MODELLISTICA E CALCOLO SCIENTIFICO

MODELING AND SCIENTIFIC COMPUTING

MOX-Report No. 10/2017

Hotelling's T^2 in separable Hilbert spaces

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

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

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

Hotelling's T^2 in Separable Hilbert Spaces

Alessia Pini†
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy.
E-mail: alessia.pini@polimi.it
Aymeric Stamm
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy.
CRL, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.
Simone Vantini
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy.

Summary. We address the problem of finite-sample null hypothesis significance testing on the mean element of a random variable that takes value in a generic separable Hilbert space. For this purpose, we propose a (re)definition of Hotelling's T^2 statistic that naturally expands to any separable Hilbert space that we further embed within a permutation inferential approach. In detail, we present a unified framework for making inference on the mean element of Hilbert populations based on Hotelling's T^2 statistic, using a permutation-based testing procedure of which we prove finite-sample exactness and consistency; we showcase the explicit form of Hotelling's T^2 statistic in the case of some famous spaces used in functional data analysis (i.e., Sobolev and Bayes spaces); we propose simulations and a case study that demonstrate the importance of the space into which one decides to embed the data; we provide an implementation of the proposed tools in the R package `fdahotelling`.

Keywords: Hilbert spaces; functional data; high-dimensional data; permutation test; non-parametric inference, Hotelling's T^2.

1. Introduction

The statistical analysis of data embedded in infinite-dimensional spaces is an active research area in modern statistics. For example, the outstanding advances that measuring instruments undergo constantly enable the collection and storage of high-resolution data. This kind of data can often be represented as continuous functions (e.g. in time or space) that can be embedded in infinite-dimensional functional spaces. This is the basis of functional data analysis (FDA) (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006; Hsing and Eubank, 2015). However, FDA is just one example of statistical analysis of complex data. Other remarkable examples include the statistical analysis of densities (e.g., Menafoglio et al., 2014; van den Boogaart et al., 2014) or

†Alessia Pini, MOX - Department of Mathematics, Politecnico di Milano. Piazza Leonardo da Vinci, 32, 20133 Milano, Italy. E-mail: alessia.pini@polimi.it
Authors are ordered alphabetically.

function derivatives (e.g., Poyton et al., 2006; Sangalli et al., 2009; Dalla Rosa et al., 2014). These applications stimulated the recent statistical literature to think of a general Hilbert framework that would cover the whole new variety of complex data (see e.g., Ramsay and Dalzell, 1991; Horváth and Kokoszka, 2012; Bongiorno et al., 2014; Goia and Vieu, 2016; Menafoglio and Petris, 2016; Menafoglio and Secchi, 2016). Moving in this direction, there is an urgent need for novel statistical methodologies that can accommodate the analysis of such data.

In this work, we tackle the problem of inference on the mean element of a random variable defined in an infinite-dimensional separable Hilbert space, with extension to inference on the difference between the means of two random variables defined in such a space. In classic multivariate statistical analysis, the traditional strategy for making inference on the mean pertains to resorting to Hotelling's T^2 statistic. As long as the sample size is larger than the dimension of the space into which data belong, this statistic is fundamentally the Mahalanobis distance between the sample mean and the true mean. To the best of our knowledge, in the literature, this statistic is not defined for random variables taking values into infinite-dimensional spaces. In such a case, no matter how many samples from the random variable one can collect, information will always be insufficient to fully characterize an underlying generative model of the data. There has been a growing literature on this matter especially in the field of FDA. In particular, both parametric (Fan and Lin, 1998; Spitzner et al., 2003; Cuevas et al., 2004; Shen and Faraway, 2004; Schott, 2007; Horváth and Kokoszka, 2012) and non-parametric (Hall and Tajvidi, 2002; Hall and Van Keilegom, 2007; Cardot et al., 2007; Cox and Lee, 2008; Cuesta-Albertos and Febrero-Bande, 2010; Corain et al., 2014; Pini and Vantini, 2016, 2017; Pini et al., 2017) inferential procedures have been proposed with a focus on the statistical analysis of functional data embedded in the $L^2(T)$ space of square-integrable functions on the compact set $T \subseteq \mathbb{R}$. These works can be divided into two macro-methodologies. The first approach pertains to projecting the data onto a low-dimensional subspace using a truncated basis expansion and evaluating Hotelling's T^2 statistic in such a subspace (e.g., Spitzner et al., 2003; Cuesta-Albertos and Febrero-Bande, 2010; Horváth and Kokoszka, 2012; Pini and Vantini, 2016), as if we were dealing with a classical multivariate statistical analysis problem. The second approach pertains to analyzing the data in its original space but trading Hotelling's T^2 statistic for other statistics based on the L^2 norm (e.g., Hall and Tajvidi, 2002; Hall and Van Keilegom, 2007).

In this work, we propose to follow and further develop the second line of research in which inference is carried out in the original space of the data because we believe that information thrown away by projecting the data on subspaces of lower dimension could actually be relevant for the ongoing inference. In particular, our goal is to demonstrate that Hotelling's T^2 statistic is actually well defined in any separable Hilbert spaces and that it can be used for inference on the mean element in separable Hilbert spaces of infinite dimension. Traditional Hotelling's T^2 statistic in $\mathbb{R}^p$ has already been extended to the case when the dimension p exceeds the sample size (Srivastava, 2007). From a parametric standpoint, under the assumption of normality, it follows a Fisher distribution when $p < n$ and a χ^2 distribution p-asymptotically (Secchi et al., 2013). Hence, inference can be carried out quite easily. However, cases of normal data are actually scarce and assessment of normality is still an open question in high-dimensional and infinite-dimensional spaces. Fortunately, even though Hotelling's T^2 statistic has mainly be used for parametric inference under the assumption of normality, we would like to emphasize that it is not limited to parametric inference. One could perform inference on the mean in a non-parametric fashion – such as in bootstrap or permutation testing – using Hotelling's T^2 statistic as well. In this work, we rely in particular on the latter.

This manuscript is outlined as follows. Section 2 starts with the traditional definition of Hotelling's T^2 statistic and introduces a novel alternative but equivalent definition from which it is straightforward to grasp why Hotelling's T^2 statistic can in fact be defined on any separable Hilbert spaces. Section 3 develops this idea by defining random variables in Hilbert spaces and the related concepts of mean and covariance operator. The section ends with the general definition that we propose for Hotelling's T^2 statistic, valid in any separable Hilbert space. Section 4 showcases three classes of separable Hilbert spaces. The first example is $\mathbb{R}^p$, used in multivariate and high-dimensional data analysis, while the other two examples feature spaces widely used in functional data analysis (i.e., Sobolev and Bayes spaces). In detail, we define Hotelling's T^2 statistic as the maximization of a ratio involving linear operators. Section 7 presents the two statistical tests that we propose in separable Hilbert spaces using Hotelling's T^2 statistic: a one-population test on the mean element of a population and a two-population test on the difference between the mean elements of two populations. Finite-sample exactness and consistency of the proposed tests are proven independently from data distribution. Sections 8 and 9 propose simulations and a case study respectively that demonstrate (i) the practicality and efficiency of the proposed procedure based on Hotelling's T^2 statistic and (ii) the importance of the space into which one decides to embed the data. Section 10 finally summarizes the main results achieved in this manuscript and discusses some important aspects.

2. Hotelling's T^2 revisited

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n *i.i.d.* $\mathbb{R}^p$-valued random variables with mean $\mathbf{m}$ and covariance matrix Σ. Traditional multivariate statistical theory teaches that when the number n of statistical units is strictly greater than the dimension p of the data (i.e. $n > p$), it is possible to make inference on the mean $\mathbf{m}$ using the well-known Hotelling's T^2 statistic usually defined in the following way:

$$T^2 := n\left(\mathbf{m}_n - \mathbf{m}\right)^\top \Sigma_n^{-1}\left(\mathbf{m}_n - \mathbf{m}\right), \tag{1}$$

where $\mathbf{m}_n$ and Σ_n are the sample mean vector and sample covariance matrix respectively:

$$\mathbf{m}_n := \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \quad \text{and} \quad \Sigma_n := \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \mathbf{m}_n)(\mathbf{X}_i - \mathbf{m}_n)^\top. \tag{2}$$

Multivariate Hotelling's T^2 statistic is closely connected to the univariate Student's t statistic. In effect, for any direction $\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$, the following Student's t statistic can be computed:

$$t_{\mathbf{a}} = \sqrt{n}\frac{\mathbf{a}^\top (\mathbf{m}_n - \mathbf{m})}{\sqrt{\mathbf{a}^\top \Sigma_n \mathbf{a}}},$$

and an alternative definition of Hotelling's T^2 statistic can be formulated as follows:

$$T^2 := \max_{\mathbf{a}\in\mathbb{R}^p\setminus\{\mathbf{0}\}} t_{\mathbf{a}}^2 = n \max_{\mathbf{a}\in\mathbb{R}^p\setminus\{\mathbf{0}\}} \frac{\left[\mathbf{a}^\top (\mathbf{m}_n - \mathbf{m})\right]^2}{\mathbf{a}^\top \Sigma_n \mathbf{a}}. \tag{3}$$

Since $n > p$, the sample covariance matrix Σ_n is a.s. of full rank p. Therefore, the column space of Σ_n entirely spans $\mathbb{R}^p$, i.e. $\mathrm{Im}(\Sigma_n) = \mathbb{R}^p$. Hence, independently from the sample size n, Hotelling's T^2 statistic can be defined in its most general form as follows:

Definition 2.1. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n i.i.d. $\mathbb{R}^p$-valued random variables with mean $\mathbf{m}$ and covariance matrix Σ. Hotelling's T^2 statistic is defined as:*

$$T^2 := n \max_{\mathbf{a}\in\mathrm{Im}(\Sigma_n)\setminus\{\mathbf{0}\}} \frac{\langle \mathbf{a}, D_n\mathbf{a}\rangle_{\mathbb{R}^p}}{\langle \mathbf{a}, \Sigma_n\mathbf{a}\rangle_{\mathbb{R}^p}}, \tag{4}$$

where $D_n = (\mathbf{m}_n - \mathbf{m})(\mathbf{m}_n - \mathbf{m})^\top$, Σ_n is the sample covariance matrix and $\langle \cdot, \cdot \rangle_{\mathbb{R}^p}$ is the natural inner product on $\mathbb{R}^p$.

Definition 2.1 boils down to the traditional definition of Hotelling's T^2 statistic given by eq. (1) when $n > p$. Moreover, this definition is very convenient since it provides a straightforward extension of Hotelling's T^2 statistic to the $p \geq n$ and high-dimensional cases because the maximization problem that defines Hotelling's T^2 according to eq. (4) is still well posed. In particular, for multivariate $p \geq n$ settings, it has been shown in Secchi et al. (2013) that:

$$T^2 = n(\mathbf{m}_n - \mathbf{m})^\top \Sigma_n^+ (\mathbf{m}_n - \mathbf{m}),$$

where Σ_n^+ is the Moore-Penrose inverse of the sample covariance matrix Σ_n.

Likewise, it is possible to exploit eq. (4) for defining Hotelling's T^2 statistic on any separable Hilbert space. This will be the object of the next section.

3. Hotelling's T^2 statistic in separable Hilbert spaces

Let $\mathbb{H}$ be a separable Hilbert space. Recall that a Hilbert space is a complete vector space endowed with an inner product. Hence, let us start by introducing notations for the operations of addition, scalar multiplication and inner product in $\mathbb{H}$:

- The **addition** of two elements $f, g \in \mathbb{H}$ is denoted $f \oplus g$;
- The **scalar multiplication** of an element $f \in \mathbb{H}$ by a scalar $\lambda \in \mathbb{R}$ is denoted by $\lambda \odot f$;
- The **subtraction** of two elements $f, g \in \mathbb{H}$ is completely defined by the operations of addition and scalar multiplication. However, since it will be widely used in this manuscript, we introduce the notation $\ominus := \oplus(-1)\odot$ for denoting the subtraction in $\mathbb{H}$;

- The **inner product** between two elements $f, g \in \mathbb{H}$ is denoted $\langle \cdot, \cdot \rangle_{\mathbb{H}}$.

The goal of this section is to extend Definition 2.1 of Hotelling's T^2 statistic from $\mathbb{R}^p$ to the general case of separable Hilbert spaces. For this purpose, we shall define the concept of random element and its mean and covariance in Hilbert spaces.

DEFINITION 3.1. *An* $\mathbb{H}$***-valued random variable*** *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a mapping* $\chi : \Omega \to \mathbb{H}$ *such that* $\langle \chi, f \rangle_{\mathbb{H}}$ *is measurable for all* $f \in \mathbb{H}$.

From now on, let χ be an $\mathbb{H}$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The concept of mean for χ relies upon a notion of integration on the probability space for objects belonging to separable Hilbert spaces. In this manuscript, following Hsing and Eubank (2015), we adopt the Bochner integral which can be viewed as the natural extension of the traditional Lebesgue integration used in multivariate statistical analysis. For simplicity and clarity, we hereby formulate a simple definition of Bochner integral for $\mathbb{H}$-valued random variables. Its general definition requires further specific assumptions on the structure of the Hilbert space which are always valid in the case of separable Hilbert spaces and whose specifications are out of the scope of this work. The interested reader can refer to Hsing and Eubank (2015) for the complete general definition of Bochner integrals.

DEFINITION 3.2. *Let* χ *be an* $\mathbb{H}$*-valued random variable on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *The* ***Bochner integral*** *of* χ *with respect to the probability measure* $\mathbb{P}$ *is the unique element* $\mathbb{E}_{\mathbb{H}}[\chi]$ *of* $\mathbb{H}$ *such that*

$$\langle \mathbb{E}_{\mathbb{H}}[\chi], f \rangle_{\mathbb{H}} = \mathbb{E}_{\mathbb{R}} \left[\langle \chi, f \rangle_{\mathbb{H}} \right], \quad \textit{for all } f \in \mathbb{H},$$

with $\mathbb{E}_{\mathbb{R}}$ *being the expectation operator for* $\mathbb{R}$*-valued random variables.*

Before proceeding to the definition of mean and variance for $\mathbb{H}$-valued random variables, it is useful to introduce a proper notation and characterization for the product of two elements of a separable Hilbert space.

