ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On properties of percentile bootstrap confidence intervals for prediction in functional linear regression

Omid Khademnoe, S. Mohammad E. Hosseini-Nasab *

Department of Statistics, Shahid Beheshti University, Tehran, Iran

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form 4 October 2015
Accepted 4 October 2015
Available online xxxx

Keywords:
Edgeworth expansion
Functional linear regression model
Functional principal component analysis
Percentile bootstrap confidence interval
Prediction

ABSTRACT

We consider a functional linear regression model with scalar response and functional covariate. For this model bootstrap confidence intervals for prediction using the residual resampling method have been already studied. In this paper, we use the paired resampling method to construct bootstrap confidence intervals for prediction in the functional linear regression model. We develop Edgeworth expansions for distribution of the prediction and apply the results to obtain coverage errors of percentile equal-tailed bootstrap confidence intervals for prediction. We carry out a simulation study to illustrate the numerical performance of the paired bootstrap confidence intervals and compare the results with those obtained by the residual resampling method.

1. Introduction

Advances in technology facilitate collecting and storing the data that are essentially in the form of curves. Because in practice, one may collect the values of the curves at a finite set of points, the multivariate methods can be applied to these kinds of data. However, the number of points observed per function may be much larger than the total number of functions. On the other hand, the existence of strong correlations between the values of a function at consecutive points can make the multivariate methods inefficient. Therefore, some theoretical justification is needed to provide the required definitions and concepts regarding the essential nature of the data. Ramsay and Silverman (2005) gave the theoretical and methodological development of functional data analysis (FDA). Many of the nonparametric techniques for FDA can be also found in Ferraty and Vieu (2006).

One of the important parts in FDA is functional regression analysis, used for modelling of a response of interest based on a set of functional predictors. The functional regression problem with a functional response was studied by Fan and Zhang (2000) and Lin and Ying (2001). Yao et al. (2005) constructed asymptotic pointwise confidence bands for response trajectories. The weak convergence in the functional autoregressive model was investigated by Mas (2007). When both the predictor and response variables are functional, He et al. (2010) used canonical expansions to estimate the coefficient functions and compared the method with functional principal component analysis (FPCA). Ferraty et al. (2012) derived the pointwise asymptotic normality for a nonparametric regression model in which the response variable is a function.

In this paper, we are interested in a functional linear regression model, relating the functional covariate to a scalar response. The problems of estimating the slope function and of prediction have been addressed in several works. Cardot et al. (1999) used FPCA to estimate the regression model. They also proved convergence of this estimator both in probability

* Corresponding author.
E-mail address: m_hosseininasab@sbu.ac.ir (S.M.E. Hosseini-Nasab).

http://dx.doi.org/10.1016/j.jspi.2015.10.001

and almost surely. A spline approach to estimate the parameter function was used by Cardot et al. (2003a). Cardot et al. (2003b) studied hypotheses testing about the relationship between a scalar response and a functional predictor. Cai and Hall (2006) investigated the rate of convergence for prediction in functional regression models. Hall and Horowitz (2007) showed that the approach based on FPCA to estimating the slope function, achieved optimal convergence rates. They also suggested another approach based on Tikhonov, or quadratic regularization, and showed that this approach is also able to attain the optimal convergence rates. Li and Hsing (2007) applied the adopted complex Fourier basis to approximate the true functional predictor and used the penalized least-squares criterion to estimate the regression weight function. They also derived the convergence rates for these methods. Cardot et al. (2007) proved a CLT for prediction in functional linear regression. James et al. (2009) proposed an approach to obtain interpretable, flexible and accurate estimate of the slope function. Li and Hsing (2010) considered a functional linear model, where only a finite number of the predictor projections affects the response variable. They focused on the principal component analysis to determine the number of effective projections. Under the assumption that the predictor is elliptically contoured distributed, Li and Hsing (2010) proposed two sequential $\mathcal{X}^2$ testing procedures for dimension selection. Yao and Müller (2010) expressed functional quadratic regression models as an extension of the common functional linear models. Zhou et al. (2013) studied estimation of the slope function, when its values are zero within certain sub-regions. For a detailed review of functional linear models, we refer the readers to Ramsay and Silverman (2002, 2005) and many articles discussed and cited there.

