Sensitivity Analysis in Functional Principal Component Analysis

Norihiro Oide*, Yoshihiro Yamanishi† and Yutaka Tanaka‡

***Abstract* — Functional principal component analysis (PCA) enables us to investigate some patterns of data over time. However, There is a problem that this analysis is sensitive to a few influential observations as in the case of multivariate PCA. In this paper, we propose a method of sensitivity analysis in functional PCA. Empirical influence function (ELFs) for eigenvalues and eigenfunctions are derived in both cases where the smoothing parameter is fixed and unfixed. And using these influence functions, we proposed a sensitivity analysis procedure to detect singly and jointly influential observations. In dealing on the eigenfunctions, we use following two methods. 1) Based on EIF for coefficient vectors of the basis function expansion. 2) Based on the sampled vectors of the functional EIF.**

1 Introduction

Sensitivity analysis has been developed in many fields to investigate the relation between observations and the result of analysis. However, in the case of functional data, few researches have been done. On the other hands, functional PCA has been developed by many person (Besse and Ramsay, 1986; Ramsay and Dalzel, 1991; Ramsay and Silverman, 1997). The objective of this paper is to propose a method of sensitivity analysis in functional PCA, or more precisely a method to detect singly and jointly influential observations in functional PCA. In section 2, we introduce penelized functional principal component analysis. Next sensitivity analysis in functional PCA is shown in section 3.

2 Functional principal component analysis

2.1 Ordinary functional principal component analysis

Suppose we have a set of functional data $\{x_i(s)\}_{i=1}^N$.Then define sample mean vector and covariance matrix as $\bar{x}(t) = N^{-1}\sum_{i=1}^N x_i(t)$, $v(s,t) = N^{-1}\sum_{i=1}^N \{x_i(s)-\bar{x}(s)\}\{x_i(t)-\bar{x}(t)\}$. In functional PCA a functional linear combination is introduced $f_i = \int \xi(s)x_i(s)ds$ using weight function $\xi(s)$. The weight function $\xi(s)$ is chosen in such a way that it maximizes the variance

$$PCASV = \int\int \xi(s)v(s,t)\xi(t)dsdt, \qquad (1)$$

under the constraints $\int \xi_l(t)^2 dt = 1$. This leads to an integral eigenproblem as follows:

$$\int v(s,t)\xi(t)dt = \rho\xi(t). \qquad (2)$$

2.2 Penalized functional principal component analysis

2.2.1 Roughness penalty

We introduce penalty function to incorporate smoothing into the principal components. Here we penalize the roughness of ξ by its integrated squared second derivative i.e., $PEN_2(\xi) =< D^2\xi, D^2\xi >=\| D^2\xi \|^2$, where $< \cdot,\cdot >$ and $\| \cdot \|^2$indicate the inner product and the squared norm, respectively. In this case the penalized variance can be expressed by

$$PCAPSV = \frac{PCASV}{\| \xi \|^2 + \lambda \times PEN_2(\xi)}, \qquad (3)$$

where λ is a smoothing parameter. This expression means that the trade-off between maximizing the sample variance and keeping the smoothness of ξ is controlled by a smoothing parameter λ. Under natural boundary conditions, $\| D^2\xi \|^2 =< \xi, D^4\xi >$, so maximizing problem leads the following eigenproblem.

$$\int v(s,t)\xi(t)dt = \rho(I + \lambda D^4)\xi(s), \qquad (4)$$

2.2.2 Algorithm

A data function $x_i(s)$ and a weight function $\xi(s)$ can be expanded as

$$\begin{aligned} x_i(s) &= \sum_{k=1}^K \mathbf{C}_{ik}\phi_k(s) = \mathbf{C}_i^T\phi(s), \\ \xi(s) &= \sum_{k=1}^K y_k\phi_k(s) = \mathbf{y}^T\phi(s), \qquad (5) \end{aligned}$$

*Graduate School of Natural Science and Technology, Okayama University, Japan 700-0081 E-mail:nohi@f7.ems.okayama-u.ac.jp, Tel: +81-86-251-8835

