Modes of Variation

Yidong Zhou, Hans Müller

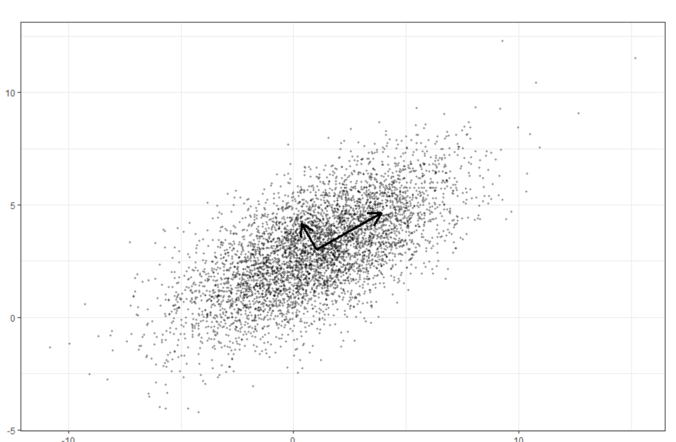
Department of Statistics, University of California, Davis

Background

With the advance of modern technology, more and more data are being recorded continuously during a time interval or intermittently at several discrete time points. Functional data analysis (FDA) deals with the analysis and theory of such function-form data. The most popular technique of FDA is functional principal component analysis (FPCA), which is an important dimensionality reduction tool and can be used to impute sparse data. Modes of variation, as a by-product of FPCA, provide a visualization of the decomposition of variation around the mean. Analogously, one can define modes of variation for principal component analysis (PCA).

Introduction

In statistics, **modes of variation**[1] are a continuously indexed set of vectors or functions that are centered at a mean and are used to depict the variation in a population or sample. Typically, variation patterns in the data can be decomposed in descending order of eigenvalues with the directions represented by the corresponding eigenvectors or eigenfunctions. Both in PCA and in FPCA, modes of variation play an important role in visualizing and describing the variation in the data contributed by each eigencomponent[2].



The vectors shown are the eigenvectors of the covariance matrix scaled by the square root of the corresponding eigenvalue, and shifted so their tails are at the mean.

Fig. 1:PCA of A Multivariate Gaussian Distribution

Formulation

- Modes of variation in PCA

  If a random vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)^\intercal$ has the mean vector $\boldsymbol{\mu}$, and the covariance matrix $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and corresponding orthonormal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p$, by eigendecomposition of a real symmetric matrix, the covariance matrix $\boldsymbol{\Sigma}$ can be decomposed as $\boldsymbol{\Sigma} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\intercal$, where $\mathbf{Q}$ is an orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix whose entries are the eigenvalues of $\boldsymbol{\Sigma}$. Then the k-th mode of variation of $\mathbf{X}$ is the set of vectors, indexed by α,

  $$\mathbf{m}_{k,\alpha} = \boldsymbol{\mu} \pm \alpha\sqrt{\lambda_k}\mathbf{e}_k, \alpha \in [-A, A],$$

  where A is typically selected as 2 or 3.

Formulation

- Modes of variation in FPCA

  For a square-integrable random function $X(t), t \in \mathcal{T} \subset R^p$, where typically $p = 1$ and $\mathcal{T}$ is an interval, denote the mean function by $\mu(t) = \mathrm{E}(X(t))$, and the covariance function by

  $$G(s,t) = \mathrm{Cov}(X(s), X(t)) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(s)\varphi_k(t),$$

  where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\{\varphi_1, \varphi_2, \cdots\}$ are the orthonormal eigenfunctions. By the Karhunen-Loève theorem, one can express the centered function in the eigenbasis,

  $$X(t) - \mu(t) = \sum_{k=1}^{\infty} \xi_k \varphi_k(t),$$

  where $\xi_k = \int_{\mathcal{T}}(X(t) - \mu(t))\varphi_k(t)dt$ is the k-th principal component with the properties

  $$\mathrm{E}(\xi_k) = 0, \mathrm{Var}(\xi_k) = \lambda_k \text{ and } \mathrm{E}(\xi_k\xi_l) = 0 \text{ for } k \neq l.$$

  Then the k-th mode of variation of $X(t)$ is the set of functions, indexed by α,

  $$m_{k,\alpha}(t) = \mu(t) \pm \alpha\sqrt{\lambda_k}\varphi_k(t), t \in \mathcal{T}, \alpha \in [-A, A]$$

  that are viewed simultaneously over the range of α, usually for $A = 2$ or 3[2].