DEFINITION 3.3. *Let* f *and* g *be two elements of a separable Hilbert space* $\mathbb{H}$. *The* ***tensor product*** *operator* $(f \otimes g)_{\mathbb{H}} : \mathbb{H} \to \mathbb{H}$ *is defined as:*

$$\begin{array}{rccc} (f \otimes g)_{\mathbb{H}} : & \mathbb{H} & \to & \mathbb{H} \\ & h & \mapsto & \langle f, h \rangle_{\mathbb{H}} \odot g \end{array}$$

Definitions 3.1 to 3.3 provide the basic fundamental elements for giving a proper definition of mean and covariance of an $\mathbb{H}$-valued random variable.

DEFINITION 3.4. *Assume that* $\mathbb{E}_{\mathbb{R}} \left[\|\chi\|_{\mathbb{H}}^2 \right] < +\infty$. *Then, the* ***mean*** m *of* χ *is the element of* $\mathbb{H}$ *given by the Bochner integral of* χ*:*

$$m := \mathbb{E}_{\mathbb{H}}(\chi),$$

and the ***covariance operator*** *of* χ *is the element of* $\mathcal{B}_{\mathrm{HS}}(\mathbb{H})$ *given by the Bochner integral of* $(\chi \ominus m) \otimes_{\mathbb{H}} (\chi \ominus m)$*:*

$$\mathcal{K} := \mathbb{E}_{\mathcal{B}_{\mathrm{HS}}(\mathbb{H})} \left[(\chi \ominus m) \otimes_{\mathbb{H}} (\chi \ominus m) \right],$$

where $\mathcal{B}_{\mathrm{HS}}(\mathbb{H})$ *denotes the space of Hilbert-Schmidt operators on* $\mathbb{H}$.

Now that we formally defined the mean m and covariance operator $\mathcal{K}$ of an $\mathbb{H}$-valued random variable χ, we shall address the problem of estimating them given a random sample drawn from the distribution of χ. By analogy to the traditional estimators used in multivariate statistical analysis and recalled in eq. (2), using standard argumentation we can construct unbiased estimators of the mean and covariance operators as follows:

Definition 3.5. *Let $\chi_1, \ldots, \chi_n$ be a sample of n i.i.d. $\mathbb{H}$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}\left[\|\chi_i\|_{\mathbb{H}}^2\right] < +\infty$ for all $i = 1, \ldots, n$. Let the mean m and covariance operator $\mathcal{K}$ of the χ_i's be given by Definition 3.4. We define:*

- *the **sample mean** m_n as:*
$$m_n := \frac{1}{n} \odot \bigoplus_{i=1}^{n} \chi_i,$$
- *the **sample covariance operator** $\mathcal{K}_n$ as:*
$$\mathcal{K}_n := \frac{1}{n-1} \odot \bigoplus_{i=1}^{n} (\chi_i \ominus m_n) \otimes_{\mathbb{H}} (\chi_i \ominus m_n).$$

The quantities m_n and $\mathcal{K}_n$ are respectively an $\mathbb{H}$-valued random variable and a $\mathcal{B}_{\mathrm{HS}}(\mathbb{H})$-valued random variable. They are both unbiased for their respective population counterparts.

Finally, recall that our aim is to apply eq. (4) to define Hotelling's T^2 statistic in Hilbert spaces. While m_n and $\mathcal{K}_n$ naturally play the roles of $\mathbf{m}_n$ and Σ_n in eq. (4) respectively, we shall define a novel operator that captures the squared error loss of the sample mean m_n in estimating the true mean m for replacing D_n in eq. (4).

Definition 3.6. *Using assumptions and notations in Definition 3.5, we define the **sample mean squared-error loss operator** $\mathcal{D}_n$ as:*
$$\mathcal{D}_n := (m_n \ominus m) \otimes_{\mathbb{H}} (m_n \ominus m).$$
The sample mean squared-error loss operator is a $\mathcal{B}_{\mathrm{HS}}(\mathbb{H})$-valued random variable.

We now have introduced all the key ingredients to extend Definition 2.1 to the general case of separable Hilbert spaces.

Definition 3.7. *Let $\chi_1, \ldots, \chi_n$ be n i.i.d. $\mathbb{H}$-valued random variables with mean m and covariance operator $\mathcal{K}$. We naturally define Hotelling's T^2 statistic as:*
$$T^2 := n \max_{f \in \mathrm{Im}(\mathcal{K}_n) \setminus \{0\}} \frac{\langle f, \mathcal{D}_n f \rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f \rangle_{\mathbb{H}}}, \tag{5}$$
or equivalently as:
$$T^2 := n \langle m_n \ominus m, \mathcal{K}_n^+ (m_n \ominus m) \rangle_{\mathbb{H}}, \tag{6}$$
where $\mathcal{K}_n$ is the sample covariance operator given in Definition 3.5 and $\mathcal{K}_n^+$ is the Moore-Penrose generalized inverse operator of the sample covariance operator $\mathcal{K}_n$, and $\mathcal{D}_n$ is the sample mean squared-error loss operator given by Definition 3.6.

Both definitions – given by eqs. (5) and (6) – have intrinsic values. Indeed, the former highlights the meaning of the statistic as the largest squared univariate Mahalanobis distance between univariate projections of the sample mean and true mean onto the random space $\mathrm{Im}(\mathcal{K}_n) = \mathrm{span}\{\chi_i \ominus m_n\}_{i=1,\ldots,n}$. The latter definition instead is far more explicit and provides a direct link with its Euclidean ancestor. In details, Lemma B.1, reported in Appendix B, proves the equivalence of the two definitions and guarantees that the maximum of $n\frac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}}$ (i.e., the value T^2) always exists and is finite, and it is reached at random directions proportional to $f = \mathcal{K}_n^+(m_n \ominus m)$. The proof is based on Cauchy-Schwarz inequality in the Hilbert space $\mathbb{H}$ and on the fact that, given a sample $\chi_1, \ldots, \chi_n$ of n *i.i.d.* $\mathbb{H}$-valued random variables with finite total variance, the sample covariance operator $\mathcal{K}_n$ is a positive semi-definite self-adjoint Hilbert-Schmidt operator. Moreover, it is easy to prove that if the maximization in eq. (5) was not constrained to the random subspace $\mathrm{Im}(\mathcal{K}_n) \setminus \{0\}$, one would trivially have $T^2 \overset{a.s}{=} +\infty$ independently from the random sample. Finally, while in multivariate statistical analysis the closed-form expression of T^2 given by eq. (6) is straightforward to compute, statistical analysis in generic Hilbert spaces of infinite dimension always instead requires approximations which makes eq. (6) much less useful from a practical point of view. In Section 5, we rather propose a solution for numerically approximating T^2 that relies upon a sequence of finite-dimensional statistics and we rely instead on eq. (5) to prove its almost sure convergence to T^2.

Remark 3.8 (Extension to semi-Hilbert spaces). *Hotelling's T^2 statistic can also be defined in semi-Hilbert spaces, i.e., Hilbert spaces for which the inner product is only positive semi-definite. In this case, the inner product only defines a semi-distance for which the identity of indiscernibles does not hold. One can then focus on the quotient set induced on the space by the related semi-distance, which defines a Hilbert space in the usual sense where Definition 3.7 holds.*

The coming section will be dedicated to examples of widely used separable Hilbert spaces for the statistical analysis of multivariate and functional data with a focus on how Hotelling's T^2 statistic can be easily defined in these particular spaces using Definition 3.7.

4. Hotelling's T^2 statistic for high-dimensional and functional data

In this section, we aim at introducing Hotelling's T^2 statistic in some of the most widely used separable Hilbert spaces in the context of high-dimensional and functional data analysis. Tables 1 to 3 summarize the key ingredients required for the proper definition of Hotelling's T^2 statistic for three classes of separable Hilbert spaces, along with their actual analytic expressions using only the operations of addition and scalar multiplication in $\mathbb{R}$. The key ingredients can be divided into two categories:

- the operations of addition, scalar multiplication and inner product that confer the Hilbert structure to the each specific space $\mathbb{H}$ that we explicit using two generic elements $f, g \in \mathbb{H}$;

- the sample mean, sample covariance operator and sample mean squared-error loss operator that we explicit using a sample of n *i.i.d.* $\mathbb{H}$-valued random variables $\chi_1, \dots, \chi_n$.

4.1. *Hotelling's T^2 statistic in $\mathbb{R}^p$*

The first class of separable Hilbert spaces that we feature is the trivial case in which $\mathbb{H} = \mathbb{R}^p$ summarized in Table 1. This is the natural space in which we traditionally embed our data for performing multivariate (including low- and high-dimensional) data analysis. As it can be seen from Table 1, our general Definition 3.7 of Hotelling's T^2 statistic in separable Hilbert spaces boils down to the traditional expression of Hotelling's T^2 in $\mathbb{R}^p$ with $p < n$. Moreover, we can see that Hotelling's T^2 statistic in itself does not depend upon the ordering between p and n. In effect, what changes in a situation where $p \geq n$ with respect to $p < n$ is that we do not have a sufficiently large amount of data to fully characterize the underlying generative model for our data. As a result, Hotelling's T^2 statistic becomes blind to any mean differences that lie in the kernel of the sample covariance matrix Σ_n. We refer the interested reader to Secchi et al. (2013) for an in-depth discussion on the matter.

4.2. *Hotelling's T^2 statistic in $H^k(T)$*

The second class of separable Hilbert spaces featured in this section are the Sobolev spaces $H^k(T)$ that are massively used in functional data analysis for modelling curves (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012). In the case of one-dimensional functions defined over one-dimensional domains, these spaces are made of (classes of equivalence of) k-differentiable functions on a compact set $T \subseteq \mathbb{R}$ with square-integrable derivatives up to the order k. Table 2 shows how the key ingredients for the definition of Hotelling's T^2 statistic can be easily expressed in terms of the addition and scalar multiplication in $\mathbb{R}$ by introducing the notation $f^{(j)}$ for the j-*th* derivative of f. The extension to the case of multivariate functions defined on a multivariate domain is trivial and readily deducible from the basic case.

The most widely used separable Hilbert space for functional data analysis is the Sobolev space $H^0(T)$ of (classes of equivalence of) square-integrable functions on T, which is traditionally indicated as $L^2(T)$. Given the popularity of this space in FDA, some statistics already emerged in the literature (mainly two) for making inference on the mean function of a population of curves or on the difference between the mean functions of two populations of curves, when curves are seen as $L^2(T)$-valued random variables. We hereby briefly state their definitions and we will critically compare them to Hotelling's T^2 statistic in the simulations reported in Section 8 and in the case study developed in Section 9. They have both been introduced by Hall and Tajvidi (2002) as:

- the L^2 *distance*:

$$T_I^2 = \int_T (m_n(t) - m(t))^2 \, \mathrm{d}t, \tag{7}$$

Table 1. The space $\mathbb{R}^p$. Hotelling's T^2 statistic in the Hilbert space $(\mathbb{R}^p, \oplus, \odot, \langle\cdot,\cdot\rangle_{\mathbb{R}^p})$ of real vectors of dimension p.

Component	Analytic expression
$f \oplus g$	$(f_1 + g_1, \ldots, f_p + g_p)^\top$
$\lambda \odot f$	$(\lambda f_1, \ldots, \lambda f_p)^\top$
$\langle f, g\rangle_{\mathbb{H}}$	$\sum_{\ell=1}^{p} f_\ell g_\ell$
m_n	$\left(\frac{1}{n}\sum_{i=1}^{n}\chi_{i1}, \ldots, \frac{1}{n}\sum_{i=1}^{n}\chi_{ip}\right)^\top$
$\mathcal{K}_n f$	$\begin{pmatrix} \frac{1}{n-1}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{p}(\chi_{i\ell} - m_{n\ell})f_\ell\right)(\chi_{i1} - m_{n1}) \\ \vdots \\ \frac{1}{n-1}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{p}(\chi_{i\ell} - m_{n\ell})f_\ell\right)(\chi_{ip} - m_{np}) \end{pmatrix}$
$\mathcal{D}_n f$	$\begin{pmatrix} \left(\sum_{\ell=1}^{p}(m_{n\ell} - m_\ell)f_\ell\right)(m_{n1} - m_1) \\ \vdots \\ \left(\sum_{\ell=1}^{p}(m_{n\ell} - m_\ell)f_\ell\right)(m_{np} - m_p) \end{pmatrix}$
T^2	$n \max_{f \in \mathrm{Im}(\mathcal{K}_n)\setminus\{0\}} \frac{\left[\sum_{\ell=1}^{p}(m_{n\ell} - m_\ell)f_\ell\right]^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left[\sum_{\ell=1}^{p}(\chi_{i\ell} - m_{n\ell})f_\ell\right]^2}$

Table 2. The space $H^k(T)$. Hotelling's T^2 statistic in the Hilbert space $\left(H^k(T), \oplus, \odot, \langle\cdot,\cdot\rangle_{H^k}\right)$ of k-differentiable functions on T with square-integrable derivatives up to the order k, also known as Sobolev space.

Component	Analytic expression
$f \oplus g$	$\begin{array}{ccc} T & \to & \mathbb{R} \\ t & \mapsto & f(t)+g(t) \end{array}$
$\lambda \odot f$	$\begin{array}{ccc} T & \to & \mathbb{R} \\ t & \mapsto & \lambda f(t) \end{array}$
$\langle f, g\rangle_{\mathbb{H}}$	$\displaystyle\sum_{j=0}^{k}\int_T f^{(j)}(t) g^{(j)}(t)\mathrm{d}t$
m_n	$\begin{array}{ccc} T & \to & \mathbb{R} \\ t & \mapsto & \displaystyle\frac{1}{n}\sum_{i=1}^{n}\chi_i(t) \end{array}$
$\mathcal{K}_n f$	$\begin{array}{ccc} T & \to & \mathbb{R} \\ t & \mapsto & \displaystyle\frac{1}{n-1}\sum_{i=1}^{n}\left(\sum_{j=0}^{k}\int_T (\chi_i^{(j)}(s) - m_n^{(j)}(s)) f^{(j)}(s)\mathrm{d}s\right)(\chi_i(t) - m_n(t)) \end{array}$
$\mathcal{D}_n f$	$\begin{array}{ccc} T & \to & \mathbb{R} \\ t & \mapsto & \displaystyle\left(\sum_{j=0}^{k}\int_T \left[m_n^{(j)}(s) - m^{(j)}(s)\right] f^{(j)}(s)\mathrm{d}s\right)[m_n(t) - m(t)] \end{array}$
T^2	$\displaystyle n \max_{f\in \mathrm{Im}(\mathcal{K}_n)\setminus\{0\}} \frac{\left[\sum_{j=0}^{k}\int_T f^{(j)}(t)\left(m_n^{(j)}(t) - m^{(j)}(t)\right)\mathrm{d}t\right]^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left[\sum_{j=0}^{k}\int_T f^{(j)}(t)\left(\chi_i^{(j)}(t) - m_n^{(j)}(t)\right)\mathrm{d}t\right]^2}$

- the *standardized L^2 distance*:
$$T^2_{D_\sigma} = \int_T \frac{(m_n(t) - m(t))^2}{\sigma_n^2(t)} \mathrm{d}t, \tag{8}$$
where $\sigma_n^2 : T \to \mathbb{R}$ is the pointwise sample variance function naturally defined by: $(n-1)\sigma_n^2(t) = \sum_{i=1}^n (\chi_i(t) - m(t))^2$.

