Functional Data Analysis: Part II
Functional Principal Component Analysis (FPCA)

Jane-Ling Wang and Hans-Georg Müller

University of California Davis

August 26-27, 2016

Outline

Outline

Dimension Reduction for Functional Data

- Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.

Dimension Reduction for Functional Data

- Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.

- There are two types of dimension reduction for functional data:

  - one on the data themselves and

  - another on the statistical modeling of such data.

Dimension Reduction for Functional Data

- Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.
- There are two types of dimension reduction for functional data:

  - one on the data themselves and

  - another on the statistical modeling of such data.
- The latter is the topic of functional regression (Part III).

Dimension Reduction for Functional Data

- Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.
- There are two types of dimension reduction for functional data:

  - one on the data themselves and

  - another on the statistical modeling of such data.
- The latter is the topic of functional regression (Part III).
- In Part II we focus on the first one:

  - dimension reduction on the data.

Review of Principal Component Analysis

- Principal component analysis for multivariate (p-dim) data is a dimension reduction tool to transform (linearly) the data to orthogonal (p-dim) data so that the first few (k) of them explains most of the variation.

Review of Principal Component Analysis

- Principal component analysis for multivariate (p-dim) data is a dimension reduction tool to transform (linearly) the data to orthogonal (p-dim) data so that the first few (k) of them explains most of the variation.

- The first eigenfunction $\phi_1 = \underset{\phi \in \mathcal{R}^p,\, \|\phi\|=1}{\operatorname{argmax}} \text{var}(\langle X - \mu), \phi \rangle)$

  - The first principal direction ϕ_1, represents the direction of the data with the largest variation.

Review of Principal Component Analysis

- Principal component analysis for multivariate (p-dim) data is a dimension reduction tool to transform (linearly) the data to orthogonal (p-dim) data so that the first few (k) of them explains most of the variation.

- The first eigenfunction $\phi_1 = \underset{\phi \in \mathcal{R}^p,\, \|\phi\|=1}{\operatorname{argmax}} \; \text{var}(\langle X - \mu), \phi \rangle)$

  - The first principal direction ϕ_1, represents the direction of the data with the largest variation.

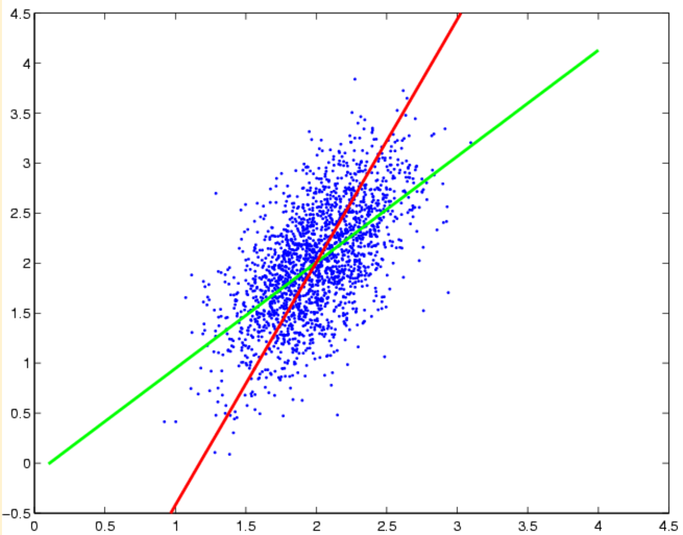
- The second eigenfunction $\phi_2 = \underset{\phi \in \mathcal{R}^p,\, \|\phi\|=1,\, \phi \perp \phi_1}{\operatorname{argmax}} \; \text{var}(\langle (X - \mu), \phi \rangle)$

  .

  .

  .

Which one is the least squares line?



Review of PCA

- The least squares line minimizes the vertical squared distance, but the first PC line minimized the perpendicular squared distance.
- In this example the first PC explains 75% of the variations.

How would you extend PCA to functional data?

- The first eigenvector $\phi_1 = \underset{\phi \in \mathcal{R}^p, \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\langle X - \mu, \phi \rangle)$

  - The first principal direction ϕ_1, represents the direction of the data with the largest variation.

- The second eigenvector $\phi_2 = \underset{\phi \perp \phi_1, \phi \in \mathcal{R}^p, \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\langle X - \mu, \phi \rangle)$

How would you extend PCA to functional data?

- The first eigenvector $\phi_1 = \underset{\phi \in \mathcal{R}^p, \|\phi\|=1}{\operatorname{argmax}} \ \text{var}(\langle X - \mu, \phi \rangle)$

  - The first principal direction ϕ_1, represents the direction of the data with the largest variation.

- The second eigenvector $\phi_2 = \underset{\phi \perp \phi_1, \phi \in \mathcal{R}^p, \|\phi\|=1}{\operatorname{argmax}} \ \text{var}(\langle X - \mu, \phi \rangle)$

- All we have to do is to change the inner product from a vector to a function in a Hilbert space (this explains why we need a Hilbert space structure for functional data).

  e.g. $\langle f, g \rangle = \int_I f(t) g(t) dt,$ for any functions f and g in $L_2(I)$.

Definition of FPCA

- The first eigenfunction ϕ_1
  $= \underset{\phi \in L^2(I), \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\langle X - \mu, \phi \rangle$
  $= \underset{\phi \in L^2(I), \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\int [X(t) - \mu(t)]\phi(t)dt)$
  The first principal direction ϕ_1, represents the direction of the data with the largest variation.