†Graduate School of Science, Kyoto University, Japan

‡Department of Environmental Science and Technology, Okayama University, Japan E-mail:tanaka@ems.okayama-u.ac.jp , Tel: +81-86-251-8835

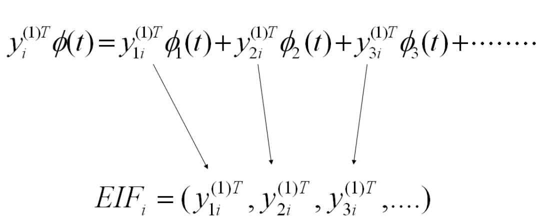

Figure 1: EIF based on coefficients vector

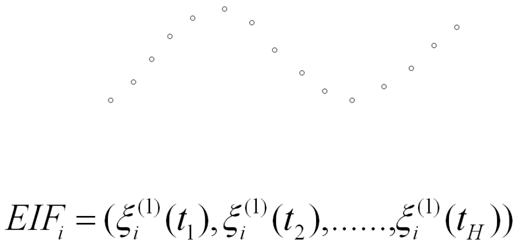

Figure 2: EIF based on the discrete functions

using basis functions $\phi(s) = (\phi_1(s), \cdots, \phi_K(s))^T$, where K is the number of basis functions. Define $\mathbf{V}$ as the covariance matrix of coefficient $\mathbf{C}_i$ and let $\mathbf{J}_\phi = \int \phi(s)\phi(s)^T ds$ and $\mathbf{K}_\phi = \int (D^2\phi(s))(D^2\phi(s))^T ds$. Then the functional eigenproblem (5) is transformed to the following matrix generalized eigenproblem

$$(\mathbf{J}_\phi \mathbf{V} \mathbf{J}_\phi)y = \rho(\mathbf{J}_\phi + \lambda \mathbf{K}_\phi)y. \tag{6}$$

By applying Cholesky factorization $\mathbf{L}\mathbf{L}^T = \mathbf{J}_\phi + \lambda\mathbf{K}_\phi$, the above generalized eigenproblem leads to an eigenproblem of a symmetrical matrix as

$$(\mathbf{S}\mathbf{J}_\phi \mathbf{V} \mathbf{J}_\phi \mathbf{S}^T)(\mathbf{S}^{-T}y) = \rho(\mathbf{S}^{-T}y), \tag{7}$$

where $\mathbf{S} = \mathbf{L}^{-1}$ and $\mathbf{S}^{-T} = (\mathbf{S}^{-1})^T$. λ is determind by cross-validation.

3 Sensitivity analysis in functional PCA

3.1 Influence function

To evaluate the influence of each individual we use the idea of influence function. Here we define the EIF by the partial derivative of the estimated parameter function or vector with respect to a perturbation parameter. Let us introduce weights $n\tilde{w}_\alpha / \sum_\beta \tilde{w}_\beta$ for the α-th observation as perturbation parameters, and define the first order partial derivative of parameter vector $\hat{\theta}$ with respect to w_α as the influence function of $\hat{\theta}$ for the α-th observation. Or in other words, we define the influence function for the α-th observation by the first derivative with respect to ϵ defined as

$$\begin{aligned} w_\alpha &= 1 \text{ for all } \alpha \longrightarrow w_\alpha = n\tilde{w}_\alpha / \sum_\beta \tilde{w}_\beta, \\ &\text{where} \quad \tilde{w}_\beta = \begin{cases} 1 & (\beta \neq \alpha) \\ 1+\epsilon & (\beta = \alpha). \end{cases} \end{aligned} \tag{8}$$

It is known from the perturbation theory of eigenproblems, when matrix $\mathbf{SJVJS}^T$ in equation (7) can be expanded as a convergent power series of ϵ. When the smoothing parameter λ is fixed,

$$\left(\mathbf{SJVJS}^T\right)^{(k)} = \mathbf{SJV}^{(k)}\mathbf{JS}^T \quad (k = 1, 2). \tag{9}$$