Observe that the statistic T_I^2 does not account for either the pointwise variance of the data or its auto-covariance structure. In effect, it gives equal weight to compact sets of equal measure in T. The statistic $T^2_{D_\sigma}$ can instead be seen as a weighted version of the statistic T_I^2, in which the pointwise estimate $\sigma_n^2(t)$ of the variance of the data is used for standardization but the auto-correlation structure is still discarded.

4.3. Hotelling's T^2 statistic in $B^2(T)$

The third class of separable Hilbert spaces that we include in this section are the Bayes spaces $B^k(T)$ of (classes of equivalence of) absolutely continuous positive functional compositions on a compact set $T \subseteq \mathbb{R}$ with k-th power integrable logarithm (Egozcue et al., 2006). In particular, we focus on the space $B^2(T)$ and we use the element integrating to 1 as the representative of a class of equivalence. In this setting, $B^2(T)$ can be seen as the natural space to embed data points that are densities (Hron et al., 2016). It has been shown by Egozcue et al. (2006) that $B^2(T)$ endowed with the operations of addition, scalar multiplication and inner product defined in Table 3 is a separable Hilbert space. As a result, it is isomorphic to $L^2(T)$ and van den Boogaart et al. (2014) have provided an isometric isomorphism between $B^2(T)$ and $L^2(T)$ called the centered log-ratio transform and denoted clr in Table 3. We thus naturally define the sample mean, sample covariance operator and sample mean squared-error loss operator in $B^2(T)$ as the inverse clr transform of the sample mean, sample covariance operator and sample mean squared-error loss operator in $L^2(T)$ of the clr-transformed version of the sample of densities, respectively. The definition of Hotelling's T^2 statistic in $B^2(T)$ then becomes straightforward using Definition 3.7 and we can therefore make inference on the mean density of a population of densities or on the difference between the mean densities of two populations of densities. The analytic formulas are summarised in Table 3.

5. Computation of Hotelling's T^2 statistic

Definition 3.7 is useful in guaranteeing the existence of Hotelling's T^2 statistic in separable Hilbert spaces and we have just illustrated its applicability to specific Hilbert spaces widely used in real-life situations. However, it does not provide the means to compute such a statistic in the practice. In the current section, we address this problem by providing a Theorem that introduces a sequence of finite-dimensional T^2 statistics that converges almost surely to Hotelling's T^2 statistic.

Since $\mathbb{H}$ is a separable Hilbert space, there exists a sequence of non-empty subspaces $\{V_p\}_{p\geq 1}$ such that $V_p \subset \mathbb{H}$, $\dim(V_p) = p$ and
$$\lim_{p\to\infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0, \quad \forall h \in \mathbb{H}. \tag{9}$$

Table 3. The space $B^2(T)$**.** Hotelling's T^2 statistic in the Hilbert space $\big(B^2(T), \oplus, \odot, \langle\cdot,\cdot\rangle_{B^2}\big)$ of densities on T with square-integrable logarithm, also known as Bayes linear space.

Component	Analytic expression
$f \oplus g$	$\begin{array}{rcl} T & \to & \mathbb{R} \\ t & \mapsto & \dfrac{f(t)g(t)}{\int_T f(s)g(s)\mathrm{d}s} \end{array}$
$\lambda \odot f$	$\begin{array}{rcl} T & \to & \mathbb{R} \\ t & \mapsto & \dfrac{f(t)^\lambda}{\int_T f(s)^\lambda \mathrm{d}s} \end{array}$
$\langle f, g\rangle_{\mathbb{H}}$	$\displaystyle\int_T \mathrm{clr}(f)(t)\mathrm{clr}(g)(t)\mathrm{d}t$
m_n	$\begin{array}{rcl} T & \to & \mathbb{R} \\ t & \mapsto & \mathrm{clr}^{-1}\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^n \mathrm{clr}(\chi_i)(t)\right] \end{array}$
$\mathcal{K}_n f$	$\begin{array}{rcl} T & \to & \mathbb{R} \\ t & \mapsto & \mathrm{clr}^{-1}\left(\dfrac{1}{n-1}\displaystyle\sum_{i=1}^n \left(\int_T [\mathrm{clr}(\chi_i)(s) - \mathrm{clr}(m_n)(s)]\, \mathrm{clr}(f)(s)\mathrm{d}s\right)[\mathrm{clr}(\chi_i)(t) - \mathrm{clr}(m_n)(t)]\right) \end{array}$
$\mathcal{D}_n f$	$\begin{array}{rcl} T & \to & \mathbb{R} \\ t & \mapsto & \mathrm{clr}^{-1}\left(\left(\displaystyle\int_T [\mathrm{clr}(m_n)(s) - \mathrm{clr}(m)(s)]\, \mathrm{clr}(f)(s)\mathrm{d}s\right)[\mathrm{clr}(m_n)(t) - \mathrm{clr}(m)(t)]\right) \end{array}$
T^2	$n \displaystyle\max_{f \in \mathrm{Im}(\mathcal{K}_n)\setminus\{0\}} \dfrac{\left[\int_T \mathrm{clr}(f)(t)\Big(\mathrm{clr}(m_n)(t) - \mathrm{clr}(m)(t)\Big)\mathrm{d}t\right]^2}{\frac{1}{n-1}\sum_{i=1}^n \left[\int_T \mathrm{clr}(f)(t)\Big(\mathrm{clr}(\chi_i)(t) - \mathrm{clr}(m_n)(t)\Big)\mathrm{d}t\right]^2}$

The last property essentially means that, as the dimensionality p of the space V_p goes to infinity, the space V_p tends to cover $\mathbb{H}$ entirely. Given the existence of such a sequence of subspaces, it is then possible to define a sequence of statistics $\{T_p^2\}_{p\geq 1}$ as follows:

$$T_p^2 := n \max_{f \in \mathrm{Im}(\mathcal{K}_n)\cap V_p \setminus\{0\}} \frac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}}. \tag{10}$$

The sequence of statistics given by eq. (10) has two key properties that are summarized in the following:

THEOREM 5.1. *Let $\mathbb{H}$ be a separable Hilbert space and $\{V_p\}_{p\geq 1}$ a sequence of subspaces such that $V_p \subset \mathbb{H}$, $\dim(V_p) = p$ and $\lim_{p\to\infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0$, for all $h \in \mathbb{H}$. Let also $\chi_1, \ldots, \chi_n$ be a sample of n i.i.d. $\mathbb{H}$-valued random variables with mean element m and covariance operator $\mathcal{K}$ such that $\mathbb{E}_{\mathbb{R}}[\|\chi_i\|^2_{\mathbb{H}}] < \infty$. Then, the sequence of statistics $\{T_p^2\}_{p\geq 1}$ defined in eq. (10) has the following two key properties:*

(a) If $(e_1, \ldots, e_p)$ is a basis set of V_p, let W be the symmetric invertible $p \times p$ matrix

such that $W_{jk} := \langle e_j, e_k \rangle_{\mathbb{H}}$. *Let also:*

$$\begin{aligned}
\boldsymbol{\chi}_i &:= W^{-1} \left(\langle \chi_i, e_1 \rangle_{\mathbb{H}}, \ldots, \langle \chi_i, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\boldsymbol{m}_n &:= W^{-1} \left(\langle m_n, e_1 \rangle_{\mathbb{H}}, \ldots, \langle m_n, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\boldsymbol{m} &:= W^{-1} \left(\langle m, e_1 \rangle_{\mathbb{H}}, \ldots, \langle m, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\Sigma_n &:= \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{\chi}_i - \boldsymbol{m}_n)(\boldsymbol{\chi}_i - \boldsymbol{m}_n)^\top.
\end{aligned}$$

Then, $T_p^2 = n(\boldsymbol{m}_n - \boldsymbol{\mu})^\top W^{1/2} (W^{1/2} \Sigma_n W^{1/2})^+ W^{1/2} (\boldsymbol{m}_n - \boldsymbol{\mu})$.

(b) $T_p^2 \xrightarrow[p \to \infty]{a.s.} T^2$, *where* T^2 *is Hotelling's* T^2 *statistic defined on the original sample* $\chi_1, \ldots, \chi_n$ *of i.i.d.* $\mathbb{H}$*-valued random variables.*

The proof can be found in Appendix B. In essence, Theorem 5.1 provides a mean to numerically approximate Hotelling's T^2 statistic in a separable Hilbert space $\mathbb{H}$ using a sequence of T^2-like statistics in $\mathbb{R}^p$. In detail, convergence is achieved for any sequence of subspaces $V_p \subset \mathbb{H}$ satisfying eq. (9), independently from their nested (e.g., wavelets with increasing frequencies) or non-nested nature (e.g., B-splines with increasing knots).

6. Properties of Hotelling's T^2 statistic in separable Hilbert spaces

Hotelling's T^2 statistic in separable Hilbert spaces has a number of desirable properties that makes it particularly appealing for inferential purposes:

T^2 is a semi-distance between m and m_n. It is important to keep in mind that, although Definition 3.7 of Hotelling's T^2 statistic boils down to the traditional Hotelling's T^2 statistic as introduced in any textbook on multivariate statistical analysis, the two statistics fundamentally differs in their mathematical implications. The multivariate $p < n$ Hotelling's T^2 statistic is defined as the maximum of the squared t statistics formed from all possible one-dimensional projections of the multivariate data. However, Hotelling's T^2 statistic in an infinite-dimensional separable Hilbert space $\mathbb{H}$ is defined as the maximum over the space $\mathrm{Im}\,(\mathcal{K}_n) \setminus \{0\}$ which is an $(n-1)$-dimensional random subspace of $\mathbb{H}$. As a result, the corresponding T^2 statistic can be viewed as a distance between m and m_n only in that random space, but it actually is only a semi-distance in $\mathbb{H}$, for which the identity of indiscernibles does not hold.

T^2 is invariant under similarity transformations. Hotelling's T^2 statistic in Definition 3.7 is invariant under similarity transformations of the data, i.e., under affine transformations $\chi \mapsto a \odot (\mathcal{O}\chi) \oplus f$, where $a \in \mathbb{R}^+$, $f \in \mathbb{H}$ and $\mathcal{O}$ is an orthogonal linear limited operator on $\mathbb{H}$, i.e., $\mathcal{O}$ satisfies $\langle \mathcal{O}f, \mathcal{O}g \rangle_{\mathbb{H}} = \langle f, g \rangle_{\mathbb{H}}$ for any $f, g \in \mathbb{H}$. Lehmann and Romano (2006) have shown that this type of invariance is the largest family of invariance transformations that one can achieve in the multivariate framework $p \geq n$. In this sense, Hotelling's T^2 statistic is invariant-optimal.

Under $\mathbb{H}$-Gaussianity the distribution of $n\frac{\langle f,\mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f,\mathcal{K}_n f\rangle_{\mathbb{H}}}$ is known. It is well known that $\mathbb{R}^p$-Gaussianity, more traditionally termed multivariate Gaussianity, is defined as follows: the $\mathbb{R}^p$-valued random variable χ follows a multivariate Gaussian distribution if and only if, for all $f \in \mathbb{R}^p$, the $\mathbb{R}$-valued random variable $\langle \chi, f\rangle_{\mathbb{R}^p}$ follows a univariate Gaussian distribution. Likewise, it is possible to extend the definition of Gaussian distribution to $\mathbb{H}$-valued random variables: an $\mathbb{H}$-valued random variable χ follows a Gaussian distribution if and only if, for all $f \in \mathbb{H}$, the $\mathbb{R}$-valued random variable $\langle \chi, f\rangle_{\mathbb{H}}$ follows a univariate Gaussian distribution. Under such an assumption and using the generalization of Cochran's Theorem to separable Hilbert spaces (see Lemma A.1 and Proposition A.2 in Appendix A), it can be shown that, for any fixed $f \in \mathbb{H}$:

$$n\frac{\langle f,\mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f,\mathcal{K}_n f\rangle_{\mathbb{H}}} \sim F(1, n-1)\,, \tag{11}$$

where $n\frac{\langle g,\mathcal{D}_n g\rangle_{\mathbb{H}}}{\langle g,\mathcal{K}_n g\rangle_{\mathbb{H}}}$ coincides with the square of Student's t statistic computed from the projections $\langle \chi_1, f\rangle_{\mathbb{H}}, \dots, \langle \chi_n, f\rangle_{\mathbb{H}}$.

Equation (11) provides, for a fixed $f \in \mathbb{H}$, the distribution of the ratio involved in the definition of Hotelling's T^2 statistic. However the distribution of its maximum over all elements $f \in \mathrm{Im}\,(\mathcal{K}_n) \setminus \{0\}$ (i.e., T^2) is not easy to elicit without introducing very strong assumptions on the covariance operator $\mathcal{K}$. In addition, it has already been acknowledged that multivariate normality is very difficult to assess because one should technically check the univariate normality of an infinite number of real-valued random variables. The problem persists and becomes even less tractable in infinite-dimensional separable Hilbert spaces.

For all these reasons, in the coming section, we give our take on the problem of inference on the mean element of a population or on the difference between the mean elements of two populations from a non-parametric perspective. In particular, we propose permutation-based inferential tools that rely on minimal and intuitive distributional assumptions of the data.

