{n\geq 1}\{\boldsymbol{E}[|Y_{j,n}|]\} \to_j 0;$$

(b) note that $X_{j,n}$ can be written as $X_{j,n} = \sum_{k=1}^{j} Z_{k,n}^2 h_k(p)$ where

$$Z_{k,n} \;:=\; \left(\frac{\langle \bar{X}_n - m, \varphi_k \rangle}{\sqrt{\lambda_k / n}} \right) \;=\; n^{-1/2} \sum_{i=1}^{n} \left(\frac{\langle X_i - m, \varphi_k \rangle}{\sqrt{\lambda_k}} \right).$$

Since $X_1, .., X_n$ are i.i.d., using the multivariate CLT, the j-vector $\boldsymbol{Z_n} := (Z_{1,n}, .., Z_{j,n})'$ converges in distribution to a standard Gaussian vector. Hence, the processes $Z_{1,n}, .., Z_{j,n}$ are asymptotically independent and, for each $k = 1, .., j$, $Z_{k,n}^2 \xrightarrow{\mathcal{D}}_n \chi_1^2$; as a consequence, $X_{j,n} \xrightarrow{\mathcal{D}}_n X_j$;

(c) we can prove $X_j \xrightarrow{\mathcal{D}}_j X$ by showing that $|X - X_j| \xrightarrow{p}_j 0$ as follows

$$\boldsymbol{E}\left[|X - X_j| \right] = \boldsymbol{E}\left[\sum_{k=j+1}^{\infty} \chi_{1,k}^2 h_k(p) \right] = \sum_{k=j+1}^{\infty} h_k(p) \longrightarrow_j 0;$$

(d) to prove this condition, call μ_1 and μ_{-1} the probability laws of $\chi_{1,1}^2 h_1(p)$ and $\sum_{k \geq 2} \chi_{1,k}^2 h_k(p)$, respectively; note that X is the sum of these two independent variables. Since $\chi_{1,1}^2 h_1(p)$ is a continuous random variable, we can denote by f_1 its density and, for any interval $I \subset \mathbb{R}$, we can write

$$\boldsymbol{P}(X \in I) \;=\; \int_I \left(\int_0^{\infty} f_1(x - y) \mu_{-1}(dy) \right) dx.$$

Now, calling $M := \max_{x \in \mathbb{R}} \{ f_1(x) \}$ we obtain

$$\boldsymbol{P}(X \in I) \;\leq\; \int_I \left(\int_0^{\infty} M \mu_{-1}(dy) \right) dx \;=\; M|I|.$$

This concludes the proof of (3.3).

Now, we focus on proving (3.4). First, note that

$$\left(d_p\left(\bar{X}_n, m_0 \right) - d_p\left(m, m_0 \right) \right) = \left(\frac{d_p^2\left(\bar{X}_n, m_0 \right) - d_p^2\left(m, m_0 \right)}{d_p\left(\bar{X}_n, m_0 \right) + d_p\left(m, m_0 \right)} \right);$$

then, since $d_p\left(\bar{X}_n, m_0 \right) \xrightarrow{a.s.} d_p\left(m, m_0 \right)$ from Theorem 3.1, by applying Slutsky's Theorem we have that equation (3.4) can be obtain by establishing that

$$\sqrt{n} \cdot \left(d_p^2\left(\bar{X}_n, m_0 \right) - d_p^2\left(m, m_0 \right) \right) \;\xrightarrow{\mathcal{D}}\; 2\widetilde{d}_p\left(m, m_0 \right) \cdot Z. \tag{B.6}$$

Moreover, by using decomposition (B.2) we can show (B.6) by establishing that

$$\sqrt{n} \cdot \sum_{k=1}^{\infty} \left(\langle \bar{X}_n - m, \varphi_k \rangle \langle m - m_0, \varphi_k \rangle \right) \frac{h_k(p)}{\lambda_k} \;\xrightarrow{\mathcal{D}}\; \widetilde{d}_p\left(m, m_0 \right) \cdot Z, \tag{B.7}$$

since by applying Slutsky's Theorem to (3.3) we have that $\sqrt{n} \cdot d_p^2\left(\bar{X}_n, m \right) \xrightarrow{p} 0$.

