STAPRO: 8198 Model 3G pp. 1–6 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Optimal weighting schemes for longitudinal and functional data

Xiaoke Zhang [a,*], Jane-Ling Wang [b]

[a] Department of Statistics, George Washington University, 801 22nd St NW, 7th Floor, Washington, DC 20052, USA
[b] Department of Statistics, University of California, Davis, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 31 May 2017
Received in revised form 6 March 2018
Accepted 11 March 2018
Available online xxxx

MSC:
62-G20
62-G08
62-G05

Keywords:
Local linear smoothers
Non-dense functional data
Dense functional data
Ultra-dense functional data
Mixture weighting schemes

ABSTRACT

We propose optimal weighting schemes for both mean and covariance estimations for functional data based on local linear smoothing such that the L^2 rate of convergence is minimized. These schemes can self-adjust to the sampling plan and lead to practical improvements.

1. Introduction

In the past few decades, substantial efforts have been made that significantly advanced the field of functional data analysis (FDA). Representative monographs include Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Hsing and Eubank (2015), Ramsay and Silverman (2005) and Zhang (2013). Recent developments in FDA, e.g., regression, classification, clustering and manifold learning, are illustrated by a few survey articles (e.g., Cuevas, 2014; Marron and Alonso, 2014; Wang et al., 2016; Reiss et al., 2017) and special journal issues (e.g., González-Manteiga and Vieu, 2007; Goia and Vieu, 2016; Kokoszka et al., 2017). As an indispensable ingredient of many advanced methods, the estimation of mean and covariance functions is fundamental in FDA. A typical functional dataset arises from a sample of functions $\{X_i : i = 1, \ldots, n\}$ collected from n subjects, which are often assumed to be independent and identically distributed (i.i.d.) copies of a L^2 stochastic process X defined on a compact time domain $\mathcal{I}$. The mean and covariance functions are defined by $\mu(t) = E\{X(t)\}, t \in \mathcal{I}$, and $\gamma(s,t) = \mathrm{Cov}\{X(s), X(t)\}, s, t \in \mathcal{I}$, respectively. In practice, the measurements of each X_i are only available at N_i discrete time points $T_{ij} \in \mathcal{I}, j = 1, \ldots, N_i$, and may contain noise e_{ij} at T_{ij}, where $\{e_{ij} : i = 1, \ldots, n; j = 1, \ldots, N_i\}$ are often assumed to be i.i.d. with zero mean and constant variance. Therefore, the observed functional data are often represented as $\{(Y_{ij}, T_{ij}) : i = 1, \ldots, n; j = 1, \ldots, N_i\}$ where $Y_{ij} = X_i(T_{ij}) + e_{ij}$.

Various nonparametric methods have been applied to the estimation of μ and γ (e.g., Rice and Silverman, 1991; Cardot, 2000; Zhang and Chen, 2007; Paul and Peng, 2009; Xiao et al., 2013). Although not always emphasized, each method usually

* Corresponding author.
E-mail addresses: xkzhang@gwu.edu, xkzhang@email.gwu.edu (X. Zhang).

https://doi.org/10.1016/j.spl.2018.03.007

adopts a particular weighting scheme in the estimation procedure, i.e., a strategy of allocating weights to observations in the objective function. In the FDA literature, the most commonly used weighting schemes are the equal-weight-per-observation (OBS) scheme (e.g., Yao et al., 2005) and equal-weight-per-subject (SUBJ) scheme (e.g., Li and Hsing, 2010). See Section 2 for details. The two schemes are identical when $\{N_i : i = 1, \ldots, n\}$ are all equal but this rarely happens in longitudinal studies involving human subjects. This raises a natural, pragmatic but open question: Which weighting scheme is better and is there an optimal weighting scheme?

Recently Zhang and Wang (2016) made the first attempt to tackle this issue. Under both OBS and SUBJ schemes, they provided a comprehensive analysis of the mean and covariance function estimators and partitioned functional data into three types in terms of the rate of convergence: non-dense, dense, and ultra-dense. It was shown that the OBS scheme leads to a more efficient estimator for non-dense data while the SUBJ scheme is superior for ultra-dense data. Moreover, the authors proposed a new class of weighting schemes by considering a convex combination of the OBS and SUBJ weights and derived the optimal convex combination scheme, which we call the "ZW scheme" hereafter, for the L^2 rate of the mean or covariance function estimator. Asymptotically the ZW scheme is as good as, if not better than, both OBS and SUBJ schemes, but this is not always guaranteed numerically.

