Computational Statistics and Data Analysis 124 (2018) 15–26

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Testing the equality of several covariance functions for functional data: A supremum-norm based test

Jia Guo [a], Bu Zhou [b,*], Jin-Ting Zhang [b]

[a] ACEMS and School of Mathematics and Statistics, The University of Melbourne, Melbourne, Parkville VIC 3010, Australia
[b] Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore

HIGHLIGHTS

- A supremum-norm based test for the equality of several covariance functions is proposed.
- Asymptotic random expression of the test statistic is derived under the null hypothesis.
- The new test is shown to be root-n consistent under a local alternative.
- The new test is more powerful than L^2-norm type tests for highly correlated data or with some local spikes.
- Simulations and real data examples show the new test outperforms some competitors.

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 11 November 2017
Accepted 15 February 2018
Available online 26 February 2018

Keywords:
Equal covariance function testing
Functional data
Non-parametric bootstrap
Supremum-norm based test

ABSTRACT

Testing the equality of covariance functions is crucial for solving functional ANOVA problems. Available methods, such as the recently proposed L^2-norm based tests work well when functional data are less correlated but are less powerful when functional data are highly correlated or with some local spikes, which are often the cases in real functional data analysis. To overcome this difficulty, a new test for the equality of several covariance functions is proposed. Its test statistic is taken as the supremum value of the sum of the squared differences between the estimated individual covariance functions and the pooled sample covariance function. The asymptotic random expressions of the test statistic under the null hypothesis and under a local alternative are derived and a non-parametric bootstrap method is suggested. The root-n consistency of the proposed test is also obtained. Intensive simulation studies are conducted to demonstrate the finite sample performance of the proposed test. The simulation results show that the proposed test is indeed more powerful than several existing L^2-norm based competitors when functional data are highly correlated or with some local spikes. The proposed test is illustrated with three real data examples collected in a wide scope of scientific fields.

1. Introduction

Functional data arise in a wide scope of scientific fields such as biology, medicine, ergonomics among others. They are collected in a form of curves or images instead of traditional scalar or vector observations. Nowadays, this type of data is widely regarded as a basis element in data analysis. Compared with traditional data, functional data are worth

* Corresponding author.
E-mail addresses: jia.guo@unimelb.edu.au (J. Guo), bu.zhou@u.nus.edu (B. Zhou), stazjt@nus.edu.sg (J.-T. Zhang).

https://doi.org/10.1016/j.csda.2018.02.002
0167-9473/

exploring since they contain much more structural information. A lot of useful and effective tools have been developed over the past two decades to extract those information hidden in functional data. The reader is referred to Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Zhang (2013) and references therein for more details.

In recent decades, much attention has been paid to hypothesis testing of the mean functions for functional data. A general and direct testing procedure is the so-called pointwise t-test described by Ramsay and Silverman (2005). However, this pointwise t-test fails to give a global conclusion which is often needed in real data analysis. To overcome this drawback, Zhang et al. (2010) proposed an L^2-norm based test whose test statistic is obtained as a squared L^2-norm of the mean function differences of two functional samples. For one-way ANOVA problems for functional data, several interesting tests have been proposed in the literature, see the ANOVA test proposed in Cuevas et al. (2004) and the GPF test proposed in Zhang and Liang (2013). Both of their methods are extensions of the classical F-test. When conducting the above mean function testing procedures, one may first need to check the equality of the covariance functions of the functional samples involved, as many mean function testing procedures are based on the assumption that the functional samples involved have a common covariance function. Motivated by the need of testing the equality of the covariance functions of two functional samples, Panaretos et al. (2010) and Fremdt et al. (2013) proposed dimension reduction based approaches to address this issue. However, their methods are based on functional principal component analysis so that their results often strongly depend on the number of selected functional principal components. More discussions about the drawbacks of the dimension reduction based methods can be found in Paparoditis and Sapatinas (2016) and Guo et al. (2016).