7. Permutation tests in separable Hilbert spaces based on Hotelling's T^2

Object-oriented non-parametric statistical inference has been addressed in the literature using either the permutation framework or the bootstrap theory (e.g., Cuevas et al., 2006; Ferraty et al., 2010). The latter does not make any assumptions on the distribution of the data but is only asymptotically valid, i.e., when the sample size n goes to infinity. Conversely, at the expense of minimal assumptions on the distribution of the data, the permutation-based approach can generate exact statistical tests even for small sample size n (e.g., Pesarin and Salmaso, 2010). Moving towards this latter direction, we hereby propose a permutation-based statistical testing procedure using Hotelling's T^2 statistic for making inference on the mean element in any separable Hilbert space.

In detail, in this section, we propose two derived inferential tools for the mean element of an $\mathbb{H}$-valued random variable (Section 7.1) and the difference between the mean elements of two $\mathbb{H}$-valued random variables (Section 7.2), respectively. Precisely, we will report the distributional assumptions, formulate the null and the alternative hypotheses, introduce the test statistic (derived from Hotelling's T^2), show how to compute the p-value, and describe the theoretical properties of the test.

7.1. One-Population Test

Distributional Assumptions. Let $\chi_1, \dots, \chi_n$ be a sample of n *i.i.d.* $\mathbb{H}$-valued random variables with mean $m \in \mathbb{H}$ and covariance operator $\mathcal{K} \in \mathcal{B}_{\mathrm{HS}}(\mathbb{H})$ such that $\chi_i = m + \varepsilon_i$. We then assume that the distribution of the χ_i's is symmetric around m (i.e., that the distribution of ε_i is the same as the one of $-1 \odot \varepsilon_i$) and that $\mathbb{E}_{\mathbb{R}}\left[\|\varepsilon_i\|^2_{\mathbb{H}}\right] < +\infty$, for all $i \in \{1, \dots, n\}$.

Null & Alternative Hypotheses. The proposed procedure aims at testing the null hypothesis $H_0 : m = m_0$ against the alternative hypothesis $H_1 : m \neq m_0$, with $m_0 \in \mathbb{H}$.

Test Statistic. Similarly to the traditional multivariate case, Hotelling's T_0^2 test statistic can be defined as:

$$T_0^2 = n \max_{f \in \mathrm{Im}(\mathcal{K}_\mathrm{n}) \setminus \{0\}} \frac{\langle f, \mathcal{D}_{n0} f \rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f \rangle_{\mathbb{H}}}, \tag{12}$$

where m_n is the sample mean, $\mathcal{K}_n$ is the sample covariance operator and $\mathcal{D}_{n0}$ is the sample mean squared-error loss operator in which the true unknown mean m is replaced by m_0: $\mathcal{D}_{n0} := (m_n \ominus m_0) \otimes_{\mathbb{H}} (m_n \ominus m_0)$. In eq. (12) the maximum is achieved for $f = \mathcal{K}_n^{+}(m_n \ominus m_0)$ which, in the case of $\mathbb{H}$-Gaussian data or large sample size, corresponds to the direction in $\mathbb{H}$ where the strongest evidence in favor of H_1 is observed.

P-value computation. Since we assume that the distribution of the χ_i's is symmetric, its center of symmetry under H_0 is m_0. Thus equally likely samples under H_0 are trivially obtained by reflecting one or more of the original realizations of $\mathbb{H}$-valued random variables with respect to m_0, i.e., replacing χ_i with χ_i^* as follows:

$$\chi_i \mapsto \chi_i^* = m_0 \oplus (-1)^{c_i} \odot (\chi_i \ominus m_0), \quad c_i \in \{0, 1\}. \tag{13}$$

The number of possible reflections (and of equally likely samples under H_0) is thus equal to 2^n, independently from the nature of $\mathbb{H}$. Pursuing the permutation-based approach (e.g., Pesarin and Salmaso, 2010), inference is carried out within the equivalence class of all samples which are equally likely to the original sample under H_0. This conditioning makes the random sample distribution under H_0 within the equivalence class (i.e., random sample permutational distribution) a discrete uniform over the 2^n elements of the equivalence class. Thus, the permutational p-value can be simply computed as the proportion of samples (among the 2^n possible ones) associated to a value of the test statistic (12) that exceeds the value associated to the original sample. Formally, it reads:

$$p_{value} = \frac{1}{2^n} \sum_{b=1}^{2^n} \mathbb{I}\left[T_0^2(\chi^*_{1_b}, \chi^*_{2_b}, \dots, \chi^*_{n_b}) \geq T_0^2(\chi_1, \chi_2, \dots, \chi_n)\right], \tag{14}$$

with b indexing the 2^n samples obtained from (13).

Exactness & Consistency of the Test. The test above is **exact**. In detail, due to the discrete distribution of p-value, exactness reads:

$$\mathbb{P}_{H_0}\left[p_{value} \leq \alpha\right] \begin{cases} = \alpha & \forall \alpha \in \mathcal{A} \\ < \alpha & \forall \alpha \in (0,1] \backslash \mathcal{A}, \end{cases}$$

where $\mathcal{A} = \{1/2^n, 2/2^n, \ldots, 2^n/2^n\}$ is the set of all attainable exact levels.

Moreover, the test is also **consistent**, which reads:

$$\lim_{n \to \infty} \mathbb{P}_{H_1}\left[p_{value} \leq \alpha\right] = 1, \quad \forall \alpha \in (0,1].$$

For the proof, see Theorem C.1 of Appendix C.

7.2. Two-Population Test

Distributional Assumptions. Let $\{\chi_{i1}\}_{i=1,\ldots,n_1}$ and $\{\chi_{i2}\}_{i=1,\ldots,n_2}$ be two independent samples of respectively n_1 and n_2 *i.i.d.* $\mathbb{H}$-valued random variables with respective means $m_1 \in \mathbb{H}$ and $m_2 \in \mathbb{H}$ and common covariance operator $\mathcal{K} \in \mathcal{B}_{\mathrm{HS}}(\mathbb{H})$, such that $\chi_{ij} = m_j + \varepsilon_{ij}$. We assume that $\mathbb{E}_{\mathbb{R}}[\|\varepsilon_{ij}\|_{\mathbb{H}}^2] < +\infty$, for all $i \in \{1, n_j\}$, $j \in \{1, 2\}$.

Null & Alternative Hypotheses. The proposed two-population statistical test aims at testing the null hypothesis of $H_0 : m_1 = m_2$ against the alternative $H_1 : m_1 \neq m_2$.

Test Statistic. Hotelling's T^2 statistic can be defined as:

$$T_0^2 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \max_{f \in \mathrm{Im}\left(\mathcal{K}_{n_{\mathrm{pooled}}}\right) \backslash \{0\}} \frac{\langle f, \mathcal{D}_{n_0} f \rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_{n_{\mathrm{pooled}}} f \rangle_{\mathbb{H}}}, \tag{15}$$

where m_{n_1} and m_{n_2} are the sample means, $\mathcal{D}_{n_0}$ is the sample mean squared-error loss operator under the null hypothesis, i.e., $\mathcal{D}_{n0} = (m_{n_1} \ominus m_{n_2}) \otimes_{\mathbb{H}} (m_{n_1} \ominus m_{n_2})$, and $\mathcal{K}_{n_{\mathrm{pooled}}}$ is the pooled sample covariance operator defined as:

$$\begin{aligned} \mathcal{K}_{n_{\mathrm{pooled}}} &: \mathbb{H} \to \mathbb{H} \\ & f \mapsto \frac{n_1 - 1}{n_1 + n_2 - 2} \odot (\mathcal{K}_{n_1} f) \oplus \frac{n_2 - 1}{n_1 + n_2 - 2} \odot (\mathcal{K}_{n_2} f) \end{aligned}$$

where $\mathcal{K}_{n_1}$ and $\mathcal{K}_{n_2}$ are the sample covariance operators of the two samples. In eq. (15) the maximum is achieved for $f = \mathcal{K}^{+}_{n_{\mathrm{pooled}}}(m_{n_1} \ominus m_{n_2})$ which, in the case of $\mathbb{H}$-Gaussian data or large sample sizes, corresponds to the direction in $\mathbb{H}$ where the strongest evidence in favor of H_1 is observed.

P-value computation. Since the two samples have the same covariance operator, under the null hypothesis they have the same distribution and are independent. They are therefore exchangeable. Thus equally likely samples under H_0 are obtained by all possible rearrangement of the values χ_{ij} across the units:

$$(\chi_{11}, ..., \chi_{n_1 1}, \chi_{12}, ..., \chi_{n_2 2}) \mapsto (\chi^*_{11}, ..., \chi^*_{n_1 1}, \chi^*_{12}, ..., \chi^*_{n_2 2}). \tag{16}$$

The number of possible rearrangements (or permutations) leading to a different allocation in the two groups is then $\binom{n_1+n_2}{n_1}$. Analogously to what is done in the one-sample case, the permutational p-value can be then computed as the proportion of samples (among the $\binom{n_1+n_2}{n_1}$ possible ones) associated to a value of the test statistic (15) that exceeds the value associated to the original sample. Formally, it reads:

$$p_{value} = \frac{1}{\binom{n_1+n_2}{n_1}} \sum_{b=1}^{\binom{n_1+n_2}{n_1}} \mathbb{I}\left[T_0^2(\chi^*_{1_b}, \chi^*_{1_b}, ..., \chi^*_{n_b}) \geq T_0^2(\chi_1, \chi_2, ..., \chi_n)\right] \tag{17}$$

with b indexing the $\binom{n_1+n_2}{n_1}$ samples obtained from (16).

Exactness & Consistency of the Test. The test above is **exact**. In detail, due to the discrete distribution of p-value, exactness reads:

$$\mathbb{P}_{H_0}\left[p_{value} \leq \alpha\right] \begin{cases} = \alpha & \forall \alpha \in \mathcal{A} \\ < \alpha & \forall \alpha \in (0,1] \backslash \mathcal{A} \end{cases}$$

where $\mathcal{A} = \{1/\binom{n_1+n_2}{n_1}, 2/\binom{n_1+n_2}{n_1}, \ldots, \binom{n_1+n_2}{n_1}/\binom{n_1+n_2}{n_1}\}$ is the set of all attainable exact levels. Moreover, the test is also **consistent**, which reads:

$$\lim_{n_1, n_2 \to \infty} \mathbb{P}_{H_1}\left[p_{value} \leq \alpha\right] = 1, \quad \forall \alpha \in (0,1].$$

For the proof, see Theorem C.2 of Appendix C.

8. Simulation study

The simulation study aims at comparing the performances of the tests proposed in Section 7 with permutation tests based on other state-of-the-art test statistics. The simulation is carried out in the space $L^2(T)$, which is undoubtedly the most used Hilbert space in the field of functional data analysis. In detail, the goal is to assess the finite-sample statistical power associated to Hotelling's T^2 statistic w.r.t. the ones associated to existing statistics in $L^2(T)$ that either fully ignore the variances and covariances in the data as does T_I^2 (eq. 7) or only account for the pointwise variance but ignore auto-correlation structure as does $T_{D_\sigma}^2$ (eq. 8). This is accomplished by generating simulated data in $L^2(T)$ under different scenarios fro the mean functions and/or the covariance operator. In detail, we hereby propose a comparison based on the statistical power of the two-population permutation tests induced by each statistic, estimated by means of Monte Carlo simulations where the same random data sets and the same random permutations (Pesarin and Salmaso, 2010) are used for all comparisons.

8.1. Data generation process

We aim at estimating the statistical power of the two-population test. We simulate $M = 10,000$ pairs of independent random samples of size $n_1 = n_2 = 20$, drawn from two Gaussian distributions on $L^2([0,1])$ with equal covariance operator $\mathcal{K}$. We purposely explore different scenarios of mean and covariance. The integrals involved in the computation of the three test statistics are obtained by using the rectangle quadrature rule based on a uniform grid of 100 points. In each scenario, we consider increasing magnitudes of the difference between the two mean functions, by varying the maximum mean difference $\Delta = \max_{t\in[0,1]} |\mu_1(t) - \mu_2(t)|$ and evaluate the rate of rejection as a function of Δ. For each randomly generated data set, we used $B = 1,000$ permutations for evaluating the p-values. In details, the proposed scenarios read:

Mean Difference Scenarios. The mean function μ_1 of the first functional data set was fixed to 0 everywhere on the interval $[0,1]$ while the mean function μ_2 of the second functional data set was designed to accommodate two scenarios of mean differences of increasing complexity:

Scenario 1. Constant mean difference over the whole domain:

$$\mu_2(t) = \Delta \quad \forall t \in [0,1],$$

Scenario 2. Sign-changing constant difference with continuously differentiable transition:

$$\mu_2(t) = \Delta \left[\mathbb{I}_{[0,\frac{2}{5}]}(t) + \cos(2\pi t - 2\pi)\mathbb{I}_{[\frac{2}{5},\frac{3}{5}]}(t) - \mathbb{I}_{[\frac{3}{5},1]}(t) \right] \quad \forall t \in [0,1];$$

Covariance Operator Scenarios. Similarly, the covariance operator was designed to simulate differentiable trajectories in the stationary and non-stationary scenarios:

Scenario A. For stationary data, we used:

$$(\mathcal{K}g)(t) = \int_0^1 e^{-2(s-t)^2} g(s)\mathrm{d}s \quad \forall t \in [0,1],$$

Scenario B. For non-stationary data, we used:

$$(\mathcal{K}g)(t) = \int_0^1 e^{-2(s-t)^2} (s+0.5)(t+0.5) g(s)\mathrm{d}s \quad \forall t \in [0,1].$$

This design of experiment led to the definition of a total of 4 mean-covariance scenarios (i.e., 1.A, 1.B, 2.A, 2.B).

8.2. Results: comparison of the statistical power

Figure 1 reports the results of the simulation study. The rows exhibit the two mean scenarios while columns distinguish the two covariance scenarios. Panels in the first row provide a visual representation of the two covariance structures in which we plot an example of randomly generated data set for the first group ($\mu_1 = 0$). Panels in the first column instead depict the mean functions of the two populations for $\Delta = 1$. The remaining panels show the estimated rate of rejection of the three tests based on T_I^2 (black), $T_{D_\sigma}^2$ (red), and T^2 (green) as a function of Δ. The left-most point in each graph corresponds to $\Delta = 0$ and thus is to be interpreted as the estimated level of the test (the nominal level was set to 5%). The remaining points are to be interpreted as estimated statistical powers.