When the smoothing parameter λ isn't fixed,

$$\begin{aligned} \left(\mathbf{SJVJS}^T\right)^{(1)} &= \mathbf{S}^{(1)}\mathbf{JVJS}^T + \mathbf{SJV}^{(1)}\mathbf{JS}^T \\ &\quad + \mathbf{SJVJ}(\mathbf{S}^{(1)})^T, \end{aligned} \tag{10}$$

where covariance matrix $\mathbf{V}^{(1)} = (\mathbf{C} - \bar{\mathbf{C}})(\mathbf{C} - \bar{\mathbf{C}})^T - \mathbf{V}$, influence function of matrix S can be calculated as $\mathbf{S}^{(1)} = \frac{\partial \lambda}{\partial \epsilon}\left(\frac{\partial \mathbf{S}}{\partial \lambda}\right) = -\mathbf{L}^{-1}\mathbf{L}^{(1)}\mathbf{L}^{-1}$ (see, Tanaka and Tarumi, 1989, p.18-19, for the derivation of $L^(1)$). Therefore, we can define the EIFs for eigenvalues and eigenfucntions respectively as follows:

$$\begin{aligned} EIF \quad & (x; \rho_s) = \rho_s^{(1)} \\ = \quad & (\mathbf{S}^{-T}y)_r^T \left(\mathbf{SJVJS}^T\right)^{(1)} (\mathbf{S}^{-T}y)_s, \end{aligned} \tag{11}$$

$$\begin{aligned} EIF(x; \xi_s(t)) &= (y_s^{(1)})^T \phi(t) = \Big[\mathbf{S}^T(\mathbf{S}^{-T}y)_s^{(1)} \\ &+ (\mathbf{S}^{(1)})^T(\mathbf{S}^{-T}y)_s\Big]^T \phi(t). \end{aligned} \tag{12}$$

When the number of the basis function is fixed, the influence function for eigenvalue and eigenfunction can be expressed as

$$EIF(x; \rho_s) = \rho_s^{(1)} = (\mathbf{S}^{-T}y)_r^T \left(\mathbf{SJVJS}^T\right)^{(1)} (\mathbf{S}^{-T}y)_s, \tag{13}$$

$$\begin{aligned} EIF(x; \xi_s(t)) &= (y_s^{(1)})^T \phi(t) = \Big[\mathbf{S}^T(\mathbf{S}^{-T}y)_s^{(1)} \\ &+ (\mathbf{S}^{(1)})^T(\mathbf{S}^{-T}y)_s\Big]^T \phi(t). \end{aligned} \tag{14}$$

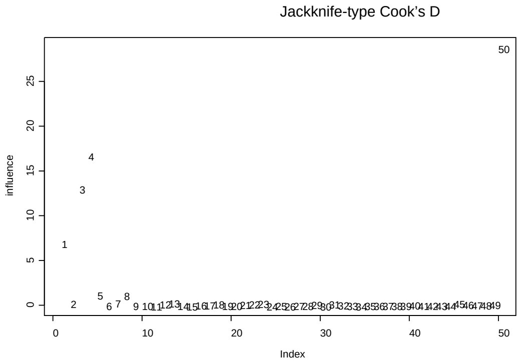

Figure 3: Cook's D based on $\widehat{acov}_{\mathbf{JK}}$: Coefficients

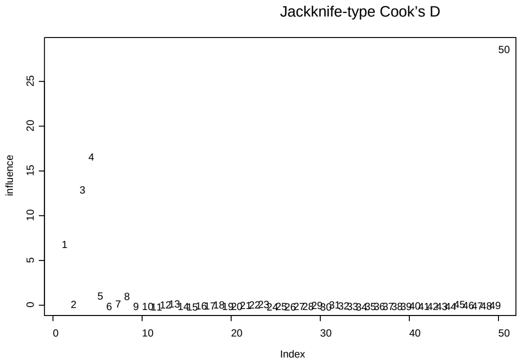

Figure 4: Cook's D based on $\widehat{acov}_{\mathbf{JK}}$: Discrete

3.1.1 Single-case diagnostics

To evaluate the influence of each individual, generalized Cook's D is used here. $D_i = [EIF_i]^T[\widehat{acov}]^{-1}[EIF_i]$, where $\widehat{acov}$ is the asymptotic covariance matrix for the parameters of interest. To avoid the theory of probability in functional context, we define the generalized Cook's D in functional data analysis in the following two ways.

