SENSITIVITY ANALYSIS IN FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

YAMANISHI, Yoshihiro
Graduate School of Natural Science and Technology, Okayama University, Tsushima, Okayama 700-8530, Japan.
E-mail: yoshi@stat.ems.okayama-u.ac.jp

TANAKA, Yutaka
Department of Environmental and Mathematical Science, Okayama University, Tsushima, Okayama 700-8530, Japan.
E-mail: tanaka@stat.ems.okayama-u.ac.jp

1 Introduction

In functional data analysis, we treat the data which consist of functions not of vectors (Ramsay and Silverman, 1997). This implies that functional data can be considered as the limit of multivariate data as the number of dimensions tends to infinity. As in the case of finite dimensional multivariate data, it is impotant to detect influential observations which have extraordinarily large effect on the result of functional data analysis. The purpose of this study is to develop sensitivity analysis in functional principal component analysis (PCA).

2 Functional Principal Component Analysis

2.1 Ordinary Functional Principal Component Analysis

Suppose we have a set of functional data $x(s)$. Weight function $\xi(s)$ is chosen in such a way that it maximizes the variance $PCASV =< \xi(s), \int v(s,t)\xi(t)dt >$, where $v(s,t)$ indicates a variance-covariance function based on the functional data set and $<,>$ represents the inner product of functions. The maximization of PCASV under the constraints $\| \xi \|^2 = 1, < \xi_k, \xi_m >= 0\,(k < m)$ leads to the eigenequation as follows:

$$\int v(s,t)\xi(t)dt = \rho\xi(t). \tag{1}$$

2.2 Penalized Functional Principal Component Analysis

Here penalty function is introduced to incorporate smoothing into the principal components (PCs). The most popular form of the penalty for ξ is given by $PEN_2(\xi) =\| D^2\xi \|^2 =< \xi, D^4\xi >$, and in this case the penalized variance can be expressed by $PCAPSV = \frac{PCASV}{\|\xi\|^2 + \lambda \times PEN_2(\xi)}$, where λ is a smoothing parameter. The solution $\xi(t)$ is obtained as the eigenfunction associated with the largest eigenvalue of the following penalized eigenequation

$$\int v(s,t)\xi(t)dt = \rho(I + \lambda D^4)\xi(s). \tag{2}$$

2.3 Choice of the Smoothing Parameter λ by Cross-Validation

Let M be the number of PCs, and define the cross-validation score as

$$CV(\lambda) = \sum_{i=1}^{N} \| x_i - \sum_{m=1}^{M}\sum_{l=1}^{M}(\mathbf{G}^{-1})_{ml} < \xi_m^{[i]}, x_i > \xi_l^{[i]} \|^2, \tag{3}$$

