MOX–Report No. 1/2015

Hotelling s T^2 in functional Hilbert spaces

Pini, A.; Stamm, A.; Vantini, S.

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

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

Hotelling's T^2 in Functional Hilbert Spaces

A.Pini[a], A.Stamm[b], S. Vantini[a]

[a] MOX– Department of Mathematics, Politecnico di Milano, Milan, Italy
[b] CRL, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
alessia.pini@polimi.it
aymeric.stamm@childrens.harvard.edu
simone.vantini@polimi.it

Abstract

The field of statistics is at the cusp of a revolution in the way data is collected by measuring instruments. Massive information is retrieved in real-time and/or spatially-referenced, hence producing new kind of data: functional data. Statistical inference for functional data is particularly challenging as it is an extreme case of high-dimensional data for which, no matter how large the sample is, information will always be insufficient to fully characterize the underlying model. In detail, after a historical excursus over the test statistics introduced for approaching the problem of testing the mean, we provide a generalization of Hotelling's T^2 on any functional Hilbert space, naturally dubbed functional Hotelling's T^2. We discuss a nonparametric permutational framework that enables statistical testing for the mean function of a population as well as for the difference between the mean functions of two populations. Within this framework, we show how a number of state-of-the-art test statistics can be seen as approximations of functional T^2 statistic hereby proposed.
Keywords: Hotelling's T square, Functional Data, Inference, Permutation Test.

1 State of the art

The tremendously fast technological developments pertaining to measuring instruments have brought the field of Statistics at the cusp of a revolution, with real-time and/or spatially-referenced continuous information as the elementary datum to be analyzed. Various constraints (time, economical or ethical issues) on the other hand often prevent data analysts from collecting large samples. This brings the statistician out of his comforting zone where enough information is available to fully characterize all the variables under study and urges the demand for new inferential procedures that make the most out of the available information to provide the best possible inference. Traditionally, the number of variables under study is referred to as the *dimension* of the problem and often denoted p, while the number of observations of these variables is referred to as the *sample size* and often denoted n. Hence, traditional samples with more observations than variables are termed small p large n data while modern samples with more variables than observations are termed large p small n data. Functional data is an extreme case of large p small n data with $p \to \infty$. In this paper, we propose a chronological overview and evolution of the statistical approach to the inference for the mean from the early works of De Moivre and Gauss back at the beginning of the XX century to the most recent advances. We will show how this evolution is tightly related to the sample characteristics and we will address this specific problem for functional data (extreme case of high-dimensional setting) by introducing a new test statistic.

z-test. In the XIX century, the German mathematician and astronomer Carl Friedrich Gauss, while trying to measure distances between stars, realized that he could not obtain perfectly reproducible measurements (Gauss 1809). Rather, his measurements were clustered around a central value, with more frequently close to this value and less frequently further away. He named this distribution of measurements the Normal distribution, also named after his name nowadays. As a matter of fact, this distribution was introduced 60 years before by the French mathematician Abraham de Moivre in the privately circulated pamphlet "Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi" (De Moivre 1733) in response to the Bernoulli brothers' paper 23 years earlier where he derived a simple approximation to the Bernoulli distribution. In this work, de Moivre unveils the mathematical expression of the Normal distribution curve, well known as the "Bell curve". French mathematician and astronomer Pierre-Simon Laplace further formalized the introduction of the Normal distribution in the "Théorie analytique des probabilités" (Laplace 1820).

Almost a century later, the English statistician and geneticist Sir Ronald Aylmer Fisher publishes "Statistical methods for research workers" (Fisher 1925*b*), in which he formalizes the use of the Normal distribution for statistical inference using elementary one-dimensional data. Let $(x_1, \ldots, x_n)$ be a sample of n independent measurements following the Normal distribution with mean μ and standard deviation (SD) σ. Fisher interprets the area under de Moivre's curve as a measure of probability. Hence, if σ is known and the hypothesis $\mu = \mu_0$ is formulated, he defines the so-called z-score $z_0 = \sigma^{-1}(\overline{x} - \mu_0)/\sqrt{n}$, where $\overline{x}$ is the sample mean and shows that z_0 follows a centered Normal distribution with unit standard deviation under the null hypothesis. Subsequently, he argues that the farther away from 0 the z-score z_0, the more evidence there is against the hypothesis $\mu = \mu_0$ since it implies that the occurrence of such a z-score was very unlikely under this assumption. This is the basis of the z-test, which enables for the first time to make inference for the mean of one-dimensional data.

However, most of one-dimensional data are not normally distributed and the above theory relies on the cornerstone that z_0 follows a centered Normal distribution with unit SD. The validity of this assumption is somehow guaranteed by the Central Limit Theorem (CLT). Hence, most of the inferential procedures proposed in the early 1900s pertain to large samples. We refer to this period as the $1 = p < n = +\infty$ age of Statistics (see Figure 1).

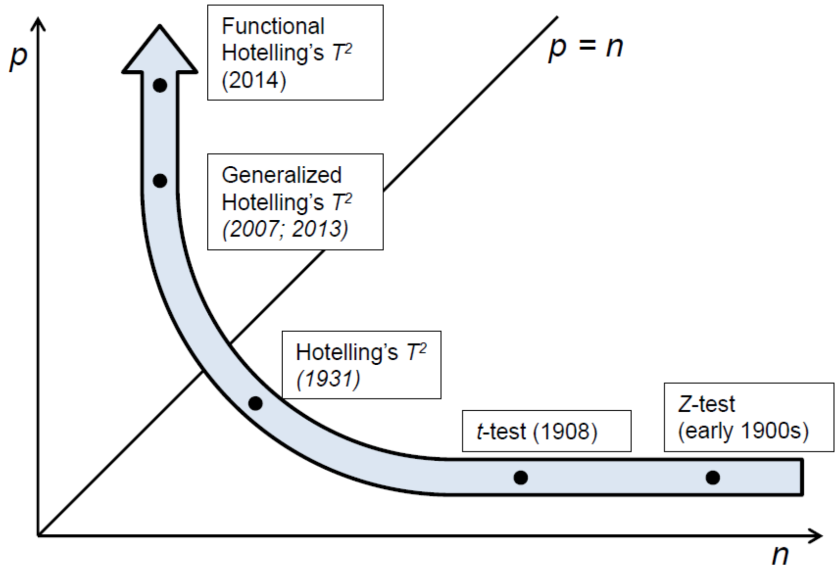

Figure 1: Timeline of the principal results on statistical inference for the mean.

***t*-test.** Eventually, in many fields of applied statistics, it turned out that large samples were not feasible, mainly for time, economical or ethical reasons. This became problematic for applying the z-test for two reasons: (i) the SD of the measurement distribution is never known in practical situations but large sample theory provides an unbiased estimator of it, which is not accurate for small finite sample sizes and (ii) the measurement distribution is rarely Normal but large sample theory ensures that the sample mean is Normal (CLT), which is not guaranteed anymore in the small finite sample size setting.

The English statistician William Sealy Gosset was the first scientist to acknowledge this fact. At that time, as reported in Fisher Box (1981, 1987), he was working on a study about breweries and had only a few observations for making inference. Hence, he circumvented this issue by introducing the t-distribution under the pen-name *Student* in his work "The Probable Error of a Mean" (Gosset 1908). He accounted for the variability of the sample standard deviation in the z-statistic, which becomes non-negligible at low sample sizes. To avoid confusion, he labels it as the t-statistic and characterizes its distribution under the assumption of normality of the data. Fisher further studies the t-distribution in "Applications of 'Student's' distribution" (Fisher 1925a).

Jointly working together, Gosset and Fisher thus introduced the Student's t-distribution and formulated the corresponding t-test, which enables inference for the mean of one-dimensional Normal data using small samples. We refer to this period as the $1 = p < n < +\infty$ age of Statistics (see Figure 1).

Hotelling's T^2 test. A few years later, a growing interest arose in studying multiple features (variables) associated to the same underlying statistical unit (observation). A simple example of this can be formulated as the following question: what are the averaged height and weight of the US population? One can obviously treat the two questions separately but would not account for the obvious correlation between the two variables by doing so. In other words, the scientific community was in need of inferential procedures for jointly distributed multi-dimensional data. Building on Indian statistician Prasanta Chandra Mahalanobis's work "Analysis of Race Mixture in Bengal" (Mahalanobis 1927) where the distance named after him is introduced, American statistician Harold Hotelling introduces the T^2-statistic as a multivariate generalization of the t-statistic in "The generalization of Student's ratio" (Hotelling 1931). In essence, the T^2-statistic is the Mahalanobis distance between the multivariate sample mean and a multivariate hypothesized mean. Hotelling derives the statistical distribution of the T^2-statistic under the assumption of multivariate normality with dimension p smaller than the sample size n, which provided the scientific community with adequate inferential procedures for simultaneously testing for the mean of multiple features.