3.2 Cook's D based on the coefficients of the basis function expansion

When the number of the basis functions is fixed, each functional observation $x_i(s)$ can be determined by the corresponding coefficient vector $\mathbf{C}_i$. Then, by regarding $\mathbf{C}_i$ as the original multivariate observation, we can define Cook's D in the same way as in the case of ordinary multivariate analysis (Figure 1). To evaluate $\widehat{acov}$ we use the Jacknife,

$$\widehat{acov}_{\mathbf{JK}} := \frac{1}{N(N-1)} \sum_{i=1}^{N} JIF_i JIF_i^T \qquad (15)$$

where $JIF(x_i; \hat{\theta}) = (N-1)(\hat{\theta}^{[-i]} - \hat{\theta})$.

3.2.1 Cook's D based on the discrete functions

We define a discrete vector $(\xi_i^{(1)}(t_1), \cdots, \xi_i^{(1)}(t_H))$ by sampling from a functional influence function $(\xi_i^{(1)}(t)$ at approriate grid point, and define Cook's D for these multivariate data. The multivariate influence function vector is defined as $EIF_i = \Phi^T y_i^{(1)}$, where Φ is a $K \times H$ matrix whose (k,h)th element is given by $\phi_k(t_h)$. Figure 1,2 shows the image of these two definitions of EIF. The asymptotic covariance matrix of $\Phi^T y_i$ can be obtained by $acov(\Phi^T y_i) = \Phi^T \mathbf{V}_y \Phi$, using the asymptotic covariance matrix $\mathbf{V}_y$ of y. Thus we can define Cook's D as

$$D_i = y_i^{(1)T} \Phi (\Phi^T \mathbf{V}_y \Phi)^{-} \Phi^T y_i^{(1)}. \qquad (16)$$

3.3 Multiple-case diagnostics

Tanaka has proposed to apply principal component analysis (PCA) to the influence functions for detecting influential subsets of observations in multivariate methods (see e.g., Tanaka, 1994). In short, the aim of multiple-case diagnostics is to investigate the pattern of the influence and detect the influential subsets. As in the single-case diagnostics, we will try PCA based on the coefficient and discrete function.

3.4 Numerical example

We shall apply to the mean daily temperature data of 50 wether station in Japan in 1999. For single-case diagnostics, two method is applied to calculate Cook's D. Figure 3 and 4 shows that Cook's D based on the coefficients of the basis function expansion and discrete functions have same result. For multiple-case diagnostics, functional PCA is applied to the functional EIFs. And this shows same result as single-case. Figure 5 shows functional PCA based on coefficient of the basis function expansion and Figure 6 shows the result based on discrete functions.

3.5 Mathematical relation

In single-case diagnostics we defined Cook's D in two different ways. One is based on the EIF for the coefficient vector of the basis function expansion, i.e., $D_{1i} = y_i^{(1)T} \mathbf{V}_y^{-1} y_i^{(1)}$, where $\mathbf{V}_y$ is an estimate for the asymptotic covariance matrix of y_i. The other is based on the sampled vector of the functional EIF $y_i^{(1)T}\phi(t)$. In this case Cook's D is defined as $D_{2i} = y_i^{(1)T} \Phi(\Phi^T \mathbf{V}_y \Phi)^{-} \Phi^T y_i^{(1)}$.

