Journal of Sciences, Islamic Republic of Iran 19(4): 347-356 (2008) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Asymptotic Distributions of Estimators of Eigenvalues and Eigenfunctions in Functional Data

P. Hall[1] and S.M.E. Hosseini-Nasab[2,*]

[1]Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
[2]Department of Statistics, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Islamic Republic of Iran

Received: 26 November 2007 / Revised: 5 November 2008 / Accepted: 16 November 2008

Abstract

Functional data analysis is a relatively new and rapidly growing area of statistics. This is partly due to technological advancements which have made it possible to generate new types of data that are in the form of curves. Because the data are functions, they lie in function spaces, which are of infinite dimension. To analyse functional data, one way, which is widely used, is to employ principal component analysis, allowing finite dimensional analysis of the problem. The authors gave stochastic expansions of estimators eigenvalues and eigenfunctions, providing not only a new understanding of the effects of truncating to a finite number of principal components, but also pointing to new methodology, such as simultaneous and individual bootstrap confidence statements for eigenvalues and eigenfunctions. The expansions explicitly include terms of sizes $n^{-1/2}$, n^{-1}, and a remainder of order $n^{-3/2}$, where n denotes sample size. The terms of size $n^{-1/2}$ are related to limit theory. Because for many situations, the exact statistical properties of the eigenvalues and eigenfunctions estimators are not directly obtainable, the way by which we can approximate their distributions is of interest in practice. In this paper, we discuss asymptotic results for eigenvalues and eigenfunctions. The work shows that eigenvalue spacings have only a second-order effect on properties of eigenvalue estimators, but a first-order effect on properties of eigenfunction estimators.

Keywords: Eigenfunction; Eigenvalue; Functional data analysis; Hilbert-Schmidt operator; Stochastic expansion

Introduction

In recent years, there has been substantial interest in research on functional data analysis. Application of new technologies allows data analysts to access a new kind of data which are functions. Similar to classical principal component analysis, PCA for FDA produces a small number of constructed variables from the original data that are uncorrelated and account for most of the variation in the original data set. Thus, it is a popular tool for dimension reduction, which results in understanding the underlying structure of the data.

* Corresponding author, Tel.: +98(21)82884480, Fax: +98(21)88009730, E-mail: m.hosseininasab@modares.ac.ir

Statistical methodologies which are finite-dimensional in conventional statistical settings, become infinite-dimensional in the context of functional data. Therefore, they challenge classical methods of data analysis and need theoretical justification.
We may describe PCA for FDA as a generalization of classical PCA so that matrix-based arguments are changed to operator-based ones in function spaces such as Hilbert spaces. However, theoretical treatments for FDA are not as easy as we are thinking about in that way. High dimensionality of the problems in this field, alter substantially both numerical and theoretical properties of statistical methodology. See, for example, [16,17]. Some theoretical justification for PCA in functional data analysis is provided by limit theory. See, for example, [14] and [8].

1. What Are Functional Data?

One of the aspects of functional data is to show changes over time (the "age" effect), separated from the effects caused by differences among subjects which are chosen from the population for the study. This is due to the nature of the collected data, consisting of repeated measurements of subjects through time. Unlike cross-sectional studies, in which we measure a single quantity for each object, here we are able to use the capacity of data to explore the "age" effect by analyzing the data. Separation of changes over time within objects from those among them can be beneficial for revealing useful characterizations of the population from which the sample was drawn. In a sample like $X_1(t),\cdots,X_n(t)$, the former variation refers to the variable t, time, which belongs to an interval, say $[a,b]$, and the second can be seen through the essential randomness of X, in which for a certain time, we have different values of X when running from the first individual ($X_1(t)$) to the last one in the sample ($X_n(t)$). In cross-sectional data, however, we can see the differences among individuals by measuring a quantity over sampled individuals, showing $X_1,X_2,\cdots,X_n$.

Figure 1 shows the heights of ten boys, obtained by measuring each boy at 29 different points of time ([25], page 2). For each boy, measurement was begun at age two and continued annually until age ten, after which it was done biannually for all boys. Therefore, we have 29 records for each person, which can be assumed as a continuous function due to the nature of growth. As the graph shows, it can be easily recognized that the sign of almost all boys' height accelerations tend to change at some points (ages), especially at 12, 14, and 16. This might be due to pubertal effects on their growth. However, the effect is not the same for all boys, and differs in the timing and the intensity. To explore the "age" effect, one can be benefited by using tools for investigating behavior of functions, such as obtaining the acceleration curves by estimating the $D^2Height_i$ from the data. Thus, thinking of records as curves rather than vectors of observations in discrete time enables us to employ derivatives for investigating the "age" effect in functional data ([25], page 2).

2. Principal Component Analysis for Functional Data

Classical principal component analysis (PCA) is amongst the oldest of the multivariate statistical methods of data reduction. A Multivariate Analysis problem could start out with a substantial number of correlated variables. In such situations, using PCA, we can reduce the number of the variables to a lesser number of constructed variables from the original data that are uncorrelated and account for most of the variation in the original data set. It helps us to understand the underlying structure of the data. For this reason, PCA has found application in fields such as signal processing, face recognition, image compression and so on ([21]). Similarly, PCA is widely used in the study of functional data, since it allows finite-dimensional analysis of a problem that is intrinsically infinite-dimensional.

Early work on PCA for FDA includes that of [4,24,27,23,29,30]. Accounts in monographs include those of [25], especially Chapter 6, and [26]. Work of [14,7,3,19,22], for example, addresses empirical basis function approximation and approximations of covariance operators.

Recent work includes contributions to techniques for functional PCA (see e.g. [1,2,6,9-13,15,18,20].