Hotelling thus introduced the T^2-statistic, which follows a Fisher distribution under the assumption of multivariate normality with $p < n$. We refer to this period as the $1 < p < n < +\infty$ age of Statistics (see Figure 1).

High-dimensional tests. At the end of the XX Century, probably one of the most dramatic changes of paradigm in the history of modern statistics occurred. So far, due to technological limitations, it was a luxury to be able to measure multiple features at the same time (and so p easily remained smaller than n). The major breakthroughs that measuring instruments underwent during the second half of the XX century yielded data with more features than observations (and thus it became usual that p exceeds n at least by an order of magnitude). In other words, statistical research translated from a world with enough information to fully characterize the features of interest ($p < n$) to a world with insufficient information to do so ($p \geq n$). DNA micro-arrays for gene expression are one of the most famous examples of such data. They are characterized by thousands of variables being evaluated on only a few replicates.

Due to the increasing number of such large p small n data, many efforts have been made to extend Hotelling's result to the $p > n$ case for enabling inference for the mean of multi-dimensional data which dimension exceeds the sample size. The work of Srivastava (2007) is pioneering in this direction. He proposes a generalized T^2-statistic and shows that it follows a Fisher distribution for each n

and p, with $n < p < +\infty$, under the assumption of multivariate normality and of proportionality of the variance-covariance matrix to the identity (which implies the independence among components). In Secchi et al. (2013), a generalized T^2-statistic is presented in a less stringent framework, i.e., without relying on the assumption of independence among components (even though still requiring multivariate normality). Under some conditions on the trace of the variance-covariance matrix, they show that it follows a χ^2 distribution with $n-1$ degrees of freedom in the $p \to \infty$ regime. We refer to this period as the $1 < n < p < +\infty$ age of Statistics (see Figure 1).

Functional tests. Some areas of applied statistics are interested in a particular kind of data: they aim at making inference for a single variable acquired in a continuous fashion by cutting-edge measuring instruments. The occurrence of such functional data is growing rapidly in these areas and raises the demand for appropriate inferential tools. Functional data analysis (FDA) has been one of the focuses of statisticians in the XXI century (Ramsay and Silverman 2002, 2005; Ferraty and Vieu 2006). The curve describing the continuous variable can be viewed as an infinity of points, or variables, and is thus the obvious extreme case of large p small n data. In addition, each "variable" describing a given point on the observed curves cannot be assumed independent from the other points on the same curves. We are thus entering in a new age of Statistics at the antipodes with respect to the beginning of the XX century that we shall refer to as the $1 < n < p = +\infty$ age of Statistics.

A commonality between the different inferential procedures provided during the last two centuries is the normality assumption of the data. This yielded parametric tests that are particularly appealing because (i) they generally achieve great statistical power and (ii) they only require the computation of a single test statistic, which is computationally easy and the comparison with tabulated critical values. In contrast, nonparametric approaches to the problem of inference, such as permutation tests, also introduced during the XX century (Fisher 1936), were not widely used because available technologies back in these days could not cope with the high computational burden that these procedures generated.

This was not really a concern during the $1 = p < n = +\infty$ age. Indeed, after Russian mathematician Aleksandr Mikhailovich Lyapunov proved the CLT under very wide assumptions in the "Nouvelle forme du théoreme sur la limite de probabilité" (Lyapunov 1901), the z-test could be easily applied to non-normal data. During the $1 = p < n < +\infty$ age, even though asymptotic normality of the sample mean was not sufficient anymore to ensure that the t-statistic follows the t-distribution, inferential procedure for testing the assumption of normality of one-dimensional data already existed and was thus not a debated point.

Debates really started with the $1 < p < n < +\infty$ age. Indeed, Hotelling's T^2 test strongly relies on the assumption of multivariate normality, which can be assessed in the bivariate case but becomes more and more challenging to assess as the dimension p increases. This is known as the "curse of dimensionality" (Hastie et al. 2009). These concerns grew even more during the $1 < n < p < +\infty$ age as most statistical procedures proposed for their analysis, such as the tests proposed in Srivastava (2007); Secchi et al. (2013), also strongly rely on the assumption of multivariate normality and, in addition, have been shown not to be robust with respect to violation of this assumption (Secchi et al. 2013). Similar concerns remain now that we enter the $1 < n < p = +\infty$ age with FDA.

Consequently, in this work, following the approach pioneered by Fisher (Fisher 1936), we propose a nonparametric permutational framework for the inference on the mean of functional data. This framework does not rely on either multivariate normality or pre-specified variance-covariance structures. As such, it offers an appealing alternative to parametric procedures, the validity of which remains unclear in the new settings we find ourselves into.

In detail, in the present work, we propose a L^2 generalization of the Hotelling's T^2 statistic. We refer to it as functional Hotelling's T^2. We define the statistic and discuss its properties in Section 2. In Subsection 2.3 we discuss how to compute the functional Hotelling's T^2, and show how its finite-dimensional approximation is related with the multivariate large p small n generalization of T^2 provided in Secchi et al. (2013). In Section 3 we discuss a possible application of the functional Hotelling's T^2 statistic to the problem of inference for the mean in FDA, by means of nonparametric permutation tests. In Section 4 we compare it with other L^2-based test statistics presented in literature to test for functional data. Finally, in Section 5 we extend functional Hotelling's T^2 to any functional Hilbert space. All proofs are reported in the Appendix.

2 Hotelling's T^2 in L^2

2.1 Theoretical Framework

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on the space $L^2(T)$ of all real-valued squared-integrable functions on the domain T (where T is an interval of $\mathbb{R}$ of the form (a, b)). The space $L^2(T)$, endowed with its natural inner product $(\xi_1, \xi_2) = \int_T \xi_1(t)\xi_2(t)dt$ for any $\xi_1, \xi_2 \in L^2(T)$, and associated norm $\|\xi\|_{L^2} = \sqrt{\int_T \xi^2(t)dt}$ (for any $\xi \in L^2(T)$), is a Hilbert space. Let $\mathbb{E}$ denote the integration with respect to the probability measure $\mathbb{P}$. The elementary datum in functional data analysis (FDA) is a random function of which we shall give a proper mathematical definition. Following (Tarpey 2003), we state:

Definition 2.1. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random L^2-function, or L^2-valued random variable, is a measurable surjective map from the sample space Ω to $L^2(T)$.*

Note that, unlike Tarpey (2003), we here require the random function to be a surjective map from Ω to $L^2(T)$. This assumption means that the random function is nondegenerate, in the sense that it covers the whole $L^2(T)$ space.

Similarly to real-valued one-dimensional (resp., multi-dimensional) variables, for a given random L^2-function ξ, we can define the concepts of mean and dispersion around the mean. In traditional discrete cases, the former is a real one-dimensional (resp., multi-dimensional) vector and the latter is summarized by the variance (resp., variance-covariance matrix). The mean of a random L^2 function on the other hand is a function and its dispersion is characterized by a covariance operator. They are given by the following.

Definition 2.2. *Let ξ be a random L^2-function. The mean function $\mu : Tto\mathbb{R}$ and covariance operator $V : L^2(T) \to L^2(T)$ of the random L^2-function ξ are: respectively given by:*

$$\mu(t) = \mathbb{E}[\xi(t)], \text{ and } (Vf)(t) = \int_T \sigma(t,s) f(s) ds,$$

where $\sigma : T \times T \to \mathbb{R}$ is the covariance function of ξ:

$$\sigma(t,s) = \mathbb{E}\left[(\xi(t) - \mu(t))(\xi(s) - \mu(s))\right], \text{ for any } (t,s) \in T \times T\,.$$

In the current work, similarly to (Horváth and Kokoszka 2012; Bosq 2000), we restrict ourselves to random L^2-functions with finite total variance, i.e., such that:

$$\mathbb{E}\left[\|\xi\|_{L^2}^2\right] < +\infty \text{ (finite total variance).}$$

This covers a huge number of practical situations and confers convenient properties to the covariance operator such as the *spectral theorem decomposition*. Indeed, the covariance function σ can then be shown to belong to $L^2(T \times T)$. As a result, the covariance operator V is an Hilbert-Schmidt operator, i.e., it belongs to the subspace $\mathrm{HS}(L^2(T))$ of the space of limited linear operators $\mathcal{L}(L^2(T))$ on $L^2(T)$ (Arveson 2002).

At this point and for the rest of the paper, we will assume that we collected a random sample of n independent and identically distributed (iid) random L^2-functions $\xi_1, \ldots, \xi_n$, with common mean function μ and covariance operator V, satisfying the finite total variance assumption.

