Journal of Multivariate Analysis 124 (2014) 150–165

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Series expansion for functional sufficient dimension reduction

Heng Lian [a,*], Gaorong Li [b]

[a] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371, Singapore

[b] College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 15 June 2012
Available online 5 November 2013

AMS subject classification:
62H12

Keywords:
Functional principal component analysis
Polynomial splines
Sliced average variance estimation
Sliced inverse regression

ABSTRACT

Functional data are infinite-dimensional statistical objects which pose significant challenges to both theorists and practitioners. Both parametric and nonparametric regressions have received attention in the functional data analysis literature. However, the former imposes stringent constraints while the latter suffers from logarithmic convergence rates. In this article, we consider two popular sufficient dimension reduction methods in the context of functional data analysis, which, if desired, can be combined with low-dimensional nonparametric regression in a later step. In computation, predictor processes and index vectors are approximated in finite dimensional spaces using the series expansion approach. In theory, the basis used can be either fixed or estimated, which include both functional principal components and *B*-spline basis. Thus our study is more general than previous ones. Numerical results from simulations and a real data analysis are presented to illustrate the methods.

1. Introduction

There has recently been increased interest in the statistical modeling of functional data. In many experiments, functional data appear as the basic unit of observations. As a natural extension of the multivariate data analysis, functional data analysis provides valuable insights into these problems. Compared with the discrete multivariate analysis, functional analysis takes into account the smoothness of the high dimensional covariates, and often suggests new approaches to the problems that have not been discovered before. Even for nonfunctional data, the functional approach can often offer new perspectives on the old problem.

The literature contains an impressive range of functional analysis tools for various problems including exploratory functional principal component analysis, canonical correlation analysis, classification and regression. Two major approaches exist. The more traditional approach, masterfully documented in the monograph [29], typically starts by representing functional data by an expansion with respect to a certain basis, and subsequent inferences are carried out on the coefficients. The most commonly utilized basis include *B*-spline basis for nonperiodic data and Fourier basis for periodic data. Another line of work by the French school [16], taking a nonparametric point of view, extends the traditional nonparametric techniques, most notably the kernel estimate, to the functional case. Some recent advances in the area of functional regression include Cardot et al. [5]; Cai and Hall [4]; Aneiros-Perez and Vieu [3]; Preda [28]; Ait-Saidi et al. [2]; Aguilera et al. [1]; Wong et al. [30]; Yao et al. [32]; Ait-Saidi et al. [2]; Crambes et al. [13].

As an extension of classical linear regression, parametric functional linear regression has achieved exclaimed success in many real problems, although it can be argued that the structural constraint is too stringent. On the other hand,

* Corresponding author.
E-mail address: henglian@ntu.edu.sg (H. Lian).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.10.019

nonparametric functional regression is more flexible but typically suffers from poor convergence rate [17]. To address these problems, Chen et al. [6] studied functional single-index and multiple-index models.

Here we consider an alternative semiparametric approach based on sufficient dimension reduction. In functional context, we assume

$$Y = g(\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle, \epsilon), \tag{1}$$

where $\langle , \rangle$ is the usual inner product in $L_2[0, 1]$. Thus the response Y only depends on the predictor through K indices obtained by projecting onto K directions. Since g is unknown, the K directions, referred to as dimension reduction directions, are not identifiable. In the multivariate case, the space spanned by them (referred to as a dimension reduction subspace, or drs) is identifiable under mild assumptions, however such assumptions are not known in the functional context yet. Thus we will not use the concept of the central space which is popularly used in the dimension reduction literature [7,34]. The reason is that for functional data there is no corresponding theory for the existence and uniqueness of the central space. This may be due to that density for functional data is a tricky concept to work with. In the literature of functional SIR, researchers typically work with a drs, even though there might be multiple drs's. Although a unique drs is generally not identifiable, useful methodology is still possible. The approach of dimension reduction is particularly useful in an exploratory stage of statistical analysis since very few structural assumptions are imposed in (1). In particular, it is not necessary to assume the different indices act additively as usually assumed in multiple-index models, and the error also is not necessarily additive on mean, or homogeneous. After the dimension reduction directions are found, in particular if there are only a small number of significant directions, one can use traditional nonparametric approaches to study the relationships between responses and the few indices. This second stage typically involves additional structural assumptions such as additive errors.