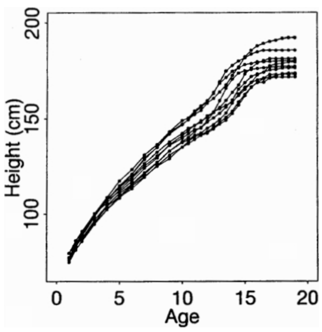

Figure 1. The heights of 10 Swiss boys measured at 29 ages. The points indicate the unequally spaced ages of measurement (See [25], page 2).

2.1. Kernel for Principal Component Analysis

Let X denote a random function, or equivalently a stochastic process, defined in the interval $\mathrm{I}=[0,1]$ and satisfying $\int_I E(X^2)<\infty$. Put $\mu=E(X)$, a conventional function. The principal component expansion of $X-\mu$ may be constructed via the covariance function,

$$K(u,v)=E\left[\{X(u)-\mu(u)\}\{X(v)-\mu(v)\}\right].$$

We interpret K as the kernel of a mapping, or operator (also denoted by K), on the space $L_2(I)$ of square-integrable functions from I to the real line. It takes $\psi\in L_2(I)$ to $K\psi$, where

$$(K\psi)(u)=\int_I K(u,v)\psi(v)dv. \tag{1}$$

We may write

$$K(u,v)=\sum_{j=1}^{\infty}\theta_j\psi_j(u)\psi_j(v), \tag{2}$$

where $\theta_1\geq\theta_2\geq\cdots\geq 0$ is an enumeration of the eigenvalues of K, and the corresponding orthonormal eigenfunctions are $\psi_1,\psi_2,\ldots$.

The ψ_j's are sometimes called the principal component functions of K.

2.2. Empirical Principal Components

Suppose we are given a set $X=\{X_1,\ldots,X_n\}$ of independent random functions, all distributed as X. The standard empirical approximation to $K(u,v)$ is

$$\hat{K}(u,v)=\frac{1}{n}\sum_{i=1}^{n}\{X_i(u)-\bar{X}(u)\}\{X_i(v)-\bar{X}(v)\},$$

where $\bar{X}=n^{-1}\sum_i X_i$.

Analogously to (1) and (2), we can write the kernel operator $\hat{K}$, on the space $L_2(I)$ as follows:

$$(\hat{K}\hat{\psi})(u)=\int_I\hat{K}(u,v)\hat{\psi}(v)dv,$$

and the empirical spectral decomposition

$$\hat{K}(u,v)=\sum_{j=1}^{\infty}\hat{\theta}_j\hat{\psi}_j(u)\hat{\psi}_j(v),$$

where $\hat{\theta}_1\geq\hat{\theta}_2\geq\cdots\geq 0$ is an enumeration of the eigenvalues of $\hat{K}$, and the corresponding orthonormal eigenfunctions are $\hat{\psi}_1,\hat{\psi}_2,\ldots$.

2.3. Karhunen-Loève Expansion

An expansion of the function $X-\mu$ with respect to the orthonormal basis ψ_j (in $L_2(I)$ sense) is its Karhunen-Loève expansion:

$$X(u)-\mu(u)=\sum_{j=1}^{\infty}\xi_j\psi_j(u), \tag{3}$$

where the principal components $\xi_1,\xi_2,\ldots$ are given by $\xi_j=\int_{\mathrm{I}}(X-\mu)\psi_j$. As regards the kernel $K(u,v)$, it follows that

$$E(\xi_j\,\xi_k)=\int_{\mathrm{I}}\int_{\mathrm{I}}\psi_j(u)K(u,v)\psi_k(v)du\,dv=\theta_j\,\delta_{jk}, \tag{4}$$

where δ_{jk} is the Kronecker delta (recall that the $\theta_j^{1/2}$ are eigenvalues and $\psi_j^{1/2}$ are the corresponding orthonornal eigenfunctions of the operator K). Equation implies that the random variables ξ_j are uncorrelated, with zero means and variance $\theta_j=E(\xi_j^2)$. Moreover, $\int_{\mathrm{I}}E(X-\mu)^2=\sum_{j\geq 1}\theta_j<\infty$. We call the expansion the Karhunen-Loève expansion of $X-\mu$. It is also known that if the kernel $K(u,v)$ is a continuous function on $\mathrm{I}\times\mathrm{I}$, then the series on the right-hand side of converges uniformly to $X(u)-\mu(u)$ (Theorem 1.5 of [8]). However, the L_2 convergence of the series to $X(u)-\mu(u)$, satisfied by the condition $\int_{\mathrm{I}}E[(X-\mu)^2]<\infty$.

Also, in regard to $\hat{K}(u,v)$ as the standard empirical approximation to $K(u,v)$, we write

$$X_i-\bar{X}=\sum_{j=1}^{\infty}\hat{\xi}_{ij}\hat{\psi}_j,$$

where $\hat{\xi}_{ij}=\int_{\mathrm{I}}(X_i-\bar{X})\hat{\psi}_j$ is the j th empirical principal component of X_i. In analogy to (4),

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{ij}\,\hat{\xi}_{ik}=\int_{\mathrm{I}}\int_{\mathrm{I}}\hat{\psi}_j(u)\hat{K}(u,v)\hat{\psi}_k(v)du\,dv=\hat{\theta}_j\,\delta_{jk}.$$

[16,17] developed rigorous arguments, based on arguments from operator theory, that derive stochastic expansions for eigenvalues and eigenfunctions estimators as follows.

2.4. Stochastic Expansions for Eigenvalues and Eigenfunctions

In this Section we give expansions for eigenvalues and eigenfunctions which explicitly include terms of sizes $n^{-1/2}$ and n^{-1}, where n denotes sample size, and a remainder of order $n^{-3/2}$. This work shows that eigenvalue spacings have only a second-order effect on properties of eigenvalue estimators, but a first-order effect on properties of eigenfunction estimators.