Unbiased estimators for μ and V are given by the following

Definition 2.3. *The sample mean function $\overline{\xi} : T \to \mathbb{R}$ is an unbiased estimator of the mean function μ and is given by:*

$$\overline{\xi}(t) = \frac{1}{n} \sum_{i=1}^{n} \xi_i(t)\,.$$

The sample covariance operator $\widehat{V} : L^2(T) \to L^2(T)$ is an unbiased estimator of the covariance operator V and is given by:

$$\left(\widehat{V} f\right)(t) = \int_T \mathcal{S}(t,s) f(s) \mathrm{d}s,$$

where $\mathcal{S}$ is the sample covariance function defined as:

$$\mathcal{S}(t,s) = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i(t) - \overline{\xi}(t))(\xi_i(s) - \overline{\xi}(s)) \text{ for any } (t,s) \in T \times T\,.$$

The proof of unbiasness of these random variables as estimators of the mean function and covariance operator respectively is straightforwardly obtained by replicating the proof of unbiasness of their multivariate counterparts. Note that $\overline{\xi}$ is a random L^2-function and $\widehat{V}$ is a random $\mathrm{HS}(L^2)$-operator.

2.2 Definition of Hotelling's T^2 in L^2

Similarly to the multivariate case, it is possible to break down the total variance in the original functional dataset into two components, one of which only depends on the data. The following theorem states such a decomposition of variance and introduces some useful operators.

Theorem 2.1. *Consider a sample of n iid random functions with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$. Then, the following variance decomposition holds:*

$$(n-1)\hat{V} + n\bar{V} = \tilde{V},$$

or, equivalently, $\forall g \in L^2(T)$:

$$(n-1)(g,\hat{V}g) + n(g,\bar{V}g) = (g,\tilde{V}g),$$

where:

- *$\hat{V}$ is the sample covariance operator, with kernel $\mathcal{S}$, that describes the dispersion of data ξ_i around the sample mean $\bar{\xi}$;*
- *$\bar{V}$ is the random operator with kernel $\big(\bar{\xi}(t) - \mu(t)\big)\big(\bar{\xi}(s) - \mu(s)\big)$, that describes the distance between the sample mean $\bar{\xi}$ and the mean μ;*
- *$\tilde{V}$ is the random operator with kernel $\sum_{i=1}^{n}\big(\xi_i(t) - \mu(t)\big)\big(\xi_i(s) - \mu(s)\big)$, that describes the dispersion of data ξ_i around the mean μ.*

The random operators $\overline{V}$ and $\widehat{V}$ introduced in Theorem 2.1 are the key concepts for generalizing Hotelling's T^2 statistic to the functional case. The following definition formally introduces this statistic:

Definition 2.4. *The functional Hotelling's T^2-statistic is defined as the L^2 distance between the sample mean function and the true mean function "standardized" to the sample covariance operator. Similarly to the multivariate case, it reads:*

$$T^2 = n \max_{g \in \mathrm{Im}(\widehat{V})} \frac{(g,\overline{V}g)}{(g,\widehat{V}g)}. \tag{1}$$

The functional T^2-statistic has a number of desirable properties that makes it particularly appealing for inferential purposes:

T^2 is a semi-distance between μ and $\overline{\xi}$. It is important to keep in mind that, although the formulation of the functional T^2-statistic proposed in Definition 2.4 is closely related to the multivariate T^2-statistic that one can find in many textbooks on introduction to multivariate analysis, the two statistics fundamentally differs in their mathematical implications. The multivariate T^2-statistic is defined as the maximum of the squared t-statistics associated to all possible one-dimensional projections of the multi-dimensional data. Differently, the functional T^2-statistic is defined as the maximum over the space $\mathrm{Im}(\widehat{V})$ spanned by the sample covariance operator $\widehat{V}$, which is an $(n-1)$-dimensional random subspace of $L^2(T)$. As a result, T^2 is a distance between μ and $\overline{\xi}$ in the random space $\mathrm{Im}(\widehat{V})$ but is only a semi-distance in $L^2(T)$, for which the identity of indiscernibles does not hold.

T^2 is invariant under similarity transformations. Functional Hotelling's T^2-statistic is invariant under similarity transformations of the data, i.e., under affine transformations $\xi \mapsto aO\xi + \mathbf{f}$, where $a \in \mathbb{R}^+$, $\mathbf{f} \in L^2(T)$ and O is an orthogonal linear limited operator on $L^2(T)$, i.e., O satisfies $(Og_1, Og_2)_{L^2} = (g_1, g_2)_{L^2}$ for any $g_1, g_2 \in L^2(T)$. Lehmann and Romano (2006) have shown that this type of invariance is the largest family of invariance transformations that one can achieve in the framework $p \geq n$. In this sense, the functional T^2-statistic is invariant-optimal.

T^2 "marginal" distributions under functional normality are known. The notion of *functional normality* has been introduced in Tarpey (2003) and stipulates that a random L^2-function is normally distributed if and only if, for all $u \in L^2(T)$, the real-valued one-dimensional random variable (ξ, u) is normally distributed. If we further assume functional normality of our dataset, Theorem 2.1 combined with Cochran's Theorem yields the following (see Proposition .1 in the Appendix):

$$n\frac{(g, \overline{V}g)}{(g, \widehat{V}g)} \sim F(1, n-1)\,. \tag{2}$$

Equation 2 provides the distribution of the ratios involved in the T^2 statistic. However the distribution of its maximum over all functions of $\mathrm{Im}(\widehat{V})$ is not easy to elicit without introducing very strong assumptions on the covariance operator V. In addition, functional normality may be too stringent for many applications and hard to defend and/or prove. For all these reasons, we will tackle the problem of inference for the mean function within a nonparametric permutational framework, based on minimal distributional assumption.

2.3 A Finite-Dimensional Approximation of Hotelling's T^2 in L^2

With Definition 2.4, we gave a formal definition of functional Hotelling's T^2 statistic. However, expressed as a maximization problem, T^2 is of little practical interest here. Indeed, permutation tests rely on the evaluation of a sufficient statistic over an enormous number of permuted datasets, which might become computationally too heavy if, for each evalution, a maximization problem has to be solved. Furthermore, in practical scenarios, analytic expressions of the observed functions ξ_i's are often not provided. Rather, finite high-dimensional approximations are available.

Let us consider a countable set of basis functions $\{\phi_k\}_{k\geq 1}$ of $L^2(T)$. It is possible to project the original n observed functions onto the first p elements of such a basis. Let $\boldsymbol{\xi}_i = ((\phi_1, \xi_i), \ldots, (\phi_p, \xi_i))$ be the vector of the scores of the *i-th* observed function ξ_i projected onto the first p elements of the basis. Then, we can define the p-dimensional random vector $\overline{\boldsymbol{\xi}}$ as the sample mean of the individual scores and the $p \times p$ matrix S as their sample variance-covariance matrix. Similarly, the mean function μ can be projected into a p-dimensional vector $\boldsymbol{\mu}$ of mean scores. At this point, the finite-dimensional approximation of functional Hotelling's T^2 can be computed directly without solving any maximization problem, as shown by the following.

Theorem 2.2. *Consider a sample of n iid random functions with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$. Let $\{\phi_k\}_{k\geq 1}$ be a countable set of basis functions of $L^2(T)$. Then, for any $p \geq 1$, the following identity holds:*

$$T_p^2 = n \max_{g \in \mathrm{Im}(\widehat{V}) \cap \{\phi_1,\ldots,\phi_p\}} \frac{(g, \overline{V}g)}{(g, \widehat{V}g)} = n(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^\top W^{1/2} S^+ W^{1/2} (\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}) \,, \qquad (3)$$

where $W \in \mathbb{R}^{p\times p}$ is the matrix of inner products between the basis functions $[W]_{i,j} = (\phi_i, \phi_j)$ and S^+ is the Moore-Penrose generalized inverse (Rao and Mitra 1971) of the sample variance-covariance matrix S. In addition:

$$T_p^2 \xrightarrow[p\to\infty]{a.s.} T^2 \,.$$

Theorem 2.2 states that, if the basis used to project the data is orthonormal (i.e., $W = I$), if we limit the search for the maximum in the functional T^2 definition to those functions in $\mathrm{Im}(\widehat{V})$ that are spanned by the first p elements of any basis of $L^2(T)$, then the resulting maximum can be formulated as a high-dimensional T^2 statistic as introduced in (Secchi et al. 2013). In the case of non-orthonormal basis, this finite-dimensional approximation is still related to the high-dimensional generalization provided in (Secchi et al. 2013), but the generalized inverse of the covariance matrix is rescaled, by considering the inner products between the basis functions. In addition, as $p \to \infty$, the sequence of such statistics converges almost surely to the functional T^2 statistic.

Note that, with the basis of principal components of $\widehat{V}$, we have the equality $T^2 = T^2_{n-1}$, i.e., the functional Hotelling's T^2 can be exactly evaluated by means of the first $n-1$ sample principal components.