- The second eigenfunction ϕ_2
  $= \underset{\phi \perp \phi_1, \phi \in L^2(I), \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\int [X(t) - \mu(t)]\phi(t)dt)$

Properties of FPCA

- For vector X, the first eigenvector ϕ_1
  $= \underset{\phi \in \mathcal{R}^p, \|\phi\|=1}{\operatorname{argmax}} \ \mathsf{var}(\langle X - \mu, \phi \rangle)$

  is equivalent to the eigenvector corresponding to the largest eigenvalue of the covariance matrix cov (X).

Properties of FPCA

- For vector X, the first eigenvector ϕ_1
$$= \underset{\phi \in \mathcal{R}^p, \|\phi\|=1}{\text{argmax}} \ \text{var}(\langle X - \mu, \phi \rangle)$$
is equivalent to the eigenvector corresponding to the largest eigenvalue of the covariance matrix cov (X).

$\Longrightarrow$ $\text{cov}(X)\phi_1 = \lambda_1 \phi_1$ and $\text{cov}(X) = \sum_{k=1}^{p} \lambda_k \phi_k^T \phi_k$.

- This concept can be extended to function data but we need to define what "$\text{cov}(X)\phi_1$" means.

Properties of FPCA

- This leads to the definition of a covariance operator.

  Covariance function: $\Sigma(s,t) = \text{cov}(X(s), X(t)), s$ & $t \in I$.

  We 'll use the same notation Σ for the covariance function and its operator and define the covariance operator Σ
  (from $L^2(I)$ to $L^2(I)$) as:

  $$\Sigma(f) = \int_I \Sigma(s,t) f(s) ds, \text{ for any } f \in L^2(I).$$

Properties of FPCA

- This leads to the definition of a covariance operator.

  Covariance function: $\Sigma(s,t) = \text{cov}(X(s), X(t)), s \ \& \ t \in I$.

  We 'll use the same notation Σ for the covariance function and its operator and define the covariance operator Σ
  (from $L^2(I)$ to $L^2(I)$) as:

$$\Sigma(f) = \int_I \Sigma(s,t) f(s) ds, \quad \text{for any } f \in L^2(I).$$

- FPCA = spectral decomposition of the covariance operator:

  $\Sigma(\phi_k) = \lambda_k \phi_k,$

  λ_k & ϕ_k are the eigenvalues and eigenfunctions of Σ.

Properties of FPCA

- Mercer's theorem (under mild assumption) implies that

  $$\Sigma(s,t) = \Sigma_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

  the convergence above is uniform over s and t.

Properties of FPCA

- Mercer's theorem (under mild assumption) implies that

  $\Sigma(s, t) = \Sigma_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$,

  the convergence above is uniform over s and t.

- This leads to the Karhunen-Loève decomposition:

  $X(t) = \mu(t) + \Sigma_{k=1}^{\infty} A_k \phi_k(t)$,

  var$(A_k) = \lambda_k$, the k-th largest eigenvalue of Σ,

  $A_k = \int_I [X(t) - \mu(t)] \phi_k(t) dt$, are uncorrelated PC (scores).

  $\Rightarrow$ Isometry between $X(\cdot)$ and $\{A_k : k \geq 1\}$.

Steps to FPCA (Yao, Müller & W., 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s,t)$.
   -This usually involves smoothing as was done in Part I.

Steps to FPCA (Yao, Müller & W., 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s,t)$.
   -This usually involves smoothing as was done in Part I.

2. Estimate the eigenvalues and eigenfunctions of $\Sigma(s,t)$.
   - This is done through discretizing the covariance function on a dense time grid, then perform spectral analysis on the discretized covariance matrix.

Steps to FPCA (Yao, Müller & W., 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s,t)$.
   -This usually involves smoothing as was done in Part I.

2. Estimate the eigenvalues and eigenfunctions of $\Sigma(s,t)$.
   - This is done through discretizing the covariance function on a dense time grid, then perform spectral analysis on the discretized covariance matrix.

3. Estimate the PC scores $A_k = \int (X(t) - \mu(t))\phi_k(t)dt$.
   - Numerical integration is used to approximate the integral but this requires intense $(n_i \to \infty)$ measurements of the functional data.

Steps to FPCA (Yao, Müller & W, 2005)

1. When functional data are observed at a few time points, the numerical integration method does not work.

Steps to FPCA (Yao, Müller & W, 2005)

1. When functional data are observed at a few time points, the numerical integration method does not work.

2. Yao, M. and W. (2005) proposed PACE (principal analysis via conditional mean) to resolve this issue.

$$\hat{A}_{ik} = \hat{E}(A_{ik}|\ Y_i) = \hat{\lambda}_k\ \hat{\phi}_k^T\ \hat{\Sigma}_{Y_i}^{-1}(Y_i - \mu_i),$$

where $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ is the observed n_i-dim data (possibly with measurement errors) for the ith subject.

Recover the Functional Data through FPCA

- Use Karhunen Loève decomposition to recover the latent curve

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \phi_k(t)$$

$$\Downarrow$$

$$\hat{X}_{ik}(t) = \hat{\mu}(t) + \sum_{k=1}^{\infty} \hat{A}_{ik} \hat{\phi}_k(t)$$

Convergence Rates for FPCA: Dense Case

- Both the mean and covariance functions can be estimated at the $\sqrt{n}-$rate.

Convergence Rates for FPCA: Dense Case

- Both the mean and covariance functions can be estimated at the $\sqrt{n}-$rate.
- Perturbation theory then implies that the eigen-values and eigen-functions can be estimated at the $\sqrt{n}-$ rate.