Take I to be the unit interval. Put $A = X - E(X)$ and assume that

(a) for all $C > 0$ and some $\varepsilon > 0$,

$$\sup_{t\in \mathrm{I}} E\,|X(t)|^C < \infty,$$

$$\sup_{s,t\in \mathrm{I}} E[\{|s-t|^{-\varepsilon}|X(s)-X(t)|\}^C] < \infty; \qquad (5)$$

(b) for each integer $r\geq 1$,

$\theta_j^{-r} E(\int_I A\psi_j)^{2r}$ is bounded uniformly in j.

For example, (5) holds for Gaussian processes with Hölder-continuous sample paths.

Recall that the eigenvalues of the covariance operator K are ordered so that $\theta_1 \geq \theta_2 \geq \ldots \geq 0$. Let $\zeta_j \in (0,1)$ denote the infimum of $1-(\theta_k/\theta_j)$ over K such that $\theta_k < \theta_j$, and let $v_j \in (0,1)$ denote the infimum of $(\theta_k/\theta_j)-1$ over k such that $\theta_k > \theta_j$. Define $\|\hat{K}-K\|^2 = \int(\hat{K}-K)^2$, and put $\rho_j = \min_{k\neq j}|\theta_j - \theta_k|$, and $s_j = \sup_u |\psi_j(u)|$.

Theorem 1. If (5) holds, then for each j for which

$$\left\|\left|\hat{K}-K\right|\right\| \leq \frac{1}{2}\theta_j \min(\zeta_j, v_j), \qquad (6)$$

$$\hat{\psi}_j(t)-\psi_j(t) = n^{-1/2}\sum_{k:k\neq j}(\theta_j-\theta_k)^{-1}\psi_k(t)\int Z\psi_j\psi_k$$
$$-12n^{-1}\psi_j(t)\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_k)^2$$
$$+n^{-1}\sum_{k:k\neq j}\psi_k(t)\{(\theta_j-\theta_k)^{-1}$$
$$\times\sum_{\ell:\ell\neq j}(\theta_j-\theta_\ell)^{-1}(\int Z\psi_j\psi_\ell)(\int Z\psi_k\psi_\ell)$$
$$-(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_j)(\int Z\psi_j\psi_k)\} + O_p(n^{-3/2}), \qquad (7)$$

$$\hat{\theta}_j - \theta_j = n^{-1/2}\int Z\psi_j\psi_j + n^{-1}$$
$$\times\sum_{k:k\neq j}(\theta_j-\theta_k)^{-1}(\int Z\psi_j\psi_k)^2 + O_p(n^{-3/2}), \qquad (8)$$

where the absolute values of the "$O_p(n^{-3/2})$" remainders on the right-hand sides of (7) and (8) are each bounded above by $n^{-3/2}U_{nj}(1-\zeta_j)^{-1/2}\rho_j^{-3}\theta_j^{-1/2}s_j$, where the random variables U_{nj} satisfy $\sup_{n,j\geq 1}E(U_{nj}^C) < \infty$ for each $C > 0$. In the case of (7), this bound is also valid uniformly in t. Moreover, the "$O_p(n^{-3/2})$" remainders on the right-hand sides of (8) are bounded above in the L_2 metric by $n^{-3/2}U_{nj}\rho_j^{-3}$, where the U_{nj} have the same properties as before.

Proof: See [16].

Using the stochastic expansions given in and, and their properties, here we discuss the weak convergence results for estimators of eigenvalues and eigenfunctions.

Lemma 1. If the random process $X(t)$ satisfies the following condition:

for all $C > 0$ and some $\varepsilon > 0$,

$$\sup_{t\in \mathrm{I}} E\,|X(t)|^C < \infty,$$

$$\sup_{s,t\in \mathrm{I}} E[\{|s-t|^{-\varepsilon}|X(s)-X(t)|\}^C] < \infty, \qquad (9)$$

then $Z(u,v) = n^{1/2}(\hat{K}-K)(u,v) \to \zeta(u,v)$ in distribution, where ζ is a Gaussian process.

Proof: See the appendix.

The stochastic expansions for eigenvalues and eigenfunctions given in (7) and (8) are of intrinsic interest. They can be used as the foundation for theory in a particularly wide range of settings, for example bootstrap methods for confidence intervals for eigenvalues and eigenfunctions ([16]). However, in order to develop informative theory about the performance of such methodologies, we need a concise account of the accuracy to which $\hat{\theta}_j$ and $\hat{\psi}_j$ approximate θ_j and ψ_j, respectively. That account can be easily provided using properties of the expansions. Moreover, the problem of determining estimator accuracy, uniformly over many components, prompts consideration of explicit uniform bounds that are obtainable via the mathematical theory of infinite-dimensional operators ([17]).

The weak limit of Z, ζ, is a bivariate Gaussian process with mean zero and the covariance function

$$C(u,v,s,t) = \mathrm{cov}\{\zeta(u,v),\zeta(s,t)\}$$
$$= E\{Y(u)Y(v)Y(s)Y(t)\}$$
$$-K(u,v)K(s,t),$$

where $Y(.) = X(.) - \eta(.)$ denotes a generic Y_i. Using the Karhunen-Loève expansion $Y(u)=\sum_{j=1}^{\infty}\xi_j \psi_j(u)$, where $\xi_j=\int_{I} Y\,\psi_j$, we have:

$$C(u,v,s,t)$$

$$=\sum_{j_1=1}^{\infty}\cdots\sum_{j_4=1}^{\infty}E(\xi_{j_1}\ldots\xi_{j_4})\psi_{j_1}(u)\psi_{j_2}(v)\psi_{j_3}(s)\psi_{j_4}(t)$$

$$-\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\theta_{j_1}\theta_{j_2}\psi_{j_1}(u)\psi_{j_1}(v)\psi_{j_2}(s)\psi_{j_2}(t), \qquad (10)$$

where we have used the fact that $E(\xi_{j_k}^2)=\theta_{j_k}$. If the absence of correlation among the ξ_j's is replaced by independence, only for the purpose of calculating the expected values of products of four of the variables ξ_j, then the first series on the right-hand side of (10) can be written as