3 Permutation test in L^2 based on Hotelling's T^2

The problem of inference for functional data has been addressed in the literature from both a parametric and a nonparametric perspective. The former approach commonly relies on distributional assumptions on functional data and on asymptotic results (Horváth and Kokoszka 2012; Spitzner et al. 2003; Cuevas et al. 2004; Fan and Lin 1998; Schott 2007). The latter approach relies instead on permutation or bootstrap techniques, which are computationally intensive (Hall and Tajvidi 2002; Cardot et al. 2007; Cuesta-Albertos and Febrero-Bande 2010; Pini and Vantini 2013; Hall and Van Keilegom 2007). The method that we propose for testing functional data relies on this latter approach.

in detail, we now show how functional Hotelling's T^2 can be used in nonparametric permutation procedures for making inference on the mean of a random L^2 function (Section 3.1) and on the difference between the means of two random L^2 functions (Section 3.2).

3.1 One-Population Test

Let $(\xi_1, \dots, \xi_n)$ be n i.i.d. random L^2-functions with mean function μ and covariance operator V that satisfy the finite total variance assumption $(\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$, for all $i \in \{1, n\})$.

Assuming that we want to test the following null hypothesis on the mean function:

$$H_0 : \mu = \mu_0, \text{ vs. } H_1 : \mu \neq \mu_0\,, \text{ with } \mu_0 \in L^2(T)\,, \tag{4}$$

one can compute, under the null hypothesis H_0, the functional T^2 statistic (Definition 2.4):

$$T_0^2 = n \max_{g \in \mathrm{Im}(\widehat{\mathrm{V}})} \frac{(g, \overline{V_0} g)}{(g, \widehat{V} g)}\,, \tag{5}$$

where $\overline{V_0}$ is the random operator with kernel $\overline{\sigma_0}(t,s) = (\bar{\xi}(t) - \mu_0(t))(\bar{\xi}(s) - \mu_0(s))$ for any $t, s \in T$ and $\widehat{V}$ is the sample covariance operator with kernel $\mathcal{S}$.

One can use the T_0^2 statistic in a permutational framework for testing the null hypothesis H_0. Instead of the normality assumption often required in this framework (see for instance Horváth and Kokoszka 2012), we make in a permutation framework the much weaker assumption of symmetry of the distribution of the data around the mean. Then, a permutation test can be constructed by evaluating the test statistic (5) over all possible reflections of data with respect to the center of symmetry under H_0, i.e., the transformations $\xi_i(t) \mapsto \xi_i^*(t) = \mu_0(t) + (-1)^{c_i}(\xi_i(t) - \mu_0(t))$, with $i = 1, ..., n$, and $c_i \in \{0, 1\}$. The p-value of test (4) is the proportion of permuted $T_0^2(\xi_1^*, \xi_2^*, ..., \xi_n^*)$ exceeding the value $T_0^2(\xi_1, \xi_2, ..., \xi_n)$ evaluated on the original data set.

3.2 Two-Population Test

Let $(\xi_{11},\dots,\xi_{n_1 1})$ and $(\xi_{12},\dots,\xi_{n_2 2})$ be two independent samples of size n_1 and n_2 respectively. Let $(\xi_{11},\dots,\xi_{n_1 1})$ be i.i.d. random L^2-functions with mean function μ_1 and covariance operator V and let $(\xi_{12},\dots,\xi_{n_2 2})$ be i.i.d. random L^2-functions with mean function μ_2 and covariance operator V. In addition, we assume that the assumption of finite total variance is met in the two samples.

Assuming that we want to test the following null hypothesis:

$$H_0 : \mu_1 = \mu_2, \text{ vs. } H_1 : \mu_1 \neq \mu_2 \,, \tag{6}$$

one can compute, under H_0, the functional T^2 statistic (Definition 2.4):

$$T_0^2 = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1} \max_{g\in \mathrm{Im}(\widehat{\mathrm{V}_{\text{pooled}}})} \frac{(g,\overline{V_0}g)}{(g,\widehat{V_{\text{pooled}}}g)}\,, \tag{7}$$

where $\overline{V_0}$ is the random operator with kernel

$$\overline{\sigma_0}(t,s) = \left[\overline{\xi_1}(t) - \overline{\xi_2}(t)\right[\ \left[\overline{\xi_1}(s) - \overline{\xi_2}(s)\right]\ , \text{ for any } t,s \in T$$

with $\overline{\xi_1}$ and $\overline{\xi_2}$ being the sample mean functions of the first and the second populations respectively and $\widehat{V_{\text{pooled}}}$ is the pooled sample covariance operator with pooled covariance function $\mathcal{S}_{\text{pooled}}$ defined as:

$$\begin{aligned}\mathcal{S}_{\text{pooled}}(t,s) = \frac{1}{n_1+n_2-2}&\left[\sum_{i=1}^{n_1}\left(\xi_{i1}(t)-\overline{\xi_1}(t)\right)\left(\xi_{i1}(s)-\overline{\xi_1}(s)\right)\right.\\ &\left.+\sum_{i=1}^{n_2}\left(\xi_{i2}(t)-\overline{\xi_2}(t)\right)\left(\xi_{i2}(s)-\overline{\xi_2}(s)\right)\right]\ , \text{ for any } t,s\in T\,.\end{aligned}$$

A permutation test can then be constructed by evaluating the test statistic (7) over all permutations of the data over the sample units $(\xi_{11},...,\xi_{n_1 1},\xi_{12},...,\xi_{n_2 2}) \mapsto (\xi^*_{11},...,\xi^*_{n_1 1},\xi^*_{12},...,\xi^*_{n_2 2})$. The p-value of the corresponding test is then the proportion of $T_0^2(\xi^*_{11},...,\xi^*_{n_1 1},\xi^*_{12},...,\xi^*_{n_2 2})$ exceeding $T^2_{f,0}(\xi_{11},...,\xi_{n_1 1},\xi_{12},...,\xi_{n_2 2})$ evaluated on the original data set.

4 Other L^2-based test statistics

To perform a permutation test on the mean of one functional population (or two functional populations), we only need to define a distance or semi-distance between the sample mean function (or difference between the sample mean functions) and the mean function under the null hypothesis H_0 (or difference between the two sample mean functions). In the literature of permutation testing, the following distances have been proposed for random L^2 functions:

The L^2 distance.

$$\Delta^2_{L^2} = \int_T \left(\bar{\xi}(t) - \mu(t)\right)^2 dt\,.$$

This test statistic and associated permutation test have been proposed in Hall and Tajvidi (2002); Hall and Van Keilegom (2007). It is also possible to derive parametric or asymptotic tests based on the same statistic under the assumption of functional normality (see for instance Horváth and Kokoszka 2012). The statistic Δ_{L^2} can be expressed as the norm of an appropriate operator in L^2 as:

$$\Delta^2_{L^2} = n \max_{g \in L^2(T)} \frac{(g, \overline{V} g)}{(g, g)}\,.$$

Hence, $\Delta^2_{L^2}$ can be seen as an approximation of the functional Hotelling's T^2, where the sample covariance operator $\widehat{V}$ is assumed to be the identity operator. Note that this statistic neither accounts for the point-wise variance of the data nor its covariance structure. It instead gives equal weight to equally-long intervals of the domain T.

The standardized L^2 distance. (i.e., the L^2 distance between standardized data)

$$\Delta^2_{L^2_t} = \int_T \frac{\left(\bar{\xi}(t) - \mu(t)\right)^2}{\mathcal{S}(t,t)} dt\,,$$

where $\mathcal{S}(t,t)$ is the point-wise sample variance. This test statistic has been introduced in Hall and Tajvidi (2002) and can be seen as a weighted version of the L^2 statistic. Similarly to the L^2 statistic, the statistic $\Delta^2_{L^2_t}$ can be expressed as the norm of an appropriate operator in L^2 as:

$$\Delta^2_{L^2_t} = n \max_{g \in L^2(T)} \frac{(g, \overline{V} g)}{(g, D_\sigma g)}\,.$$

Hence, $\Delta^2_{L^2_t}$ can be seen as a more sophisticated approximation of the functional Hotelling's T^2 statistic. The sample covariance operator is indeed assumed to be "diagonal" and reads $(D_\sigma g)(t) = \mathcal{S}(t,t) g(t)$. The $\Delta^2_{L^2_t}$ statistic thus makes use of the point-wise estimates $\mathcal{S}(t,t)$ of the variance of the data but does not account for its auto-correlation structure $\mathcal{S}(t,s)$.

Note that, unlike the functional T^2 statistic, the $\Delta^2_{L^2}$ and $\Delta^2_{L^2_t}$ statistics are distances in $L^2(T)$ (and not semi-distances). On the other hand, they share no commonality with traditional test statistics used for null hypothesis statistical testing in multivariate analysis and they are not invariant under similarity transformations.

5 Hotelling's T^2 in functional Hilbert spaces

In the previous sections we presented the functional Hotelling's T^2 in the L^2 geometry as the natural extension of finite-dimensional Euclidean geometry to the space $L^2(T)$. Nevertheless, functional Hotelling's T^2 can be extended to every functional Hilbert space. Indeed, its definition only requires the evaluation of mean function and covariance operator, which directly derive from the notion of inner product.