To prove (B.7), we apply Lemma A.1 with the following notation:

$$
\begin{aligned}
W_n &:= \sqrt{n} \cdot \sum_{k=1}^{\infty} \left(\langle \bar{X}_n - m, \varphi_k \rangle \langle m - m_0, \varphi_k \rangle \right) \frac{h_k(p)}{\lambda_k}, \\
X_{j,n} &:= \sqrt{n} \cdot \sum_{k=1}^{j} \left(\langle \bar{X}_n - m, \varphi_k \rangle \langle m - m_0, \varphi_k \rangle \right) \frac{h_k(p)}{\lambda_k}, \\
Y_{j,n} &:= \sqrt{n} \cdot \sum_{k=j+1}^{\infty} \left(\langle \bar{X}_n - m, \varphi_k \rangle \langle m - m_0, \varphi_k \rangle \right) \frac{h_k(p)}{\lambda_k}, \\
X_j &:= \sqrt{\sum_{k=1}^{j} \left(\langle m - m_0, \varphi_k \rangle \right)^2 \frac{h_k^2(p)}{\lambda_k}} \cdot Z, \\
X &:= \widetilde{d}_p\left(m, m_0\right) \cdot Z,
\end{aligned}
$$

where Z denotes a standard normal variable. Then, once we prove that the assumptions of Theorem A.1 are satisfied, from Theorem A.1 we have that $W_n \xrightarrow{\mathcal{D}} X$. This result is equal to (B.7) and the proof of (3.4) is so concluded. We now show that the conditions (a)-(b)-(c)-(d) of Theorem A.1 hold:

(a) using Cauchy-Schwartz inequality to $Y_{j,n}$ and since $h_k(p)/\lambda_k \leq p$, we obtain

$$
\boldsymbol{E}\left[|Y_{j,n}|\right] \;\leq\; p \cdot \sqrt{\sum_{k=j+1}^{\infty} \boldsymbol{E}\left[\left(\sqrt{n} \langle \bar{X}_n - m, \varphi_k \rangle \right)^2 \right]} \cdot \sqrt{\sum_{k=j+1}^{\infty} \left(\langle m - m_0, \varphi_k \rangle \right)^2},
$$

for any $n \geq 1$. Then, since $\boldsymbol{E}\left[\left(\sqrt{n} \langle \bar{X}_n - m, \varphi_k \rangle \right)^2 \right] = \lambda_k$, we have that

$$
\sup_{n \geq 1} \{ \boldsymbol{E}[|Y_{j,n}|] \} \;\leq\; p \cdot \sqrt{\sum_{k=j+1}^{\infty} \lambda_k} \cdot \| m - m_0 \| \;\longrightarrow_j\; 0;
$$

(b) note that $X_{j,n}$ can be written as

$$
X_{j,n} = \sum_{k=1}^{j} Z_{k,n} \left(\langle m - m_0, \varphi_k \rangle \right) \frac{h_k(p)}{\sqrt{\lambda_k}},
$$

where

$$
Z_{k,n} \;:=\; \left(\frac{\langle \bar{X}_n - m, \varphi_k \rangle}{\sqrt{\lambda_k / n}} \right) \;=\; n^{-1/2} \sum_{i=1}^{n} \left(\frac{\langle X_i - m, \varphi_k \rangle}{\sqrt{\lambda_k}} \right).
$$

Since $X_1, .., X_n$ are i.i.d., using the multivariate CLT, the j-vector $\boldsymbol{Z_n} := (Z_{1,n}, .., Z_{j,n})'$ converges in distribution to a standard Gaussian vector. Hence, the processes $Z_{1,n}, .., Z_{j,n}$ are asymptotically independent and, for each $k = 1, .., j$, $Z_{k,n} \xrightarrow{\mathcal{D}}_n Z$; as a consequence, $X_{j,n} \xrightarrow{\mathcal{D}}_n X_j$;

(c) the proof of this condition follows by applying Slutsky's Theorem and by noting that

$$
\sqrt{\sum_{k=1}^{j} \left(\langle m - m_0, \varphi_k \rangle \right)^2 \frac{h_k^2(p)}{\lambda_k}} \rightarrow_j \widetilde{d}_p\left(m, m_0\right);
$$

(d) since X is a continuous random variable, the proof of this condition follows by taking M equal to the maximum of the density of X, i.e. $M := (2\pi \widetilde{d}_p^2(m, m_0))^{-1/2}$.