Although some recent works have shed some light on how to test the equality of the covariance functions of two functional samples, how to test the equality of the covariance functions of several functional samples receives relatively little attention. Paparoditis and Sapatinas (2016) described a general bootstrap procedure for the problem but only studied a two-sample test. Zhang (2013) proposed an L^2-norm based test for the multi-sample equal-covariance function testing problem with the Welch–Satterthwaite χ^2-approximation and a random permutation approximation. Guo et al. (2016) further studied the asymptotic power and finite sample performance of the L^2-norm based test. However, as shown in Guo et al. (2016), these L^2-norm based tests are less powerful when the functional data are highly correlated in the sense that the dependence of discrete measured values in a single observed functional curve is high, which is a common feature of functional samples due to their inherent smoothness. The smoothness of functional data arises from the fact that the discrete measured values in a single functional observation are obtained sequentially over a continuum such as time. In addition, the above L^2-norm based tests have limited ability to perceive covariance function differences with some local spikes because the L^2-norm is a global measure that tends to obscure local characteristics. To surmount these challenges, in this paper, we develop a supremum-norm based test which is good at detecting the covariance function differences when the functional data are moderately or highly correlated. It is also powerful in detecting the covariance function differences with some local spikes. The asymptotic random expressions of the proposed test statistic under the null hypothesis and a local alternative are derived. A nonparametric bootstrap method is proposed to approximate the associated null distribution. The proposed test is demonstrated via intensive simulation studies and three real data applications.

The rest of the paper is organized as follows. We present the main results in Section 2 and intensive simulations in Section 3. We illustrate the methodologies using three real data examples in Section 4. The technical proofs of the main results are given in Appendix A.

2. Main results

Throughout this paper, let $\mathrm{SP}(\eta, \gamma)$ and $\mathrm{GP}(\eta, \gamma)$ respectively denote a stochastic process and a Gaussian process with mean function $\eta(\mathbf{t})$ and covariance function $\gamma(\mathbf{s}, \mathbf{t})$ where $\mathbf{s}$ and $\mathbf{t}$ can be scalars or vectors. Let $y_{i1}(t), y_{i2}(t), \ldots, y_{in_i}(t)$, $i = 1, 2, \ldots, k$ be independent functional samples over a given finite time interval $\mathcal{T} = [a, b]$, $-\infty < a < b < \infty$, which satisfy

$$\begin{aligned} & y_{ij}(t) = \eta_i(t) + v_{ij}(t),\ j = 1, 2, \ldots, n_i, \\ & v_{i1}(t), v_{i2}(t), \ldots, v_{in_i}(t) \overset{i.i.d.}{\sim} \mathrm{SP}(0, \gamma_i);\ i = 1, 2, \ldots, k, \end{aligned} \tag{1}$$

where $\eta_1(t), \eta_2(t), \ldots, \eta_k(t)$ model the unknown group mean functions of the k samples, $v_{ij}(t), j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k$ denote the subject-effect functions, and $\gamma_i(s, t),\ i = 1, 2, \ldots, k$ are the associated covariance functions. Throughout this paper, we assume that $\mathrm{tr}(\gamma_i) = \int_{\mathcal{T}} \gamma_i(t, t)\mathrm{d}t < \infty$ and $\eta_i(t) \in \mathcal{L}^2(\mathcal{T}), i = 1, 2, \ldots, k$, where $\mathcal{L}^2(\mathcal{T})$ denotes the Hilbert space formed by all the squared integrable functions over $\mathcal{T}$ with the inner-product defined as $\langle f, g\rangle = \int_{\mathcal{T}} f(t)g(t)\mathrm{d}t,\ f,\ g \in \mathcal{L}^2(\mathcal{T})$. It is often of interest to test the equality of the k covariance functions:

$$H_0 : \gamma_1(s, t) \equiv \gamma_2(s, t) \equiv \cdots \equiv \gamma_k(s, t), \quad \text{for all } s, t \in \mathcal{T}. \tag{2}$$

For convenience, we refer to the above problem as the k-sample equal-covariance function (ECF) testing problem for functional data.