Estimation

- Modes of variation in PCA

  Suppose the data $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ represent n independent drawings from some p-dimensional population $\mathbf{X}$ with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. These data yield the sample mean vector $\overline{\mathbf{x}}$, and the sample covariance matrix $\mathbf{S}$ with eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \cdots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$. Then the k-th mode of variation of $\mathbf{X}$ can be estimated as

  $$\hat{\mathbf{m}}_{k,\alpha} = \overline{\mathbf{x}} \pm \alpha\sqrt{\hat{\lambda}_k}\hat{\mathbf{e}}_k, \alpha \in [-A, A].$$

- Modes of variation in FPCA

  Consider n realizations $X_1(t), X_2(t), \cdots, X_n(t)$ of a square-integrable random function $X(t), t \in \mathcal{T}$ with the mean function $\mu(t) = \mathrm{E}(X(t))$ and the covariance function $G(s,t) = \mathrm{Cov}(X(s), X(t))$. The estimation method for $\mu(t)$ and $G(s,t)$ depends on the sampling schedule. For same time schedule across subjects, one can first estimate $\mu(t)$ and $G(s,t)$ pointwise and then apply smooth interpolation. For different time schedules, one can combine all subjects to apply nonparametric smoother directly. Substituting estimates for the unknown quantities yields that the k-th mode of variation of $X(t)$ can be estimated as

  $$\hat{m}_{k,\alpha}(t) = \hat{\mu}(t) \pm \alpha\sqrt{\hat{\lambda}_k}\hat{\varphi}_k(t), t \in \mathcal{T}, \alpha \in [-A, A].$$

Applications

Modes of variation are useful tools to visualize the variation patterns in the data sorted by the eigenvalues[3]. In real-world applications, the eigencomponents and associated modes of variation aid to interpret complex data, especially in exploratory data analysis (EDA).

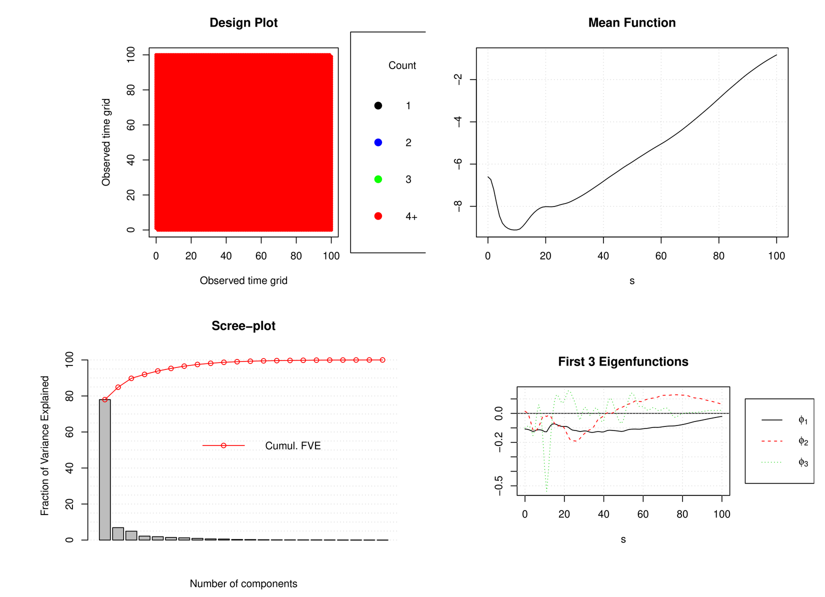


Fig. 2:FPCA for Female Mortality Data (41 Countries, 2003)

