Re-sampling techniques for estimating the distribution of descriptive statistics of functional data

Han Lin Shang[1],

ESRC Centre for Population Change, University of Southampton,
Southampton, SO17 1BJ, United Kingdom

Abstract

Re-sampling methods for estimating the distribution of descriptive statistics of functional data are considered. Through Monte-Carlo simulations, we compare the performance of several re-sampling methods commonly used for estimating the distribution of descriptive statistics. We introduce two re-sampling methods that rely on functional principal component analysis, where the scores were randomly drawn from multivariate normal distribution and Stiefel manifold. Illustrated by one-dimensional Canadian weather station data and two-dimensional bone shape data, the re-sampling methods provide a way of visualizing the distribution of descriptive statistics for functional data.

Keywords: bootstrap validity, functional mean, trimmed functional mean, functional median, functional variance, smoothed bootstrap

1. Introduction

Recent computer technology facilitates the presence of functional data, whose graphical representations are in the form of curve (Hyndman and Shang, 2010), image (Locantore et al., 1999), or shape (Epifanio and Ventura-Campos, 2011). The monographs by Ramsay and Silverman (2002, 2005) have greatly popularized the functional data analysis (FDA), offering a number of appealing case studies and presenting many parametric statistical methods. The book by Ferraty and Vieu (2006) is an excellent reference on the introduction

Email address: H.Shang@soton.ac.uk (Han Lin Shang)
[1]Tel: +44 (0) 2380 595 796; Fax: +44 (0) 2380 593 858

of many nonparametric statistical methods, such as functional kernel estimator in density estimation and regression models. A collective book by Ferraty and Romain (2011) collected the latest development of functional data analysis, covering the topics of classification, regression and theory of operator-based statistics. Additional information on FDA can be found in the web sites `http://www.psych.mcgill.ca/misc/fda/` and `http://www.math.univ-toulouse.fr/staph/npfda/`.

The increasing popularity of FDA demands new development of statistical techniques that can accommodate the continuous and smooth nature of functional data. Despite considerable progress made on probability theory in function spaces (see for example, Hall and Horowitz, 2007) and in estimation of infinite-dimensional parameter (Mas and Pumo, 2011), the methodology of FDA is far from complete as many topics remain un-explored. Nonetheless, some theoretical, methodological and practical developments have been mainly focused on functional principal component analysis (Hall and Hosseini-Nasab, 2006, 2009; Hall, 2011), functional partial least squares analysis (Preda and Saporta, 2005; Reiss and Ogden, 2007; Delaigle and Hall, 2012), functional regression models (Cardot et al., 2003; Ferraty and Vieu, 2006; Müller and Yao, 2008), functional clustering (Hall et al., 2001; Tarpey and Kinateder, 2003; Delaigle et al., 2012). These lines of research have been frequently appeared in many statistical journals, such as the special issues of *Statistica Sinica* (vol 14, issue 3), *Computational Statistics & Data Analysis* (vol 51, issue 10), *Computational Statistics* (vol 22, issue 3), and *Journal of Multivariate Analysis* (vol 101, issue 2).

In FDA, a main task is to make inference about the distribution of descriptive statistics, such as functional mean or median, from a set of realized samples. Not only it is important to obtain a consistent estimator, we are also interested in estimating the variability associated with the descriptive statistics and constructing its confidence intervals (CIs). When such a problem arises, re-sampling methodology especially bootstrapping turns out to be the only practical alternative (Cuevas et al., 2006; McMurry and Politis, 2011). Pioneering work by Mahalanobis (1946), Hartigan (1969), Efron (1979), Hall (1992), Simon (1993)

and Efron and Tibshirani (1993) show the use of bootstrap techniques to compute valid statistical measures, such as CIs, by means of computational-intensive simulations using only the observed data. In contrast to multivariate data analysis, there is comparably less work that has been done on bootstrapping functional data. Nonetheless, for instance, Giné and Zinn (1990) established consistency for the mean of Banach-space-valued random variable via bootstrapping; Politis and Romano (1994) investigated the use of bootstrap for the mean of Hilbert-space-valued time series; Cuevas et al. (2006) have proposed bootstrap methods and conducted a large scale of simulation studies. Hall and Vial (2006), Bathia et al. (2010) and Poskitt and Sengarapillai (2013) used the bootstrap methods to determine the optimal number of functional principal components. Hall et al. (2009) proposed tie-respecting bootstrap methods for estimating distributions of eigenvalue estimators. Ferraty et al. (2010) established the validity of bootstrap in nonparametric functional regression, while the asymptotic validity of a componentwise bootstrap procedure has been studied by Ferraty et al. (2012) in nonparametric functional regression with functional predictor and functional response. Ferraty and Vieu (2011) applied a residual-based bootstrap method to construct CIs of regression function in nonparametric functional regression, while González-Manteiga and Martínez-Calvo (2011) applied a residual-based bootstrap method to construct CIs of regression coefficient in functional linear regression. In addition, González-Manteiga et al. (2012) proposed bootstrap independence test for functional linear models.

We use the re-sampling methods to study the distribution of descriptive statistics of functional data. Following the early work by Cuevas et al. (2006) and McMurry and Politis (2011), we extend the scale of previous empirical study and presents two novel re-sampling techniques for visualizing the distribution of descriptive statistics of functional data. This paper by no means tries to single out the best bootstrap method, instead, it aims to contribute to the area of explanatory analysis for functional data.

The rest of this paper is given as follows. The descriptions of re-sampling methods are given in Section 2. Through Monte Carlo simulations, the bootstrap performance is compared

based on their empirical coverage probability and width of CIs in Section 3. In Sections 4 and 5, we apply the bootstrap methods to the one-dimensional Canadian weather station data and two-dimensional bone shape data, for visualizing the CIs of their descriptive statistics. Conclusions are presented in Section 6, along with some thoughts on how the methods developed here might be further extended.

2. Some re-sampling techniques

2.1. Notation

It is commonly assumed that random functions are sampled from a second-order stochastic process X in $L^2[0, \tau]$, where $L^2[0, \tau]$ is the Hilbert space of square-integrable functions on the bounded interval $[0, \tau]$. The stochastic process X satisfies the condition $\int_0^\tau \mathrm{E}(X^2(t))dt < \infty$, inner product $< f, g >= \int_0^\tau f(t)g(t)dt$ for any two functions, f and $g \in L^2[0, \tau]$ and induced squared norm $\| \cdot \|^2 =< \cdot, \cdot >$.

Although the space of its trajectories of the stochastic process X is infinite dimensional on the interval $[0, \tau]$, it is seldom observed at infinitely many data points. Instead, a finite number of data points are observed. Here, we presume that each curve is observed on a common and dense grid of data points $(t_1, t_2, \ldots, t_T)$ with $0 \leq t_1 < t_2 \cdots < t_T \leq \tau$.

2.2. Bootstrap functions

Denote n realizations of a stochastic process X evaluated at $t \in [0, \tau]$ as $\{X_1(t), X_2(t), \ldots, X_n(t)\}$. When functional data are independent and identically distributed (iid), we can obtain bootstrap sample $\{X_1^*(t), X_2^*(t), \ldots, X_n^*(t)\}$ directly, by randomly sampling with replacement from the original functions $\{X_1(t), X_2(t), \ldots, X_n(t)\}$ with the fixed t. In practice, we can only observe and evaluate X at discretized data points $0 \leq t_1 < t_2 \cdots < t_T \leq \tau$, thus the discretized bootstrap samples are obtained as $\{\boldsymbol{X}^*(t_u) = [X_1^*(t_u), X_2^*(t_u), \ldots, X_n^*(t_u)]^\top; u = 1, 2, \ldots, T\}$.