Convergence Rates for FPCA: Dense Case

- Both the mean and covariance functions can be estimated at the $\sqrt{n}-$rate.

- Perturbation theory then implies that the eigen-values and eigen-functions can be estimated at the $\sqrt{n}-$ rate.

  - **Life is easy!**

Convergence Rates for FPCA: Non-dense Case

- The mean function can be estimated at the one-dim nonparametric rate.

Convergence Rates for FPCA: Non-dense Case

- The mean function can be estimated at the one-dim nonparametric rate.
- The covariance function can be estimated at the 2-dim nonparametric rate.
- Perturbation theory implies that the eigen-values and eigen-functions can be estimated at the 2-dim nonparametric rate, **but this rate is suboptimal!**

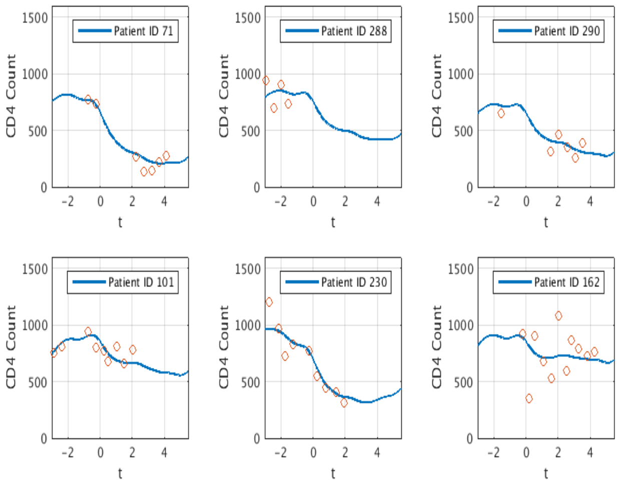
Convergence Rates for FPCA: Non-dense Case

- The mean function can be estimated at the one-dim nonparametric rate.
- The covariance function can be estimated at the 2-dim nonparametric rate.
- Perturbation theory implies that the eigen-values and eigen-functions can be estimated at the 2-dim nonparametric rate, **but this rate is suboptimal!**
- Hall, M. and W. (2006) showed that by undersmoothing the covariance function:

  the first K (finite) eigen-values can be estimated at the $\sqrt{n}$-rate ,

  the first K (finite) eigen-functions can be estimated at the one-dim nonparametric rate.

AIDS CD4: 6 Randomly Selected Subjects



http://www.stat.ucdavis.edu/PACE/

Outline

Outline

What Does FPCA Offer?

- The Karhunen-Loève decomposition is useful to impute functional data.

What Does FPCA Offer?

- The Karhunen-Loève decomposition is useful to impute functional data.

- FPCA provides a way to project an infinite dimensional function onto a finite K-dimensional subspace,
  the space spanned by the first K eigenfunctions.

  $\Longrightarrow$ The main information of functional data can be summarized by finitely many (K) PC components.

What Does FPCA Offer?

- The Karhunen-Loève decomposition is useful to impute functional data.

- FPCA provides a way to project an infinite dimensional function onto a finite K-dimensional subspace,
  the space spanned by the first K eigenfunctions.

  $\Longrightarrow$ The main information of functional data can be summarized by finitely many (K) PC components.

  - The proportion of variation explained by the first K PC components is the ratio $\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}$.

What Does FPCA Offer?

- This process transfers functional data to K-dim multivariate data consisting of the first K PC scores, so any existing method for multivariate data can be applied to these scores providing off-the-shelf methods for functional data.

  Examples: Clustering and classification of functional data.

What Does FPCA Offer?

- This process transfers functional data to K-dim multivariate data consisting of the first K PC scores, so any existing method for multivariate data can be applied to these scores providing off-the-shelf methods for functional data.

  Examples: Clustering and classification of functional data.

- **Caveat:** In theory $K = K_n \to \infty$, so the post-FPCA theory still falls into the nonparametric paradigm.

What Does FPCA Offer?

- While it is possible to expand functional data with any basis functions, such as B-splines, Fourier bases, wavelets, etc., FPCA provides the most parsimonious way to do so.

  By virtue of its definition, FPCA requires less components than other basis functions.

What Does FPCA Offer?

- While it is possible to expand functional data with any basis functions, such as B-splines, Fourier bases, wavelets, etc., FPCA provides the most parsimonious way to do so.

  By virtue of its definition, FPCA requires less components than other basis functions.

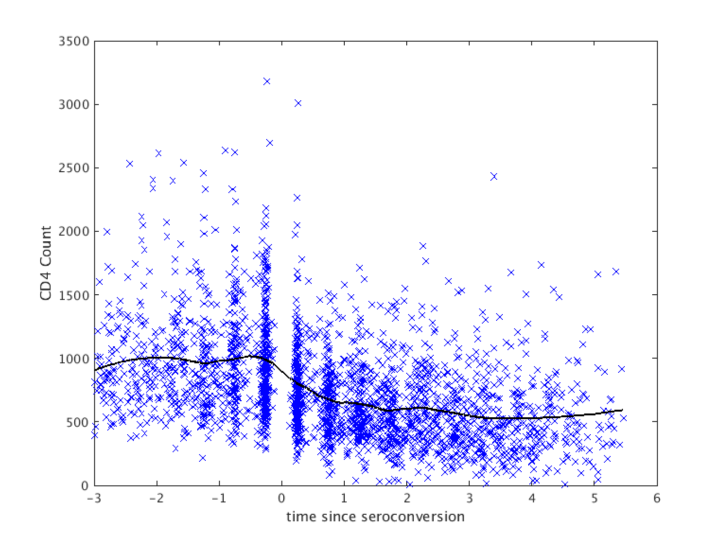
- However, the basis functions for FPCA needs to be estimated, which makes the theory a little harder than for preselected basis functions.