This concludes the proof of (3.4). ∎

Proof. [Proof of Theorem 3.4] In this proof, all the asymptotic results are considered for $N = n_1 + n_2$ going to infinity, where we assume that $N \to \infty$ implies that both n_1 and n_2 increase to infinity. Moreover, we present the proofs by assuming that condition (c2) is satisfied in Assumption 1. In this case, we recall that $\{\lambda_k; k \geq 1\}$ and $\{\varphi_k; k \geq 1\}$ denote the sequences of the eigenvalues and the eigenfunctions, respectively, of the covariance function defined as $v = (1-c)v_1 + cv_2$, where $c \in (0,1)$ is such that $c_N = n_1/N \to_N c$, which is guaranteed by (c2). These eigenvalues and eigenfunctions generate the distance d_p defined in (2.4). If (c2) is not satisfied, then (c1) holds in Assumption 1 and the proofs of the same results are obtained through slight modifications of the proofs presented here.

The convergence result (3.6) is a generalization to two functional populations of the result (3.1) shown in Theorem 3.1. Hence, the proof of (3.6) follows the same arguments used in the proof of Theorem 3.1, where we replace $\bar{X}_n$ with $(\bar{X}_{1n_1} - \bar{X}_{2n_2})$, m with $(m_1 - m_2)$, m_0 with 0, n with $(1/n_1 + 1/n_2)^{-1}$ and we frequently use the relation $d_p(x, y) \equiv d_p(x - y, 0)$ for any $x, y \in L^2(T)$.

Analogously to the previous case, results (3.7) and (3.8) are generalizations for two functional populations of the results (3.3) and (3.4) shown in Theorem 3.2. Hence, the proofs of (3.7) and (3.8) follow the same arguments used in the proof of Theorem 3.2, where the notation have been changed similarly as we have done above to prove (3.6). Therefore, we only discuss the parts of the proof where the modifications from the proof of Theorem 3.2 are more relevant. Specifically, we present how to verify that the assumptions (a) and (b) of Theorem A.1 hold in the cases $m_1 = m_2$ and $m_1 \neq m_2$.

First, consider the proof of (3.7), i.e. $m_1 = m_2$; we can show the following results:

(a) Consider

$$Y_{j,N} := \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \cdot \sum_{k=j+1}^{\infty} \left(\frac{\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \varphi_k \rangle}{\sqrt{\lambda_k}}\right)^2 h_k(p).$$

We that remind condition (a) in this case is $\sup_{N \geq 1}\{\boldsymbol{E}[|Y_{j,N}|]\} \longrightarrow_j 0$. Let us introduce the following notation: $\nu_{1,k} = \langle v_1 \varphi_k, \varphi_k \rangle$ and $\nu_{2,k} = \langle v_2 \varphi_k, \varphi_k \rangle$. Then, we have that $\boldsymbol{E}[|Y_{j,N}|] = \sum_{k=j+1}^{\infty} \rho_{N,k} h_k(p)$, where

$$\rho_{N,k} := \frac{(1/n_1 + 1/n_2)^{-1} \boldsymbol{E}[(\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \varphi_k \rangle)^2]}{\lambda_k};$$

since the covariance function of $(1/n_1 + 1/n_2)^{-1/2} \cdot (\bar{X}_{1n_1} - \bar{X}_{2n_2})$ is $(1 - c_N)v_1 + c_N v_2$, we have that

$$\rho_{N,k} = \frac{(1-c_N)\nu_{1,k} + c_N \nu_{2,k}}{\lambda_k} = \frac{(1-c_N)\nu_{1,k} + c_N \nu_{2,k}}{(1-c)\nu_{1,k} + c\nu_{2,k}},$$

because $\lambda_k = \langle v\varphi_k, \varphi_k \rangle$ and c is defined in Assumption 1. Since for any $k \geq 1$ we have that

$$\rho_{N,k} \leq \frac{1-c_N}{1-c} + \frac{c_N}{c} \to_N 2,$$

we obtain that $\sup_{N\geq 1}\{\boldsymbol{E}[|Y_{j,N}|]\} \longrightarrow_j 0$ and condition (a) is verified;

