Journal of Statistical Planning and Inference xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Q1 On two sample inference for eigenspaces in functional data analysis with dependent errors

Q2 Jan Beran *, Haiyan Liu, Klaus Telkmann

Department of Mathematics and Statistics, University of Konstanz, Germany

ARTICLE INFO

Article history:
Received 14 February 2015
Received in revised form 5 January 2016
Accepted 1 February 2016
Available online xxxx

Keywords:
Two sample problem
Functional data analysis
Long-range dependence
Repeated time series
Eigenspaces
Bootstrap

ABSTRACT

We consider functional data analysis for randomly perturbed repeated time series with a general dependence structure of the error process. Specifically, the question of testing for equality of subspaces spanned by a finite number of eigenfunctions is addressed. The asymptotic distribution of standardized residual processes based on projections of eigenfunctions is derived. A two-sample test based on the residual processes is proposed together with a nuisance parameter free bootstrap procedure. Simulations illustrate finite sample properties.

1. Introduction

In many applications, repeated time series Y_{ij} $(i = 1, \ldots, n, j = 1, \ldots, N)$ can be assumed to be of the form

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (1)$$

where $X_i(t) \in L^2[0, 1]$ are independent randomly generated smooth functions, $t_j = jN^{-1} \in [0, 1]$ is rescaled time and $\epsilon_i(j)$ $(j \in \mathbb{N})$ denote independent error processes (independent in i). In functional data analysis (FDA) one is mainly interested in the covariance operator $\mathbf{C}(y)(t) = \int C(s, t)y(s)ds$ $(y \in L^2[0, 1])$ and its spectral representation. Here, $\mu(t) = E[X(t)]$ and $C(s, t) = \mathrm{Cov}(X(s), X(t)) = E[(X(s) - \mu(s))(X(t) - \mu(t))]$. The spectral representation of $C(s, t)$ is of the form

$$C(s, t) = \mathrm{cov}\left(X(t), X(s)\right) = \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\,\phi_l(t) \quad (t, s \in [0, 1]) \quad (2)$$

where the eigenfunctions $\phi_l(t)$ $(l \in \mathbb{N}, t \in [0, 1])$ build an orthonormal $L^2[0, 1]$-basis and the eigenvalues $\lambda_l \geq 0$ are such that $\sum_l \lambda_l < \infty$ and $\lambda_l \geq \lambda_{l+1}$. In classical functional data analysis, the functions X_i are assumed to be observed directly, i.e. $\epsilon_i(j) \equiv 0$ (see e.g. Ramsay and Silverman, 2002, 2005, Bosq, 2000, Ferraty and Vieu, 2006, Horváth and Kokoska, 2012). The asymptotic distribution of estimated eigenfunctions and eigenvalues can be found in Dauxois et al. (1982) and Hall and Hosseini-Nasab (2006, 2009). The situation with non-zero i.i.d. errors is considered for instance in Hall et al. (2006), Yao (2007), Staniswalis and Lee (1998) and Yao et al. (2003, 2005). Related results on nonparametric estimation of $\mu(t) = E(X_i(t))$ for repeated time series are discussed in Hart and Wehrly (1986), Lin and Carroll (2000), Sererini and Staniswalis (1994), Staniswalis and Lee (1998), Verbyla et al. (1999) and Wild and Yee (1996), among others. Q3 Q4

* Corresponding author.
E-mail address: jan.beran@uni-konstanz.de (J. Beran).

http://dx.doi.org/10.1016/j.jspi.2016.02.001

Please cite this article in press as: Beran, J., et al., On two sample inference for eigenspaces in functional data analysis with dependent errors. J. Statist. Plann. Inference (2016), http://dx.doi.org/10.1016/j.jspi.2016.02.001

Kernel estimation of $\mu(t)$ in repeated series with strongly dependent errors is considered in Ghosh (2001) and Beran and Liu (2014a,b). Beran and Liu (2014a,b) derive functional limit theorems for kernel estimators of $\mu(t)$ and the covariance operator $\mathbf{C}$ under short- and long-memory assumptions. For further references on FDA see e.g. Bosq (2000), Ramsay and Silverman (2002, 2005), Clarkson et al. (2005), Ferraty and Vieu (2006), Ramsay et al. (2009), Ferraty and Romain (2011) and Horváth and Kokoska (2012). Nonparametric regression for single long-memory time series observations is discussed for instance in Hall and Hart (1990), Csörgö and Mielniczuk (1995), Ray and Tsay (1997), Robinson (1997) and Beran and Feng (2002a,b,c). For a general overview on statistical inference for long-memory processes see e.g. Beran (1994), Giraitis et al. (2012), Beran et al. (2013) and references therein.