In particular, let H be a functional Hilbert space, endowed with the inner product $(\cdot,\cdot)_H$ and associated norm $\|\cdot\|_H$. Let $(\xi_1,\ldots,\xi_n)$ be n i.i.d. H-valued random variables with mean $\mu \in H$ and covariance operator $V \in \mathcal{L}(H)$. A sample estimate of the mean in H is the Fréchet mean: $\bar{\xi} = \text{argmin}_{\xi\in H} \sum_{i=1}^{n} \|\xi_i - \xi\|_H^2$. Hence, functional Hotelling's T^2 can be defined in the space H as:

$$T^2 = n \max_{g\in \text{Im}(\hat{V})} \frac{(g,\bar{V}g)_H}{(g,\hat{V}g)_H}. \tag{8}$$

where:

- $\hat{V} \in \mathcal{L}(H)$ is the sample covariance operator in the space H, (defined according to the scalar product in H), describing the dispersion of data ξ_i around the Fréchet mean $\bar{\xi}$. Indeed, $\hat{V}$ is such that $(g,\hat{V}g)_H$ is the sample variance of the scores of the orthogonal projections of ξ_i on g, with respect to the inner product in H.
- $\bar{V} \in \mathcal{L}(H)$ is a random operator associated to the distance between the Fréchet mean $\bar{\xi}$ and the mean μ. Indeed, $\bar{V}$ is such that $(g,\bar{V}g)_H$ is the square distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ over g, with respect to the inner product in H.

In the following, we report two concrete examples of geometry where we explicit the definition of these operators: (i) the Sobolev space $H^k(T)$ of k-differentiable squared-integrable real functions with squared-integrable derivatives (Section 5.1) and (ii) the Bayes linear space $B^2(T)$ of non-negative real functions on T with squared-integrable logarithm (Boogaart et al. 2014) (Section 5.2).

5.1 Example: Hotelling's T^2 in Sobolev Spaces

Consider the Sobolev space $H^k(T)$, that is, the space of k-differentiable functions $g \in L^2(T)$ such that, for $j \le k$, $D^j g \in L^2(T)$ (where $D^j g$ denotes the *j-th* derivative of g). The space $H^k(T)$ is a Hilbert space, endowed with the following inner product:

$$(f,g)_{H^k} = \sum_{j=0}^{k} (D^j f, D^j g)_{L^2} = \sum_{j=0}^{k} \int_T (D^j f)(t) \cdot (D^j g)(t) dt\,. \tag{9}$$

Let $(\xi_1, \dots, \xi_n)$ be n i.i.d. $H^k(T)$-valued random variables with mean $\mu \in H^k(T)$ defined as $\mu = \text{argmin}_{m \in H^k(T)} \mathbb{E}\left[\|\xi_i - m\|^2_{H^k}\right]$. The functional Hotelling's T^2 in $H^k(T)$ then reads:

$$T^2 = n \max_{g \in \text{Im}(\widehat{V})} \frac{(g, \overline{V}g)_{H^k}}{(g, \widehat{V}g)_{H^k}}, \tag{10}$$

where the operators $\widehat{V}$ and $\overline{V}$ can be explicitly defined using the inner product in $H^k(T)$ given by Eq.(9). In details,

- the operator $\hat{V} \in \mathcal{L}(H^k)$ is defined as:

$$\left(\hat{V}f\right)(t) = \int_T \sum_{j=0}^{k} \mathcal{S}_{0j}(t,s) D^j f(s) \mathrm{d}s,$$

  where $\mathcal{S}_{lj}(t,s)$ is the sample covariance between lth and jth derivatives of data ξ_i: $\mathcal{S}_{lj}(t,s) = \frac{1}{n-1}\sum_{i=1}^{n}(D^l\xi_i(t) - D^l\bar{\xi}(t))(D^j\xi_i(s) - D^j\bar{\xi}(s))$, and $\bar{\xi}(t)$ is the Fréchet mean of the ξ_i, defined as $\bar{\xi} = \text{argmin}_{m \in H^k} \sum_{i=1}^{n} \| \xi_i - m \|^2_{H^k}$;

- the operator $\bar{V} \in \mathcal{L}(H^k)$ is defined as:

$$\left(\bar{V}f\right)(t) = \int_T \sum_{j=0}^{k} (\bar{\xi}(t) - \mu(t))(D^j\bar{\xi}(s) - D^j\mu(s)) D^j f(s) \mathrm{d}s.$$

To have a better insight into the interpretation of Hotelling's T^2 in the Sobolev space $H^k(T)$, we can rely on the following identities (Lemma .1 of the Appendix):

$$\begin{aligned}(g, \hat{V}g)_{H^k} &= \widehat{Var}\left[(g, \xi_1)_{H^k}, \dots, (g, \xi_n)_{H^k}\right];\\ (g, \bar{V}g)_{H^k} &= ((g, \bar{\xi} - \mu)_{H^k})^2.\end{aligned}$$

These identities show that Hotelling's T^2 in $H^k(T)$ can be interpreted as the maximum over all elements in the image space of $\widehat{V}$ of the ratio between: (i) the squared distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ on g, with respect to the inner product in $H^k(T)$; and (ii) the sample variance of the scores of the orthogonal projections of the ξ_i's on g, with respect to the inner product in $H^k(T)$.

5.2 Example: Hotelling's T^2 in the Bayes Linear Space

Another example of functional Hilbert space recently introduced in the Functional Data Analysis literature is the Bayes linear space $B^2(T)$, that is, the space of absolutely continuous density functions on the compact set T with squared-integrable logarithm. The interested reader can find detailed descriptions of

Bayes spaces in Egozcue et al. (2006); Egozcue and Pawlowsky-Glahn (2006); Menafoglio et al. (2013); Boogaart et al. (2014); Hron et al. (2014). As shown by Egozcue et al. (2006), $B^2(T)$ is a functional Hilbert space when proper addition $\oplus$, scalar multiplication $\odot$ and inner product $(\cdot,\cdot)_{B^2}$ operations are defined. In details, for any $f, g \in B^2(T)$ and $\alpha \in \mathbb{R}$:

$$\begin{gathered}(f \oplus g)(t) = \frac{f(t)g(t)}{\int_T f(s)g(s)ds}\,, \qquad (\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_T f(s)^\alpha ds}\,, \\ (f,g)_{B^2} = \frac{1}{2|T|}\iint_{T\times T} \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)} dtds\,,\end{gathered} \tag{11}$$

where $|T|$ is the measure of the compact set T.

An isometric isomorphism between $B^2(T)$ and $L^2(T)$ is defined by the centred log-ratio (clr) transformation (Boogaart et al. 2014; Menafoglio et al. 2013):

$$\operatorname{clr}(f)(t) = \ln f(t) - \frac{1}{|T|} \ln f(s) \mathrm{d}s. \tag{12}$$

Using both the Hilbert geometry conferred from the addition, scalar multiplication and inner product proposed by Egozcue et al. (2006) and the isomorphism in Eq.(12), we can provide a functional Hotelling's T^2 statistic in $B^2(T)$ useful for making inference on the mean of populations of density functions on a compact support.

Let $(\xi_1, \ldots, \xi_n)$ be n i.i.d. $B^2(T)$-valued random variables with mean $\mu \in B^2(T)$ defined as $\mu = \operatorname{argmin}_{m\in B^2(T)} \mathbb{E}\left[\|\xi_i - m\|^2_{B^2}\right]$. The functional Hotelling's T^2 in $H^k(T)$ then reads:

$$T^2 = n \max_{g\in \operatorname{Im}(\widehat{V})} \frac{(g, \overline{V}g)_{B^2}}{(g, \widehat{V}g)_{B^2}}\,, \tag{13}$$

where the operators $\widehat{V}$ and $\overline{V}$ can be explicitly defined using the inner product in $B^2(T)$ given by Eq.(11) and the isomorphism given by Eq.(12). In details,

- the sample covariance operator $\hat{V} \in \mathcal{L}(B^2)$ is defined as:

  $$\left(\hat{V} f\right)(t) = \operatorname{clr}^{-1}\left(\int_T \mathcal{S}_c(t,s)\operatorname{clr}(f)(s)\mathrm{d}s\right),$$

  where clr^{-1} is the inverse centered log-ratio transformation, and $\mathcal{S}_c(t,s)$ is the sample covariance between clr-transformed data:
  $\mathcal{S}_c(t,s) = \frac{1}{n-1}\sum_{i=1}^n \left(\operatorname{clr}(\xi_i)(t) - \operatorname{clr}(\bar{\xi})(t)\right)\left(\operatorname{clr}(\xi_i)(s) - \operatorname{clr}(\bar{\xi})(s)\right)$.

- the operator $\bar{V} \in \mathcal{L}(B^2)$ is defined as:

  $$\left(\bar{V} f\right)(t) = \operatorname{clr}^{-1}\left(\int_T \left(\operatorname{clr}(\bar{\xi})(t) - \operatorname{clr}(\mu)(t)\right)\left(\operatorname{clr}(\bar{\xi})(s) - \operatorname{clr}(\mu)(s)\right)\operatorname{clr}(f)(s)\mathrm{d}s\right),$$

  where $\bar{\xi}(t)$ is the Fréchet mean of the ξ_i, defined as $\bar{\xi} = \operatorname{argmin}_{m\in B^2}\sum_{i=1}^n \| \xi_i - m \|^2_{B^2}$.