To avoid the possible appearance of repeated curves in the bootstrap samples, Cuevas et al. (2006) replaced the standard iid bootstrap samples by the so-called smooth bootstrap

samples, which are drawn from a smooth empirical distribution function. This can be achieved by adding a white noise to the bootstrap sample, expressed as

$$\boldsymbol{X}^0(t_u) = \boldsymbol{X}^*(t_u) + \boldsymbol{z}(t_u), \quad u = 1, 2, \ldots, T,$$

where $\{\boldsymbol{z}(t_u) = [z_1(t_u), z_2(t_u), \ldots, z_n(t_u)]^\top\}$; and $[\boldsymbol{z}(t_1), \boldsymbol{z}(t_2), \ldots, \boldsymbol{z}(t_T)]$ is normally distributed with mean $\boldsymbol{0}$ and covariance matrix $\beta \boldsymbol{\Sigma_X}$; and $\boldsymbol{\Sigma_X}$ is the covariance matrix of $[X(t_1), X(t_2), \ldots, X(t_T)]$. Although Cuevas et al. (2006) suggest to set $\beta = 0.05$ customarily, we chose an optimal value of β based on a holdout (training) sample in our simulation study in hope to obtain a more accurate empirical coverage probability in the testing sample.

In this paper, we put forward some other sampling techniques by using the continuous Karhunen-Loève decomposition (also known as functional principal component analysis), and then compare the sample performance among several bootstrap methods considered.

2.3. Karhunen-Loève decomposition

Among many techniques for modeling the variability of functional data, methods based on Karhunen-Loève decomposition are quite popular (see Cardot et al., 2003; Cai and Hall, 2006; Hall and Hosseini-Nasab, 2006, 2009; Hall and Horowitz, 2007; Delaigle and Hall, 2012, among others), and we have also considered this method in this paper. A Karhunen-Loève expansion of X evaluated at $t \in [0, \tau]$ is expressed by

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j \rho_j(t), \tag{1}$$

with the mean function $\mu(t) = \mathrm{E}[X(t)]$ and the basis functions $(\rho_1(t), \rho_2(t), \ldots)$ are the orthonormal eigenfunctions of the covariance kernel $\Gamma(s, t) = \mathrm{Cov}[X(s), X(t)]$. The covariance

kernel $\Gamma(s,t)$ can be expressed as

$$\int_0^\tau \Gamma(s,t)\rho_j(t)dt = \lambda_j \rho_j(s),$$

$$\Gamma(t,s) = \sum_{j=1}^\infty \lambda_j \rho_j(t)\rho_j(s),$$

where $\lambda_j \geq 0$ is a set of eigenvalues in a decreasing order, and the condition $\int_0^\tau \mathrm{E}(X^2(t))dt < \infty$ entails $\sum_{j=1}^\infty \lambda_j < \infty$. The coefficient ξ_j in (1) is given by the projection of $X - \mu$ in the direction of the jth eigenfunction ρ_j, i.e., $\xi_j =< X - \mu, \rho_j >$. The coefficients $(\xi_1, \xi_2, \dots)$ consist of uncorrelated sequences of random variables with zero mean and finite variance. Asymptotically, the distribution of ξ_j can be captured by $N(0, \lambda_j^2)$.

In practice, we will presume that each curve is observed on a grid of T points with $0 \leq t_1 < t_2 \cdots < t_T \leq \tau$. Thus, the raw data set $\{X_1(t), X_2(t), \dots, X_n(t)\}$ of n observations on X will consist of an $n \times T$ data matrix $\boldsymbol{X}$, where

$$X_s(t_u) = \mu(t_u) + \sum_{j=1}^\infty \xi_j \rho_j(t_u),$$

where $s = 1, 2, \dots, n$ and $u = 1, 2, \dots, T$. The mean centered data matrix is given by $\mathbf{X} = \boldsymbol{X} - \sum_{s=1}^n X_s/n$. Then, a standard approach to estimate the covariance kernel is to take

$$\widehat{\Gamma}(t_u, t_v) = \frac{1}{n}\sum_{s=1}^n [X_s(t_u) - \bar{X}(t_u)][X_s(t_v) - \bar{X}(t_v)],$$

as an estimator of $\Gamma(t_u, t_v)$, where $\bar{X}(t_u) = \frac{1}{n}\sum_{s=1}^n X_s(t_u)$.

By applying the singular value decomposition, $\mathbf{X}$ can be decomposed into

$$\mathbf{X} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{R}^\top, \tag{2}$$

where $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{I}_m$ and $m = \mathrm{rank}(\mathbf{X}) = \min(n, T)$, the diagonal matrix $\boldsymbol{L} = \mathrm{diag}(\sqrt{l_1}, \sqrt{l_2}, \dots, \sqrt{l_m})$ is a set of non-negative singular values, and $\{l_1, l_2, \dots, l_m\}$ lists the

positive eigenvalues of $\mathbf{XX}^\top/n$ in decreasing order. The columns of $\boldsymbol{U}$ are the normalized eigenvectors of $\mathbf{XX}^\top/n$, and the columns of $\boldsymbol{R}$ are the normalized eigenvectors of $\mathbf{X}^\top\mathbf{X}/n$. Equation (2) provides a discrete counterpart to the Karhunen-Loève expansion given in (1) in that a curve in X can be written as

$$X_s(t_u) = \bar{X}(t_u) + \sum_{j=1}^{m} u_{sj}\sqrt{l_j}r_j(t_u). \tag{3}$$

Poskitt and Sengarapillai (2013, Lemmas 1-3) respectively showed that $\bar{X}$ is a uniformly consistent estimate of μ; l_j provides a uniformly consistent estimate of λ_j; and r_j provides a uniformly consistent estimate of ρ_j. Similar types of derivations and theoretical arguments can also be found in Yao et al. (2005, Section 3).

2.4. Bootstrap functional principal component scores

From (2) and (3), the re-sampling procedure first holds the mean $\bar{X}(t_u)$, $\boldsymbol{L}$ and $\boldsymbol{R}$ fixed at their realized values, the following re-sampling methods differ mainly by the ways of re-sampling $\boldsymbol{U}$.

a) Obtain the re-sampled singular column matrix $\boldsymbol{U}^* = [\boldsymbol{u}_1^*, \boldsymbol{u}_2^*, \ldots, \boldsymbol{u}_{\min(n,T)}^*]$ by randomly sampling with replacement from the original principal component scores $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{\min(n,T)}]$.

b) To avoid the possible appearance of repeated values in $\boldsymbol{U}^*$, we adapt a smooth bootstrap procedure by adding a white noise component to the bootstrap $\boldsymbol{U}$. More precisely, $\boldsymbol{U}^0 = \boldsymbol{U}^* + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix of $\beta\boldsymbol{\Sigma_U}$. Note that $\boldsymbol{\Sigma_U} = \boldsymbol{I}_{\min(n,T)}$ is the covariance matrix of $\boldsymbol{U}$.

c) Because $\boldsymbol{U}$ follows a standard multivariate normal distribution asymptotically, we can randomly draw $\boldsymbol{U}^*$ from a multivariate normal distribution with mean vector and covariance matrix of $\boldsymbol{U}$.

d) Recall that $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_{\min(n,T)}$, therefore $\boldsymbol{U}$ can be considered as points on the Stiefel manifold $\boldsymbol{V}_{n,T}$. As defined in Chikuse (2003), the (compact) Stiefel manifold $\boldsymbol{V}_{n,T}$ is the space whose points (denoted by $u_1,\ldots,u_n$) are n orthonormal vectors in $R^T (n \leq T)$, that is, $< u_i, u_j > = \delta_{ij}, i,j = 1,\ldots,n$ and δ_{ij} is the Kronecker product. James (1976, Chapter 2) provides a good introduction to the geometry of the Stiefel manifold, while Chikuse (2003) put forward gradient methods on the Stiefel manifold in the context of eigenvalue problems. Following the early work by Hoff (2009), we propose to sample $\boldsymbol{U}^*$ from the Stiefel manifold. A computational algorithm in ® language (R Development Core Team, 2012) is presented in the appendix.

For $s = 1, 2, \ldots, n$, the realization $X_s^*(t_u), u = 1, 2, \ldots, T$ is constructed as in (2) and (3) by replacing $\boldsymbol{U}$ with the bootstrapped $\boldsymbol{U}^*$ or smoothed bootstrapped $\boldsymbol{U}^0$. In this paper, we utilize all components of $\boldsymbol{U}$.