A first consideration is that the performances of the significance test based on $T^2_{D_\sigma}$ are always close to or in-between the ones achieved by the tests based on T^2_I and T^2. This result is coherent with the theory, since the L^2 distance (which T^2_I is based on) completely disregards the covariance structure, the standardized L^2 distance (which $T^2_{D_\sigma}$ is based on) incorporates only information on the sample variance function while the functional Mahalanobis semi-distance (which T^2 is based on) fully incorporates the information from the sample covariance structure. A second consideration is that in the four scenarios the test based on $T^2_{D_\sigma}$ outperforms the one based on T^2_I. Indeed, in the stationary case, the statistical power of the test based on T^2_I and on $T^2_{D_\sigma}$ are basically identical while, in the non-stationary case, the test based on $T^2_{D_\sigma}$ achieves a statistical power larger than the test based on T^2_I. This suggests that, at least in the case of Gaussian data, using information provided by the sample variance function improves the test performances. This is consistent with the optimality of the Student's t test statistic in the one-dimensional Gaussian case.

The comparison between the tests based on T^2 and $T^2_{D_\sigma}$ is more complex and insightful. Power performances depend indeed on both the mean and the covariance functions. In all scenarios but 1.A, the test based on T^2 is more powerful than the one based on $T^2_{D_\sigma}$ (and on T^2_I as well). This fact proves that – as in the standard multivariate Euclidean case – also in the case of more complex Hilbert spaces (e.g., $L^2(T)$), it is not possible to prove that for Gaussian data the test based on Hotelling's T^2 is the uniformly most powerful one. Scenario A.1 depicts indeed a paradigmatic example in which the test based on the ergodic means of functional data (i.e., which is as a matter of fact the test based on T^2_I) is more powerful than the one based on Hotelling's T^2.

9. Case study: statistical analysis of aneurysms

This section presents a case study in which the elements of the separable Hilbert space are smooth functions. The scope is two-fold: (i) show the possible use of Hotelling's T^2 in a non-trivial real application (making also a comparison with other possible choices of the test statistic); (ii) show the importance of properly selecting the Hilbert space into which data are embedded in the light of the specific research question under investigation. In particular, we hereby perform the analysis of the Aneurisk data set described in Passerini et al. (2012). The long-term objective of the Aneurisk project (Sangalli et al., 2009, 2014) is to find predictors of the formation of a cerebral aneurysm. This boils down to eliciting biomarkers that perform aneurysmal risk assessment for the individual subject.

The Aneurisk dataset collected to answer these questions includes 65 subjects hospitalized at Ospedale Niguarda Ca' Granda Milano from September 2002 to October 2005 for suspicion of the presence of aneurysm along the internal carotid artery (ICA). The ICA is a major artery that handles blood supply to the brain. The upper part of the ICA (u-ICA) seats within the skull and provides blood to the Circle of Willis while the lower part (l-ICA) stands outside the skull and takes blood from the Common Carotid

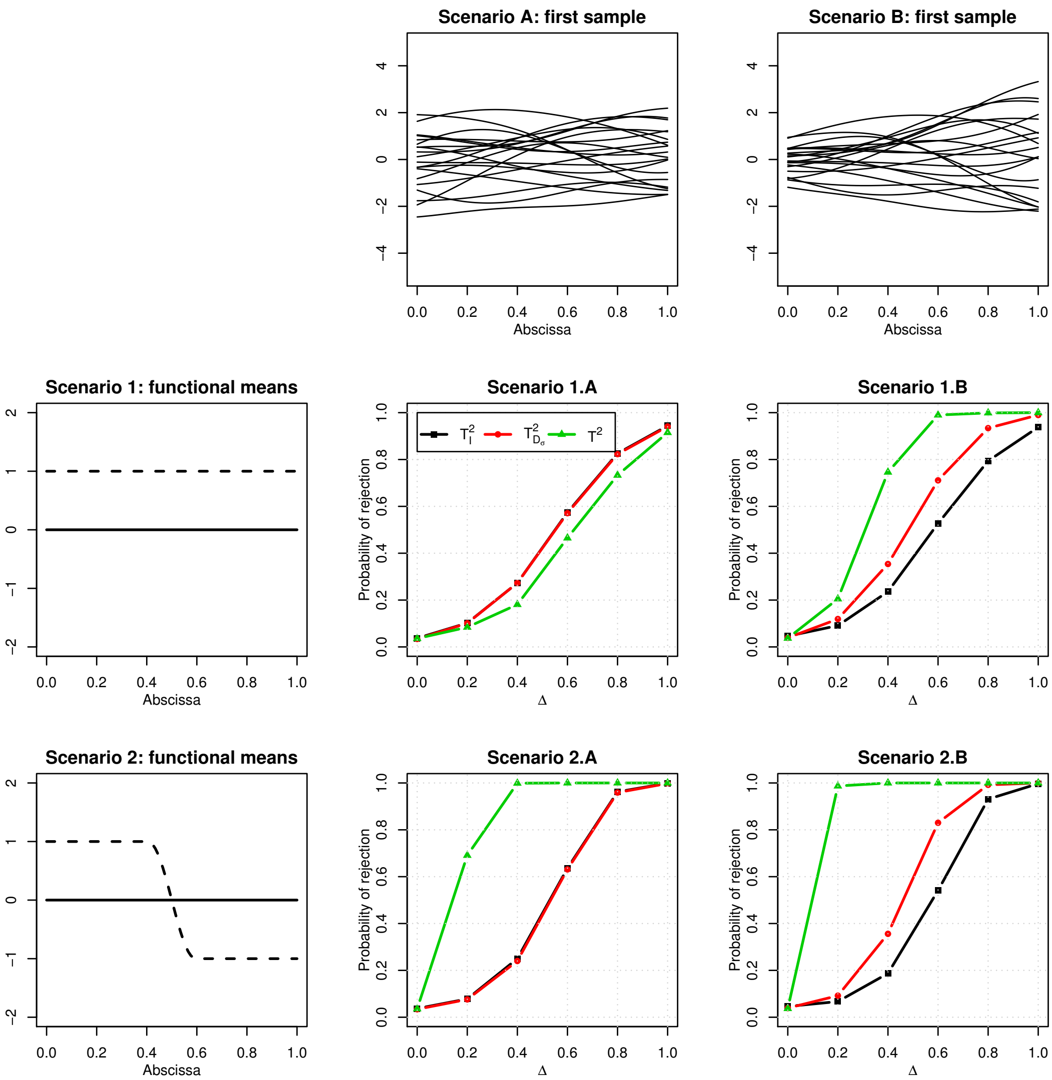

Fig. 1. Estimated Rates of Rejection. Estimated statistical powers as a function of the maximum difference Δ between the means. Left-most point shows the estimated level of each test. Compared tests are based on T_I^2 (black), $T_{D_\sigma}^2$ (red) and T^2 (green). Mean scenarios by row; covariance scenarios by column.

Artery. Among the subjects, nearly half of them had an aneurysm either in the Circle of Willis or in the u-ICA while the other half had either no aneurysm or an aneurysm in the l-ICA. This distinction in the localization of the aneurysm allows us to label subjects into two categories:

- the **High-Risk** group: when the aneurysm seats within the skull, its rupture would often be fatal leading to permanent or lethal brain tissue damage;
- the **Low-Risk** group: when there is no aneurysm or if the aneurysm seats outside the skull, its possible rupture is not directly affecting brain tissues.

In addition to associating each patient to one of these two groups, for a subset of 50 subjects, the dataset also contains a number of geometrical and hemodynamical features of the last 5 cm of the ICA that are believed to be relevant in predicting the localization of aneurysms and thus in predicting the group status (high-risk vs low-risk) of a patient (Passerini et al., 2012). In this case study, we focus on the local maximal inscribed sphere **radius** as a function of the arc-length along the ICA centerline. This feature was modelled in the literature as functional data. Details about the preprocessing (including smoothing and registration) can be found in Sangalli et al. (2009) and Passerini et al. (2012). Data pertaining to the 50 subjects used for performing our analysis are available as part of the `fdahotelling` package.

The clinical objective in this case study is to determine whether the ICA radius is a good biomarker for assessing aneurysmal localization. This goal is however not precise enough for designing an appropriate statistical test that could be helpful. In effect, assume that the ICA radius somehow discriminates high-risk from low-risk patients. It might be because high-risk subjects have a larger radius in the last 5 cm of the ICA w.r.t. low-risk subjects that leads to an increased stress on the Circle of Willis and on the u-ICA, subsequently facilitating the formation of the aneurysm there (**Clinical Question 1**). However, it might instead be that not enough spatial *variation* in radius along the last 5 cm of the ICA, i.e. not enough wiggliness, prevents the blood flow from being slowed down at the entrance of the Circle of Willis (**Clinical Question 2**). Statistically, Clinical Question 1 can be tested by analyzing the radius functions themselves, which boils down to immersing the sample of radius curves into the Hilbert space $L^2(T)$ endowed with its usual inner product. Differently, Clinical Question 2 requires to work on radius derivatives instead, which boils down to immersing the sample of radius curves into the semi Hilbert space $\widetilde{H}^1(T)$ of real-valued functions with squared integrable first derivatives endowed with the inner product $\int_T f'(t)g'(t)\mathrm{d}t$ (see Remark 3.8). In this application, the compact set T is an interval of $\mathbb{R}$ mapping the last 5 cm of the ICA. Figure 2 exhibits the sample of radius curves (left) and their first derivatives (right).

We applied the permutation test for two populations presented in Section 7.2 in the separable Hilbert space $L^2(T)$ for answering Clinical Question 1 and in the separable semi-Hilbert space $\widetilde{H}^1(T)$ for answering Clinical Question 2 using Hotelling's T^2 statistic from eq. (15). We also performed the same permutation test using the L^2 and standardized L^2 distances T_I^2 and $T_{D_\sigma}^2$ introduced in Hall and Tajvidi (2002) and recalled in eqs. (7) and (8), respectively (adapted for two populations). Furthermore, we applied eqs. (7) and (8) to the sample of radius derivatives, which effectively defines

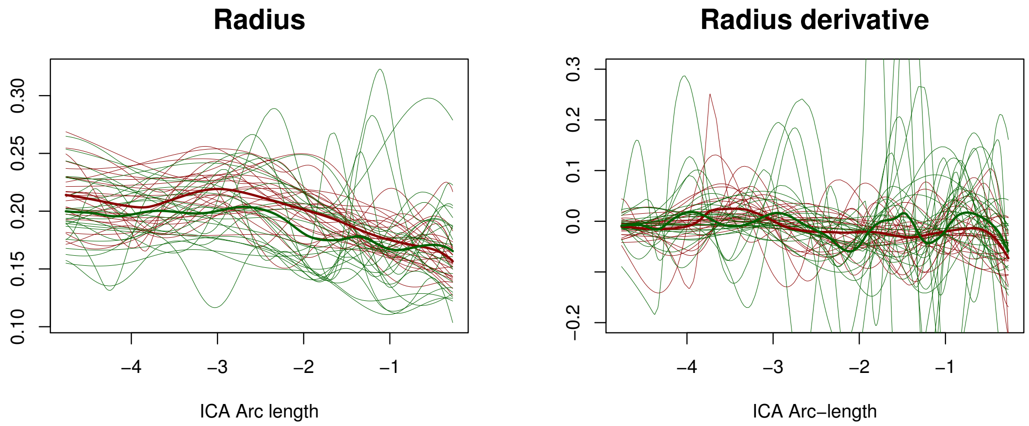

Fig. 2. The Aneurisk dataset. The sample of radius curves (left) and their first derivatives (right). Green (resp. red) curves correspond to the low-risk (resp. high-risk) group. Mean curves are displayed with a bold line.

the statistics T_I^2 and $T_{D_\sigma}^2$ in the Hilbert space $\widetilde{H}^1(T)$ and allows us to compare them to Hotelling's T^2 statistic in that space as well. In summary, we performed a total of six permutation tests (three different test statistics, two different Hilbert spaces). For each of them, we report the p-value of the test as defined in eq. (17), estimated by means of a conditional Monte-Carlo algorithm (Pesarin and Salmaso, 2010) using $B = 1,000$ random permutations. The same permutations were used for all tests in order to provide a fair comparison between test statistics and Hilbert spaces.

We summarise the output of the six tests (which are exact, see Section C) by means of their p-values in Table 4. The first observation that we can make is that the same inferential approach (e.g., a permutation test based on T^2) can be used to answer different research questions by simply changing the Hilbert space into which immersing the data. For example, focusing on Hotelling's T^2 statistic, the radius of the last 5 cm of the ICA is not found to be significantly different between high-risk and low-risk patients when the analysis is carried out in $L^2(T)$ but it is when the analysis is carried out in $\widetilde{H}^1(T)$. This has critical clinical implications in terms of interpretation. Indeed, it suggests that the radius itself is not a good biomarker for discriminating high- from low-risk patients but its derivative is, which means that the data better supports Clinical Question 2 w.r.t. Clinical Question 1. This is a message of paramount importance for any statistician who deals with complex data for which a unique natural Hilbert space into which immersing the data can hardly be identified. In effect, this simple example shows that the choice of the space in which data is embedded plays a key role in the statistical analysis and thus it deserves a very careful attention and should not be neglected.

A second observation from the results in Table 4 is that, for both $L^2(T)$ and $\widetilde{H}^1(T)$, the p-values of tests based on T_I^2 are larger than the ones based on $T_{D_\sigma}^2$. This fact is in line with what we observed in the simulation study, i.e., that the permutation test that relies on the standardized distance between the means is equally or more powerful than the one relying on the non-standardized one. As confirmed also by simulations,

Table 4. Statistical tests for the difference between high-risk and low-risk groups. For each clinical question, data is embedded into a specific Hilbert space for properly answering that question and the p-value of the two-population permutation test for the difference between the mean functions of the high- and low-risk groups is reported, using three different test statistics: the simple distance between the means in the corresponding Hilbert space, the same distance standardized using the pointwise variance in that space and our proposed Hotelling's T^2 statistic.