Similarly to we did in Sobolev spaces, to have a better insight into the interpretation of Hotelling's T^2 in the Bayes space $B^2(T)$, we can rely on the following identities (Lemma .2 of the Appendix):

$$(g, \hat{V}g)_{B^2} = \widehat{Var}[(g, \xi_1)_{B^2}, \ldots, (g, \xi_n)_{B^2}]$$
$$(g, \bar{V}g)_{B^2} = ((g, \bar{\xi} - \mu)_{B^2})^2$$

Hence, Hotelling's T^2 in $B^2(T)$ is the the maximum over all elements in the image space of $\widehat{V}$ of the ratio between: (i) the squared distance between the scores of the orthogonal projections of $\overline{\xi}$ and μ on g, with respect to the inner product in $B^2(T)$; and (ii) the sample variance of the scores of the orthogonal projections of the ξ_i's on g, with respect to the inner product in $B^2(T)$.

6 Conclusions

After a historical excursus on how the problem of inference for the mean evolved in the statistical research, from the early works of De Moivre and Gauss back at the beginning of the XX century to the most recent advances, we presented a generalization of Hotelling's T^2 (functional Hotelling's T^2) in functional Hilbert spaces, and demonstrated how it can be used for hypothesis testing for the mean of functional data within a permutational framework.

The functional Hotelling's T^2 is presented as a natural extension of Euclidean geometry to the functional L^2 space. It is a semi-distance based on a semi-metric in L^2. In essence, the functional T^2 statistic maximizes the ratio of an operator that assesses the distance between the sample mean of an i.i.d. functional dataset and its actual mean to another operator that assesses the variability of such a functional dataset around its sample mean. We presented a practical way of computing this statistic without resorting to optimization algorithms by projecting the dataset onto any basis of the image space of the sample covariance operator.

For inferential purposes, we set up a permutational framework for making inference on the mean (or difference between means) of functional data. We discussed the advantage of our proposed functional T^2 statistic, which, unlike all other statistics proposed in the literature, fully accounts for the covariance structure of the input data. Moreover, we showed that already existing test statistics recently presented in the literature are in fact approximations of functional Hotelling's T^2, where the variance and/or correlation of the data is ignored.

Finally, even though we presented functional Hotelling's T^2 in the L^2 geometry, as the natural functional extension of Euclidean geometry, we also showed how our functional T^2 statistic can be defined and used in virtually any Hilbert space. Examples included in this work are the Sobolev spaces $H^k(T)$ and the Bayes linear space $B^2(T)$.

An interesting and challenging future development of this work would be the extension of T^2 to the larger family of functional metric spaces (e.g., Banach spaces), following the direction of some lively and very recent areas of statistical research, such as object-oriented data analysis and shape analysis (see for instance Marron and Alonso 2014). This extension requires a definition of T^2 exclusively based on a metric that relies neither on the notion of inner product nor on the one of vector space.

Appendix: Proofs

Proof. [Theorem 2.1] Note that, by their definition, these 3 operators have respectively 1, $n-1$ and n degrees of freedom. Moreover, we have that, $\forall \omega \in \Omega$, $\forall t, s \in T$:

$$
\begin{aligned}
(n-1)\mathcal{S}(\omega)(t,s) &= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \bar{\xi}(\omega)(t)\big)\big(\xi_i(\omega)(s) - \bar{\xi}(\omega)(s)\big) \\
&= \sum_{i=1}^{n} \Big[\big(\xi_i(\omega)(t) - \mu(t) + \mu(t) - \bar{\xi}(\omega)(t)\big) \\
&\times \big(\xi_i(\omega)(s) - \mu(s) + \mu(s) - \bar{\xi}(\omega)(s)\big)\Big] \\
&= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&+ \sum_{i=1}^{n} \big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
&- \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
&- \sum_{i=1}^{n} \big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&- n\big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big).
\end{aligned}
$$

Hence, we have:

$$
\begin{aligned}
(n-1)\mathcal{S}(\omega)(t,s) &+ n\big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
&= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big), \qquad (14)
\end{aligned}
$$

and the thesis follows. □

Proposition .1. *Consider a sample of n iid random functions $\xi_1, ..., \xi_n$ with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$, and $\mathrm{Im}(V) = L^2(T)$. Let the random functions be normally distributed, i.e., $\forall u \in L^2(T)$, (ξ, u) is a real univariate gaussian random variable. Then, we have:*

$$
n\frac{(g, \bar{V}g)}{(g, \hat{V}g)} \sim F(1, n-1).
$$

Proof. Let $g \in \mathrm{Im}(\hat{V})$. Under the normality assumption we have:

- $(g, \tilde{V}g) \sim (g, Vg)\chi^2(n)$;
- $n(g, \bar{V}g) \sim (g, Vg)\chi^2(1)$.

Indeed, for the first one, we have:

$$(g, \tilde{V}g) = \sum_{i=1}^{n} \Big(\int_T \big(\xi_i(t) - \mu(t)\big) g(t) dt \Big)^2$$

We know that the random functions $\xi_i - \mu$, $i = 1, \ldots, n$, are independent and identically distributed as $\mathcal{N}_\infty(0, V)$ (Gaussian random function with mean 0 and covariance operator V). Thus, the random variables

$$\int_T \big(\xi_i(t) - \mu(t)\big) g(t) dt, \ i = 1, \ldots, n$$

are independent and identically distributed as $\mathcal{N}_1\big(0, (g, Vg)\big)$, thanks to the definition of gaussian random function. The thesis follows immediately by definition of the χ^2 distribution.

The second statistic can be written:

$$n(g, \bar{V}g) = \Big(\int_T \sqrt{n}\big(\bar{\xi}(t) - \mu(t)\big) g(t) dt \Big)^2$$

Similar arguments give the distribution of $N(g, \bar{V}g)$.

This result put us in the conditions to use Cochran's theorem (J. and Wichern 2007). It leads then to

- $(n-1)(g, \hat{V}g) \sim (g, Vg)\chi^2(n-1)$;
- $n(g, \bar{V}g)$ and $(n-1)(g, \hat{V}g)$ are independent.

These 2 points carry with them the following consequence: given $\hat{V}$, $\forall g \in \mathrm{Im(V)} \cap \mathrm{Im(\hat{V})}$, i.e., $\forall g \in \mathrm{Im(\hat{V})}$,

$$n \frac{(g, \bar{V}g)}{(g, \hat{V}g)} \sim F(1, n-1) \tag{15}$$

Finally, we know that $\mathrm{Ker}(\hat{V})$ has null measure in $L^2(T)$. Hence, $\mathbb{P}[g \notin \mathrm{Im}(\hat{V})] = 0$. This last condition leads to the thesis. □

Proof. [Theorem 2.2] For the first part of the statement it is sufficient to note that T_p^2 is a monotonic increasing sequence which converges to the functional statistic T_f^2 defined in (1), as the basis $\{\phi_k\}_{k \geq 1}$ is dense in L^2.

Now, at p fixed, we aim at finding the expression of T_p^2. It requires first to write the decomposition of each function involved on the basis $\{e_k\}_{k \geq 1}$, and project them on the space generated by the first p basis components. We have:

$$g = \sum_{k=1}^{p} g_k \phi_k \qquad \xi_{i,p} = \sum_{k=1}^{p} \xi_{ik} \phi_k \qquad \bar{\xi}_p = \sum_{k=1}^{p} \bar{\xi}_k \phi_k \qquad \mu_p = \sum_{k=1}^{p} \mu_k \phi_k.$$

Note that we are now working with finite-dimensional approximations $\xi_{i,p}, \bar{\xi}_p, \mu_p$ of the functions $\xi_i, \bar{\xi}, \mu$, and that all approximations converge to the respective function for $p \to \infty$.