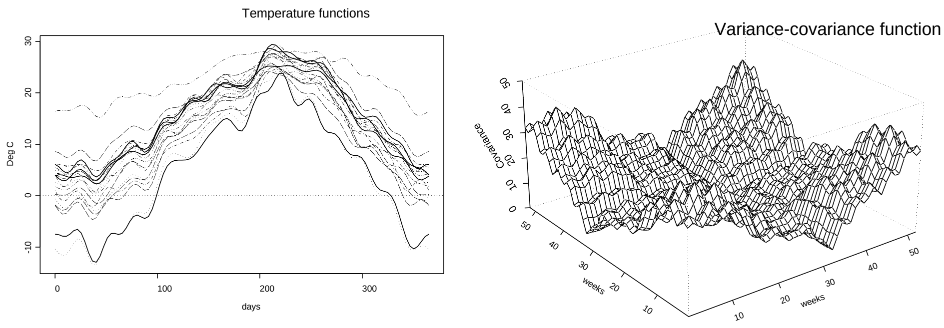

Figure 1: Daily temperature data and their variance-covariance function

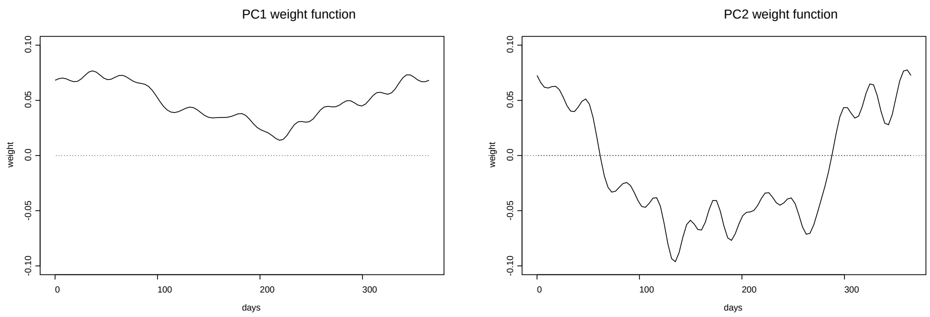

Figure 2: Weight functions for PC1 and PC2

where $\mathbf{G}$ is the $M \times M$ matrix whose (i, j) element is the inner product $< \xi_i, \xi_j >$ and the superscript $[i]$ means the omission of the $i-$th observation. Choose λ which minimize $CV(\lambda)$.

2.4 Numerical Algorithm

Suppose we use a set of basis functions $\{\phi(s)\}$. Then a functional observation $x(s)$ and a weight function $\xi(s)$ can be expanded as $x(s) = c^T\phi(s)$, $\xi(s) = y^T\phi(s)$. Define $\mathbf{V}$ as the variance-covariance matrix of c_i and let $\mathbf{J} = \int \phi\phi^T, \mathbf{K} = \int (D^2\phi)(D^2\phi)^T, \mathbf{LL}^T = \mathbf{J} + \lambda\mathbf{K}, \mathbf{S} = \mathbf{L}^{-1}$. Then the functional eigenequation (2) is transformed to a matrix eigenequation as follows:

$$(\mathbf{SJVJS}^T)((\mathbf{S}^{-1})^T y) = \rho((\mathbf{S}^{-1})^T y). \tag{4}$$

2.5 Numerical Example

The penalized functional PCA is applied to the mean daily temperature data of the 20 weather stations in Japan in 1999. Figure 1 shows the observed temperature curves and their variance-covariance function. The stations include those located at Hokkaido, Tokyo, Kyoto, and Osaka. Figure 2 shows the weight function for PC1 and PC2, respectively. Looking at these figures, we can interpret the PCs as follows: The PC1 is a measure of overall temperatures throughout the year, while the PC2 represents the contrast between the temperatures in summer and in winter.

3 Sensitivity Analysis in Functional PCA

3.1 Influence function

To evaluate the influence of each individual we make use of the idea of influence function, in particular, empirical influence function (EIF). Formally the EIF for the estimated parameter $\hat{\theta}$, which is a functional of the empirical comulative distribution function (cdf) $\hat{F}$, is defined by

$$EIF(x_i;\hat{\theta}) = \hat{\theta}_i^{(1)} = \lim_{\epsilon\to 0}\left\{\theta\left[(1-\epsilon)\hat{F}+\epsilon\delta(x_i)\right]-\theta(\hat{F})\right\}/\epsilon, \tag{5}$$

where $\delta(x_i)$ is the cdf of a unit point mass at x. Now we introduce the following case-weight perturbation.

$$w_\alpha = 1 \,\text{for all}\, \alpha \;\longrightarrow\; w_\alpha = n\tilde{w}_\alpha/\sum_\beta \tilde{w}_\beta,\ \text{where}\ \tilde{w}_\beta = \begin{cases} 1 & (\beta\neq\alpha) \\ 1+\epsilon & (\beta=\alpha). \end{cases} \tag{6}$$

Then it is easily verified that the first derivative of $\hat{\theta}$ with respect to ϵ defined in (6) is equal to a constant times the EIF (Tanaka, 1994). We shall call this derivative as influence function.

There is another sample version of influence function called sample influence function (SIF). The SIF, which is obtained by omitting "lim" and setting $\epsilon = 1/(n-1)$ in EIF, is defined as

$$SIF(x_i;\hat{\theta}) = -(n-1)(\hat{\theta}^{[i]}-\hat{\theta}). \tag{7}$$

It is the most valid mesure for evaluating the influence of each individual, although it is time-consuming.