(b) Consider

$$X_{j,N} := \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1} \cdot \sum_{k=1}^{j} \left(\frac{\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \varphi_k\rangle}{\sqrt{\lambda_k}}\right)^2 h_k(p).$$

We remind condition (b) is $X_{j,N} \xrightarrow{\mathcal{D}}_N X_j$, where $X_j = \sum_{k=1}^{j} \chi^2_{1,k} h_k(p)$. We note that $X_{j,N}$ can be written as $X_{j,N} = \sum_{k=1}^{j} Z^2_{k,N} h_k(p)$, where

$$Z_{k,N} := \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1/2} \left(\frac{\langle \bar{X}_{1n_1} - \bar{X}_{2n_2}, \varphi_k\rangle}{\sqrt{\lambda_k}}\right).$$

Note that, since $m_1 = m_2$, for any $k \in \{1,..,j\}$, $Z_{k,N}$ can also be written as follows

$$(1-c_N)^{-1/2} \cdot n_1^{-1/2} \cdot \sum_{i=1}^{n_1} \left(\frac{\langle X_{1i} - m_1, \varphi_k\rangle}{\sqrt{\lambda_k}}\right) + c_N^{-1/2} \cdot n_2^{-1/2} \cdot \sum_{i=1}^{n_2} \left(\frac{\langle X_{2i} - m_2, \varphi_k\rangle}{\sqrt{\lambda_k}}\right),$$

which is a linear combination of two independent sequence of random variables. Thus, by using the multivariate CLT, the j-vector $\boldsymbol{Z_N} := (Z_{1,N},..,Z_{j,N})'$ converges in distribution to a Gaussian vector with zero mean. Now, we need to determine the asymptotic covariance matrix of $\boldsymbol{Z_N}$. From case (a), we have that $\boldsymbol{Var}[Z_{k,N}] = \rho_{N,k} \longrightarrow_N 1$ for any $k = 1,..,j$. Moreover, calling for $i \neq k$ $\nu_{1,ik} := \langle v_1\varphi_i, \varphi_k\rangle$ and $\nu_{2,ik} := \langle v_2\varphi_i, \varphi_k\rangle$, we obtain

$$\boldsymbol{Cov}[Z_{i,N}, Z_{k,N}] = (\lambda_i\lambda_k)^{-1/2}[(1-c_N)\nu_{1,ik} + c_N\nu_{2,ik}].$$

Since $\langle v\varphi_i, \varphi_k\rangle = (1-c)\nu_{1,ik} + c\nu_{2,ik} = 0$ by definition of eigenfunctions, we can write

$$\boldsymbol{Cov}[Z_{i,N}, Z_{k,N}] = (\lambda_i\lambda_k)^{-1/2}\nu_{1,ik}\left[-c\left(\frac{1-c_N}{1-c}\right) + c_N\right] \longrightarrow_N 0.$$

Hence, $\boldsymbol{Z_N}$ converges in distribution to a standard Gaussian vector, so that $Z^2_{k,N} \xrightarrow{\mathcal{D}}_N \chi^2_1$; as a consequence, $X_{j,N} \xrightarrow{\mathcal{D}}_N X_j$.

Then, we can apply Theorem A.1 and the proof of (3.7) is concluded.

The proof of (3.8), i.e. when $m_1 \neq m_2$, follows analogous arguments of the proof of (3.4) shown in Theorem 3.2. In the case $m_1 \neq m_2$, the validity of conditions (a) and (b) are obtained by following the proof of (3.4) combined with similar modifications to those reported above in the proof of (3.7). ■

Proof. [Proof of Theorem 3.5] All the results proposed in Theorem 3.5 can be proved by applying Lemma A.2 to the results presented in Theorem 3.1 and 3.2. Here, we show how to use Lemma A.2 to obtain such results.