The main contribution of this article is threefold. First, we develop the optimal weighting schemes for both mean and covariance function estimations. These weighting schemes are optimal in terms of the L^2 rate of convergence among all possible weighting schemes, and are always adaptive to the design in the sense that they self-adjust to the sampling plan regardless of whether the data are non-dense, dense, or ultra-dense. This adaptive feature is particularly important since in practice one would not know for sure which one of the three design settings a particular dataset belongs to. Moreover, the proposed optimal scheme is superior to the ZW scheme, which is restricted to be a convex combination of the OBS and SUBJ weights. Simulation studies also demonstrate that the empirical advantage of the optimal scheme over the ZW scheme may be substantial. Second, we establish two new corollaries that shed light on the relationship between the optimal schemes and OBS/SUBJ schemes. For either mean or covariance estimation, the OBS and SUBJ schemes are shown to be asymptotically equivalent to the optimal one when data are sufficiently non-dense or ultra-dense respectively. The new results provide theoretical insights for the satisfactory performance of the OBS and SUBJ schemes in the two special scenarios. Lastly, the approach used in this article to finding optimal weighting schemes is general and may be broadly applicable to other frameworks. Typically it is straightforward to find the optimal scheme as long as the L^2 rate of convergence that depends on weights can be achieved. For example, by applying this approach, we are able to derive the optimal weighting scheme for the coefficient estimators of varying-coefficient models (Huang et al., 2002).

2. Estimation

We employ local linear smoothers as in Zhang and Wang (2016). Denote $K_h(\cdot) = K(\cdot/h)/h$ for a one-dimensional kernel K and a bandwidth h.

The estimator of μ, denoted by $\hat{\mu}$, is given by $\hat{\mu}(t) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} \left\{ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right\}^2 K_{h_\mu}(T_{ij} - t). \tag{1}$$

Here w_i represents the general weight assigned to each observation for the ith subject. For normalization we require $\sum_{i=1}^n N_i w_i = 1$.

Assume $N_i \geq 2$ for all $i = 1, \ldots, n$ when estimating γ. With the mean estimator $\hat{\mu}$, one could obtain the residuals $Y_{ij} - \hat{\mu}(T_{ij})$ and smooth over the within-subject off-diagonal products of the residuals $R_{ijl} = \{Y_{ij} - \hat{\mu}(T_{ij})\}\{Y_{il} - \hat{\mu}(T_{il})\}$ to achieve $\hat{\gamma}$, the estimator of γ, which is given by $\hat{\gamma}(s, t) = \hat{\beta}_0$, where

$$\begin{aligned}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_0,\beta_1,\beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} v_i \sum_{1\le j\ne l\le N_i} \Big\{ & R_{ijl} - \beta_0 - \beta_1(T_{ij} - s) \\ & - \beta_2(T_{il} - t) \Big\}^2 K_{h_\gamma}(T_{ij} - s) K_{h_\gamma}(T_{il} - t). \end{aligned} \tag{2}$$

Similarly $\sum_{i=1}^n N_i(N_i - 1)v_i = 1$ for normalization.

The OBS and SUBJ weighting schemes are two special cases. The OBS scheme allocates the same weight $w_i^{obs} = 1/\sum_{i=1}^n N_i$ to each observation Y_{ij} and the same weight $v_i^{obs} = 1/\sum_{i=1}^n N_i(N_i - 1)$ to each R_{ijl}, so a subject with larger N_i receives a heavier weight. The corresponding estimators are denoted by $\hat{\mu}^{obs}$ and $\hat{\gamma}^{obs}$ respectively. The SUBJ scheme assigns different weights, $w_i^{subj} = 1/(nN_i)$ and $v_i^{subj} = 1/\{nN_i(N_i - 1)\}$, to observations from different subjects but the same total weights $1/n$ to all subjects. The corresponding estimators are denoted by $\hat{\mu}^{subj}$ and $\hat{\gamma}^{subj}$ respectively. To conform with Zhang and Wang (2016), all weights and $\{N_i : i = 1, \ldots, n\}$ are regarded as non-random but may vary with the sample size n.

3. Optimal weighting schemes

For simplicity and without loss of generality we assume that the domain of the functional data is $\mathcal{I} = [0, 1]$. Denote the L^2 norm by $\|\phi\|_2 = [\int \phi(t)^2\, dt]^{1/2}$ for any univariate function $\phi(\cdot) \in [0, 1]$, and the Hilbert–Schmidt norm by $\|\Sigma\|_{\mathrm{HS}} = [\iint \Sigma^2(s, t)\, ds\, dt]^{1/2}$ for any bivariate function $\Sigma(\cdot, \cdot) \in [0, 1]^2$.

Please cite this article in press as: Zhang X., Wang J.-L., Optimal weighting schemes for longitudinal and functional data. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.03.007.