3. Simulation study

In this paper, the re-sampling techniques are designed to estimate distribution of descriptive statistics of functional data. In Section 3.1, we introduce a number of descriptive statistics that characterize a set of functional data. Section 3.2 introduces two simulation examples of Gaussian process, while the evaluation metrics are given in Section 3.3. Section 3.4 presents the simulation results, where the performances of different sampling techniques are evaluated and compared, based on their empirical coverage probability and width of CIs.

3.1. Descriptive statistics of functional data

Similar to univariate point estimation, we seek an estimator of functional median, which also allows us to rank a sample of curves based on their location depth, that is, the distance from the functional median (the deepest curve). This leads to the notion of functional depth (see, for example, Fraiman and Muniz, 2001; Cuevas et al., 2006, 2007; Febrero et al., 2007, 2008; Lopéz-Pintado and Romo, 2009; Cuesta-Albertos and Nieto-Reyes, 2010; Gervini, 2012).

In what follows, we shall briefly describe three functional depth measures considered, namely the Fraiman and Muniz (2001) depth, Cuevas et al. (2007) depth based on random projection, and Gervini (2012) depth based on small ball property.

The Fraiman and Muniz (2001) depth was the oldest functional depth measure. For each $t \in [0, \tau]$, let $F_{n,t}$ be the empirical sample distribution of $\{X_1(t), X_2(t), \ldots, X_n(t)\}$ and let $Z_i(t)$ be the univariate depth of function $X_i(t)$, given by

$$I_i = \int_0^\tau Z_i(t)dt = \int_0^\tau 1 - \left| \frac{1}{2} - F_{n,t}(X_i(t)) \right| dt, \tag{4}$$

and the values of I_i provide a way of ranking curves from inward to outward. Thus, the functional median is the deepest curve with the maximum I_i value.

Cuevas et al. (2007) considered a random direction a (typically simulated from standard Gaussian distribution) and project the data along this direction by the inner product. The functional depth of $X_i(t)$, for $i = 1, \ldots, n$, is then defined as the univariate depth given in (4) of the one-dimensional projection $< a, X_i >$. This functional depth by random projection gives a random measure of depth, as it relies on the rank of the projections along a random direction (Cuevas et al., 2007). In order to reduce variability, the functional depth of each function can be evaluated by averaging the functional depths over a number (typically 50) of random directions.

Furthermore, Gervini (2012) considered the set of interdistances between any two functions, denoted by $\{d(X_i, X_j)\}$. An observation X_i is an outlier if it is far from most other observations. Given $\alpha \in [0, 1]$, Gervini (2012) defined α-radius r_i as the distance between $X_i(t)$ and the $\lceil \alpha n \rceil^{\text{th}}$ closest observations, where $\lceil x \rceil$ denotes the integer closest to x from above. The rank of r_i provides a measure of outlyingness of $X_i(t)$. The smaller the r_i is, the more dense the X_i is. The functional median has the smallest r_i.

In addition to the functional median, we also consider sample versions of functional mean

and functional variance, given by

$$\bar{X}(t) = \frac{1}{n} \sum_{s=1}^{n} X_s(t),$$

$$\widehat{V}(t) = \frac{1}{n-1} \sum_{s=1}^{n} [X_s(t) - \bar{X}(t)]^2,$$

and α-trimmed functional mean that is the mean function of the $100(1 - \alpha)\%$ most deepest curves. The functional trimmed mean is expressed as

$$\frac{1}{n - \lceil \alpha n \rceil} \sum_{i=1}^{n - \lceil \alpha n \rceil} X_i(t), \qquad \alpha \in [0, (n-1)/n],$$

where α is the amount of trimming, and $(X_{(1)}, X_{(2)}, \ldots, X_{(n - \lceil \alpha n \rceil)})$ are the ordered sample curves based on their increasing location depth.

3.2. Simulation setup

A Monte Carlo simulation study is implemented to evaluate and compare the performance of the bootstrap methods, for estimating the distribution of descriptive statistics of functional data in Section 3.1. For comparison, we consider the same two functional models as previously studied in Cuevas et al. (2006):

(a) A Wiener process with trend, expressed by $X(t) = m(t) + B(t)$, where $B(t)$ is a standard Brownian motion, $m(t) = 0.95 \times 10t(1-t) + 0.05 \times 30t(1-t)$ and $\mathrm{Var}(X(t)) = \mathrm{Var}(B(t))$.

(b) A Gaussian process $X(t)$ with mean $m(t) = 0.95 \times 10t(1 - t) + 0.05 \times 30t(1 - t)$, $\mathrm{Cov}(X(t_i), X(t_j)) = \exp(-|t_i - t_j|/0.3)$, and $\mathrm{Var}(X(t)) = 1$.

In practice, we can only evaluate and compare the bootstrap methods on a common set of grid points. In our simulation study, we have taken 101 equally-spaced grid points between 0 and 1, for two different sample sizes $n = 25$ and $n = 100$.

Since both models (a) and (b) are Gaussian processes, computationally, we can draw samples from a multivariate normal distribution with mean $[m(t_1), m(t_2), \ldots, m(t_T)]^\top$ and

covariance matrix $\text{Cov}(X(t_i), X(t_j)) = \min\{t_i, t_j\} = t_i$ for model (a) and $\text{Cov}(X(t_i), X(t_j)) = \exp(-|t_i - t_j|/0.3) = 1$ for model (b).

3.3. Simulation evaluation

To evaluate the performance of the bootstrap methods, we calculate the bootstrap CIs of a descriptive statistics. Given the raw data $\{X_1, X_2, \ldots, X_n\}$, we draw $R = 200$ replications of bootstrap samples; and the same random seed is used for all the methods in order to give the same simulation randomness. For each replication, the $100(1 - \alpha)\%$ bootstrap CIs of a descriptive statistics $T(X_1, X_2, \ldots, X_n)$, is defined by calculating the cut-off value $D(X_1, X_2, \ldots, X_n)$, such that the $100(1 - \alpha)\%$ of the bootstrap repetitions $T(X_1^*, X_2^*, \ldots, X_n^*)$ are within a distance smaller than $D(X_1, X_2, \ldots, X_n)$. In our simulation study, the performance of bootstrap CIs is evaluated through the 200 replications, the corresponding CIs are constructed based on $B = 150$ repetitions for each replication.

We calculate the coverage probability that the target function at the population level lies within the CIs. Two distance metrics, L^2 and L^∞, are used for constructing CIs, these are

$$\|T_{\text{boot}} - T_{\text{sample}}\|_2 = \left(\int (T_{\text{boot}} - T_{\text{sample}})^2 \right)^{\frac{1}{2}},$$

$$\|T_{\text{boot}} - T_{\text{sample}}\|_\infty = \sup |T_{\text{boot}} - T_{\text{sample}}|.$$

Apart from the coverage probability, the performance of the bootstrap methods is also assessed by the width of bootstrap CIs, which is the range of distances covered for each bootstrap CIs (see also Cuevas et al., 2006). For a given confidence level (customarily $\alpha = 0.05$), a smaller value of the width corresponds to a more informative of the CIs.

3.4. Simulation results

With two models (a) and (b), L^2 and L^∞ are the metrics used to measure the distance between the bootstrapped and sample descriptive statistics of simulated functional data. For simplicity, we shall use the abbreviations of the bootstrap methods as shown in Table 1.

[Table 1 about here.]

In Table 2, we report the coverage probability of bootstrap CIs for different descriptive statistics of functional data and bootstrap methods. As the sample size increases from $n = 25$ to 100, the empirical coverage probability generally improves for all the bootstrap methods. This is not surprising, as the validity of bootstrapping relies on a moderate or large sample size (see also, McMurry and Politis, 2011). Subject to the same random seed, the St and StU methods give the same empirical coverage probability. For estimating functional median, the higher empirical coverage probabilities of the bootstrap methods reflect the robustness of functional median. For estimating the 5% trimmed functional mean, the empirical coverage probabilities of the smoothed bootstrap methods outperform their non-smoothed counterparts, using three different depth measures. This is reflected in model (a), where the smoothing parameter β selected from the training set is large (close to 1). On the contrary, when β is small, the differences in coverage probability between the un-smoothed and smoothed bootstrap methods are small (as shown in model (b)).