Clinical Question	Hilbert Space	Statistic		
		T_I^2	$T_{D_\sigma}^2$	T^2
1	$L^2(T)$	0.034	0.025	0.902
2	$\widetilde{H^1}(T)$	0.115	0.019	0.012

the comparison with Hotelling's T^2 statistic is not that irrevocable. The comparison of the p-values associated to T_I^2 and T^2 is paradigmatic in this sense. Indeed, from Table 4, we observe that the permutation tests with Hotelling's T^2 statistic detect significant differences in the mean radius derivatives but does not reject the null hypothesis of equal mean radius curves themselves. The contrary is instead observed for the permutation tests with T_I^2 statistic. These opposite conclusions are consistent with simulations that show that neither of the two tests is uniformly most powerful, and that the test statistic T_I^2 seems to be more powerful in detecting simple violations of H_0 (e.g., a simple vertical shift of the two populations) while Hotelling's T^2 is more powerful when "more complex" violation of H_0 take place.

10. Conclusions

In this manuscript, we provide a (re)definition of Hotelling's T^2 statistic that shows that, contrary to common belief, it can actually be defined in any separable Hilbert spaces of any dimension. We believe that this definition (Definition 3.7) could replace the traditional but misleading form – eq. (1) – of Hotelling's T^2 statistic, only valid in $\mathbb{R}^p$ for $p < n$. We showcase, as illustrative specifications, three particular classes of Hilbert spaces widely used in practice: (i) $\mathbb{R}^p$ in which all problems of multivariate statistical analysis are typically carried out, (ii) Sobolev spaces where most problems pertaining to Functional Data are settled and (iii) Bayes spaces, which are convenient to handle statistical analysis of continuous probability density functions. We provide a key theorem (Theorem 5.1) for the practical computation of Hotelling's T^2 statistic in any separable Hilbert space. Finally, we design a null hypothesis significance test for the mean element of an $\mathbb{H}$-valued random variable (for the difference of mean elements of two $\mathbb{H}$-valued random variables) embedding Hotelling's T^2 statistic within a permutation testing framework. We provide proofs of exactness and consistency of the proposed inferential procedures in any separable Hilbert space.

We report a simulation study to compare - within the permutation testing procedure - the statistical power of the test based on Hotelling's T^2 with the one of other state-of-the-art test statistics proposed for performing inference on the mean of functional data defined in the $L^2(T)$ space. Results suggest that, in the presence of a non-trivial difference between the means, an inference based on Hotelling's T^2 statistic provides larger statistical power.

We subsequently conduct a case study on the Aneurisk data set – containing functions describing the geometry of arteries – to answer the question of whether the radius of an artery matters for predicting the location of a possible aneurysm. The aims of this case study were two-fold: (i) to compare the inference produced by the use of Hotelling's T^2 statistic within the permutation testing procedure with the inference produced by the use of other statistics previously proposed in the literature and (ii) to assess the impact of the Hilbert space chosen to embed the data into. Results of the comparison between the different statistics are in line with what we already observed in the simulation study. More importantly perhaps, for a given test statistic, results can be so different when data is embedded in different spaces as to provide opposite conclusions w.r.t. rejection of the corresponding null hypothesis, which puts the choice of the space into which data is analyzed as a key component of the statistical analysis itself.

All the analyses were performed using our R package `fdahotelling` that implements permutation test and power calculation functions in $L^2(T)$ for a number of statistics including Hotelling's T^2. In details, seven test statistics are implemented in the package provided as supplementary material and the user can also implement its own test statistic and plug it into the test function.

An interesting and challenging future development of this work pertains to the extension of Hotelling's T^2 statistic to the larger family of metric 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 and the discussion therein). This extension would require a purely metrical definition of T^2 that neither relies on the notion of inner product nor on the one of vector space.

References

Bongiorno, E. G., Salinelli, E., Goia, A. and Vieu, P. (2014) *Contributions in infinite-dimensional statistics and related topics.* Società Editrice Esculapio.

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

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

Corain, L., Melas, V. B., Pepelyshev, A. and Salmaso, L. (2014) New insights on permutation approach for hypothesis testing on functional data. *Advances in Data Analysis and Classification*, **8**, 339–356.

Cox, D. D. and Lee, J. S. (2008) Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika*, **95**, 621–634.

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

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

— (2006) On the use of the bootstrap for estimating functions with functional data. *Computational statistics & data analysis*, **51**, 1063–1074.

Dalla Rosa, M., Sangalli, L. M. and Vantini, S. (2014) Principal differential analysis of the aneurisk65 data set. *Advances in Data Analysis and Classification*, **8**, 287–302.

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**, 1175–1182.

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

Ferraty, F., Van Keilegom, I. and Vieu, P. (2010) On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*, **37**, 286–306.

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

Goia, A. and Vieu, P. (2016) An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, **146**, 1–6. Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.

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

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

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

Hron, K., Menafoglio, A., Templ, M., Hrzov, K. and Filzmoser, P. (2016) Simplicial principal component analysis for density functions in bayes spaces. *Computational Statistics & Data Analysis*, **94**, 330–350.

Hsing, T. and Eubank, R. (2015) *Theoretical foundations of functional data analysis, with an introduction to linear operators.* John Wiley & Sons.

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

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

Marron, J. S. and Alonso, A. M. (2014) Overview of object oriented data analysis. *Biometrical Journal*, **56**, 732–753.

Menafoglio, A., Guadagnini, A. and Secchi, P. (2014) A kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. *Stochastic environmental research and risk assessment*, **28**, 1835–1851.

Menafoglio, A. and Petris, G. (2016) Kriging for Hilbert-space valued random fields: The operatorial point of view. *Journal of Multivariate Analysis*, **146**, 84–94.

Menafoglio, A. and Secchi, P. (2016) Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research*, **258**, 401–410.

Passerini, T., Sangalli, L. M., Vantini, S., Piccinelli, M., Bacigaluppi, S., Antiga, L., Boccardi, E., Secchi, P. and Veneziani, A. (2012) An integrated statistical investigation of internal carotid arteries of patients affected by cerebral aneurysms. *Cardiovascular Engineering and Technology*, **3**, 26–40.

Pesarin, F. and Salmaso, L. (2010) *Permutation tests for complex data: theory, applications and software.* John Wiley & Sons Inc, Chichester.

Pini, A. and Vantini, S. (2016) The interval testing procedure: A general framework for inference in functional data analysis. *Biometrics*, **72**, 835–845.

— (2017) Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, in press.

Pini, A., Vantini, S., Colosimo, B. M. and Grasso, M. (2017) Domain-selective functional ANOVA for supervised statistical profile monitoring of signal data. *Journal of the Royal Statictical Society Series C (Applied Statistics)*, in press.

Poyton, A. A., Varziri, M. S., McAuley, K. B., McLellan, P. J. and Ramsay, J. O. (2006) Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & chemical engineering*, **30**, 698–708.

Ramsay, J. O. and Dalzell, C. J. (1991) Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**, 539–572.

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

— (2005) *Functional data analysis.* Springer, New York.

Sangalli, L. M., Secchi, P. and Vantini, S. (2014) Aneurisk65: A dataset of three-dimensional cerebral vascular geometries. *Electronic Journal of Statistics*, **8**, 1879–1890.

Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009) A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery. *Journal of the American Statistical Association*, **104**, 37–48.

Schott, J. R. (2007) Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis*, **98**, 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.

Shen, Q. and Faraway, J. (2004) An f test for linear models with functional responses. *Statistica Sinica*, **14**, 1239–1257.

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**, 263–272.

Srivastava, M. S. (2007) Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, **37**, 53–86.

A. Cochran's Theorem in separable Hilbert spaces

LEMMA A.1. *Let $\chi_1, \ldots, \chi_n$ be a sample of n i.i.d. $\mathbb{H}$-valued random variables with mean m and covariance operator $\mathcal{K}$ s.t. $\mathbb{E}\left[\|\chi_i\|^2_{\mathbb{H}}\right] < +\infty$. Let $\mathcal{K}_n$ and $\mathcal{D}_n$ be the sample covariance and sample mean squared-error loss operators respectively and $\mathcal{V}_n := \bigoplus_{i=1}^n (\chi_i \ominus m) \otimes_{\mathbb{H}} (\chi_i \ominus m)$. The following variance decomposition holds:*

$$(n-1) \odot \mathcal{K}_n + n \odot \mathcal{D}_n = \mathcal{V}_n,$$

or, equivalently, $\forall f \in \mathbb{H}$:

$$(n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} + n\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}} = \langle f, \mathcal{V}_n f\rangle_{\mathbb{H}}.$$

PROOF. Note that, by their definition, the 3 operators $\mathcal{K}_n$, $\mathcal{D}_n$, and $\mathcal{V}_n$ have respectively $n-1$, 1 and n degrees of freedom. Moreover, $\forall f \in \mathbb{H}$:

$$\begin{aligned}
(n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} &= \langle f, \bigoplus_{i=1}^{n} \langle \chi_i \ominus m_n, f\rangle_{\mathbb{H}} \odot (\chi_i \ominus m_n)\rangle_{\mathbb{H}} \\
&= \sum_{i=1}^{n} \langle \chi_i \ominus m_n, f\rangle^2_{\mathbb{H}}; \\
\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}} &= \langle f, \langle m_n \ominus m, f\rangle_{\mathbb{H}} \odot (m_n \ominus m)\rangle_{\mathbb{H}} \\
&= \langle m_n \ominus m, f\rangle^2_{\mathbb{H}}.
\end{aligned}$$

Finally:

$$\begin{aligned}
\langle f, \mathcal{V}_n f\rangle_{\mathbb{H}} &= \langle f, \bigoplus_{i=1}^{n} \langle \chi_i \ominus m, f\rangle_{\mathbb{H}} \odot (\chi_i \ominus m)\rangle_{\mathbb{H}} \\
&= \sum_{i=1}^{n} \langle \chi_i \ominus m, f\rangle_{\mathbb{H}}^2 \\
&= \sum_{i=1}^{n} \langle \chi_i \ominus m_n \oplus m_n \ominus m, f\rangle_{\mathbb{H}}^2 \\
&= \sum_{i=1}^{n} \left(\langle \chi_i \ominus m_n, f\rangle_{\mathbb{H}}^2 + 2\langle \chi_i \ominus m_n, f\rangle_{\mathbb{H}} \langle m_n \ominus m, f\rangle_{\mathbb{H}} + \langle m_n \ominus m, f\rangle_{\mathbb{H}}^2 \right) \\
&= (n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} + 2\langle m_n \ominus m, f\rangle_{\mathbb{H}} \langle \bigoplus_{i=1}^{n} (\chi_i \ominus m_n), f\rangle_{\mathbb{H}} + n\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}} \\
&= (n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} + n\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}.
\end{aligned}$$

PROPOSITION A.2. *Let $\chi_1, ..., \chi_n$ be a sample of n i.i.d. $\mathbb{H}$-valued Gaussian random variables with mean m and covariance operator $\mathcal{K}$ s.t. $\mathbb{E}\left[\|\chi_i\|_{\mathbb{H}}^2\right] < +\infty$ and $\mathrm{Im}(\mathcal{K}) = \mathbb{H}$. The Gaussianity in $\mathbb{H}$ is intended as the Gaussianity in $\mathbb{R}$ of all the possible projections of the $\mathbb{H}$-valued random variable, i.e. $\forall f \in \mathbb{H}$, $\langle \chi, f\rangle_{\mathbb{H}}$ is a real-valued Gaussian random variable. Then, we have:*

$$n\frac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}} \sim F(1, n-1) \quad \forall f \in \mathbb{H}.$$

PROOF. Let $f \in \mathbb{H}$ and $\mathcal{V}_n := \bigoplus_{i=1}^{n} (\chi_i \ominus m) \otimes_{\mathbb{H}} (\chi_i \ominus m)$. We can write:

$$\langle f, \mathcal{V}_n f\rangle_{\mathbb{H}} = \sum_{i=1}^{n} \langle \chi_i \ominus m, f\rangle_{\mathbb{H}}^2.$$

Since χ_i is an $\mathbb{H}$-valued Gaussian random variable with mean m and covariance operator $\mathcal{K}$, then the random variables

$$\langle \chi_i \ominus m, f\rangle_{\mathbb{H}}^2, \quad i = 1, \ldots, n$$

are *i.i.d.* real-valued Gaussian random variables with zero mean and variance $\langle f, \mathcal{K} f\rangle_{\mathbb{H}}$. Hence we have

$$\langle f, \mathcal{V}_n f\rangle_{\mathbb{H}} \sim \langle f, \mathcal{K} f\rangle_{\mathbb{H}} \chi^2(n).$$

Similar arguments also lead to

$$n\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}} \sim \langle f, \mathcal{K} f\rangle_{\mathbb{H}} \chi^2(1).$$

Using the decomposition of Lemma A.1, we can thus apply Cochran's theorem in $\mathbb{R}$ (Johnson and Wichern, 2007) to get:

(a) $(n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} \sim \langle f, \mathcal{K} f\rangle_{\mathbb{H}} \chi^2(n-1)$,

(b) $n\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}$ and $(n-1)\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}$ are independent.

Point 1 implies in particular that $\mathbb{P}\left[\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}} > 0\right] = 1$. Hence, the ratio in the Proposition is well-defined. Its distribution is the ratio of two independent χ^2 distributions, i.e., a Fisher distribution.

B. Finite approximation of Hotelling's T^2

Lemma B.1 (Extended Maximization Lemma.). *Let $h \in \mathbb{H}$ be an element of the Hilbert space $\mathbb{H}$ and $\mathcal{K} \in \mathcal{B}_{\mathrm{HS}}(\mathbb{H})$ be a positive semi-definite self-adjoint Hilbert-Schmidt operator on $\mathbb{H}$. Then:*

$$\max_{x \in \mathrm{Im}(\mathcal{K}) \setminus \{0\}} \frac{\langle x, h\rangle_{\mathbb{H}}^2}{\langle x, \mathcal{K}x\rangle_{\mathbb{H}}} = \langle h, \mathcal{K}^+ h\rangle_{\mathbb{H}},$$

where $\mathcal{K}^+$ is the Moore-Penrose generalized inverse of the operator $\mathcal{K}$. In addition, the maximum is reached for the direction $x^\star = \mathcal{K}^+ h$.