What Does FPCA Offer?

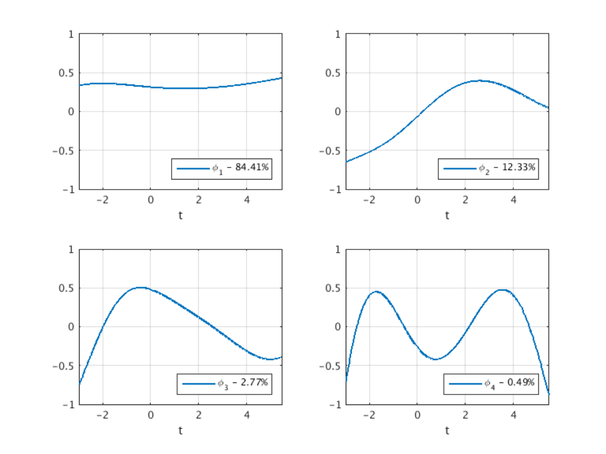
- The shape of μ and ϕ_k may help us to settle on a more parsimonious or parametric model.

  Example: CD4 counts.

Mean Curve: CD4 counts of all patients



AIDS CD4: Eigenfunctions



Model Building through Mean and Eigenfunctions

- AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only.

Model Building through Mean and Eigenfunctions

- AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only.
- The linear time trend is arguable but the random intercept is not far off as the first eigenfunction looks flat and already explains over 84% of the variation in the data.

Model Building through Mean and Eigenfunctions

- AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only.

- The linear time trend is arguable but the random intercept is not far off as the first eigenfunction looks flat and already explains over 84% of the variation in the data.

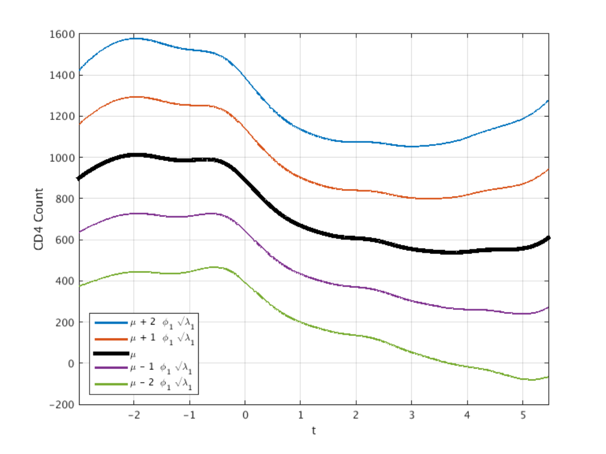
- However, we can increase the total variation explained to over 96% if a second PC is added.

  - Note that even with two components, the computational effort may be less than that for a good parametric random effects model, which may need four random effects (if a piecewise linear function is used to capture the shape of the AIDS data) and EM-algorithm.

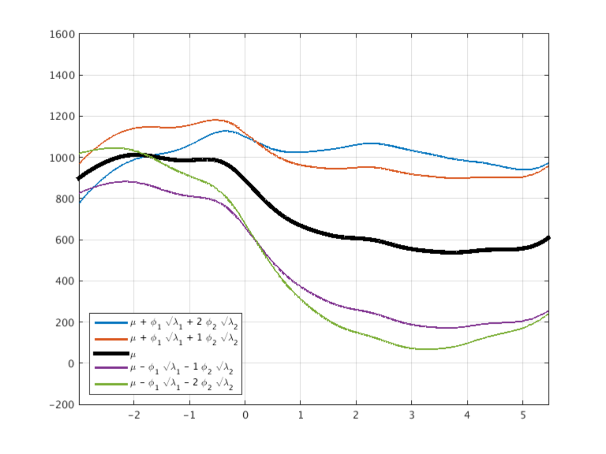
What Else can FPCA Offer?

- The principal directions ϕ_k explain the modes of variations of functional data (Rice and Jones, 1991).

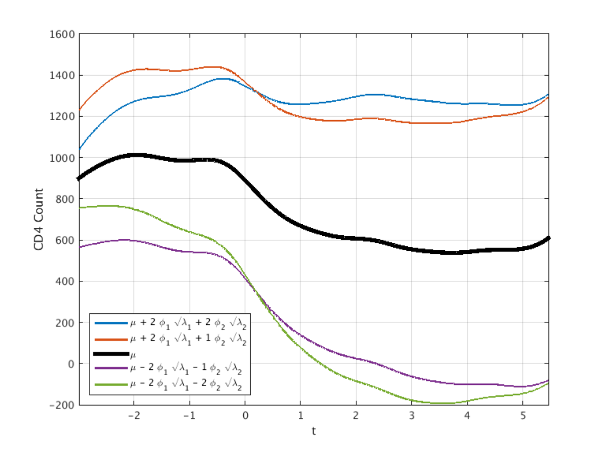
AIDS CD4: Modes of Variation



AIDS CD4: Modes of Variation



AIDS CD4: Modes of Variation



Functional Box Plot

- The above plots remind us of Tukey's Box plot for scalar data.
- Box plots have been extended to functional data, but the extension is non trivial and still open for improvements!

  Why?

Functional Box Plot

- The above plots remind us of Tukey's Box plot for scalar data.
- Box plots have been extended to functional data, but the extension is non trivial and still open for improvements!

  Why?

- There are two R-packages for functional box plots.