As noted by an anonymous referee, the performance of the trimmed mean, based on St (α-radius) is much worse for $n = 100$ than for $n = 25$ in model (a). A possible explanation is that there are some tuning parameters in the calculation of α-radius depth, and they need to be determined. The results presented in Table 2 were obtained using the default tuning parameters. A data-driven selection of these tuning parameters may improve the coverage probabilities.

[Table 2 about here.]

Apart from the comparison of coverage probability, we also calculate the width of CIs. For a given coverage probability, a smaller value of the width corresponds to a more informative of the CIs. Based on the 200 replications and 150 bootstrap repetitions within each replication, we have evaluated the width of the CIs for every considered bootstrap method, functional estimator, distance metric, sample size and underlying model. As an illustration, Figure 1

plots two histograms of the range of CIs using the standard bootstrap method under model (a) with sample size $n = 100$.

[Figure 1 about here.]

The $192(6 \times 4 \times 2 \times 2 \times 2)$ histograms for all the considered cases are summarized in Tables 3 and 4. The numbers in the cells of Table 3 represent the estimated mode value of the CI width, that is, the most frequently occurred value of the interval width. The table also displays the range of bootstrapped CIs, defined as the support of the respective histograms in model (a). Table 4 shows the estimated mode value and range of the CI width in model (b). From the simulation output, there is an improvement in precision, as the width of CIs becomes narrower when sample size increases from $n = 25$ to 100 (McMurry and Politis, 2011). Subject to the same random seed, the St and StU methods provide the same interval width. As shown in Table 3, better empirical coverage probabilities of the smoothed bootstrap methods (with large values of the smoothing parameter) are at the cost of their wider CI width. In contrast, the differences in precision between the un-smoothed and smoothed bootstrap methods are small as shown in Table 4.

[Table 3 about here.]

[Table 4 about here.]

4. Application to Meteorology

The Canadian weather station data is one of the classical functional data sets, which is available at the *fda* package (Ramsay et al., 2012). This data set has been studied by Ramsay and Silverman (2002, 2005), Ramsay et al. (2009) and James et al. (2009) in the areas of explanatory analysis and regression analysis for functional data.

Figure 2 plots the change in temperature over the course of a year, taken from 35 weather stations across Canada. The functional curves are interpolated from 365 data points, which

13

measure the daily mean temperature recorded by a weather station averaged over the period from 1960 to 1994. The colors correspond to the geographic climates of stations. The red lines show the weather stations located at the warmer regions, whereas the purple lines show the weather stations located at the colder regions.

[Figure 2 about here.]

Our aim is to utilize a bootstrap method to visualize the distribution of descriptive statistics of the Canadian weather station data. As an illustration, we used the smoothed bootstrap principal component score method with $\beta = 0.05$ to plot the 95% CIs of functional mean, functional median, 5% trimmed functional mean and functional variance.

In Figure 3, we plot the distribution of descriptive statistics for the Canadian weather station. While the sample estimates are shown in red, their 95% CIs are shown in blue. There is a 95% chance that the true population mean, median, 5% trimmed mean and variance are within the constructed CIs.

[Figure 3 about here.]

5. Application to Paleopathology

While Section 4 focus on the one-dimensional functional curves, this section considers two-dimensional functional shapes in the field of paleopathology. Paleopathology is the study of disease in human history, incorporating information that can be gathered from human skeletal remains. Ramsay and Silverman (2002, Chapter 4) analyzed the shapes of 68 bones from hundreds years ago, in an attempt to gain insights into a possible relationship between bone shape and osteoarthritis of the knee. The paleopathologists attempted to identify every person in the sample with definite signs of osteoarthritis of the knee, as evidenced by eburnation — polished bone surface caused by complete cartilage loss. This left 16 eburnated femora and 52 non-eburnated femora for analysis.

14

We concentrate on images of the knee end of the femur (the upper leg bone). These images were constructed from the x and y coordinates of 12 landmarks. The shape of a bone can be reasonably well approximated by interpolating through the coordinates of the 12 landmarks. These landmarks have been centered. Because the size and orientation of the bones are of no particular interest, we have performed a procrustes rotation to eliminate size and orientation variabilities, so that various bones are fitted together as closely as possible (Ramsay and Silverman, 2002, Chapters 4 and 8). The bone shapes characterized by the positions of 12 landmarks are plotted in Figure 4, for the 16 eburnated femora and 52 non-eburnated femora.

[Figure 4 about here.]

Our aim is to apply a bootstrap method to visualize the distribution of descriptive statistics of the bone shape data. As an illustration, we implemented the standard bootstrap score method to plot the 95% CIs of the functional mean and functional median, for both the eburnated and non-eburnated femora.

In Figure 5, we plot the distribution of some descriptive statistics for the bone shape data. While the sample estimates are shown in red, their 95% confidence region is shown by the upper and lower black circles in two dimensions. There is a 95% chance that the true population mean and median are within the constructed confidence region, for both the eburnated femora and non-eburnated femora.

[Figure 5 about here.]

6. Conclusion

In this paper, we present some re-sampling procedures to visualize the distribution of descriptive statistics of functional data. Through Monte-Carlo simulation, we demonstrate the better empirical coverage probability and narrower width of CIs, as the sample size increases from 25 to 100. This phenomenon applies to all the bootstrap procedures investigated, regardless of their differences in bootstrapping original functions or bootstrapping

functional principal component scores. However, as the dimension $T \to \infty$, bootstrapping smoothed principal component scores outperforms bootstrapping smoothed functions in terms of empirical coverage probability, which may due to better numerical stability in estimating the covariance structure. By truncating the number of retained principal components, the methods of bootstrapping principal component scores and smoothed principal component scores are also capable of eliminating possible noise in data, and this feature has not been explored in this paper.

Illustrated by the two empirical applications, it is shown that the bootstrap procedures provide an explanatory tool to the functional data analysis toolbox. It is expected that the bootstrap methods will receive increasing popularity in functional data analysis, where the object of interest is on the distribution of functional estimators.

The present study is limited to the bootstrap methods applied to iid functional data. However, the presence of functional time series is not uncommon, such as in modeling and forecasting ozone concentration (Damon and Guillas, 2002), demographic rates (Hyndman and Shang, 2009), and term structure of the Eurodollar futures rate (Kargin and Onatski, 2008). In some recent surveys, Bühlmann (2002), Politis (2003) and Kreiss and Paparoditis (2011) revisited some bootstrap methods, such as moving block bootstrap and sieve bootstrap, which are mainly applied to univariate or multivariate dependent data. In the future research, we intend to extend bootstrap methods for functional time series.

Another possible research direction is to propose an efficient algorithm, for implementing a high-order bootstrapping for functional data. Advances in statistical and econometric theory show that iterating the bootstrap principle brings further refinements upon the single bootstrap (Beran, 1988; Hall, 1986, 1992). Iterating the bootstrap principle reduces the dependence between the probability distribution of the resample and the unknown data generating process. Therefore, double bootstrap has typically higher order accuracy than single bootstrap, but at a much higher computational cost. Although a fast double bootstrap algorithm has been developed by Davidson and MacKinnon (2007), its extension to functional

data remains a future research.

Acknowledgements

Appendix: Stiefel-manifold algorithm in R

```
library(MASS)

rstiefel = function(data)
{
    # data is a (n ×  p) matrix

    U = svd(data)$u

    n = dim(U)[1]

    X = mvrnorm(n, colMeans(U), cov(U))

    tmp = eigen(t(X)%*%X)

    # Let the eigenvectors (tmp$vec) be Q, and the eigenvalues (tmp$val) be S,

    # If matrix X^T X is eigen-decomposable and none of its eigenvalues is 0,

    # then (X^T X)^{-1/2} = Q% * %(diag(S)^{-1/2})% * %Q^{-1}, where Q^{-1} = Q^T.

    Ustar = X%*%(tmp$vec%*%sqrt(diag(1/tmp$val))%*%t(tmp$vec))

    return(Ustar)
}
```

References

Bathia, N., Yao, Q., and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38(6):3352–3386.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83(403):687–697.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17(1):52–72.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159–2179.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13(3):571–591.