Proof. Denote by $r \leq \infty$ the rank of the operator $\mathcal{K}$. Let $\{\lambda_i, e_i\}_{i=1,\dots,r}$ be the r pairs of positive eigenvalues and corresponding eigen-elements of $\mathcal{K}$. Let us introduce the operators $\mathcal{K}^{1/2}$ and $\mathcal{K}^{-1/2}$ such that, for all $x \in \mathbb{H}$:

$$\mathcal{K}^{1/2}x := \bigoplus_{i=1}^{r} \lambda_i^{1/2} \odot_{\mathbb{H}} \left[(e_i \otimes_{\mathbb{H}} e_i)x\right] = \bigoplus_{i=1}^{r} \left[\lambda_i^{1/2} \langle e_i, x\rangle_{\mathbb{H}}\right] \odot_{\mathbb{H}} e_i$$
$$\mathcal{K}^{-1/2}x := \bigoplus_{i=1}^{r} \lambda_i^{-1/2} \odot_{\mathbb{H}} (e_i \otimes_{\mathbb{H}} e_i)x = \bigoplus_{i=1}^{r} \left[\lambda_i^{-1/2} \langle e_i, x\rangle_{\mathbb{H}}\right] \odot_{\mathbb{H}} e_i.$$

Observe that, for all $x \in \mathbb{H}$, we have:

$$\mathcal{K}x = \mathcal{K}^{1/2}\left(\mathcal{K}^{1/2}x\right), \quad \mathcal{K}^+ x = \mathcal{K}^{-1/2}\left(\mathcal{K}^{-1/2}x\right),$$
$$\mathcal{K}^{1/2}\left(\mathcal{K}^{-1/2}x\right) = x \ominus \bigoplus_{i=r+1}^{\infty} (e_i \otimes e_i)x, \quad \mathcal{K}^+\left(\mathcal{K}\left(\mathcal{K}^+ x\right)\right) = \mathcal{K}^+ x,$$

where $\{e_i\}_{i>r}$ are the eigen-elements of $\mathcal{K}$ that span its kernel. Hence, for any $x \in \mathbb{H}$:

$$\begin{aligned}
\langle x, h\rangle_{\mathbb{H}} &= \left\langle x, \mathcal{K}^{1/2}\left(\mathcal{K}^{-1/2}h\right) \oplus \bigoplus_{i=r+1}^{\infty} (e_i \otimes e_i)h \right\rangle_{\mathbb{H}} \\
&= \left\langle x, \mathcal{K}^{1/2}\left(\mathcal{K}^{-1/2}h\right)\right\rangle_{\mathbb{H}} + \sum_{i=r+1}^{\infty} \left\langle x, (e_i \otimes e_i)h\right\rangle_{\mathbb{H}} \\
&= \langle \mathcal{K}^{1/2}x, \mathcal{K}^{-1/2}h\rangle_{\mathbb{H}} + \sum_{i=r+1}^{\infty} \langle e_i, x\rangle_{\mathbb{H}} \langle e_i, h\rangle_{\mathbb{H}}.
\end{aligned}$$

Now, if $x \in \mathrm{Im}(\mathcal{K}) \setminus \{0\}$, this expression reduces to $\langle x, h\rangle_{\mathbb{H}} = \langle \mathcal{K}^{1/2}x, \mathcal{K}^{-1/2}h\rangle_{\mathbb{H}}$, which, thanks to Cauchy-Schwarz inequality, yields $\langle x, h\rangle_{\mathbb{H}}^2 \leq \langle x, \mathcal{K}x\rangle_{\mathbb{H}}\langle h, \mathcal{K}^+h\rangle_{\mathbb{H}}$. We thus proved the following inequality:

$$\frac{\langle x, h\rangle_{\mathbb{H}}^2}{\langle x, \mathcal{K}x\rangle_{\mathbb{H}}} \leq \langle h, \mathcal{K}^+h\rangle_{\mathbb{H}} \quad \forall x \in \mathrm{Im}(\mathcal{K}) \setminus \{0\}.$$

Hence, the ratio on the left handside of the above inequality is upper-bounded for $x \in \mathrm{Im}(\mathcal{K}) \setminus \{0\}$ and the bound is reached for $x^\star \propto \mathcal{K}^+h$.

Proof (Theorem 5.1). For sake of clarity, recall that Hotelling's T^2 is defined as:

$$T^2 := n \max_{f\in\mathrm{Im}(\mathcal{K}_n)\setminus\{0\}} \frac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}}.$$

Let $\{V_p\}_{p\geq 1}$ be a sequence of subspaces of $\mathbb{H}$, with $\dim(V_p) = p$, that spans $\mathbb{H}$ when p goes to infinity, i.e., such that:

$$\lim_{p\to\infty} \inf_{w_p\in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0, \quad \forall h \in \mathbb{H}. \tag{18}$$

Let now define the sequence of statistics $\{T_p^2\}_{p\geq 1}$ such that:

$$T_p^2 := n \max_{f\in\mathrm{Im}(\mathcal{K}_n)\cap V_p\setminus\{0\}} \frac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}}.$$

The goal of this proof is two-fold: (i) to show that the sequence of T_p^2 statistics converges almost surely to T^2 and (ii) that, for a fixed $p \geq 1$, the statistic T_p^2 can be written in the usual matrix form and thus can be straightforwardly computed in practice.

Part I: Convergence of T_p^2 to T^2. Let us start by defining the functional J as follows:

$$\begin{array}{rccl} J: & \mathrm{Im}(\mathcal{K}_n) \setminus \{0\} & \to & \mathbb{R}^+ \\ & f & \mapsto & \dfrac{\langle f, \mathcal{D}_n f\rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f\rangle_{\mathbb{H}}}. \end{array}$$

Let $g \in \mathrm{Im}(\mathcal{K}_n) \setminus \{0\}$ be an element that maximizes J, i.e., such that $J(g) = T^2$. Thanks to eq. (18), it is possible to find a sequence $\{f_p\}_{p\geq 1}$, such that $f_p \in V_p$ for any $p \geq 1$ and

$$\lim_{p\to\infty} \|g \ominus f_p\|_{\mathbb{H}} = 0. \tag{19}$$

Now, since $V_p \subset \mathbb{H}$, there exist $g_p \in \mathrm{Im}(\mathcal{K}_n) \cap V_p \setminus \{0\}$ and $k_p \in \mathrm{Ker}(\mathcal{K}_n) \cap V_p$ such that $f_p = g_p \oplus k_p$. We can thus write:

$$\|g \ominus g_p\|_{\mathbb{H}}^2 = \langle g \ominus g_p, g \ominus g_p\rangle_{\mathbb{H}} = \langle g \ominus f_p, g \ominus g_p\rangle_{\mathbb{H}} + \langle k_p, g \ominus g_p\rangle_{\mathbb{H}}$$

Since both g and g_p belong to $\mathrm{Im}(\mathcal{K}_n) \setminus \{0\}$ while $k_p \in \mathrm{Ker}(\mathcal{K}_n)$, we have that $\langle k_p, g \ominus g_p \rangle_{\mathbb{H}} = 0$, which leaves us with:

$$\|g \ominus g_p\|_{\mathbb{H}}^2 = \langle g \ominus f_p, g \ominus g_p \rangle_{\mathbb{H}} \leq \|g \ominus f_p\|_{\mathbb{H}} \|g \ominus g_p\|_{\mathbb{H}} \quad (\text{Cauchy} - \text{Schwarz}).$$

Hence, provided that $g \neq g_p$ (which can be assumed without loss of generality, otherwise the proof is trivial), we obtain:

$$\|g \ominus g_p\|_{\mathbb{H}} \leq \|g \ominus f_p\|_{\mathbb{H}},$$

which, combined with eq. (19), guarantees that:

$$\lim_{p \to \infty} \|g \ominus g_p\|_{\mathbb{H}} = 0.$$

Now, let $W_p^2 := J(g_p)$. Invoking the continuity of J over $\mathrm{Im}(\mathcal{K}_n) \setminus \{0\}$ leads to:

$$\lim_{p \to \infty} W_p^2 = T^2.$$

Finally, by definition of T_p^2 and T^2 as maxima of J over increasing spaces, we have:

$$W_p^2 \leq T_p^2 \leq T^2,$$

which guarantees that $\lim_{p \to \infty} T_p^2 = T^2$.

Part II: Matrix form of T_p^2. Observe first that $\mathbb{H}$ can be viewed as the direct sum of V_p and its orthogonal space $V_p^{\perp}$. This means that any element $f \in \mathbb{H}$ can be written as $f = v_p \oplus z_p$, where $v_p \in V_p$, $z_p \in V_p^{\perp}$ and $\langle v_p, z_p \rangle_{\mathbb{H}} = 0$. In particular, it is possible to find a set of basis elements $\{e_k\}_{k \geq 1}$ for $\mathbb{H}$ such that $e_k \in V_p$ for $k \leq p$ and $e_k \in V_p^{\perp}$ for $k > p$. Then, we can write the decomposition of the different elements involved in the definition of T_p^2 as follows:

$$f = \bigoplus_{k=1}^{p} f_k \odot e_k, \quad \chi_i = \bigoplus_{k=1}^{\infty} \chi_{ik} \odot e_k,$$
$$m_n = \bigoplus_{k=1}^{\infty} m_{nk} \odot e_k, \quad m = \bigoplus_{k=1}^{\infty} m_k \odot e_k,$$

where the decomposition of f is truncated to the first p basis elements since $f \in V_p$. Let us now first rewrite the inner product $\langle f, \mathcal{D}_n f \rangle_{\mathbb{H}}$:

$$\begin{aligned} \langle f, \mathcal{D}_n f \rangle_{\mathbb{H}} = \langle m_n - m, f \rangle_{\mathbb{H}}^2 &= \left\langle \bigoplus_{k=1}^{\infty} (m_{nk} - m_k) \odot e_k, \bigoplus_{\ell=1}^{p} f_\ell \odot e_\ell \right\rangle_{\mathbb{H}}^2 \\ &= \left(\sum_{k=1}^{p} \sum_{\ell=1}^{p} (m_{nk} - m_k) f_\ell \langle e_k, e_\ell \rangle_{\mathbb{H}} \right)^2, \end{aligned}$$

where the last equality holds since $\langle e_k, e_l \rangle_{\mathbb{H}} = 0$ if $(e_k, e_\ell) \in V_p \times V_p^{\perp}$ or if $(e_k, e_\ell) \in V_p^{\perp} \times V_p$. Now, define the $p \times p$ real-valued symmetric invertible matrix W such that

$W_{k\ell} := \langle e_k, e_\ell \rangle_{\mathbb{H}}$. Furthermore, let $\boldsymbol{f} := (f_1, \ldots, f_p)^\top$, $\boldsymbol{m}_n := (m_{n1}, \ldots, m_{np})^\top$ and $\boldsymbol{m} := (m_1, \ldots, m_p)^\top$. We obtain:

$$\langle f, \mathcal{D}_n f \rangle_{\mathbb{H}} = \left[(\boldsymbol{m}_n - \boldsymbol{m})^\top W \boldsymbol{f}\right]^2.$$

Similar algebraic calculations lead to:

$$\langle f, \mathcal{K}_n f \rangle_{\mathbb{H}} = \boldsymbol{f}^\top W \Sigma_n W \boldsymbol{f},$$

where

$$\Sigma_n = \frac{1}{n-1} \sum_{i=1}^n (\boldsymbol{\chi}_i - \boldsymbol{m}_n)(\boldsymbol{\chi}_i - \boldsymbol{m}_n)^\top \quad \text{with} \quad \boldsymbol{\chi}_i := (\chi_{i1}, \ldots, \chi_{ip})^\top.$$

Thus, the statistic T_p^2 can be written as follows:

$$T_p^2 = n \max_{\boldsymbol{f} \in \mathrm{Im}(W \Sigma_n W) \setminus \{0\}} \frac{\left[(\boldsymbol{m}_n - \boldsymbol{m})^\top W^{1/2} W^{1/2} \boldsymbol{f}\right]^2}{\boldsymbol{f}^\top W^{1/2} W^{1/2} \Sigma_n W^{1/2} W^{1/2} \boldsymbol{f}}.$$

If we now operate the change of variables $\boldsymbol{g} = W^{1/2} \boldsymbol{f}$, we get:

$$T_p^2 = n \max_{\boldsymbol{g} \in \mathrm{Im}(W^{1/2} \Sigma_n W^{1/2}) \setminus \{0\}} \frac{\left[(\boldsymbol{m}_n - \boldsymbol{m})^\top W^{1/2} \boldsymbol{g}\right]^2}{\boldsymbol{g}^\top \left(W^{1/2} \Sigma_n W^{1/2}\right) \boldsymbol{g}}.$$

Lemma B.1 applied to the operator $W^{1/2} \Sigma_n W^{1/2}$ – which is a positive semi-definite self-adjoint operator – yields:

$$T_p^2 = n(\boldsymbol{m}_n - \boldsymbol{m})^\top W^{1/2} \left(W^{1/2} \Sigma_n W^{1/2}\right)^+ W^{1/2} (\boldsymbol{m}_n - \boldsymbol{m}),$$

which ends the proof.