  Hyndman and Shang (2010, JCGS) - rainbow, box, and bag plots

  Sun and Genton (2011, JCGS) - boxplots

References

- Yao, Müller and Wang (2005, JASA)
  Methods and theory
- Hall, Müller and Wang (2006, AOS)
  Improved theory on eigenfunctions and eigenvalues.

References

- Yao, Müller and Wang (2005, JASA)
  Methods and theory
- Hall, Müller and Wang (2006, AOS)
  Improved theory on eigenfunctions and eigenvalues.
- Ramsay and Silverman (2005) "Functional data Analysis"
- Hsing and Eubank (2015)
  "Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators"

References for FPCA

- Functional data

  Dauxois, Pousse & Romain (1982)
  Rice & Silverman (1991)
  Cardot (2000)
  Hall & Hosseini-Nasab (2006)

- Longitudinal data

  Shi, Weiss & Taylor(1996)
  James, Sugar & Hastie(2000)
  Rice & Wu (2001)
  Yao, Müller & Wang (2005)

End of FPCA

Outline

Covariate Adjusted FPCA

The above FPCA assumes that data come from one population.

What if we have additional information of a covariate Z (a scalar) or $Z(t)$ (a functional or longitudinal covariate)?

- This is straightforward for dense functional data with scalar covariate Z.

  Chiou, Müller & W. (2003)
  Cardot (2006)

Covariate Adjusted FPCA

The above FPCA assumes that data come from one population.

What if we have additional information of a covariate Z (a scalar) or $Z(t)$ (a functional or longitudinal covariate)?

- This is straightforward for dense functional data with scalar covariate Z.

  Chiou, Müller & W. (2003)
  Cardot (2006)

- Their methods do not work for sparse functional data or longitudinal covariates.

Chiou, Müller & W. (2003, JRSS B)

- First, pool all the data together to get the overall mean function $\mu(\bullet)$ and eigenfunctions $\phi_k(\bullet)$ of the overall covariance function.

  $\Rightarrow \qquad Y(t) = \mu(t) + \sum_k A_k \phi_k(t)$

Chiou, Müller & W. (2003, JRSS B)

- First, pool all the data together to get the overall mean function $\mu(\bullet)$ and eigenfunctions $\phi_k(\bullet)$ of the overall covariance function.

  $\Rightarrow \qquad Y(t) = \mu(t) + \sum_k A_k \phi_k(t)$

- Next, incorporate the covariate information through the conditional mean function :

$$\mu(t, Z) = E(Y(t)|Z) = \mu(t) + \sum_k E(A_k|Z)\phi_k(t)$$

Chiou, Müller & W. (2003, JRSS B)

- First, pool all the data together to get the overall mean function $\mu(\bullet)$ and eigenfunctions $\phi_k(\bullet)$ of the overall covariance function.

  $\Rightarrow \qquad Y(t) = \mu(t) + \sum_k A_k \phi_k(t)$

- Next, incorporate the covariate information through the conditional mean function :

$$\mu(t, Z) = E(Y(t)|Z) = \mu(t) + \sum_k E(A_k|Z)\phi_k(t)$$

- This requires consistent estimates of A_k, which is not feasible for sparse functional data. **(Why?)**

Cardot (2006, SJS)

- Assume the whole process can be observed without measurement errors, but the mean $\mu(t, z)$ and covariance $\Sigma(s, t, z)$ functions both vary with $Z = z$, hence the eigenfunctions and PC scores also vary with z.

$$Y(t, z) = \mu(t, z) + \sum_k A_k(z)\phi(t, z).$$

Cardot (2006, SJS)

- Assume the whole process can be observed without measurement errors, but the mean $\mu(t, z)$ and covariance $\Sigma(s, t, z)$ functions both vary with $Z = z$, hence the eigenfunctions and PC scores also vary with z.

$$Y(t, z) = \mu(t, z) + \sum_k A_k(z)\phi(t, z).$$

- Since the whole random functions are observable, one can perform one-dimensional smoothing on Z to estimate $\mu(t, z)$ (at each fixed time t) and $\Sigma(s, t, z)$ (at each fixed (s, t).

Cardot (2006, SJS)

- Assume the whole process can be observed without measurement errors, but the mean $\mu(t,z)$ and covariance $\Sigma(s,t,z)$ functions both vary with $Z=z$, hence the eigenfunctions and PC scores also vary with z.

$$Y(t,z) = \mu(t,z) + \sum_k A_k(z)\phi(t,z).$$

- Since the whole random functions are observable, one can perform one-dimensional smoothing on Z to estimate $\mu(t,z)$ (at each fixed time t) and $\Sigma(s,t,z)$ (at each fixed (s,t).

- This approach does not work for sparse or irregular dense data.

Jiang & W. (2010, AoS) : A Unified Approach

- Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA

Jiang & W. (2010, AoS) : A Unified Approach

- Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA

- Both approaches consist of two parts: A systematic part corresponding to the mean function and a stochastic part comprising the random components.

Jiang & W. (2010, AoS) : A Unified Approach

- Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA
- Both approaches consist of two parts: A systematic part corresponding to the mean function and a stochastic part comprising the random components.
- The difference between these two approaches is in the handling of the covariance structure.

Covariate adjusted FPCA: Longitudinal Data

- Suppose the data originate from a random function $X(t, z)$ with mean $\mu(t, z)$, where z is the value of a covariate Z.