One of the main objectives in FDA is to obtain a low-dimensional representation of X in terms of eigenfunctions ϕ_l ($l = 1, \ldots, m$) with the largest m eigenvalues. In this paper, we consider the following two-sample problem. Suppose we observe two independent samples $Y_{ij}^{(1)} = X_i^{(1)}(t_j) + \epsilon_i^{(1)}(j)$ ($i = 1, \ldots, n^{(1)}, j = 1, \ldots, N^{(1)}$) and $Y_{ij}^{(2)} = X_i^{(2)}(t_j) + \epsilon_i^{(2)}(j)$ ($i = 1, \ldots, n^{(2)}, j = 1, \ldots, N^{(2)}$), with $X_i^{(k)}, \epsilon_i^{(k)}$ ($k = 1, 2$) defined as above. The Karhunen–Loève expansions of $X_i^{(k)}$ ($k = 1, 2$) are given by

$$X_i^{(k)}(t) = \mu^{(k)}(t) + \sum_{l=1}^{\infty} \xi_{il}^{(k)} \phi_l^{(k)}(t) \quad (k = 1, 2).$$

Given a finite fixed dimension $m \in \mathbb{N}$, we are interested in testing whether the subspaces (of $L^2[0, 1]$) spanned by $\phi_l^{(1)}$ ($l = 1, \ldots, m$) and $\phi_l^{(2)}$ ($l = 1, \ldots, m$) are the same. A test based on residual functions is developed and its asymptotic distribution is derived under the null hypothesis of identical m-dimensional eigenspaces. Note that this null hypothesis does not imply that the eigenfunctions and eigenvalues are identical. Thus, neither tests for equality of the covariance structure (Fremdt et al., 2013; also see Horváth et al., 2009 in a functional regression context) nor tests for equality of eigenfunctions (Benko et al., 2009; Boente et al., 2011) are directly applicable.

The test developed here was motivated by questions raised in the context of event related potentials (ERP) obtained from EEG measurements. For instance, in experiments on rational decision making (see e.g. Achtziger et al. 2014), it was conjectured that different experimental conditions (or "treatments") may lead to changes in some low dimensional eigen-subspaces. Note that this differs from testing equality of covariance functions or eigenvalues and eigenfunctions themselves. Equality of certain low dimensional eigen-subspaces does not necessarily imply equality of the entire covariance operator. Moreover, equality of eigenfunctions and eigenvalues is not required under the null hypothesis, because only the spans of the eigenfunctions are compared. In many applications equality of eigenfunctions or eigenvalues is too restrictive and therefore not of direct interest.

Typically, EEG data are very noisy and the noise component is highly correlated, including the possibility of long memory (see e.g. Watters, 2000, Linkenkaer-Hansen et al., 2001, Nikulin and Brismar, 2005, Bornas et al., 2013). The model defined above takes this into account. The bootstrap method developed in Section 4 is applicable without the necessity of modeling the dependence structure of the noise component explicitly, nor is it necessary to estimate eigenfunctions and eigenvalues orthogonal to the eigenspaces considered in the test (see Section 4). Note that in a related paper, Benko et al. (2009) also consider testing equality of eigenspaces. Asymptotic results in Benko et al. (2009) are derived explicitly for the noiseless case only, with comments on possible extensions to observations with i.i.d. noise. In contrast, the method and asymptotic results discussed in the following apply to randomly perturbed FDA with a very general range of possible dependence structures in the error process.

The paper is organized as follows. Definitions and fundamental lemmas are discussed in Section 2. A residual process defined in Section 3, and its asymptotic distribution is derived under the null hypothesis. This provides the basis for defining suitable test procedures. As an example, a simple Bonferroni corrected test is proposed. An improved test together with a bootstrap procedure are discussed in Section 4. Simulations illustrate the results in Section 5. Final remarks in Section 6 conclude the paper. Proofs are given in the Appendix.

2. Definitions and auxiliary results

2.1. Estimation of μ and C

First we define the basic model for one sample. Observations are assumed to consist of n independent time series $Y_{i\cdot} = (Y_{i1}, \ldots, Y_{iN})$ ($i = 1, \ldots, n$) defined by

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (t_j = jN^{-1}, j = 1, \ldots, N) \tag{3}$$

with $t_j = j/N$ denoting rescaled time. The random curves X_i have the expansion

$$X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il} \phi_l(t), \tag{4}$$

Please cite this article in press as: Beran, J., et al., On two sample inference for eigenspaces in functional data analysis with dependent errors. J. Statist. Plann. Inference (2016), http://dx.doi.org/10.1016/j.jspi.2016.02.001