C. Exactness and consistency of permutation tests

THEOREM C.1. *Consider one sample $\chi_i = m \oplus \varepsilon_i$, $i = 1, \ldots, n$ embedded in a separable Hilbert space $\mathbb{H}$, where $m \in H$ is a fixed element and $\varepsilon_i \in \mathbb{H}$ are i.i.d. random elements with zero mean and covariance operator $\mathcal{K}$ satisfying $\mathbb{E}_{\mathbb{R}}[\|\varepsilon_i\|_{\mathbb{H}}^2] < +\infty$, for all $i \in \{1, n\}$. Assume that the distribution of ε_i is symmetric $\forall i \in \{1, \ldots, n\}$.*
*The permutation test of hypotheses $H_0 : m = m_0$ against $H_1 : m \neq m_0$ based on statistic (12) and permutations (13) is **exact**. Namely, let $\mathcal{A} = \{1/2^n, 2/2^n, \ldots, 2^n/2^n\}$ be the set of all attainable levels, and let p_{value} be the p-value of the test. Then:*

$$\mathbb{P}_{H_0}\left[p_{value} \leq \alpha\right] \begin{cases} = \alpha & \forall \alpha \in \mathcal{A} \\ < \alpha & \forall \alpha \in [0,1] \setminus \mathcal{A}. \end{cases}$$

*The same test is **consistent**, i.e., $\forall \alpha \in (0, 1]$:*

$$\lim_{n \to \infty} \mathbb{P}_{H_1}\left[p_{value} \leq \alpha\right] = 1.$$

PROOF. The permutation test is based on transformations $\chi_i \mapsto \chi_i^* = m_0 \oplus (-1)^{c_i} \odot (\chi_i \ominus m_0)$, with $i = 1, ..., n$, and $c_i \in \{0, 1\}$. Under the assumption of a symmetric distribution of χ_i, $i = 1, ..., n$, and under $H_0 : m = m_0$, transformations $(\chi_1^*(t), \dots, \chi_n^*(t))$ of the data set $(\chi_1(t), \dots, \chi_n(t))$ are likelihood-invariant. This means that under H_0 the conditional distribution of the test statistic is a discrete uniform. The p-value defined in equation (14) by means of the counting measure gives the probability - under H_0 - that the test statistic is greater or equal to its value on the observed data. Hence, the permutation test is **exact** (see Prop. 2, 3.1.1 of Pesarin and Salmaso, 2010).
For proving consistency, consider a sequence of subspaces $\{V_p\}_{p \geq 1}$ such that $V_p \subset \mathbb{H}$, $\dim(V_p) = p$, and $\lim_{p\to\infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0$ for all $h \in \mathbb{H}$. Let $(e_1, \dots, e_p)$ be a basis set of V_p and let W be the symmetric invertible $p \times p$ matrix such that $W_{jk} = \langle e_j, e_k \rangle_{\mathbb{H}}$. Let T^2_{0p}, $p > 0$ denote the finite-dimensional approximation of T_0^2:

$$T^2_{0p} = n(\boldsymbol{m}_n - \boldsymbol{m}_0)^\top W^{1/2} (W^{1/2} \Sigma_n W^{1/2})^+ W^{1/2} (\boldsymbol{m}_n - \boldsymbol{m}_0),$$

where, with the same notation of Theorem 5.1:

$$\begin{aligned}
\boldsymbol{\chi}_i &:= W^{-1} \left(\langle \chi_i, e_1 \rangle_{\mathbb{H}}, \dots, \langle \chi_i, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\boldsymbol{m}_n &:= W^{-1} \left(\langle m_n, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_n, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\boldsymbol{m} &:= W^{-1} \left(\langle m, e_1 \rangle_{\mathbb{H}}, \dots, \langle m, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\boldsymbol{m}_0 &:= W^{-1} \left(\langle m_0, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_0, e_p \rangle_{\mathbb{H}} \right)^\top, \\
\Sigma_n &:= \frac{1}{n-1} \sum_{i=1}^n (\boldsymbol{\chi}_i - \boldsymbol{m}_n)(\boldsymbol{\chi}_i - \boldsymbol{m}_n)^\top.
\end{aligned}$$

From Theorem 5.1 we have $T^2_{0p} \xrightarrow[p\to\infty]{a.s.} T_0^2$. For ease of notation, define

$$\Sigma^+_{nW} = W^{1/2} (W^{1/2} \Sigma_n W^{1/2})^+ W^{1/2}.$$

For every finite p, we have:

$$\begin{aligned}
T^2_{0p} &= n(\boldsymbol{m}_n - \boldsymbol{m} + \boldsymbol{m} - \boldsymbol{m}_0)^\top \Sigma^+_{nW} (\boldsymbol{m}_n - \boldsymbol{m} + \boldsymbol{m} - \boldsymbol{m}_0) \\
&= n(\boldsymbol{m}_n - \boldsymbol{m})^\top \Sigma^+_{nW} (\boldsymbol{m}_n - \boldsymbol{m}) \\
&\quad + n \left(2(\boldsymbol{m}_n - \boldsymbol{m})^\top \Sigma^+_{nW} (\boldsymbol{m} - \boldsymbol{m}_0) + (\boldsymbol{m} - \boldsymbol{m}_0)^\top \Sigma^+_{nW} (\boldsymbol{m} - \boldsymbol{m}_0) \right).
\end{aligned}$$

Since the sequence of subspaces V_p tends to cover the whole $\mathbb{H}$, there exists $\tilde{p} > 0$ s.t. $Im(\mathcal{K}_n) \cap V_p \backslash \{0\} \neq \emptyset$. Taking the limit of the last expression for $n \to \infty$, we have, $\forall p \geq \tilde{p}$:

$$T^2_{0p} \xrightarrow[n\to\infty]{\mathbb{P}} \infty$$

since under H_1, $\boldsymbol{m} - \boldsymbol{m}_0 \neq 0$, $\boldsymbol{m}_n \xrightarrow[n\to\infty]{a.s.} \boldsymbol{m}$, and for all $n > p$, Σ^+_{nW} is positive definite. Hence, $\forall p \geq \tilde{p}$ the test based on T^2_{0p} is consistent. For proving consistency of the test based on T_0^2, observe that $\forall n$ we have $T_0^2 \geq T^2_{0p}$. Hence $\lim_{n\to\infty} T_0^2 \geq \lim_{n\to\infty} T^2_{0p}$. This also imply that the test based on T_0^2 is **consistent**:

$$T^2_{0p} \xrightarrow[n\to\infty]{\mathbb{P}} \infty.$$

Theorem C.2. *Consider two independent samples $\chi_{ij} = m_j \oplus \varepsilon_{ij}$, $j = 1, 2$, $i = 1, \dots, n_j$ embedded in a separable Hilbert space $\mathbb{H}$, where $m_j \in H$ are fixed elements and $\varepsilon_{ij} \in \mathbb{H}$ are i.i.d. random elements with zero mean and covariance operator $\mathcal{K}$ satisfying $\mathbb{E}_{\mathbb{R}}[\|\varepsilon_{ij}\|^2_{\mathbb{H}}] < +\infty$, for all $i \in \{1, n_j\}$.*
The permutation test of hypotheses $H_0 : m_1 = m_2$ against $H_1 : m_1 \neq m_2$ based on statistic (15) *and permutations* (16) *is* ***exact****.*
Namely, let $\mathcal{A} = \{1/\binom{n_1+n_2}{n_1}, 2/\binom{n_1+n_2}{n_1}, \dots, \binom{n_1+n_2}{n_1}/\binom{n_1+n_2}{n_1}\}$ be the set of all attainable levels, and let p_{value} be the p-value of the test. Then:

$$\mathbb{P}_{H_0}\left[p_{value} \leq \alpha\right] \begin{cases} = \alpha & \forall \alpha \in \mathcal{A} \\ < \alpha & \forall \alpha \in [0,1] \backslash \mathcal{A}. \end{cases}$$

The same test is ***consistent****, i.e., $\forall \alpha \in (0, 1]$:*

$$\lim_{n_1, n_2 \to \infty} \mathbb{P}_{H_1}\left[p_{value} \leq \alpha\right] = 1.$$

Proof. The permutation test is based on all permutations of the data over the sample units. Under $H_0 : m_1 = m_2$, we have that data of the two samples are independent and identically distributed. The permutations are then likelihood-invariant. This means that under H_0 the conditional distribution of the test statistic is a discrete uniform. The p-value defined in equation (17) by means of the counting measure gives the probability - under H_0 - that the test statistic is greater or equal to its value on the observed data. Hence, the permutation test is **exact** (see Prop. 2, 3.1.1 of Pesarin and Salmaso, 2010).
For proving consistency, consider a sequence of subspaces $\{V_p\}_{p\geq 1}$ such that $V_p \subset \mathbb{H}$, $\dim(V_p) = p$, and $\lim_{p\to\infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0$ for all $h \in \mathbb{H}$. Let $(e_1, \dots, e_p)$ be a basis set of V_p and let W be the symmetric invertible $p \times p$ matrix such that $W_{jk} = \langle e_j, e_k \rangle_{\mathbb{H}}$. Let T^2_{0p}, $p > 0$ denote the finite-dimensional approximation of T^2_0:

$$T^2_{0p} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} (\boldsymbol{m}_{n_1} - \boldsymbol{m}_{n_2})^\top W^{1/2} (W^{1/2} \Sigma_{n_{\text{pooled}}} W^{1/2})^+ W^{1/2} (\boldsymbol{m}_{n_1} - \boldsymbol{m}_{n_2}),$$

where, with the same notation of Theorem 5.1:

$$\begin{aligned}
\boldsymbol{\chi}_i &:= W^{-1} \left(\langle \chi_i, e_1 \rangle_{\mathbb{H}}, \dots, \langle \chi_i, e_p \rangle_{\mathbb{H}}\right)^\top, \\
\boldsymbol{m}_{n_1} &:= W^{-1} \left(\langle m_{n_1}, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_{n_1}, e_p \rangle_{\mathbb{H}}\right)^\top, \\
\boldsymbol{m}_{n_2} &:= W^{-1} \left(\langle m_{n_2}, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_{n_2}, e_p \rangle_{\mathbb{H}}\right)^\top, \\
\boldsymbol{m}_1 &:= W^{-1} \left(\langle m_1, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_1, e_p \rangle_{\mathbb{H}}\right)^\top, \\
\boldsymbol{m}_2 &:= W^{-1} \left(\langle m_2, e_1 \rangle_{\mathbb{H}}, \dots, \langle m_2, e_p \rangle_{\mathbb{H}}\right)^\top, \\
\Sigma_{n_1} &:= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\boldsymbol{\chi}_i - \boldsymbol{m}_{n_1})(\boldsymbol{\chi}_i - \boldsymbol{m}_{n_1})^\top \\
\Sigma_{n_2} &:= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\boldsymbol{\chi}_i - \boldsymbol{m}_{n_2})(\boldsymbol{\chi}_i - \boldsymbol{m}_{n_2})^\top \\
\Sigma_{n_{\text{pooled}}} &:= \frac{(n_1 - 1)\Sigma_{n_1} + (n_2 - 1)\Sigma_{n_2}}{n_1 + n_2 - 2}.
\end{aligned}$$

From Theorem 5.1 we have $T_{0p}^2 \xrightarrow[p\to\infty]{a.s.} T_0^2$. Let $\delta := m_1 - m_2$ and $\boldsymbol{\delta} = \boldsymbol{m}_1 - \boldsymbol{m}_2$ For ease of notation, define $\Sigma^+_{n_{\text{pooled}_W}} = W^{1/2}(W^{1/2}\Sigma_{n_{\text{pooled}}}W^{1/2})^+W^{1/2}$. For every finite p, we have:

$$
\begin{aligned}
T_{0p}^2 &= \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}(\boldsymbol{m}_{n_1}-\boldsymbol{m}_{n_2}+\boldsymbol{\delta}-\boldsymbol{\delta})^\top\Sigma^+_{n_{\text{pooled}_W}}(\boldsymbol{m}_{n_1}-\boldsymbol{m}_{n_2}+\boldsymbol{\delta}-\boldsymbol{\delta})\\
&= \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\boldsymbol{\delta}^\top\Sigma^+_{n_{\text{pooled}_W}}\boldsymbol{\delta}\\
&+ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}2(\boldsymbol{m}_{n_1}-\boldsymbol{m}_{n_2}-\boldsymbol{\delta})^\top\Sigma^+_{n_{\text{pooled}_W}}\boldsymbol{\delta}\\
&+ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}(\boldsymbol{m}_{n_1}-\boldsymbol{m}_{n_2}-\boldsymbol{\delta})^\top\Sigma^+_{n_{\text{pooled}_W}}(\boldsymbol{m}_{n_1}-\boldsymbol{m}_{n_2}-\boldsymbol{\delta}).
\end{aligned}
$$

Since the sequence of subspaces V_p tends to cover the whole $\mathbb{H}$, there exists $\tilde{p} > 0$ s.t. $Im(\mathcal{K}_{n_{\text{pooled}}}) \cap V_p\backslash\{0\} \neq \emptyset$. Taking the limit of the last expression for $n_1, n_2 \to \infty$, we have, $\forall p \geq \tilde{p}$:

$$T_{0p}^2 \xrightarrow[n_1,n_2\to\infty]{\mathbb{P}} \infty$$

since under H_1, $\boldsymbol{\delta} \neq 0$, $\boldsymbol{m}_{n_1} - \boldsymbol{m}_{n_2} \xrightarrow[n_1,n_2\to\infty]{a.s.} \boldsymbol{\delta}$, and for all $n_1, n_2 > p$, $\Sigma^+_{n_{\text{pooled}_W}}$ is positive definite. Hence, $\forall p \geq \tilde{p}$ the test based on T_{0p}^2 is consistent. For proving consistency of the test based on T_0^2, observe that $\forall n_1, n_2$ we have $T_0^2 \geq T_{0p}^2$. Hence $\lim_{n_1,n_2\to\infty} T_0^2 \geq \lim_{n_1,n_2\to\infty} T_{0p}^2$. This also imply that the test based on T_0^2 is **consistent**:

$$T_{0p}^2 \xrightarrow[n_1,n_2\to\infty]{\mathbb{P}} \infty.$$

MOX Technical Reports, last issues

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

09/2017 Antonietti, P.F.; Ferroni, A.; Mazzieri, I.; Paolucci, R.; Quarteroni, A.; Smerzini, C.; Stupazzin
Numerical modeling of seismic waves by Discontinuous Spectral Element methods

08/2017 Ambrosi, D.; Beloussov, L.V.; Ciarletta, P.
Mechanobiology and morphogenesis in living matter: a survey

06/2017 Ekin, T.; Ieva, F.; Ruggeri, F.; Soyer, R.
On the Use of the Concentration Function in Medical Fraud Assessment

07/2017 Cabassi A.; Pigoli D.; Secchi P.; Carter P.A.
Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology

05/2017 Menafoglio, A.; Hron, K.; Filzmoser, P.
Logratio approach to distributional modeling

04/2017 Dede', L; Garcke, H.; Lam K.F.
A Hele-Shaw-Cahn-Hilliard model for incompressible two-phase flows with different densities

03/2017 Fumagalli, I.; Parolini, N.; Verani, M.
On a free-surface problem with moving contact line: from variational principles to stable numerical approximations

02/2017 Arena, M.; Calissano, A.; Vantini, S.
Monitoring Rare Categories in Sentiment and Opinion Analysis - Expo Milano 2015 on Twitter Platform.

01/2017 Riccobelli, D.; Ciarletta, P.
Rayleigh-Taylor instability in soft elastic layers

58/2016 Antonietti, P. F.; Bruggi, M. ; Scacchi, S.; Verani, M.
On the Virtual Element Method for Topology Optimization on polygonal meshes: a numerical study