Chikuse, Y. (2003). *Statistics on Special Manifolds*. Number 174 in Lecture Notes in Statistics. Springer, New York.

Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2010). Functional classification and the random Tukey depth. Practical issues. In Borgelt, C., Rodriguez, G. G., Trutschnig, W., Lubiano, M. A., Gil, M., Grzegorzewski, P., and Hryniewicz, O., editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77, pages 123–130. Advances in Intelligent and Soft Computing.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51(2):1063–1074.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3):481–496.

Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, 13(7):759–774.

Davidson, R. and MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics and Data Analysis*, 51(7):3259–3281.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1):322–352.

Delaigle, A., Hall, P., and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika*, 99(2):299–313.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Epifanio, I. and Ventura-Campos, N. (2011). Functional data analysis in shape analysis. *Computational Statistics and Data Analysis*, 55(9):2758–2773.

Febrero, M., Galeano, P., and Gonzalez-Manteiga, W. (2007). A functional analysis of NOx levels: location and scale estimation and outlier detection. *Computational Statistics*, 22(3):411–427.

Febrero, M., Galeano, P., and Gonzalez-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics*, 19(4):331–345.

Ferraty, F. and Romain, Y., editors (2011). *The Oxford Handbook of Functional Data Analysis.* Oxford University Press, Oxford.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*, 37(2):286–306.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109:10–28.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, New York.

Ferraty, F. and Vieu, P. (2011). Kernel regression estimation for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis.* Oxford University Press, Oxford.

Fraiman, R. and Muniz, G. (2001). Trimmed mean for functional data. *Test*, 10(2):419–440.

Gervini, D. (2012). Outlier detection and trimmed estimation for general functional data. *Statistica Sinica*, 22(4):1639–1660.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability*, 18(2):851–869.

González-Manteiga, W., González-Rodríguez, G., Martínez-Calvo, A., and García-Portugués, E. (2012). Bootstrap independence test for functional linear models. Working paper, University of Santiago de Compostella. http://arxiv.org/pdf/1210.1072v2.pdf.

González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference*, 141(1):453–461.

Hall, P. (1986). On the bootstrap and confidence intervals. *The Annals of Statistics*, 14(4):1431–1452.

Hall, P. (2011). Principal component analysis for functional data: methodology, theory, and discussion. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis.* Oxford University Press, Oxford.

Hall, P. and Horowitz, J. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, 68(1):109–126.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, 146(1):225–256.

Hall, P., Lee, Y. K., Park, B. U., and Paul, D. (2009). Tie-respecting bootstrap methods for estimating distributions of sets and functions of eigenvalues. *Bernoulli*, 15(2):380–401.

Hall, P., Poskitt, D. S., and Presnell, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics*, 43(1):1–9.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society (Series B)*, 68(4):689–705.

Hall, P. G. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

Hartigan, J. A. (1969). Using subsample values as typical values. *Journal of the American Statistical Association*, 64(328):1303–1317.

Hoff, P. D. (2009). Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18(2):438–456.

Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series (with discussions). *Journal of the Korean Statistical Society*, 38(3):199–211.

Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45.

James, G. M., Wang, J., and Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics*, 37(5A):2083–2108.

James, I. M. (1976). *The Topology of Stiefel Manifolds*. Number 24 in Lecture Note Series. Cambridge University Press, London Mathematical Society.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99(10):2508–2526.

Kreiss, J. and Paparoditis, E. (2011). Bootstrap methods for dependent data: a review (with discussion). *Journal of the Korean Statistical Society*, 40(4):357–395.

Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., and Cohen, K. (1999). Robust principal component analysis for functional data. *Test*, 8(1):1–73.

Lopéz-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.

Mahalanobis, P. C. (1946). On large-scale sample surveys. *Philosophical Transactions of the Royal Society of London, Series B*, 231(584):329–451.

Mas, A. and Pumo, B. (2011). Linear processes for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, Oxford.

McMurry, T. and Politis, D. N. (2011). Resampling methods for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, Oxford.

Müller, H. and Yao, F. (2008). Functional additive models. *Journal of the American Statistical Association*, 103(484):1534–1544.

Politis, D. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230.

Politis, D. and Romano, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 4(2):461–476.

Poskitt, D. S. and Sengarapillai, A. (2013). Description length and dimensionality reduction in functional data analysis. *Computational Statistics and Data Analysis*, 58:98–113.

Preda, C. and Saporta, G. (2005). Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 49(1):99–108.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, 2nd edition.

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2012). *fda: Functional Data Analysis*. R package version 2.3.2. http://CRAN.R-project.org/package=fda.

Reiss, P. and Ogden, T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996.

Simon, J. (1993). *Resampling: the New Statistics*. Duxbury, Belmont, CA.

Tarpey, T. and Kinateder, K. K. J. (2003). Clustering functional data. *Journal of Classification*, 20(1):93–114.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Table 1: Abbreviations of the bootstrap methods. St and Sm are described in Section 2.2, while the rest are described in Section 2.4.

St	=	Standard function bootstrap
Sm	=	Smoothed function bootstrap
StU	=	Standard score bootstrap
SmU	=	Smoothed score bootstrap
StGU	=	Standard Gaussian-distributed score bootstrap
StiefelU	=	Random Stiefel manifold score bootstrap

Table 2: Coverage probabilities for the bootstrapped CIs based on $B = 150$ repetitions and 200 replications, at the nominal coverage probability of 95%.

Estimator and bootstrap	Model (a) Distance				Model (b) Distance			
	L^2		L^∞		L^2		L^∞	
	n				n			
	100	25	100	25	100	25	100	25
Mean								
St	0.920	0.880	0.935	0.900	0.945	0.935	0.940	0.925
Sm	0.950	0.975	0.995	0.980	0.945	0.940	0.950	0.930
StU	0.920	0.880	0.935	0.900	0.945	0.935	0.940	0.925
SmU	0.980	0.980	0.990	0.975	0.945	0.920	0.960	0.925
StGU	0.915	0.885	0.945	0.915	0.945	0.935	0.950	0.945
StiefelU	0.970	0.920	0.975	0.950	0.980	0.955	0.995	0.970
Median								
St	0.970	0.940	0.985	0.975	1.000	0.995	1.000	1.000
Sm	0.965	0.955	0.985	0.990	1.000	1.000	1.000	1.000
StU	0.970	0.940	0.985	0.975	1.000	0.995	1.000	1.000
SmU	0.970	0.950	0.980	0.985	1.000	1.000	1.000	1.000
StGU	1.000	0.970	1.000	1.000	1.000	1.000	1.000	1.000
StiefelU	1.000	0.985	1.000	1.000	1.000	1.000	1.000	1.000
5% Trimmed mean								
St (Fraiman-Muniz)	0.780	0.915	0.795	0.930	0.960	0.935	0.975	0.935
St (Random projection)	0.795	0.965	0.795	0.980	0.980	0.985	0.985	0.990
St (α-radius)	0.450	0.890	0.490	0.930	0.945	0.970	0.960	0.970
Sm (Fraiman-Muniz)	0.945	0.975	0.970	0.965	0.960	0.935	0.980	0.940
Sm (Random projection)	0.940	0.925	0.975	0.940	0.975	0.970	0.980	0.980
Sm (α-radius)	0.920	0.915	0.940	0.925	0.945	0.965	0.955	0.975
StU (Fraiman-Muniz)	0.780	0.915	0.795	0.930	0.960	0.935	0.975	0.935
StU (Random projection)	0.795	0.965	0.795	0.980	0.980	0.985	0.985	0.990
StU (α-radius)	0.450	0.890	0.490	0.930	0.945	0.970	0.960	0.970
SmU (Fraiman-Muniz)	0.940	0.960	0.975	0.970	0.945	0.935	0.965	0.945
SmU (Random projection)	0.940	0.925	0.975	0.945	0.975	0.960	0.975	0.975
SmU (α-radius)	0.915	0.920	0.935	0.935	0.955	0.965	0.960	0.975
StGU (Fraiman-Muniz)	0.925	0.930	0.940	0.940	0.970	0.935	0.975	0.960
StGU (Random projection)	0.940	0.935	0.950	0.960	0.980	0.965	0.970	0.965
StGU (α-radius)	0.955	0.955	0.970	0.965	0.980	0.980	0.980	0.975
StiefelU (Fraiman-Muniz)	0.960	0.945	0.975	0.970	1.000	0.980	1.000	0.985
StiefelU (Random projection)	0.970	0.985	0.990	0.980	0.995	0.985	1.000	0.995
StiefelU (α-radius)	0.990	0.980	0.995	0.980	0.995	0.990	1.000	1.000
Variance								
St	0.905	0.900	0.910	0.910	0.925	0.860	0.980	0.960
Sm	0.950	0.985	0.935	0.990	0.940	0.930	0.980	0.970
StU	0.905	0.900	0.910	0.910	0.925	0.860	0.980	0.960
SmU	0.940	0.985	0.945	0.935	0.945	0.880	0.980	0.950
StGU	0.925	0.905	0.930	0.935	0.935	0.915	0.965	0.995
StiefelU	0.900	0.900	0.920	0.910	0.945	0.875	0.975	0.980

Table 3: Mode and range of the bootstrapped CI widths under model (a), based on $B = 150$ repetitions and 200 replications.