$$\sum_{j_1=1}^{\infty}\cdots\sum_{j_4=1}^{\infty}E(\xi_{j_1}\ldots\xi_{j_4})\psi_{j_1}(u)\psi_{j_2}(v)\psi_{j_3}(s)\psi_{j_4}(t)=$$

$$\sum_{j=1}E(\xi_j^4)\psi_j(u)\psi_j(v)\psi_j(s)\psi_j(t)$$

$$+\sum_{j_1\neq j_2}\sum\theta_{j_1}\theta_{j_2}\{\psi_{j_1}(u)\psi_{j_1}(v)\psi_{j_2}(s)\psi_{j_2}(t)$$

$$+\psi_{j_1}(u)\psi_{j_1}(s)\psi_{j_2}(v)\psi_{j_2}(t)$$

$$+\psi_{j_1}(u)\psi_{j_1}(t)\psi_{j_2}(s)\psi_{j_2}(v)\}.$$

Therefore,

$$C(u,v,s,t)$$

$$=\sum_{j=1}^{\infty}\{E(\xi_j^4)-\theta_j^2\}\psi_j(u)\psi_j(v)\psi_j(s)\psi_j(t)$$

$$+\sum_{j_1\neq j_2}\sum\theta_{j_1}\theta_{j_2}\{\psi_{j_1}(u)\psi_{j_1}(s)\psi_{j_2}(v)\psi_{j_2}(t)$$

$$+\psi_{j_1}(u)\psi_{j_1}(t)\psi_{j_2}(s)\psi_{j_2}(v)\}. \qquad (11)$$

Under the assumption that random processes X is a Gaussian processes, (i.e. the variables ξ_j are independent and jointly normally distributed, rather than merely uncorrelated, and with zero kurtosis), $E(\xi_j^4)=3\theta_j^2$. In this situation, the asymptotic covariance function is simplified as follows:

$$C(u,v,s,t)$$

$$=\sum_{j_1\neq j_2}\sum\theta_{j_1}\theta_{j_2}\{\psi_{j_1}(u)\psi_{j_2}(v)\psi_{j_1}(s)\psi_{j_2}(t)$$

$$+\psi_{j_1}(u)\psi_{j_2}(v)\psi_{j_2}(s)\psi_{j_1}(t)\}$$

$$+2\sum_{j=1}^{\infty}\theta_j^2\psi_j(u)\psi_j(v)\psi_j(s)\psi_j(t). \qquad (12)$$

Results

For many situations, the exact statistical properties of the eigenvalues and eigenfunctions estimators are not directly obtainable. Therefore, the way by which we can approximate their distributions is of interest in practice. In this section, we discuss asymptotic results for eigenvalues and eigenfunctions. The asymptotic results discussed here, can be used to a construct confidence interval for θ_j, helping to qualify the amount of variability that can be explained by the *j*th principle component. Also, an asymptotic confidence interval for ψ_j provides information about the likelihood that a bump on $\hat{\psi}_j$ reflects a similar feature in the true eigenfunction ψ_j.

Asymptotic Distribution of Eigenvalues and Eigenfunctions

It should be mentioned that accounts of asymptotic normality of eigenfunctions and eigenvectors and their projections have been given by [14] and [8]. In connection with the results discussed above, using the expansion (7) and (8) , the following shorter expansions can be derived:

$$n^{1/2}(\hat{\psi}_j(t)-\psi_j(t))$$

$$=\sum_{k:k\neq j}(\theta_j-\theta_k)^{-1}\psi_k(t)\int Z\,\psi_j\,\psi_k+o_p(1). \qquad (13)$$

$$n^{1/2}(\hat{\theta}_j-\theta_j)=\int Z\,\psi_j\,\psi_j+o_p(1). \qquad (14)$$

Similarly, it can be seen from equation (7) that:

$$n\|\hat{\psi}_j-\psi_j\|^2$$

$$=\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}\left(\int Z\,\psi_j\,\psi_k\right)^2+o_p(1), \qquad (15)$$

Results 13-15 lead directly to limit theorems for $\hat{\psi}_j$ and $\hat{\theta}_j$, as follows.

Let $P_j=\psi_j\otimes\psi_j$ and $Q_j=\sum_{k:k\neq j}(\theta_j-\theta_k)^{-1}\psi_k\otimes\psi_k$, where $x \otimes y$ is defined by $(x \otimes y)f = \langle x, f \rangle y$ for each $f \in E$. Define the operator Φ_j such that it maps $Z \in F$ to $Q_j Z P_j \in F$, where F, the space of all Hilbert-Schmidt operators on E with the inner product $\langle .,. \rangle_F$

$$\langle T_1, T_2 \rangle_F = \sum_j \langle T_1 e_j, T_2 e_j \rangle_E ,$$

and the Borel field B_F.

Theorem 2. if the ψ_j's are continuous (for each $j \ge 1$, $\psi_j \in C_{[0,1]}$), then the random function $n^{1/2}(\hat{\psi}_j(t) - \psi_j(t))$ converges weakly to a Gaussian process, $\Psi_j(t)$ say, precisely;

$$n^{1/2}(\hat{\psi}_j(t) - \psi_j(t)) \to \Psi_j(t) = \sum_{k:k \ne j} (\theta_j - \theta_k)^{-1} \psi_k(t) \int \zeta \psi_j \psi_k \text{, in distribution} \quad (16)$$

Proof: The proof is straightfoward. The operator Φ_j is linear and continuous. So, Lemma 1 implies that $\Phi_j(Z) \to \Phi_j(\zeta)$ in distribution.