Covariate adjusted FPCA: Longitudinal Data

- Suppose the data originate from a random function $X(t, z)$ with mean $\mu(t, z)$, where z is the value of a covariate Z.
- There are two ways to handle the covariance function:

  (i) Fully adjusted FPCA (fFPCA)

  - the covariance function $\Sigma(s, t, z)$ varies with the covariate z,

  (ii) Mean adjusted FPCA (mFPCA)

  - the covariance function $\Sigma(s, t)$ does not vary with the covariate.

Fully Adjusted FPCA (fFPCA)

- This approach assumes that the covariance function $\Sigma(s,t,z)$ varies with z,
  and the corresponding eigenfunctions $\phi_k(t,z)$ and eigenvalues $\lambda_k(z)$ vary with Z:

$$\Sigma(s,t,z) = \sum_k \lambda_k(z)\phi_k(s,z)\phi_k(t,z)$$

Fully Adjusted FPCA (fFPCA)

- This approach assumes that the covariance function $\Sigma(s,t,z)$ varies with z,
  and the corresponding eigenfunctions $\phi_k(t,z)$ and eigenvalues $\lambda_k(z)$ vary with Z:

  $$\Sigma(s,t,z) = \sum_k \lambda_k(z)\phi_k(s,z)\phi_k(t,z)$$

- Karhunen-Loeve expansion implies random trajectory $X(t,z)$ can be represented as:

  $$X(t,z) = \mu(t,z) + \sum_k A_k(z)\phi_k(t,z)$$

Mean Adjusted FPCA (mFPCA)

- The second approach takes the view:
  first center each subject to $Y(t) - \mu(t, z)$, then pool all the centered subjects together to get a pooled covariance function:

  $\Sigma(s, t) = \sum_k \lambda_k^* \phi_k^*(s) \phi_k^*(t)$.

Mean Adjusted FPCA (mFPCA)

- The second approach takes the view:
  first center each subject to $Y(t) - \mu(t,z)$, then pool all the centered subjects together to get a pooled covariance function:

  $\Sigma(s,t) = \sum_k \lambda_k^* \phi_k^*(s) \phi_k^*(t)$.

- Karhunen-Loeve expansion thus implies that the random trajectory $X(t,z)$ can be represented as:

  $X(t,z) = \mu(t,z) + \sum_k A_k^*(z) \phi_k^*(t)$

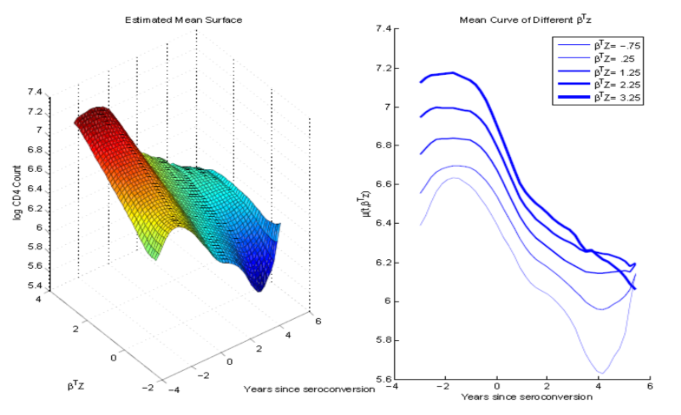
Estimation: Mean Function

The mean function for fFPCA and mFPCA are the same and can be estimated using any two-dimensional scatter-plot smoother of Y_{ij} on (T_{ij}, Z_i).

Local linear estimator: $\hat{\mu}_L(t,z) = \hat{\beta}_0$, where for $\beta = (\beta_0, \beta_1, \beta_2)$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_2\left(\frac{t - T_{ij}}{h_{\mu,t}}, \frac{z - Z_i}{h_{\mu,z}}\right) \times [Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) - \beta_2(Z_i - z)]^2$$

AIDS CD4: Mean Function



Estimation: Covariance Function

The covariance can be estimated by a scatter-plot smoother of the *raw covariances* defined as:

$$C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}, Z_i))(Y_{ik} - \hat{\mu}(T_{ik}, Z_i))$$

- fFPCA: three-dimensional smoother of C_{ijk} on (T_{ij}, T_{ik}, Z_i)
- mFPCA: two-dimensional smoother of C_{ijk} on (T_{ij}, T_{ik})

Estimation: Covariance Function

Since:

$$\begin{aligned}&\mathsf{cov}(Y_{ij}, Y_{ik}|T_{ij}, T_{ik}, Z_i)\\ =\ &\mathsf{cov}(X(T_{ij}, Z_i), X(T_{ik}, Z_i)) + \sigma^2\delta_{jk}\end{aligned}$$

where δ_{jk} is 1 if $j = k$, and 0 otherwise, the diagonal of the raw covariances C_{ijk} should not be included in the covariance function smoothing step.

Example of Covariance Estimates

- Linear local smoother for fFPCA:

$$\Sigma_L(t,s,z) = \hat{\beta}_0, \text{where:}$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{n} \sum_{1 \le j \ne k \le N_i} K_3(\frac{t - T_{ij}}{h_{G,t}}, \frac{s - T_{ik}}{h_{G,t}}, \frac{z - Z_i}{h_{G,z}})$$

$$\times [C_{ijk} - (\beta_0 + \beta_1(T_{ij} - t) + \beta_2(T_{ik} - s) - \beta_3(Z_i - z))]^2\}$$