Estimator and bootstrap	Model (a) at the nominal coverage probability of 95%							
	Distance							
	L^2				L^∞			
	n							
	100		25		100		25	
	mode	range	mode	range	mode	range	mode	range
Mean								
St	2.36	1.54-4.25	3.06	2.14-10.28	0.31	0.22-0.61	0.62	0.36-1.56
Sm	2.81	1.78-5.20	7.68	2.52-13.57	0.40	0.25-0.76	0.99	0.42-1.76
StU	2.36	1.54-4.25	3.06	2.14-10.28	0.31	0.22-0.61	0.62	0.36-1.56
SmU	3.47	1.69-6.11	8.66	2.63-15.30	0.43	0.27-0.74	0.81	0.36-1.84
StGU	2.63	1.60-4.51	4.03	2.16-10.80	0.33	0.23-0.59	0.66	0.36-1.45
StiefelU	3.32	1.70-5.42	4.35	2.27-11.72	0.36	0.24-0.73	0.89	0.40-1.53
Median								
St	3.60	3.00-7.06	5.62	3.82-13.78	1.06	0.78-1.79	1.23	0.92-2.56
Sm	3.79	3.19-7.84	6.16	3.85-13.17	0.99	0.78-2.02	1.38	1.03-2.74
StU	3.60	3.00-7.06	5.62	3.82-13.78	1.06	0.78-1.79	1.23	0.92-2.56
SmU	3.39	3.03-6.95	6.07	4.20-21.46	0.96	0.76-1.78	1.43	1.04-4.03
StGU	3.51	2.25-7.32	5.91	3.12-15.57	0.89	0.63-1.85	1.24	0.78-2.36
StiefelU	3.44	2.53-9.04	5.77	3.81-17.32	0.98	0.68-1.85	1.36	0.87-3.07
5% Trimmed mean								
St (Fraiman-Muniz)	2.34	1.35-6.28	3.83	2.65-16.95	0.34	0.21-0.93	0.88	0.41-2.41
St (Random projection)	3.62	1.80-6.57	4.27	3.32-16.34	0.45	0.25-0.89	0.65	0.50-2.21
St (α-radius)	2.71	1.68-6.01	4.73	3.07-17.33	0.35	0.24-0.81	0.81	0.47-2.45
Sm (Fraiman-Muniz)	4.15	2.47-6.88	7.27	2.95-15.77	0.57	0.33-0.91	0.99	0.49-2.05
Sm (Random projection)	4.19	2.61-8.27	5.56	4.15-19.59	0.66	0.35-1.12	0.98	0.68-2.73
Sm (α-radius)	3.28	2.76-7.39	8.09	4.20-20.41	0.58	0.38-1.03	1.10	0.75-2.90
StU (Fraiman-Muniz)	2.34	1.35-6.28	3.83	2.65-16.95	0.34	0.21-0.93	0.88	0.41-2.41
StU (Random projection)	3.62	1.80-6.57	4.27	3.32-16.34	0.45	0.25-0.89	0.65	0.50-2.21
StU (α-radius)	2.71	1.68-6.01	4.73	3.07-17.33	0.35	0.24-0.81	0.81	0.47-2.45
SmU (Fraiman-Muniz)	4.22	2.42-8.03	6.26	3.53-16.86	0.56	0.34-1.15	1.16	0.47-1.91
SmU (Random projection)	4.06	2.53-7.68	6.25	3.83-17.94	0.65	0.39-1.02	1.04	0.63-2.62
SmU (α-radius)	4.00	2.79-8.57	5.80	3.90-19.67	0.57	0.37-1.12	1.19	0.76-2.71
StGU (Fraiman-Muniz)	4.10	1.92-6.35	7.60	2.40-12.48	0.56	0.26-0.82	0.84	0.40-1.68
StGU (Random projection)	3.52	2.19-6.82	6.24	2.95-15.08	0.62	0.31-0.94	1.09	0.49-2.03
StGU (α-radius)	4.92	2.65-7.61	6.77	2.91-15.46	0.57	0.36-1.08	0.89	0.51-2.28
StiefelU (Fraiman-Muniz)	3.39	2.18-6.23	6.69	2.46-14.03	0.52	0.30-0.86	0.72	0.46-1.92
StiefelU (Random projection)	4.67	2.79-8.04	5.18	3.45-16.76	0.73	0.38-1.06	0.79	0.53-2.19
StiefelU (α-radius)	5.74	3.12-9.08	3.85	3.62-17.38	0.61	0.42-1.25	1.21	0.56-2.35
Variance								
St	1.71	0.95-4.76	2.94	1.19-8.62	0.34	0.21-0.88	0.72	0.27-1.98
Sm	2.85	1.37-4.95	5.01	2.46-14.25	0.40	0.24-0.81	1.32	0.49-3.24
StU	1.71	0.95-4.76	2.94	1.19-8.62	0.34	0.21-0.88	0.72	0.27-1.98
SmU	1.62	1.24-4.15	5.44	3.03-12.57	0.43	0.26-0.89	0.81	0.44-2.06
StGU	1.68	1.15-3.58	3.73	1.66-9.94	0.41	0.21-0.88	0.85	0.39-1.97
StiefelU	1.76	1.12-3.82	4.11	1.58-9.03	0.35	0.23-0.75	0.65	0.36-1.85

Table 4: Mode and range of the bootstrapped CI widths under model (b), based on $B = 150$ repetitions and 200 replications.