where the non-stationary Gaussian process $\Psi_j(t)$ has zero mean and covariance function

$$\tau_j(u,v) = \sum_{k1:k1 \ne j} \sum_{k2:k2 \ne j} (\theta_j - \theta_{k1})^{-1} (\theta_j - \theta_{k2})^{-1} \psi_{k1}(u) \psi_{k2}(v) \times \int C(u,v,s,t) \psi_j(u) \psi_{k1}(v) \psi_j(s) \psi_{k2}(t) du\, dv\, ds\, dt,$$

where $C(u,v,s,t)$ was introduced in (10). After some algebraic calculations, under the assumption of independence the above formula is simplified to

$$\tau_j(u,v) = \sum_{k:k \ne j} (\theta_j - \theta_k)^{-2} \theta_j \theta_k \psi_k(u) \psi_k(v).$$

In particular, the asymptotic variance of $n^{1/2}\hat{\psi}_j(t)$ equals

$$\sigma_j(t)^2 = \tau_j(t,t) = \sum_{k:k \ne j} (\theta_j - \theta_k)^{-2} \theta_j \theta_k \psi_k(t)^2. \quad (17)$$

Result (16) can be extended to a p-tuple of the $\hat{\psi}_j$. The p-tuple $n^{1/2}(\hat{\psi}_j(t) - \psi_j(t))$ for $1 \le j \le p$ converge jointly and weakly to the non-stationary Gaussian process $\Psi_1, \cdots, \Psi_p$. In particular, for $j_1, j_2 \ge 1$, the two random functions $n^{-1/2}(\hat{\psi}_{j_1}(t) - \psi_{j_1}(t))$ and $n^{-1/2}(\hat{\psi}_{j_2}(t) - \psi_{j_2}(t))$ have the asymptotic covariance function

$$\tau_{j_1,j_2}(u,v) = \delta_{j_1,j_2} \sum_{k:k \ne j_1} (\theta_{j_1} - \theta_k)^{-2} \theta_{j_1} \theta_k \psi_k(u) \psi_k(v) + (\delta_{j_1,j_2} - 1)(\theta_{j_2} - \theta_{j_1})^{-2} \theta_{j_1} \theta_{j_2} \psi_{j_2}(u) \psi_{j_1}(v),$$

where δ_{j_1,j_2} denotes the Kronecker delta. The covariance function shows that for $j_1 \ne j_2$, the two elements $n^{-1/2}(\hat{\psi}_{j_1}(t) - \psi_{j_1}(t))$ and $n^{-1/2}(\hat{\psi}_{j_2}(t) - \psi_{j_2}(t))$ are not asymptotically independent.

Therem 3. in connection to (15), if (9) holds, then we have

$$n^{1/2} \| \hat{\psi}_j - \psi_j \| \to U_j \text{, in distribution,} \quad (20)$$

where

$$U_j^2 = \sum_{k:k \ne j} (\theta_j - \theta_k)^{-2} N_{jk}^2, \quad (19)$$

and the random variables $N_{jk} = \int \zeta \psi_j \psi_k$ are jointly normally distributed with zero mean. Morever, if the random function X is a Gaussian process then $N_{j1}, N_{j2}, \ldots$ are independent as well as normally distributed.

Proof: The proof is straightforward by using lemma1 and result (15). Note that, since $\int_I E(X^4) < \infty$,

$$\begin{aligned} \sum_{k=1}^{\infty} E(N_{jk}^2) &= E(\sum_{k=1}^{\infty} N_{jk}^2) \\ &= E\{\int_I [\int_I \zeta(u,v) \psi_j(v) dv]^2 du\} \\ &\le \iint_{I^2} E(\zeta^2) \\ &= \int_I \int_I \{E[X(u)^2 X(v)^2] - K(u,v)^2\} du\, dv \\ &\le \int_I \int_I E[X(u)^2 X(v)^2] du\, dv \\ &= E(\int_I X^2)^2 \le \int_I E(X^4) < \infty, \end{aligned} \quad (20)$$

from which it follows that the series defining U_j^2 is finite provided the eigenvalue θ_j is not repeated.

Result 1. It implies that

$$n^{1/2}(\hat{\theta}_j - \theta_j) \to N_{jj}, \quad \text{in distribution,} \quad (21)$$

where *Njj* is normally distributed with mean zero and variance

$$\gamma_j^2 = E(\xi_j^4) - \theta_j^2. \quad (22)$$

An extension of the result (21) is available for any p-tuple of the $\hat{\theta}_j$. The p-tuple $n^{1/2}(\hat{\theta}_j - \theta_j)$ for $1 \le j \le p$ converge jointly to a p-variate Normal distribution. In particular, for $j_1, j_2 \ge 1$, the two random quantities $n^{1/2}(\hat{\theta}_{j_1} - \theta_{j_1})$ and $n^{1/2}(\hat{\theta}_{j_2} - \theta_{j_2})$ are asymptotically distributed as a two-variate Normal distribution with the covariance

$$\gamma_{j_1 j_2} = E(\xi_{j_1}^2 \xi_{j_2}^2) - \theta_{j_1}\theta_{j_2}.$$

Comparing result (21) with formula (19) for the limiting distribution of $n^{1/2}\|\hat{\psi}_j - \psi_j\|$, we see that spacings among the eigenvalues θ_k impact immediately on properties of $\hat{\psi}_j$, through first-order terms in its limiting distribution; but impact on $\hat{\theta}_j$ only through second-order terms. Note also , where it is clear that eigenvalue spacings affect only the term in n^{-1}, not that in $n^{-1/2}$.

Appendix

Proof of Lemma 1. We have

$$Z(u,v) = n^{1/2}(\hat{K} - K)(u,v)$$

$$= n^{1/2}[(\frac{1}{n}\sum_{i=1}^{n} Y_i(u)Y_i(v) - K(u,v))$$

$$-\{\bar{X}(u) - \eta(u)\}\{\bar{X}(v) - \eta(v)\}]$$

$$= n^{1/2}[\frac{1}{n}\sum_{i=1}^{n} W_i(u,v) - \{\bar{X}(u)$$

$$-\eta(u)\}\{\bar{X}(v) - \eta(v)\}], \quad (23)$$

where $\eta(u) = E\{X(u)\}$, $Y_i = X_i - \eta$ and $W_i(u,v) = Y_i(u)Y_i(v) - K(u,v)$. The first term on the right-hand side of is the sample mean of the n independent terms. Furthermore, by using the fact that $\bar{X}(v) - \eta(v) = O_p(n^{-1/2})$ (Theorem 3.2.5 of [29] with tightness of $X(v)$ which is discussed later), we can write $Z(u,v) = Z_n(u,v) + O_p(n^{-1/2})$, where $Z_n(u,v) = n^{-1/2}\sum_{i=1}^{n} W_i(u,v)$.