First, consider (3.10) which follows by combining (3.1) and (A.8), with $Y_n = \bar{X}_n$ and $W_n = m_0$; condition (A.7) of Lemma A.2 needed to use (A.8) is verified since in this case

$$S_k = (\langle m - m_0, \varphi_k\rangle)^2 + \sup_{n\geq 1}\left\{\frac{\lambda_k}{n}\right\} = (\langle m - m_0, \varphi_k\rangle)^2 + \lambda_k,$$

and so $\sum_k S_k < \infty$.

Same arguments can be applied for (3.11), which follows from (3.3) and (A.8), with $Y_n = \sqrt{n}\bar{X}_n$ and $W_n = \sqrt{n}m$, since

$$S_k \;=\; \sup_{n\geq 1}\left\{n\cdot\mathbb{E}\left[\left(\langle \bar{X}_n - m, \varphi_k\rangle\right)^2\right]\right\} \;=\; \lambda_k,$$

and so $\sum_k S_k < \infty$.

Then, consider (3.12) and note that

$$\sqrt{n}\cdot\left(\widehat{d}_p^2\left(\bar{X}_n, m_0\right) - \widehat{d}_p^2\left(m, m_0\right)\right) = \sqrt{n}\sum_{k=1}^{\infty}\left[\left(\langle \bar{X}_n - m_0, \widehat{\varphi}_k\rangle\right)^2 - \left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)^2\right]\frac{\widehat{h}_k(p)}{\widehat{\lambda}_k},$$

with the convention $\frac{\widehat{h}_k(p)}{\widehat{\lambda}_k} = p$ for $k \geq n$. Now, for each $k \geq 1$ we can write

$$\begin{aligned}&\sqrt{n}\left[\left(\langle \bar{X}_n - m_0, \widehat{\varphi}_k\rangle\right)^2 - \left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)^2\right] \;=\\ &\sqrt{n}\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)^2 + 2\sqrt{n}\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)\left(\langle m - m_0, \widehat{\varphi}_k\rangle\right),\end{aligned}$$

and, by adding and subtracting $n\cdot\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)^2$ and $\left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)^2$, we get

$$\begin{aligned}=&(\sqrt{n} - n)\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)^2 + n\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)^2 + 2\sqrt{n}\left(\langle \bar{X}_n - m, \widehat{\varphi}_k\rangle\right)\left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)\\ &+ \left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)^2 - \left(\langle m - m_0, \widehat{\varphi}_k\rangle\right)^2.\end{aligned}$$

From this, we can obtain the following decomposition

$$\sqrt{n}\cdot\left(\widehat{d}_p^2\left(\bar{X}_n, m_0\right) - \widehat{d}_p^2\left(m, m_0\right)\right) \;=\; A_{1,n} + A_{2,n} + A_{3,n}$$

where

$$\begin{aligned}A_{1,n} &:= (\sqrt{n} - n)\widehat{d}_p^2\left(\bar{X}_n, m\right),\\ A_{2,n} &:= \widehat{d}_p^2\left(\sqrt{n}(\bar{X}_n - m) + m, m_0\right),\\ A_{3,n} &:= -\,\widehat{d}_p^2\left(m, m_0\right).\end{aligned}$$

By using previous calculations, it can be shown that condition (A.7) of Theorem A.2 is verified by all the sequences $A_{1,n}$, $A_{2,n}$ and $A_{3,n}$. Thus, by applying Lemma A.2 to the sequences $A_{1,n}$, $A_{2,n}$ and $A_{3,n}$, we have that $\sqrt{n}\cdot\left(\widehat{d}_p^2\left(\bar{X}_n, m_0\right) - \widehat{d}_p^2\left(m, m_0\right)\right)$ and $\sqrt{n}\cdot\left(d_p^2\left(\bar{X}_n, m_0\right) - d_p^2\left(m, m_0\right)\right)$ have the same asymptotic distribution, i.e. from (B.6)

$$\sqrt{n}\cdot\left(\widehat{d}_p^2\left(\bar{X}_n, m_0\right) - \widehat{d}_p^2\left(m, m_0\right)\right) \;\xrightarrow{\mathcal{D}}\; \widetilde{d}_p\left(m, m_0\right)\cdot Z.$$