3.2 Influence function in Functional Eigenequation

Let us introduce the perturbation given by (6). Then the perturbed matrix $\mathbf{SJVJS}^T$ in the matrix eigenequation (4) can be expanded as

$$\left(\mathbf{SJVJS}^T\right)(\epsilon) = \mathbf{SJVJS}^T + \epsilon\left(\mathbf{SJVJS}^T\right)^{(1)} + (\epsilon^2/2)\left(\mathbf{SJVJS}^T\right)^{(2)} + O(\epsilon^3). \tag{8}$$

When the smoothing parameter λ is fixed, $\left(\mathbf{SJVJS}^T\right)^{(k)}$ is given by

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

On the other hand, if we take into consideration that the smoothing parameter λ, which is decided by cross-validation, varies due to a small perturbation to the case weights, it is given by

$$\begin{aligned}
\left(\mathbf{S}(\lambda)\mathbf{JVJS}^T(\lambda)\right)^{(1)} &= \lambda^{(1)}\mathbf{S}_\lambda(\lambda)\mathbf{JVJS}^T(\lambda) + \mathbf{S}(\lambda)\mathbf{JV}^{(1)}\mathbf{JS}^T(\lambda) + \mathbf{S}(\lambda)\mathbf{JVJS}_\lambda^T(\lambda)\lambda^{(1)}, \\
\left(\mathbf{S}(\lambda)\mathbf{JVJS}^T(\lambda)\right)^{(2)} &= \mathbf{S}_\lambda(\lambda)\mathbf{JV}^{(2)}\mathbf{JS}_\lambda^T(\lambda) + \\
& \quad 2\left(\lambda^{(1)}\mathbf{S}_\lambda(\lambda)\mathbf{JV}^{(1)}\mathbf{JS}^T(\lambda) + \mathbf{S}(\lambda)\mathbf{JV}^{(1)}\mathbf{JS}_\lambda^T(\lambda)\lambda^{(1)} + \lambda^{(1)}\mathbf{S}(\lambda)\mathbf{JVJS}^T(\lambda)\lambda^{(1)}\right),
\end{aligned} \tag{10}$$

where $\mathbf{S}_\lambda(\lambda) = \partial\mathbf{S}(\lambda)/\partial\lambda$ and $\lambda^{(1)} = \partial\lambda/\partial\epsilon$. It is known from the perturbation theory of eigenvalue problems, when matrix $\mathbf{SJVJS}^T$ in equation (4) can be expanded as a convergent power series of ϵ, its eigenvalues and eigenvectors can be also expanded in a convergent power series in the neighborhood of $\epsilon = 0$ as

$$\rho_s(\epsilon) = \rho_s + \epsilon\rho_s^{(1)} + (\epsilon^2/2)\rho_s^{(2)} + O(\epsilon^3), \tag{11}$$

$$u_s(\epsilon) = u_s + \epsilon u_s^{(1)} + (\epsilon^2/2)u_s^{(2)} + O(\epsilon^3). \tag{12}$$

where $u_s = ((\mathbf{S}^{-1})^T y)_s$, and we have the following formulas (see, e.g. Sibson, 1979, Tanaka, 1984):

$$\begin{cases}
\rho_s^{(1)} = a_{ss}^{(1)}, \\
u_s^{(1)} = \sum_{r\neq s}(\rho_s-\rho_r)^{-1}a_{rs}^{(1)}u_r. \\
\rho_s^{(2)} = a_{ss}^{(2)} + 2\sum_{r\neq s}(\rho_s-\rho_r)^{-1}(a_{rs}^{(1)})^2, \\
u_s^{(2)} = \sum_{r\neq s}(\rho_s-\rho_r)^{-1}\left\{a_{rs}^{(2)} + 2\sum_{t\neq s}(\rho_s-\rho_t)^{-1}a_{st}^{(1)}(a_{rt}^{(1)} - a_{ss}^{(1)}\delta_{rt})\right\}u_r - \parallel u_s^{(1)} \parallel^2 u_s.
\end{cases} \tag{13}$$