The bootstrap and Edgeworth expansions have been already used in the context of classical regression for linear and nonparametric models. See, for example, Freedman (1981), Stine (1985), Navidi (1989) and Hall (1992). Recently, bootstrap methods have been used in the context of FDA. Hall and Hosseini-Nasab (2006) used a bootstrap procedure to construct simultaneous confidence regions for an infinite number of eigenvalues, and also for individual eigenvalues and eigenfunctions of the covariance operator. Ferraty et al. (2010) proposed a bootstrap method to approximate the distribution of the estimated nonparametric functional regression model with scalar response. González-Manteiga and Martínez-Calvo (2011) used the residual bootstrap procedure to build confidence intervals for prediction in functional linear regression models. They also studied the asymptotic validity of the procedure. Ferraty et al. (2012) investigated the validity of a componentwise bootstrap procedure, in which a functional covariate is related to the functional response by a nonparametric regression model.

Edgeworth expansions for infinite dimensional problems were studied by Götze (1979, 1985, 1989) and Anderson et al. (1998). In this paper, we apply the paired bootstrap method and use the Edgeworth expansions to derive the distribution of prediction in functional linear regression models with scalar response. Based on these results, we then obtain the coverage levels of percentile bootstrap confidence intervals for prediction in the functional linear models.

The paper is organized as follows. In Section 2, functional linear models are introduced and then FPCA method for estimation of the slope function is described. Section 3 introduces the paired resampling method in the context of functional linear regression models, and contains the main theoretical results of the work. In Section 4, we conduct a simulation study to illustrate the performance of the bootstrap confidence intervals under various parameter settings. A comparison between the results obtained by the paired resampling method and those obtained by the residual resampling method is also provided via the simulation study in this section. Finally, all of the proofs are collected in the Appendix.

2. Models and estimators

Functional linear regression models expressing the relationship between a scalar response and functional predictors have been widely used. This arises, for instance, in meteorology, where prediction of total annual precipitation for weather stations from patterns of temperature variation through the year is considered. The NIR spectroscopic information of wheat can be used to predict percentages of protein and moisture content. For more details about other applications of these models, see Ramsay and Silverman (2002, 2005).

Let X be a square integrable random function defined on $\mathcal{I} = [0, T]$, and the real-valued response Y is generated by the model

$$Y = a + \int_{\mathcal{I}} b(s) X(s)\, ds + \varepsilon, \tag{1}$$

where a and b denote the intercept and unknown slope function, respectively, and the error ε is a real-valued variable. We assume that the error ε is independent of the random function X, $E[\varepsilon] = 0$, $E[\varepsilon^2] = \sigma^2 < \infty$ and $E[(\int_{\mathcal{I}} X^2(s)ds)^2] < \infty$. Without loss of generality, we can drop the intercept a in (1) and write $\int bX$ instead of $\int_{\mathcal{I}} b(s) X(s)\, ds$ for notational convenience.

The bivariate function $K(s_1, s_2) = \mathrm{cov}\{X(s_1), X(s_2)\}$ is defined on $\mathcal{I} \times \mathcal{I}$. Based on $K(s_1, s_2)$, the covariance operator of X denoted also by K is defined on $\mathcal{L}^2(\mathcal{I})$, the space of all square integrable function on $\mathcal{I}$, such that

$$Kh(s_1) = \int_{\mathcal{I}} K(s_1, s_2) h(s_2) ds_2, \quad \forall\, h \in \mathcal{L}^2(\mathcal{I}). \tag{2}$$

The spectral decomposition of K can be written as

$$K(s_1, s_2) = \sum_{j=1}^{\infty} \theta_j \varphi_j(s_1) \varphi_j(s_2), \quad s_1,\ s_2 \in \mathcal{I}, \tag{3}$$