Theorem 4. Let P_n and P be probability measures on $(C_{[0,1]}, \mathrm{F})$, where $C_{[0,1]}$ is the space of continuous functions on [0,1] with the uniform metric $\rho(x,y) = \sup_t |x(t) - y(t)|$, for each $x, y \in C_{[0,1]}$, and F is the σ-field constructed on $C_{[0,1]}$. If the finite-dimensional distributions of P_n converge weakly to those of P, and if $\{P_n\}$ is tight, then P_n converges weakly to P.

Proof: See Theorem 8.1 of [5]. ■

Using the above theorem, if X_n are random elements of $C_{[0,1]}$, then $\{X_n\}$ is tight if $\{P_n\}$ is tight, where P_n is the distribution of X_n, as we identify the finite-dimensional distribution of X_n with those of P_n in the above theorem. Therefore, Theorem 4 is equivalent to the following argument.

If the finite-dimensional distributions of X_n converge weakly to those of X, and $\{X_n\}$ is tight, then $X_n \to X$ in distribution. Regarding the k points $(u_1,v_1),\cdots,(u_k,v_k)$, for each point if $\int_{\mathrm{I}} E(X^4) < \infty$, then the classical Central Limit Theorem with the Slutsky Theorem (Theorem 3.3.1 and 3.4.2 of [28]) implies that, for each $j = 1,\cdots,k$,

$$Z_n(u_j,v_j) + o_p(1) \to \zeta(u_j,v_j), \text{ in distribution}, \quad (24)$$

where $\zeta(u_j,v_j)$ is the weak limit of $Z(u_j,v_j)$, for each $j = 1,\cdots,k$. Now, for the k points $(u_1,v_1),\cdots,(u_k,v_k)$ in $[0,1]\times[0,1]$, let $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}$ be the mapping that carries the point h of $C_{[0,1]\times[0,1]}$ to the point $(h(u_1,v_1),\cdots,h(u_k,v_k))$ of R^k. Since $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}$ is continuous, we have $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}(Z_n) \to \Pi_{(u_1,v_1),\cdots,(u_k,v_k)}(\zeta)$ in distribution (Corollary 1 of Theorem 5.1 of [5]), i.e.

$$(Z_n(u_1,v_1),\cdots,Z_n(u_k,v_k)) \to$$

$$(\zeta(u_1,v_1),\cdots,\zeta(u_k,v_k)), \text{in distribution}. \quad (25)$$

(in particular, (25) follows via the Cramér-Wold device.)

Theorem 5. The sequence $\{X_n\}$ is tight if it satisfies these two conditions:

(i) The sequence $\{X_n(0)\}$ is tight.

(ii) There exist constants $\gamma \ge 0$ and $\alpha > 1$ and a nondecreasing, continuous function F on [0,1] such that

$$P\{|X_n(t_2)-X_n(t_1)|\geq\lambda\}\leq\frac{1}{\lambda^{\gamma}}|F(t_2)-F(t_1)|^{\alpha}, \quad (26)$$

holds for all t_1,t_2 and n and all positive λ.

Proof: See Theorem 12.3 of [5]. ■

We know that the moment condition

$$E\{|X_n(t_2)-X_n(t_1)|^{\gamma}\}\leq|F(t_2)-F(t_1)|^{\alpha}, \quad (27)$$

implies. Furthermore, we can immediately obtain tightness of $\{X_n\}$ from condition (9), (27) and Theorem 5. Also, for some $\gamma=2k$, where $k\geq 1$ is an integer, using Rosenthal's inequality for fixed $u,v,s,t\in[0,1]$ results in

$$E[|n^{-1/2}\sum_{i=1}^{n}W_i(u,v)-n^{-1/2}\sum_{i=1}^{n}W_i(s,t)|^{\gamma}]$$

$$=n^{-k}E|\sum_{i=1}^{n}\{W_i(u,v)-W_i(s,t)\}|^{2k}$$

$$\leq C_{1\gamma}n^{-k}\{\sum_{i=1}^{n}E|W_i(u,v)-W_i(s,t)|^{2k}$$

$$+(\sum_{i=1}^{n}E|W_i(u,v)-W_i(s,t)|^{2})^{k}\}$$

$$\leq C_{2\gamma}E|W(u,v)-W(s,t)|^{\gamma}, \quad (28)$$

and

$$E|W(u,v)-W(s,t)|^{\gamma}=$$

$$E|\{Y(u)Y(v)-Y(s)Y(t)\}-\{K(u,v)-K(s,t)\}|^{\gamma}$$

$$\leq C_{3\gamma}(E|Y(u)Y(v)-Y(s)Y(t)|^{\gamma}$$

$$+E|Y(u)Y(v)-Y(s)Y(t)|^{\gamma})$$

$$\leq C_{4\gamma}E|Y(u)Y(v)-Y(s)Y(t)|^{\gamma}, \quad (29)$$

where $C_{1\gamma}$, $C_{2\gamma}$, $C_{3\gamma}$ and $C_{4\gamma}$ are constants depending only on γ, γ, and W and Y denote a generic W_i and Y_i, respectively. Also,