where $a_{rs}^{(k)} = ((\mathbf{S}^{-1})^T y)_r^T\left(\mathbf{SJVJS}^T\right)^{(k)}((\mathbf{S}^{-1})^T y)_s$ $(k=1,2,)$ and δ_{rt} is the Kronecker delta. Therefore, we can define the EIFs for eigenvalues and eigenfucntions respectively as follows:

$$EIF(x;\rho_s) = \rho_s^{(1)}, \tag{14}$$

$$EIF(x;\xi_s) = \left[\mathbf{S}^T(\lambda)((\mathbf{S}^{-1}(\lambda))^T y)_s^{(1)} + \lambda^{(1)}\mathbf{S}_\lambda^T(\lambda)(\mathbf{S}^{-1}(\lambda))^T y\right]^T\phi. \tag{15}$$

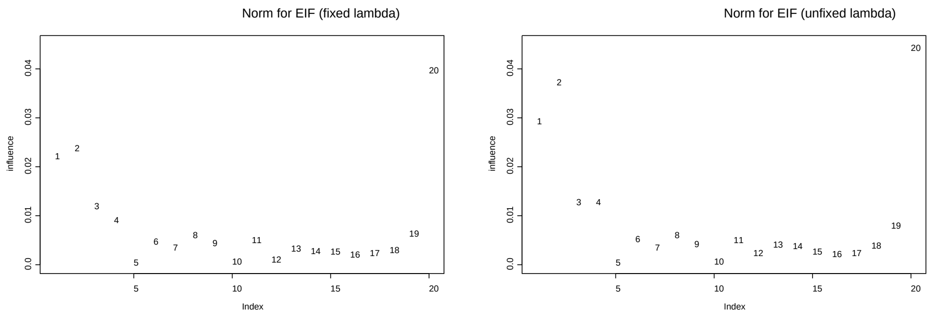

Figure 3: Euclidean norm EIFs for PC1 weight : (Left) with fixed λ, (Right) with unfixed λ

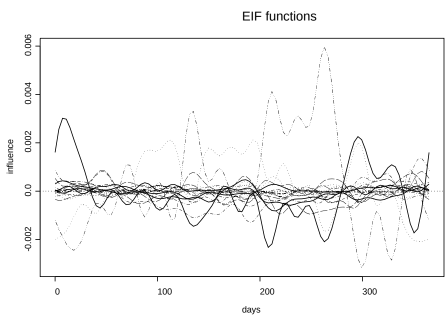

Figure 4: EIF functions

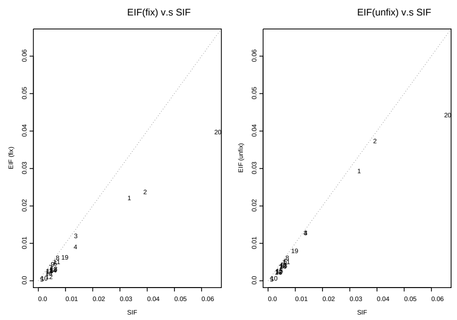

Figure 5: EIFs versus SIFs: (Left) with fixed λ, (Right) with unfixed λ

3.3 Numerical Example

Now we shall apply sensitivity analysis to the penalized functional PCA of the temperature data. For example, we evaluate the influence of each individual for the PC1 weight function. Figure 3 shows the index plots of the Euclidean norms of the EIFs, and Figure 4 shows the EIF functions for PC1 weight function. Looking at these figures we can detect observations No.1, No.2 and No.20 as candidates for highly influential observations. To confirm the validity of the derived EIF, the scatterplots are drawn between the EIFs and the SIFs, where the latter are calculated by actually omitting each observation. The figures in Figure 5 show the scatterplots of the EIFs versus the SIFs. In the left panel the EIFs are calculated by using equation (9), while in the right panel, they are calculated by using equation (10). In these figures, almost all points are located on or near the straight line through the origin. Therefore we can conclude that the EIF can be used instead of the SIF in evaluating the influence of each observation. Comparing the left and right panels, we can find the improvement of approximation by taking the influence through λ into consideration.

REFERENCES

Besse. P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions, Psychometrika, 51, 285-311.

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

Ramsay, J. O. and Silverman, B. W. (1996). Functional Data Analysis, Springer.

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