Now, the projection of the quantity $(g, \bar{V}g)$ in the p-dimensional space generated by the first p ϕ_k can be written as:

$$(g, \bar{V}g)_p = \Big(\int_T \big(\bar{\xi}_p(t) - \mu_p(t)\big) g(t)dt \Big)^2 = \Big(\int_T \sum_{k=1}^{p} \sum_{l=1}^{p} (\bar{\xi}_k - \mu_k) g_l \phi_k(t) \phi_l(t) dt \Big)^2$$

At this point, note that, by definition:

$$\int_T \phi_k(t)\phi_l(t)dt = W_{kl}$$

Thus, we obtain:

$$(g, \bar{V}g)_p = \Big(\sum_{k=1}^{p} \sum_{l=1}^{p} (\bar{\xi}_k - \mu_k) W_{kl} g_l \Big)^2 = \big((\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W \mathbf{g} \big)^2 = \big((\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W^{1/2} W^{1/2} \mathbf{g} \big)^2,$$

where

$$\mathbf{g} = (g_1, \dots, g_p)'$$
$$\bar{\boldsymbol{\xi}} - \boldsymbol{\mu} = (\bar{\xi}_1 - \mu_1, \dots, \bar{\xi}_p - \mu_p)'$$

Similarly, we have:

$$\begin{aligned}
(n-1)(g, \hat{V}g)_p &= \sum_{i=1}^{n} \Big(\int_T \big(\xi_{i,p}(t) - \bar{\xi}_p(t)\big) g(t)dt \Big)^2 \\
&= \sum_{i=1}^{n} \Big(\int_T \sum_{k=1}^{p} \sum_{l=1}^{p} (\xi_{ik} - \bar{\xi}_k) g_l \phi_k(t) \phi_l(t) dt \Big)^2 \\
&= \sum_{i=1}^{n} \Big(\sum_{k=1}^{p} \sum_{l=1}^{p} (\xi_{ik} - \bar{\xi}_k) W_{lk} g_l \Big)^2 = (n-1) \mathbf{g}' W^{1/2} S W^{1/2} \mathbf{g},
\end{aligned}$$

where

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}})'$$
$$\boldsymbol{\xi}_i - \bar{\boldsymbol{\xi}} = (\xi_{i1} - \bar{\xi}_1, \dots, \xi_{ip} - \bar{\xi}_p)'$$

Thus, we obtain the following:

$$T_p^2 = n \max_{\mathbf{g} \in \mathrm{Im(S)}} \frac{\big((\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W^{1/2} W^{1/2} \mathbf{g} \big)^2}{\mathbf{g}' W^{1/2} S W^{1/2} \mathbf{g}}.$$

It can be written in another interesting way thanks to the Maximization Lemma in J. and Wichern (2007). We get the final representation:

$$T_p^2 = n(\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W^{1/2} S^+ W^{1/2} (\bar{\boldsymbol{\xi}} - \boldsymbol{\mu}),$$

where S^+ is the Moore-Penrose inverse of S.

□

Lemma .1. *Let $\{\xi_i\}_{i=1,\dots,n}$ a set of random elements of $H^k(T)$, with common mean μ, and let $\hat{V}$ and $\bar{V}$ be the two H^k operators defined in Subsection 5.1. The two operators $\hat{V}$ and $\bar{V}$ are such that, for any $g \in H^k$:*

- $(g, \hat{V}g)_{H^k} = \widehat{Var}[(g, \xi_i)_{H^k}]$;
- $(g, \bar{V}g)_{H^k} = ((g, \bar{\xi} - \mu)_{H^k})^2$.

Proof. For any $g \in \text{Im}(\hat{V})$, we have:

$$
\begin{aligned}
(g, \hat{V}g)_{H^k} &= \sum_{l=0}^{k} (D^l g, D^l(\hat{V}g))_{L^2} \\
&= \sum_{l=0}^{k} \int_T D^l g(t) D^l \left[\int_T \sum_{j=0}^{k} \mathcal{S}_{0j}(t,s) D^j g(s) \mathrm{d}s \right] \mathrm{d}t \\
&= \sum_{l=0}^{k} \int_T D^l g(t) \int_T \sum_{j=0}^{k} \partial_t^l \mathcal{S}_{0j}(t,s) D^j g(s) \mathrm{d}s \mathrm{d}t \\
&= \sum_{l=0}^{k} \sum_{j=0}^{k} \iint_{T\times T} D^l g(t) \mathcal{S}_{lj}(t,s) D^j g(s) \mathrm{d}s \mathrm{d}t,
\end{aligned}
$$

where in the last equality, we used the fact that:

$$
\begin{aligned}
\partial_t^l \mathcal{S}_{0j}(t,s) &= \partial_t^l \frac{1}{n-1} \sum_{i=1}^{n} (\xi_i(t) - \bar{\xi}(t))(D^j \xi_i(s) - D^j \bar{\xi}(s)) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} (D^l \xi_i(t) - D^l \bar{\xi}(t))(D^j \xi_i(s) - D^j \bar{\xi}(s)) = \mathcal{S}_{lj}(t,s).
\end{aligned}
$$

Furthermore, we have:

$$
\begin{aligned}
&\sum_{l=0}^{k} \sum_{j=0}^{k} \iint_{T\times T} D^l g(t) \mathcal{S}_{lj}(t,s) D^j g(s) \mathrm{d}s \mathrm{d}t \\
=& \sum_{l=0}^{k} \sum_{j=0}^{k} \iint_{T\times T} D^l g(t) \frac{1}{n-1} \sum_{i=1}^{n} (D^l \xi_i(t) - D^l \bar{\xi}(t))(D^j \xi_i(s) - D^j \bar{\xi}(s)) D^j g(s) \mathrm{d}s \mathrm{d}t \\
=& \frac{1}{n-1} \sum_{i=1}^{n} \sum_{l=0}^{k} \sum_{j=0}^{k} \left(\int_T D^l g(t)(D^l \xi_i(t) - D^l \bar{\xi}(t)) \mathrm{d}t \right) \left(\int_T D^j g(t)(D^j \xi_i(t) - D^j \bar{\xi}(t)) \mathrm{d}t \right) \\
=& \frac{1}{n-1} \sum_{i=1}^{n} \left(\sum_{l=0}^{k} \int_T D^l g(t)(D^l \xi_i(t) - D^l \bar{\xi}(t)) \mathrm{d}t \right)^2 \\
=& \widehat{Var} \left[\sum_{l=0}^{k} \int_T D^j g(t) D^j \xi_i(t) \mathrm{d}t \right] \\
=& \widehat{Var}[(g, \xi_i)_{H^k}],
\end{aligned}
$$

i.e., $(g, \bar{V}g)_{H^k}$ is the sample variance of the scores of the orthogonal projections of ξ_i on g, $\widehat{Var}[(g, \xi_i)_{H^k}]$.

In the same way, for any $g \in H^k$, we have:

$$
\begin{aligned}
(g, \bar{V}g)_{H^k} &= \sum_{l=0}^{k} (D^l g, D^l(\bar{V}g))_{L^2} \\
&= \sum_{l=0}^{k} \int_T D^l g(t) D^l \left[\int_T \sum_{j=0}^{k} (\bar{\xi}(t) - \mu(t))(D^j \bar{\xi}(s) - D^j \mu(s)) D^j g(s) \mathrm{d}s \right] \mathrm{d}t \\
&= \sum_{l=0}^{k} \int_T D^l g(t) \int_T \sum_{j=0}^{k} \partial_t^l (\bar{\xi}(t) - \mu(t))(D^j \bar{\xi}(s) - D^j \mu(s)) D^j g(s) \mathrm{d}s\mathrm{d}t \\
&= \sum_{l=0}^{k} \sum_{j=0}^{k} \iint_{T \times T} D^l g(t)(D^l \bar{\xi}(t) - D^l \mu(t))(D^j \bar{\xi}(s) - D^j \mu(s)) D^j g(s) \mathrm{d}s\mathrm{d}t \\
&= \sum_{l=0}^{k} \sum_{j=0}^{k} \left(\int_T D^l g(t)(D^l \bar{\xi}(t) - D^l \mu(t)) \mathrm{d}t \right) \left(\int_T D^j g(t)(D^j \bar{\xi}(t) - D^j \mu(t)) \mathrm{d}t \right) \\
&= \left(\sum_{l=0}^{k} \int_T D^l g(t)(D^l \bar{\xi}(t) - D^l \mu(t)) \mathrm{d}t \right)^2 \\
&= \left((g, \bar{\xi} - \mu)_{H^k} \right)^2 ,
\end{aligned}
$$

that is, $(g, \bar{V}g)_H$ is the square distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ over g. □

Lemma .2. *Let $\{\xi_i\}_{i=1,\dots,n}$ a set of random elements of $B^2(T)$, with common mean μ, and let $\hat{V}$ and $\bar{V}$ be the two H^k operators defined in Subsection 5.2. The two operators $\hat{V}$ and $\bar{V}$ are such that, for any $g \in B^2(T)$:*

- $(g, \hat{V}g)_{B^2} = \widehat{Var}[(g, \xi_i)_{B^2}]$;
- $(g, \bar{V}g)_{B^2} = ((g, \bar{\xi} - \mu)_{B^2})^2$.