| Estimator and bootstrap method	Model (b) at the nominal coverage probability of 95% Distance							
	L^2				L^∞			
	n							
	100		25		100		25	
	mode	range	mode	range	mode	range	mode	range
Mean								
St	1.50	1.04-2.90	2.77	2.18-6.07	0.27	0.20-0.49	0.55	0.43-0.96
Sm	1.65	1.10-2.65	3.21	1.86-5.30	0.28	0.21-0.45	0.63	0.41-0.96
StU	1.50	1.04-2.90	2.77	2.18-6.07	0.27	0.20-0.49	0.55	0.43-0.96
SmU	1.67	1.08-3.12	3.14	2.06-5.59	0.29	0.22-0.43	0.61	0.46-1.04
StGU	1.42	1.15-2.95	2.78	2.34-5.59	0.27	0.21-0.48	0.63	0.44-1.01
StiefelU	1.86	1.27-2.98	3.80	2.45-6.73	0.32	0.23-0.47	0.71	0.47-1.05
Median								
St	7.30	6.35-11.36	10.33	7.70-18.36	2.14	1.60-3.99	2.75	1.89-5.46
Sm	7.74	6.19-10.84	12.22	7.68-17.59	1.88	1.62-4.03	2.69	2.11-5.41
StU	7.30	6.35-11.36	10.33	7.70-18.36	2.14	1.60-3.99	2.75	1.89-5.46
SmU	6.86	6.12-11.35	12.03	7.65-16.57	2.24	1.64-3.99	2.88	2.00-4.16
StGU	3.75	2.73-6.34	5.50	4.35-10.01	1.53	1.14-2.76	1.95	1.24-3.32
StiefelU	4.01	3.52-8.16	6.08	5.15-12.76	1.42	1.23-3.23	1.95	1.39-4.09
5% Trimmed mean								
St (Fraiman-Muniz)	2.03	1.35-3.36	3.57	2.44-7.39	0.30	0.23-0.51	0.69	0.48-1.13
St (Random projection)	2.08	1.64-3.58	5.32	2.91-8.71	0.36	0.29-0.55	0.86	0.58-1.46
St (α-radius)	2.50	1.61-3.63	4.40	3.34-9.38	0.38	0.27-0.51	0.79	0.69-1.58
Sm (Fraiman-Muniz)	1.74	1.12-3.38	2.96	2.76-7.81	0.32	0.21-0.49	0.69	0.45-1.09
Sm (Random projection)	2.25	1.58-3.63	4.04	3.13-7.75	0.36	0.29-0.54	0.85	0.61-1.20
Sm (α-radius)	2.09	1.66-3.80	4.67	3.21-8.43	0.37	0.30-0.69	0.93	0.62-1.25
StU (Fraiman-Muniz)	2.03	1.35-3.36	3.57	2.44-7.39	0.30	0.23-0.51	0.69	0.48-1.13
StU (Random projection)	2.08	1.64-3.58	5.32	2.91-8.71	0.36	0.29-0.55	0.86	0.58-1.46
StU (α-radius)	2.50	1.61-3.63	4.40	3.34-9.38	0.38	0.27-0.51	0.79	0.69-1.58
SmU (Fraiman-Muniz)	1.67	1.27-3.12	4.20	2.46-7.83	0.34	0.23-0.47	0.67	0.47-1.27
SmU (Random projection)	2.11	1.44-3.41	4.80	2.97-8.58	0.36	0.30-0.52	0.82	0.61-1.31
SmU (α-radius)	2.07	1.68-3.83	4.48	2.96-8.29	0.39	0.28-0.60	0.79	0.63-1.33
StGU (Fraiman-Muniz)	1.86	1.21-2.99	3.53	2.56-8.00	0.30	0.23-0.48	0.71	0.51-1.20
StGU (Random projection)	2.34	1.57-3.30	3.66	3.20-8.30	0.35	0.29-0.53	0.80	0.61-1.28
StGU (α-radius)	2.24	1.60-3.63	4.77	3.13-7.94	0.36	0.29-0.52	0.88	0.56-1.27
StiefelU (Fraiman-Muniz)	2.38	1.39-3.99	4.46	2.76-7.52	0.34	0.27-0.51	0.79	0.55-1.21
StiefelU (Random projection)	2.64	1.75-4.19	5.01	3.58-9.17	0.41	0.31-0.65	0.89	0.67-1.47
StiefelU (α-radius)	2.71	1.90-3.95	5.70	3.32-9.24	0.42	0.34-0.65	0.92	0.67-1.45
Variance								
St	1.90	1.02-4.09	3.08	1.76-10.42	0.41	0.24-0.93	0.82	0.57-2.37
Sm	2.09	0.98-3.92	4.08	2.07-10.95	0.48	0.30-0.86	0.79	0.55-2.49
StU	1.90	1.02-4.09	3.08	1.76-10.42	0.41	0.24-0.93	0.82	0.57-2.37
SmU	2.03	1.31-4.03	2.69	1.79-11.26	0.52	0.33-0.93	0.85	0.61-2.74
StGU	2.05	1.25-3.24	3.71	1.79-9.66	0.46	0.30-0.81	1.03	0.64-2.55
StiefelU	2.03	1.29-3.44	3.15	1.71-8.63	0.44	0.29-0.80	0.98	0.58-2.41
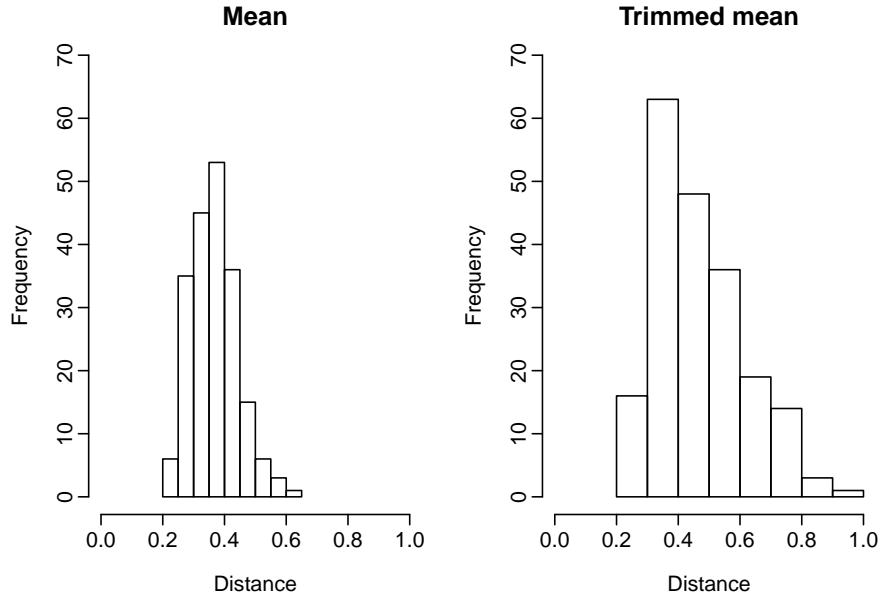
Figure 1: Histograms of the widths for the CIs with the L^∞ metric, based on the sample mean (left) and the 5%-trimmed mean (right).
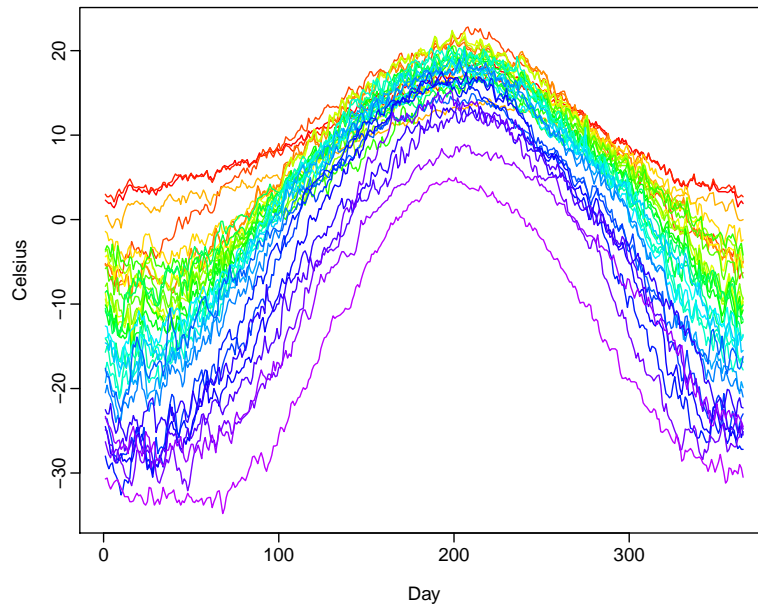
Figure 2: Averaged daily temperatures from 1960 to 1994 observed at 35 Canadian weather stations. Note that each curve represents the averaged daily temperatures at a weather station, not at a particular year.
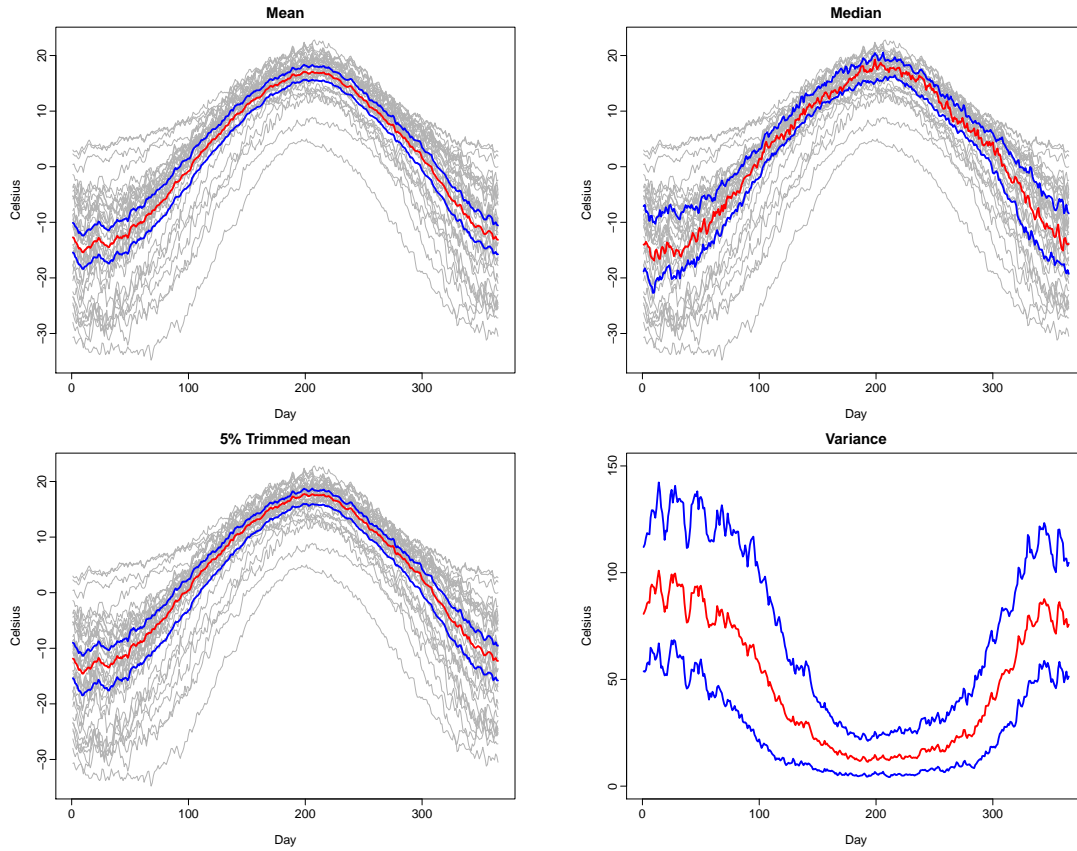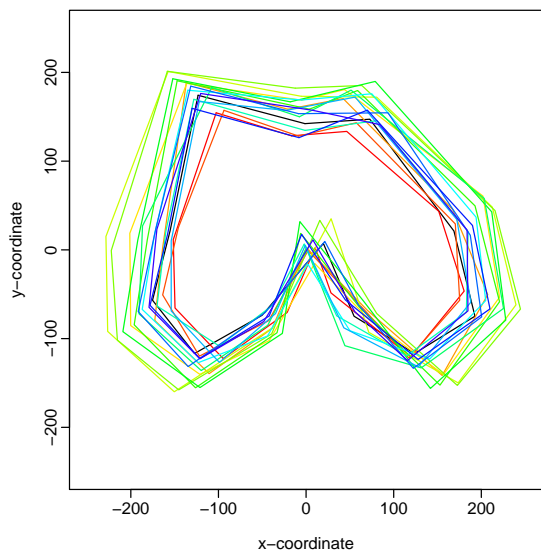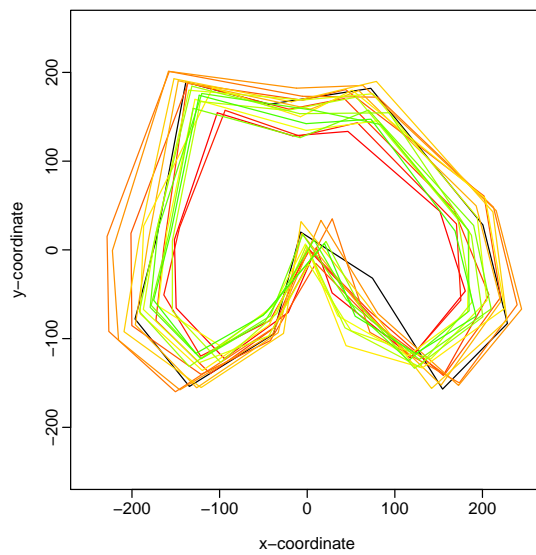
Figure 3: 95% CIs of the descriptive statistics for the Canadian weather station data, based on the smoothed bootstrap principal component score method with smoothing parameter $\beta = 0.05$ and $B = 150$ repetitions. While the sample estimates are shown in red, their 95% CIs are shown in blue.
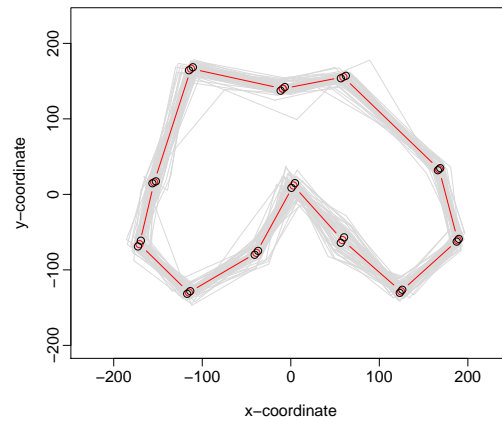
(a) 16 eburnated femora.