$$E[|Y(u)Y(v)-Y(s)Y(t)|^{\gamma}]$$

$$=E[|Y(u)(Y(v)-Y(t))+Y(t)(Y(u)-Y(s)|^{\gamma}]$$

$$\leq C_{\gamma}\{E[|Y(u)|^{\gamma}|Y(v)-Y(t)|^{\gamma}]$$

$$+E[|Y(t)|^{\gamma}|Y(u)-Y(s)|^{\gamma}]\}$$

$$\leq C_{\gamma}\{(E[|Y(u)|^{2\gamma}])^{1/2}(E[|Y(v)-Y(t)|^{2\gamma}])^{1/2}$$

$$+(E[|Y(t)|^{2\gamma}])^{1/2}(E[|Y(u)-Y(s)|^{2\gamma}])^{1/2}\}, \quad (30)$$

where C_{γ} is a constant depending only on γ. If condition (9) holds, then (28)-(30) imply that for each two points $(u,v),(s,t)\in[0,1]\times[0,1]$,

$$E[|n^{-1/2}\sum_{i=1}^{n}W_i(u,v)-n^{-1/2}\sum_{i=1}^{n}W_i(s,t)|^{\gamma}]$$

$$\leq C_{\gamma}\{|v-t|^{\alpha}+|u-s|^{\alpha}\}, \quad (31)$$

where γ can be chosen such that $\alpha=\varepsilon\gamma>2$. Hence, using Markov's inequality, for each $\lambda>0$, each two points $(u,v),(s,t)\in I\times I$ and all n we have

$$P_n(|n^{-1/2}\sum_{i=1}^{n}W_i(u,v)-n^{-1/2}\sum_{i=1}^{n}W_i(s,t)|>\lambda)$$

$$\leq C_{\gamma}\lambda^{-\gamma}\{|v-t|^{\alpha}+|u-s|^{\alpha}\}. \quad (32)$$

The proof of the above theorem, given by [5], with condition (32) instead of (26) may be followed to show that $n^{-1/2}\sum_{i=1}^{n}W_i(u,v)$ is tight. To appreciate why, fix n, δ, j and k, then for a positive integer m consider the random variables

$$\kappa_{\ell}=n^{-1/2}\sum_{i=1}^{n}(W_i(j\delta+\frac{\ell}{m}\delta,k\delta+\frac{\ell}{m}\delta)$$

$$-W_i(j\delta+\frac{(\ell-1)}{m}\delta,k\delta+\frac{(\ell-1)}{m}\delta)),$$

for $\ell=1,\cdots,m$. The random variables κ_{ℓ} with $S_k=\sum_{\ell=1}^{k}\kappa_{\ell}$ satisfy

$$S_s-S_r=\sum_{r<\ell\leq s}\kappa_{\ell}$$

$$=n^{-1/2}\sum_{i=1}^{n}(W_i(j\delta+\frac{s}{m}\delta,k\delta+\frac{s}{m}\delta)$$

$$-W_i(j\delta+\frac{r}{m}\delta,k\delta+\frac{r}{m}\delta)).$$

So, for $0\leq r\leq s\leq m$ and $\lambda>0$ we have

$$P(|S_s-S_r|\geq\lambda)=P(|\sum_{r<\ell\leq s}\kappa_{\ell}|\geq\lambda)$$

$$\leq C_{\gamma}\lambda^{-\gamma}[(\frac{(s-r)\delta}{m})^{\alpha}+(\frac{(s-r)\delta}{m})^{\alpha}]$$

$$\leq C_{\gamma}\lambda^{-\gamma}[(\sum_{r<\ell\leq s}\delta m^{-1})^{\alpha}+(\sum_{r<\ell\leq s}\delta m^{-1})^{\alpha}]$$

$$\leq C_\gamma \lambda^{-\gamma} \delta^\alpha,$$

where we have used (32) to obtain the first inequality above. By using Theorem 12.2 of [5], we have

$$P(\max_{0\leq \ell \leq m} |n^{-1/2}\sum_{i=1}^{n}(W_i(j\delta+\frac{\ell}{m}\delta, k\delta+\frac{\ell}{m}\delta)$$

$$-W_i(j\delta,k\delta))|>\varepsilon) \leq \frac{B}{\varepsilon^\gamma}\delta^\alpha,$$

where B depends on γ and α ($B = B_{\gamma,\alpha}$). Since the W_i for each $1 \leq i \leq n$ are continuous functions, if $m \to \infty$ we have

$$P(\sup_{\|(u,v)-(j\delta,k\delta)\|_E<\delta} |n^{-1/2}\sum_{i=1}^{n}(W_i(u,v)$$

$$-W_i(j\delta,k\delta))|>\varepsilon) \leq \frac{B}{\varepsilon^\gamma}\delta^\alpha. \quad (33)$$

where $\|.\|_E$ denotes the Euclidian norm in R^2. If δ^{-1} is integer, the above inequality leads to

$$\sum_{j<\delta^{-1}}\sum_{k<\delta^{-1}} P(\sup_{\|(u,v)-(j\delta,k\delta)\|_E<\delta} |n^{-1/2}\sum_{i=1}^{n}(W_i(u,v)$$

$$-W_i(j\delta,k\delta))|>\varepsilon) \leq \frac{B}{\varepsilon^\gamma}\delta^{\alpha-2}. \quad (34)$$

Define the modulus of continuity of an element x of $C_{[0,1]\times[0,1]}$ by

$$w_x^{(2)}(\delta) = \sup_{\|(u,v)-(s,t)\|_E<\delta} |x(u,v)-x(s,t)|,$$

where $0<\delta\leq 1$. Let

$$A_{s,t} = \{n^{-1/2}\sum_{i=1}^{n}W_i :$$

$$\sup_{\|(u,v)-(s,t)\|_E<\delta} |n^{-1/2}\sum_{i=1}^{n}(W_i(u,v)-W_i(s,t))| \geq \varepsilon\}.$$