Proof. For any $g \in \mathrm{Im}(\hat{V})$, exploiting the isomorphism (12), we have:

$$
\begin{aligned}
(g, \hat{V}g)_{B^2} &= (\mathrm{clr}(g), \mathrm{clr}(\hat{V}g))_{L^2} \\
&= \left(\mathrm{clr}(g), \int_T \mathcal{S}_c(t,s) \mathrm{clr}(g)(s) \mathrm{d}s \right)_{L^2} \\
&= \iint_{T \times T} \mathrm{clr}(g)(t) \mathcal{S}_c(t,s) \mathrm{clr}(g)(s) \mathrm{d}s\mathrm{d}t \\
&= (\mathrm{clr}(g), \hat{V}_c \mathrm{clr}(g))_{L^2},
\end{aligned}
$$

where $\hat{V}_c \in \mathcal{L}(L^2)$ is the integral operator of kernel $\mathcal{S}_c(t,s)$. Finally, we have:

$$
\begin{aligned}
(\mathrm{clr}(g), \hat{V}_c \mathrm{clr}(g))_{L^2} &= \widehat{Var}((\mathrm{clr}(\xi_i), \mathrm{clr}(g))_{L^2}) \\
&= \widehat{Var}((\xi_i, g))_{B^2}).
\end{aligned}
$$

In the same way, for any $g \in B^2(T)$:

$$
\begin{aligned}
(g, \bar{V}g)_{B^2} &= (\mathrm{clr}(g), \mathrm{clr}(\bar{V}g))_{L^2} \\
&= \left(\mathrm{clr}(g), \int_T (\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))\mathrm{clr}(g)(s)\mathrm{d}s\right)_{L^2} \\
&= \iint_{T\times T} \mathrm{clr}(g)(t)(\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))\mathrm{clr}(g)(s)\mathrm{d}s\mathrm{d}t \\
&= (\mathrm{clr}(g), \bar{V}_c\mathrm{clr}(g))_{L^2},
\end{aligned}
$$

where $\bar{V}_c \in \mathcal{L}(L^2)$ is the integral operator of kernel $(\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))$. Finally, we have:

$$
\begin{aligned}
(\mathrm{clr}(g), \bar{V}_c\mathrm{clr}(g))_{L^2} &= \left((\mathrm{clr}(\bar{\xi}) - \mathrm{clr}(\mu), \mathrm{clr}(g))_{L^2}\right)^2 \\
&= \left((\bar{\xi} - \mu, g)_{B^2}\right)^2.
\end{aligned}
$$

□

References

Arveson, W. (2002), *A Short Course on Spectral Theory*, Vol. 209 of *Graduate Texts in Mathematics*, first edn, Springer, New York.

Boogaart, K. G., Egozcue, J. J. and Pawlowsky-Glahn, V. (2014), Bayes hilbert spaces, *Australian & New Zealand Journal of Statistics* **56**(2), 171–194.

Bosq, D. (2000), *Linear processes in function spaces: theory and applications*, Vol. 149, Springer, New York.

Cardot, H., Prchal, L. and Sarda, P. (2007), No effect and lack-of-fit permutation tests for functional regression, *Computational Statistics* **22**(3), 371–390.

Cuesta-Albertos, J. A. and Febrero-Bande, M. (2010), A simple multiway anova for functional data, *Test* **19**(3), 537–557.

Cuevas, A., Febrero, M. and Fraiman, R. (2004), An anova test for functional data, *Computational statistics & data analysis* **47**(1), 111–122.

De Moivre, A. (1733), Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi.

Egozcue, J. J., Díaz-Barrero, J. L. and Pawlowsky-Glahn, V. (2006), Hilbert space of probability density functions based on aitchison geometry, *Acta Mathematica Sinica* **22**(4), 1175–1182.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2006), Simplicial geometry for compositional data, *Geological Society, London, Special Publications* **264**(1), 145–159.

Fan, J. and Lin, S. K. (1998), Test of significance when data are curves, *Journal of the American Statistical Association* **93**(443), 1007–1021.

Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis: theory and practice*, Springer, New York.

Fisher Box, J. (1981), Gosset, fisher, and the t distribution, *The American Statistician* **35**(2), 61–66.

Fisher Box, J. (1987), Guinness, gosset, fisher, and small samples, *Statistical Science* pp. 45–52.

Fisher, R. A. (1925*a*), Applications of 'student's' distribution, *Metron* **5**(3), 90–104.

Fisher, R. A. (1925*b*), *Statistical methods for research workers*, Genesis Publishing Pvt Ltd.

Fisher, R. A. (1936), The coefficient of racial likeness and the future of craniometry, *Journal of the Anthropological Institute of Great Britain and Ireland* pp. 57–63.

Gauss, C. F. (1809), *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.

Gosset, W. S. S. (1908), The probable error of a mean, *Biometrika* **6**, 1–25.

Hall, P. and Tajvidi, N. (2002), Permutation tests for equality of distributions in high-dimensional settings, *Biometrika* **89**(2), 359–374.

Hall, P. and Van Keilegom, I. (2007), Two-sample tests in functional data analysis starting from discrete data, *Statistica Sinica* **17**(4), 1511.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The elements of statistical learning*, Vol. 2, Springer.

Horváth, L. and Kokoszka, P. (2012), *Inference for functional data with applications*, Vol. 200, Springer, New York.

Hotelling, H. (1931), The generalization of student's ratio, *Annals of Mathematical Statistics* **2**(3), 360–378.

Hron, K., Menafoglio, A., Templ, M., Hruzova, K. and Filzmoser, P. (2014), Simplicial principal component analysis for density functions in bayes spaces, Technical Report 25/2014, MOX, Politecnico di Milano.

J., R. A. and Wichern, D. W. (2007), *Applied multivariate statistical analysis*, 6th edn, Pearson Education Inc.

Laplace, P. S. (1820), *Théorie analytique des probabilités*, Courcier.

Lehmann, E. L. and Romano, J. P. (2006), *Testing statistical hypotheses*, Springer, New York.

Lyapunov, A. M. (1901), Nouvelle forme du théoreme sur la limite de probabilité, *Acad. Imp. Sci. St. Petersbourg, Classe Phys. Math* **12**, 12.

Mahalanobis, P. C. (1927), Analysis of race mixture in bengal, *Jour. and Proc. Asiat. Soc., Bengal,(New Series) Vol. XXIII* (3), 301–333.

Marron, J. S. and Alonso, A. M. (2014), Overview of object oriented data analysis, *Biometrical Journal* .

Menafoglio, A., Guadagnini, A. and Secchi, P. (2013), A kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers, *Stochastic Environmental Research and Risk Assessment* pp. 1–17.

Pini, A. and Vantini, S. (2013), The interval testing procedure: Inference for functional data controlling the family wise error rate on intervals., Technical Report 13/2013, MOX, Politecnico di Milano.

Ramsay, J. O. and Silverman, B. W. (2002), *Applied functional data analysis: methods and case studies*, Vol. 77, Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer, New York.

Rao, C. R. and Mitra, S. K. (1971), *Generalized inverse of matrices and its applications*, Wiley Series in Probability and Statistics, John Wiley and Sons Inc.

Schott, J. R. (2007), Some high-dimensional tests for a one-way manova, *Journal of Multivariate Analysis* **98**(9), 1825–1839.

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, *Electronic Journal of Statistics* **7**, 2005–2031.

Spitzner, D. J., Marron, J. S. and Essick, G. K. (2003), Mixed-model functional anova for studying human tactile perception, *Journal of the American Statistical Association* **98**(462), 263–272.

Srivastava, M. S. (2007), Multivariate theory for analyzing high dimensional data, *J. Japan Statist. Soc* **37**(1), 53–86.

Tarpey, T. (2003), Clustering Functional Data, *Journal of Classification* **20**, 93–114.

MOX Technical Reports, last issues

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

63/2015 Pini, A.; Stamm, A.; Vantini, S.
Hotelling s T^2 in functional Hilbert spaces

62/2014 Andr, C.; Parolini, N.; Verani, M.
Using gambling simulators to foster awareness about gambling risks

61/2014 Taffetani, M.; Ciarletta, P.
Elasto-capillarity controls the formation and the morphology of beads-on-string structures in solid fibres

59/2014 Manzoni, A.; Pagani, S.; Lassila, T.
Accurate solution of Bayesian inverse uncertainty quantification problems using model and error reduction methods

57/2014 Giverso, C.; Verani, M.; Ciarletta P.;
Branching instability in expanding bacterial colonies

58/2014 Menafoglio, A.; Secchi, P.; Guadagnini, A.
A Class-Kriging predictor for Functional Compositions with Application to Particle-Size Curves in Heterogeneous Aquifers

60/2014 Signorini, M.; Micheletti, S.; Perotto, S.
CMFWI: Coupled Multiscenario Full Waveform Inversion for seismic inversion

55/2014 Antonietti, P. F.; Houston P.; Sarti, M.; Verani, M.
Multigrid algorithms for hp-version Interior Penalty Discontinuous Galerkin methods on polygonal and polyhedral meshes

56/2014 Antonietti, P. F.; Sarti, M.; Verani, M.; Zikatanov, L. T.
A uniform additive Schwarz preconditioner for the hp-version of Discontinuous Galerkin approximations of elliptic problems

54/2014 Ferrario, E.; Pini, A.
Uncertainties in renewable energy generation systems: functional data analysis, monte carlo simulation, and fuzzy interval analysis