(b) 52 non-eburnated femora.
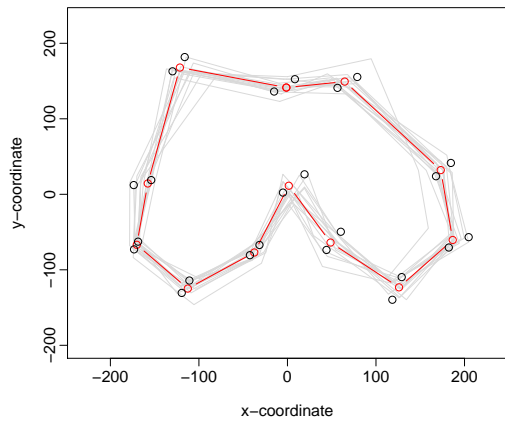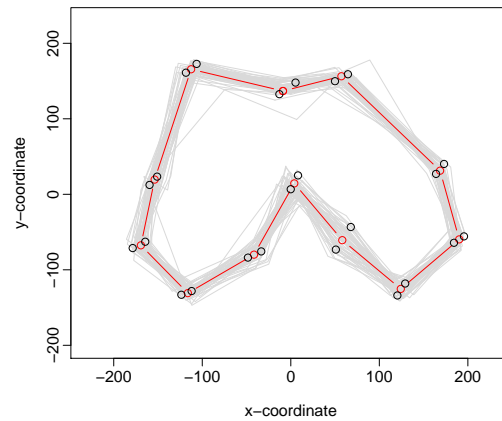
Figure 4: Raw data for the 68 bone shapes.
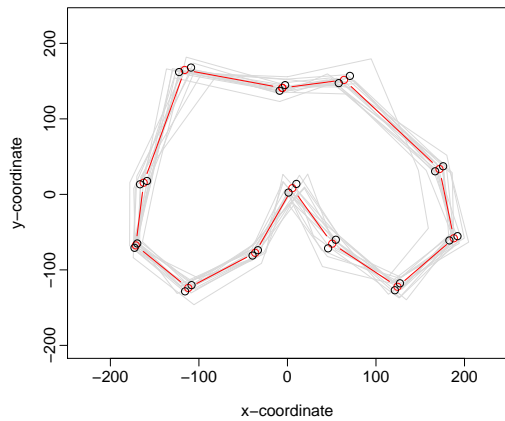
(a) 95% CI for the mean shape of eburnated bones.

(b) 95% CI for the mean shape of non-eburnated bones.

(c) 95% CI for the median shape of eburnated bones.

(d) 95% CI for the median shape of non-eburnated bones.

Figure 5: 95% CIs of the descriptive statistics for the bone shape data, based on the standard bootstrap principal component score method. The mean or median shape is shown in red, while the 95% pointwise CI of the mean or median shape based on $B = 150$ repetitions is shown by the black circles. The variability associated with the mean shape is smaller than the variability associated with the median shape.