Example of Covariance Estimates

- Linear local smoother for fFPCA:

$$\Sigma_L(t,s,z) = \hat{\beta}_0, \text{where:}$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{n} \sum_{1 \le j \ne k \le N_i} K_3(\frac{t - T_{ij}}{h_{G,t}}, \frac{s - T_{ik}}{h_{G,t}}, \frac{z - Z_i}{h_{G,z}}) \times [C_{ijk} - (\beta_0 + \beta_1(T_{ij} - t) + \beta_2(T_{ik} - s) - \beta_3(Z_i - z))]^2\}$$

- Linear local smoother for mFPCA:

$$\Sigma^*(t,s) = \hat{\beta}_0, \text{where:}$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{n} \sum_{1 \le j \ne k \le N_i} K_1(\frac{t - T_{ij}}{h_{G^*}}) K_1(\frac{s - T_{ik}}{h_{G^*}}) \times [C_{ijk} - (\beta_0 + \beta_1(T_{ij} - t) + \beta_2(T_{ik} - s)]^2\}$$

Estimation: Variance of Measurement Errors

The variance of $Y(t)$ for a given z is:

$$V(t,z) = \Sigma(t,t,z) + \sigma^2$$

$$\hat{V}(t,z) = \hat{\beta}_0, \text{where:}$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_2\left(\frac{t - T_{ij}}{h_{V,t}}, \frac{z - Z_i}{h_{V,z}}\right)$$
$$\times \left[C_{ijj} - \beta_0 + \beta_1(T_{ij} - t) + \beta_2(Z_i - z)\right)\right]^2$$

Estimation: Variance of Measurement Errors

For stability,

$$\hat{\sigma}^2 = \frac{2}{T}\int_Z \int_{T_1} \{\hat{V}(t,z) - \hat{\Sigma}_L(t,t,z)\} dtdz,$$

where:

$$T_1 = [inf\{t : t \in T\} + |T|/4, sup\{t : t \in T\} - |T|/4]$$

Estimation: Variance of Measurement Errors

For stability,

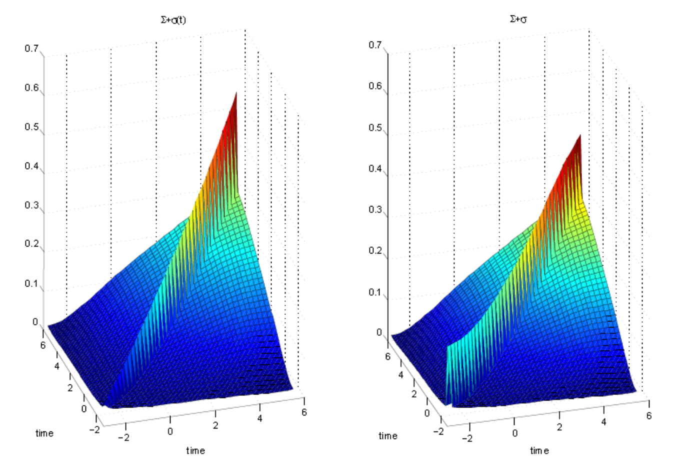
$$\hat{\sigma}^2 = \frac{2}{T} \int_Z \int_{T_1} \{\hat{V}(t,z) - \hat{\Sigma}_L(t,t,z)\} dt dz,$$

where:

$$T_1 = [inf\{t : t \in T\} + |T|/4, sup\{t : t \in T\} - |T|/4]$$

• So far, we have assumed homoscedastic errors but it is possible to allow the errors to vary with the time of measurement.

AIDS: Estimated Covariance + measurement error



Estimation: Eigenvalues and Eigenfunctions

- fFPCA: The solutions of the eigen-equations,

$$\int \hat{\Sigma}_L(t,s,z)\hat{\phi}_k(s,z)ds = \hat{\lambda}_k(z)\hat{\phi}_k(t,z),$$

  where the $\hat{\phi}_k(t,z)$ satisfies $\int \hat{\phi}_k^2(t,z)dt = 1$ and $\int \hat{\phi}_k(t,z)\hat{\phi}_m(t,z)dt = 0$ for $m < k$.

Estimation: Eigenvalues and Eigenfunctions

- fFPCA: The solutions of the eigen-equations,

$$\int \hat{\Sigma}_L(t,s,z)\hat{\phi}_k(s,z)ds = \hat{\lambda}_k(z)\hat{\phi}_k(t,z),$$

  where the $\hat{\phi}_k(t,z)$ satisfies $\int \hat{\phi}_k^2(t,z)dt = 1$ and $\int \hat{\phi}_k(t,z)\hat{\phi}_m(t,z)dt = 0$ for $m < k$.

- mFPCA: The solutions of the eigen-equations,

$$\int \hat{\Sigma}_L^*(t,s)\hat{\phi}_k^*(s)ds = \hat{\lambda}_k^*\hat{\phi}_k(t),$$

  where the $\hat{\phi}_k^*(t)$ satisfies $\int (\hat{\phi}_k^*(t))^2 dt = 1$ and $\int \hat{\phi}_k^*(t)\hat{\phi}_m^*(t)dt = 0$ for $m < k$.

Estimation: Principal Component Scores

- fFPCA:

  Use the conditional expectation (PACE) $E(A_{ik}(Z_i)|\tilde{Y}_i)$ to estimate the principal component scores, where $\tilde{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$
- Under the assumption that $\tilde{Y}_i$ is multivariate normal:

$$\hat{A}_{ik}(Z_i) = \hat{\lambda}\ \hat{\phi}_{ik}^T\ \Sigma_{\tilde{Y}_i}^{-1}(\tilde{Y}_i - \hat{\mu}_i)$$

where

$$\begin{aligned}
\hat{\mu}_i =& (\hat{\mu}(T_{i1}, Z_i), \ldots, \hat{\mu}(T_{iN_i}, Z_i))^T \\
(\hat{\Sigma}_{\tilde{Y}_i})_{j,k} =& \hat{\Sigma}_L(T_{ij}, T_{ik}, Z_i) + \hat{\sigma}^2 \delta_{jk} \\
\hat{\phi}_{ik} =& (\hat{\phi}_k(T_{ij}, Z_i), \ldots, \hat{\phi}_k(T_{iN_i}, Z_i))^T
\end{aligned}$$

Estimation: Principal Component Scores

The prediction of principal component scores in mFPCA is similar.