If we want to lie (u,v) and (s,t) each in rectangles of the form $[j\delta,(j+1)\delta]\times[k\delta,(k+1)\delta]$, then if $\|(u,v)-(s,t)\|_E<\delta$, these rectangles either coincide or abut. Therefore,

$$P(n^{-1/2}\sum_{i=1}^{n}W_i : w^{(2)}_{n^{-1/2}\sum_{i=1}^{n}W_i}(\delta) \geq 3\varepsilon)$$

$$\leq P(\cup_{j,k<\delta^{-1}} A_{j\delta,k\delta})$$

$$\leq \sum_{k<\delta^{-1}}\sum_{j<\delta^{-1}} P(A_{j\delta,k\delta}) \leq \frac{B}{\varepsilon^\gamma}\delta^{\alpha-2},$$

where we have used (34) to obtain the last inequality above. Because we can choose γ such that $\alpha=\varepsilon\gamma>2$, we may take δ as the reciprocal of a large integer, and in this way make $(B/\varepsilon^\gamma)\delta^{\alpha-2}$ very small. Moreover, for all $\alpha>0$ and adequate choice of δ we have

$$P(|n^{-1/2}\sum_{i=1}^{n}W_i(0,0)| \geq \alpha)$$

$$\leq \alpha^{-\delta} E[|n^{-1/2}\sum_{i=1}^{n}W_i(0,0)|^\delta]$$

$$\leq C_{1\delta}\alpha^{-\delta}\{E|Y(0)Y(0)|^\delta + |K(0,0)|\}^\delta$$

$$\leq C_{2\delta}\alpha^{-\delta}\{E|Y(0)\times Y(0)|^\delta + E|Y(0)\times Y(0)|^\delta\}$$

$$\leq C_{3\delta}\alpha^{-\delta}\{E|Y(0)|^{2\delta}\}^{1/2}\times\{E|Y(0)|^{2\delta}\}^{1/2}$$

$$\leq C_{4\delta}\alpha^{-\delta},$$

where $C_{1\delta}, C_{2\delta}, C_{3\delta}, C_{4\delta}$ are constants depending only on δ, and we have used condition to obtain the last inequality above. Thus,

$$Z(u,v) \to \zeta(u,v), \text{in distribution. } \blacksquare \quad (35)$$

Acknowledgements

We are grateful to two reviewers for helpful comments, and to Mr. Tazikeh for converting the original file of the article into the required format of the journal.

References

1. Aguilera, A.M., Ocana, F.A. and Valderrama, M.J. Forecasting time series by functional PCA. Discussion of several weighted approaches. *Comput. Statist.* **14**: 443-467, (1999).
2. Aguilera, A.M., Ocana, F.A. and Valderrama, M.J. Forecasting with unequally spaced data by a functional principal component approach. *Test* **8**, 233-253, (1999).
3. Besse, P. PCA stability and choice of dimensionality. *Statist. Probab. Lett.* **13**, 405-410, (1992).
4. Besse, P. and Ramsay J.O. Principal components-analysis of sampled functions. *Psychometrika* **51**, 285-311, (1986).
5. Billingsley, P. *Probability and Measure*. John Wiley, New York, (1979).
6. Boente, G. and Fraiman, R. Kernel-based functional principal components. *Statist. Probab. Lett.* **48**, 335-345, (2000).

7. Bosq, D. Propriétés des opérateurs de covariance empiriques d'un processus stationnaire hilbertien. *C. R. Acad. Sci. Paris* Sér. I **309**, 873-875, (1989).
8. Bosq, D. *Linear Processes in Function Spaces. Lect. Notes Statist.*, **149**, (2000).
9. Brumback, B.A. and Rice, J.A. Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.* **93**, 961-976, (1998).
10. Cardot, H. Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparam. Statist.* **12**, 503-538(2000).
11. Cardot, H., Ferraty, F. and Sarda, P. Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. *C.R. Acad. Sci. Paris* Sér. I **330**, 501-504, (2000).
12. Cardot, H., Ferraty, F. and Sarda, P. Spline estimators for the functional linear model. *Statist. Sinica* **13**, 571-591, (2003).
13. Cardot, H., Ferraty, F., Mas, A. and Sarda, P. Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 241-255, (2003).
14. Dauxois, J., Pousse, A. and Romain, Y. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.* **12**, 136-154, (1982).
15. Girard, S. A nonlinear PCA based on manifold approximation. *Comput. Statist.* **15**, 145-167, (2000).
16. Hall, P. and Hosseini-Nasab, M. On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **68** , pp 109-126, (2006).
17. Hall, P. and Hosseini-Nasab, M. Theory for high-order bounds in functional principal components analysis. Will be published by *Mathematical Proceedings of the Cambridge Philosophical Society.*
18. He, G.Z., Müller, H.-G. and Wang, J.-L. Functional canonical analysis for square integrable stochastic processes. *J. Multivar. Anal.* **85**, 54-77, (2003).
19. Huang, J.H.Z., Wu C.O. and Zhou L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111-128, (2002).
20. James, G.M., Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **64**, 411-432, (2002).
21. Jolliffe, I. T. Principal Component Analysis, 2nd edition. Springer, New York, (2002).
22. Mas, A. Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.* **99**, 117-135, (2002).
23. Pezzulli, S. and Silverman, B.W. Some properties of smoothed principal components analysis for functional data. *Computational Statistics* **8**, 1-16, (1993).
24. Ramsay, J.O. and Dalzell, C.J. Some tools for functional data analysis. (With discussion.) *J. Roy. Statist. Soc.* Ser. B **53**, 539-572, (1991).
25. Ramsay, J.O. and Silverman, B.W. *Functional Data Analysis.* Springer, New York, (1997).
26. Ramsay, J.O. and Silverman, B.W. *Applied Functional Data Analysis: Methods and Case Studies.* Springer, New York, (2002).
27. Rice, J.A. and Silverman, B.W. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc.* Ser. B **53**, 233-243, (1991).
28. Sen, P. K. and Singer, J. M. *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall, (1993).
29. Silverman, B.W. Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser.* **B, 57**, 673-689, (1995).
30. Silverman, B.W. Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24**, 1-24, (1996).