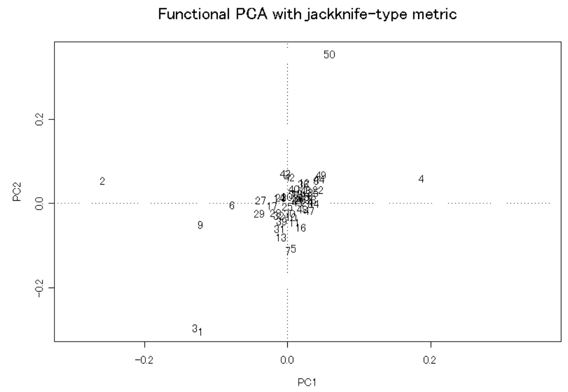

Figure 5: Functional PCA: Coefficients

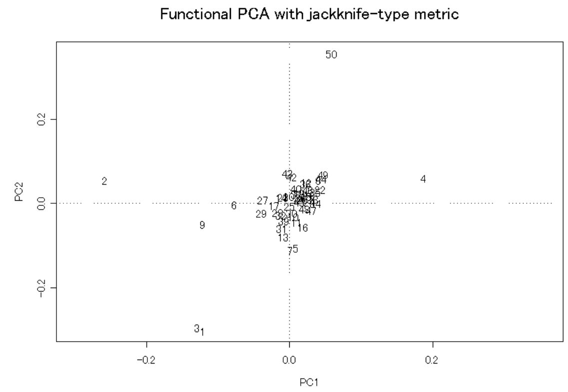

Figure 6: Functional PCA: Discrete

If we choose a $K \times H$ matrix Φ so that rank $\Phi = K$, then we can prove that the relation $\mathbf{V}_y^{-1} = \Phi(\Phi^T \mathbf{V}_y \Phi)^{-} \Phi^T$ holds (see, Searle, 1082, p.224), and therefore $D_{1i} \equiv D_{2i}$ for any i. In multiple-case diagnostics we applied PCA with metric $[\widehat{acov}]^{-}$ to the EIF to detect influential subsets of observations (see, Tanaka, 1994; Tanaka and Zhang, 1999). Here we can use both of the EIF for the coefficient vectors and the EIF for the sampled functional EIF. We obtain eigenvalue problems

$$\left(\frac{1}{N} \sum_{i=1}^{N} y_i^{(1)} y_i^{(1)T} - \nu \mathbf{V}_y \right) a = 0, \tag{17}$$

in the former formulation, and

$$\left(\frac{1}{N} \sum_{i=1}^{N} \Phi^T y_i^{(1)} y_i^{(1)T} \Phi - \nu (\Phi^T \mathbf{V}_y \Phi) \right) b = 0, \tag{18}$$

in the latter formulation. Assume that rank $\Phi = K$. Then, multiplying $(\Phi\Phi^T)^{-1}\Phi$ to equation (18) from the left we obtain an eigenproblem of Φb which is just equivalent to equation (17). Therefore, we may conclude that, though we derived two kinds of EIF, we need not develop sensitivity analysis based on the functional EIF in addition to sensitivity analysis based on the EIF for the coefficients of the basis function expansion.

References

[1] Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions, *Psychometrika, 51, 285-311.* Influence in principal component analysis. *Biometrika,* **72**, *627-636.*

[2] Japan Meteorological Agency (1999). Annual report of Automated Meteorological Data Acquisition System, *Japan Meteorological Business Support Center (JMBSC).*

[3] Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis (with discussion), *J. Royal Statist. Soc.*, **B 53**, 539-572.

[4] Ramsay, J. O. and Silverman, B.W. (1997). Functional Data Analysis, *Springer.*

[5] Sibson, R. (1979). Studies in the robustness of multidimensional scaling: Perturbation analysis of classical scaling, *J. R. Statist. Soc.* **B 41**, *217-229.*

[6] Searle, S. R. (1982). Matrix Algebra Useful for Statistics, *Wiley.*

[7] Tanaka, Y. (1994). Recent advance in sensitivity anaysis in multivariate statistical methods, *J.Japanese Soc.Comp.Statist.,7,1-25.*

[8] Tanaka,Y. and Tarumi,T.(1989). Sensitivity analysis in canonical factor analysis. *J. Japanese Soc. Comp. Statist.2, 9-20.*

[9] Tanaka, Y. and Zhang, F. (1999). R-mode and Q-mode influence analyses in statistical modelling: Relationship between influence function approach and local influence approach,

[10] Yamanishi, Y. and Tanaka, Y. (2000). Sensitivity Analysis in Functional Principal Component Analysis, *Proceedings of the Seventh Japan-China Symposium on Statistics*, 37-40

[11] Yang, W. S. and Tanaka, Y. (2000). Some modified principal component analysis and their applications, *Proceedings of the Tenth Japan and Korea Joint Conference of Statistics*, 119-124