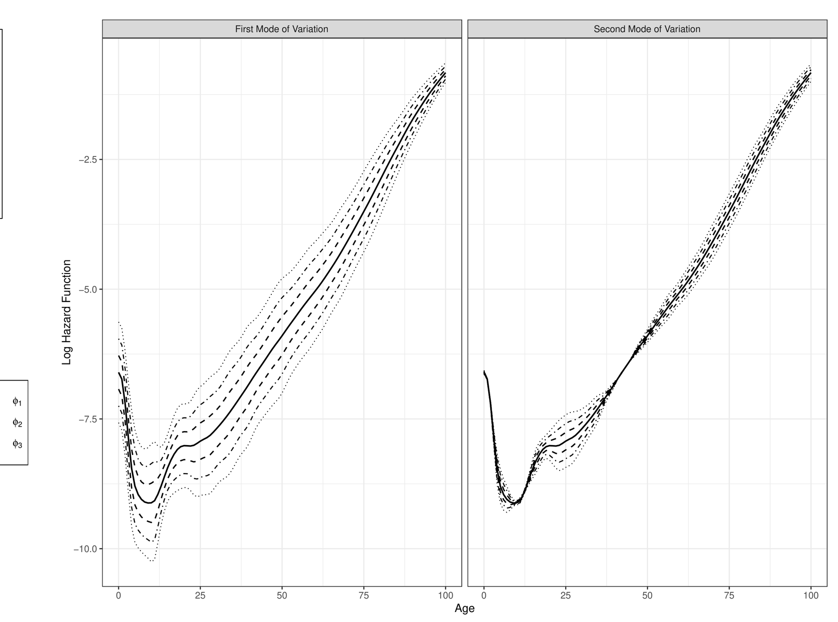


Fig. 3:1st and 2nd Modes of Variation

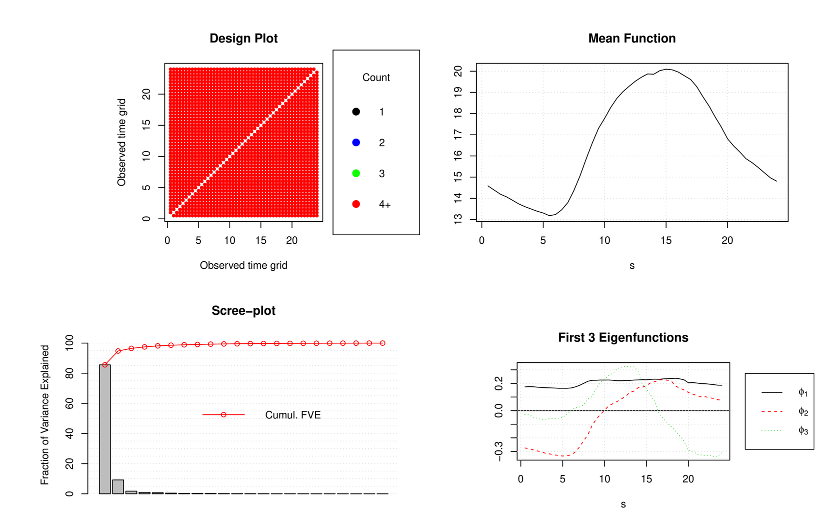


Fig. 4:FPCA for Temperatures in Adelaide Airport on Friday (07/11/1997-03/30/2007)

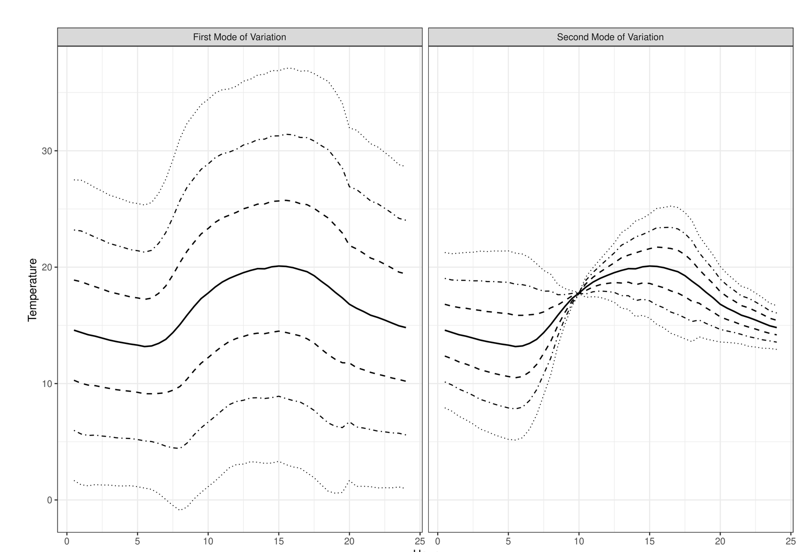


Fig. 5:1st and 2nd Modes of Variation

Acknowledgement

I would like to thank Yaqing Chen for her guidance on selection of real data and implementation of modes of variation.

References

[1] Peter E Castro, W H_ Lawton, and EA Sylvestre.
Principal modes of variation for processes with continuous sample curves.
Technometrics, 28(4):329–337, 1986.

[2] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller.
Functional data analysis.
Annual Review of Statistics and Its Application, 3:257–295, 2016.

[3] MC Jones and John A Rice.
Displaying the important features of large collections of similar curves.
The American Statistician, 46(2):140–145, 1992.