Finally, from the relation

$$\left(\widehat{d}_p\left(\bar{X}_n, m_0\right) - \widehat{d}_p\left(m, m_0\right)\right) = \frac{\left(\widehat{d}_p^2\left(\bar{X}_n, m_0\right) - \widehat{d}_p^2\left(m, m_0\right)\right)}{\left(\widehat{d}_p\left(\bar{X}_n, m_0\right) + \widehat{d}_p\left(m, m_0\right)\right)},$$

we can use Slutsky's Theorem and (3.10) to get (3.12). ■

References

[1] 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*, 73, Part. 4, 531–558.

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

[3] Degras, D.A. (2011). Simultaneous Confidence Bands for Nonparametric Regression with Functional Data. *Journal of Nonparametric Statistics*, 24, No. 2, 359–377.

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

[5] Ghiglietti, A. and Paganoni, A.M. (2014). Statistical inference for functional data based on a generalization of Mahalanobis distance. Mox Report 39/2014, Department of Mathematics, Politecnico di Milano. *Submitted*.

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

[7] Ieva, F. and Paganoni, A.M. (2013). Risk Prediction for Myocardial Infarction via Generalized Functional Regression Models. *Statistical Methods in Medical Research*, doi: 10.1177/0962280213495988

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

[9] Pini, A. and Vantini, S. (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. Mox Report 13/2013, Department of Mathematics, Politecnico di Milano. *Submitted*.

[10] Pomann, G.M., Staicu, A.M. and Ghosh, S. (2014). Two Sample Hypothesis Testing for Functional Data. *Submitted*.

[11] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [online] http:www.R-project.org, 2009.

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

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

[14] Staicu, A.M., Li, Y., Crainiceanu, C. M. and Ruppert, D. (2014). Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, **41** (4), 932–949.

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

[16] Zhang, J.-T., Liang, X. and Xiao, S. (2010). On the two-sample Behrens-Fisher problem for functional data. *Journal of Statistical Theory and Practice*, **4**, 571–587.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

48/2015 Ambrosi, D.; Pettinati, V.; Ciarletta, P.
Active stress as a local regulator of global size in morphogenesis

47/2015 Colombo, M. C.; Giverso, C.; Faggiano, E.; Boffano,C.; Acerbi, F.; Ciarletta, P.
Towards the personalized treatment of glioblastoma: integrating patient-specific clinical data in a continuous mechanical model

45/2015 Lange, M.; Palamara, S.; Lassila, T.; Vergara, C.; Quarteroni, A.; Frangi, A.F.
Improved hybrid/GPU algorithm for solving cardiac electrophysiology problems on Purkinje networks

46/2015 Giverso, C.; Verani, M.; Ciarletta P.
Emerging morphologies in round bacterial colonies: comparing volumetric versus chemotactic expansion

44/2015 Antonietti, P.F.; Houston, P.; Smears, I.
A note on optimal spectral bounds for nonoverlapping domain decomposition preconditioners for hp-Version Discontinuous Galerkin methods

43/2015 Deparis, S.; Forti, D.; Gervasio, P.; Quarteroni, A.
INTERNODES: an accurate interpolation-based method for coupling the Galerkin solutions of PDEs on subdomains featuring non-conforming interfaces

42/2015 Brugiapaglia, S.; Nobile, F.; Micheletti, S.; Perotto, S.
A theoretical study of COmpRessed SolvING for advection-diffusion-reaction problems

41/2015 Quarteroni, A.; Veneziani, A.; Vergara, C.
Geometric multiscale modeling of the cardiovascular system, between theory and practice

40/2015 Patelli, A.S.; Dedè, L.; Lassila, T.; Bartezzaghi, A.; Quarteroni, A.
Isogeometric approximation of cardiac electrophysiology models on surfaces: an accuracy study with application to the human left atrium

39/2015 Guglielmi, A.; Ieva, F.; Paganoni, A.M.; Quintana, F.A.
A semiparametric Bayesian joint model for multiple mixed-type outcomes: an Application to Acute Myocardial Infarction