There exist quite a few different methods aimed at estimating the dimension reduction space [24,12,25,37,38]. Among these sliced inverse regression (SIR) and sliced average variance estimation (SAVE) are probably the most popular. Both required linearity assumption of the predictors. However, SIR will fail when $E[X|Y] = 0$ which motivated the use of SAVE. On the other hand, SAVE requires an additional assumption on the distribution of predictors.

Adapting SIR to functional context has been proposed in [18] based on functional principal component analysis on the random predictor process. In particular the predictor process is approximated by a truncation of the Karhunen–Loève expansion, using the eigenfunctions as the basis. The basic procedure is to (i) approximate the functional predictors with series expansion using certain basis and obtain the coefficients; (ii) perform dimension reduction using the finite-dimensional coefficients as the predictors; (iii) use directions obtained in (ii) as the coefficients of the basis to finally obtain the direction in functional space. It turns out this computational procedure is correct only when the basis is orthonormal, and we will detail the general algorithm in Section 4.

In terms of theory, Ferré and Yao [18] assumes that the number of slices is fixed which works well for discrete responses, but is only an approximation for continuous responses. On the other hand, the kernel estimate used in [19] was later shown to require much stronger assumptions [9].

Our contributions in this study are summarized as follows. First, our theory for SIR allows various basis systems, either fixed in advanced or estimated from data. Second, our theory works for both categorical and continuous response Y. Third and most importantly, we extend SAVE to the functional context which has not been considered before.

2. SIR and SAVE

Let Y be a real random response and $X \in L_2[0, 1]$ the random functional predictor. In this article, we assume the entire trajectory of noise-free process X is observed. When the process is densely measured, this is a reasonable assumption. For simplicity, we assume $EX = 0$. We also assume the fourth moment of X exists, that is $E\|X\|^4 < \infty$. The (population) covariance operator of X is given by $\Gamma = E(X \otimes X)$, where for any $x, y \in L_2[0, 1]$, $x \otimes y$ denotes the linear operator $L_2[0, 1] \to L_2[0, 1]$ such that $(x \otimes y)(z) = \langle x, z\rangle y$. Using the well-known Karhunen–Loève expansion, we can write

$$X = \sum_{j=1}^{\infty} \xi_j \phi_j,$$

where $E\xi_j^2 = \lambda_j$ are the eigenvalues and ϕ_j are the eigenfunctions. We assume all the eigenvalues, $\lambda_1 > \lambda_2 > \cdots > 0$ are distinct and positive, as usually assumed in the functional data literature [22,18]. If some eigenvalues are zero, the components of β_k in the kernel space of Γ cannot be identified. We focus on the estimation of the space spanned by K linearly independent directions $\beta_1, \ldots, \beta_K$, which is called a dimension reduction subspace (drs) and denoted by $\mathcal{S}$. Let $\Gamma\mathcal{S}$ be the space spanned by $\Gamma\beta_1, \ldots, \Gamma\beta_K$.

Let $B_X = (\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle)$. The principle of SIR and SAVE is based on the following result with proofs omitted, which is a direct extension of the multivariate case.

Theorem 1. (a) [18] *Suppose for all $b \in L_2[0, 1]$, the conditional expectation $E(\langle b, X\rangle|B_X)$ is linear in $\langle\beta_1, X\rangle, \ldots, \langle\beta_K, X\rangle$. Then $E(X|Y) \in \Gamma\mathcal{S}$. Obviously, if the linearity assumption is true for all drs's, then $E(X|Y) \in \Gamma(\cap_{\mathcal{S} \text{ is a drs}} \mathcal{S})$. Note that generally, $\cap_{\mathcal{S} \text{ is a drs}} \mathcal{S}$ is not guaranteed to be a drs, but s spanning system for the intersection may still be useful in practice.*