Rate of Convergence

- If $E(N) < \infty$ the rate of convergence for the 2D mean and covariance function is $n^{1/3}$.

  This is the optimal rate of convergence for 2D smoothers with independent data.

- If $E(N) \to \infty$, the rate of convergence can be as close to $n^{2/5}$ as possible but not be equal to $n^{2/5}$.

Optimal Rates of Convergence

- The first k eigenfunctions can be estimates at the same optimal rate as a 1- or 2-dim nonparametric regression function.
- The largest k eigenvalues can be estimated at the $\sqrt{n}$ rate.

Bandwidth Selection

- Mean function $\mu(t, z)$ and covariance $\Gamma^*(s, t)$:
  Leave one subject out cross-validation

- Covariance Function $\Gamma(s, t, z)$: k-fold cross-validation Suppose that the subjects are randomly assigned to k sets(S_1, S_2,$\ldots$,S_k).

$$h = \underset{h}{\operatorname{argmin}} \sum_{l=1}^{k} \sum_{i \in S_l} \sum_{1 \leq j \neq m \leq N_i} \{C_{ijm} - \hat{\Gamma}^{-S_l}(T_{ij}, T_{im}, z_i)\}^2$$

where $\hat{\Gamma}^{-S_l}(T_{ij}, T_{im}, z_i)$ is the estimated covariance function at (T_{ij}, T_{im}, z_i) when the subjects in S_l are not used to estimate $\Gamma(t, s, z)$.

Number of Eigenfunctions

Three methods:

- AIC
- BIC
- FVE: minimum number of eigen-components needed to explained at least a specified total fraction of the variation.

Extension to a Functional or Longitudinal Covariate

- The regression function can be extended to functional/longitudinal covariates by replacing Z with $Z(t)$.

$$E(Y(t)|Z) = E(Y(t)|Z(t)) = \mu(t, Z(t))$$

Extension to a Functional or Longitudinal Covariate

- The regression function can be extended to functional/longitudinal covariates by replacing Z with $Z(t)$.

$$E(Y(t)|Z) = E(Y(t)|Z(t)) = \mu(t, Z(t))$$

- The aforementioned method to estimate $\mu(t, Z)$ still works by replacing the scatter plot $\{(t_{ij}Z_i) : \forall\ i, j\}$ with $\{(t_{ij}, Z_i(t_{ij}) : \forall\ i, j\}$

End of Covariate Adjusted FPCA

Outline

1. Introduction to FPCA
   - What Does FPCA Offer?

2. Covariate adjusted FPCA
   - FPCA with Multidimensional Covariates
   - What's Next After FPCA?

Outline

Multidimensional Covariates

- Although both fFPCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.

Multidimensional Covariates

- Although both fFPCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.

- Assume that $Z \in R^p$, and for simplicity only the mean function depends on Z (i.e. mFPCA).

Multidimensional Covariates

- Although both fFPCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.

- Assume that $Z \in R^p$, and for simplicity only the mean function depends on Z (i.e. mFPCA).

  $\Longrightarrow \mu(t,z) = \mu(t, \beta^T z) \rightarrow$ single index

  or

  $$\mu(t,z) = \underbrace{\mu(t, \beta_1^T z, \beta_2^T z, \dots, \beta_k^T z)}_{\text{multiple indices}}, k < p$$

Dimension Reduction Models

- There are many ways to estimate the indices for independent data, i.e. when there is no t, including SIR (Li, 1991) and MAVE (Xia and Li, 2002).

$$Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.$$

Dimension Reduction Models

- There are many ways to estimate the indices for independent data, i.e. when there is no t, including SIR (Li, 1991) and MAVE (Xia and Li, 2002).

$$Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.$$

- A few have been extended to functional or longitudinal data, but none for the model:

$$Y(t) = \mu(t, \beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon(t).$$

Dimension Reduction Models

- There are many ways to estimate the indices for independent data, i.e. when there is no t, including SIR (Li, 1991) and MAVE (Xia and Li, 2002).

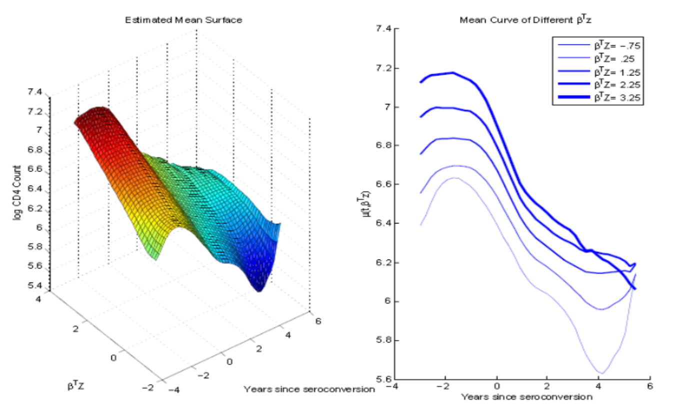
$$Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.$$

- A few have been extended to functional or longitudinal data, but none for the model:

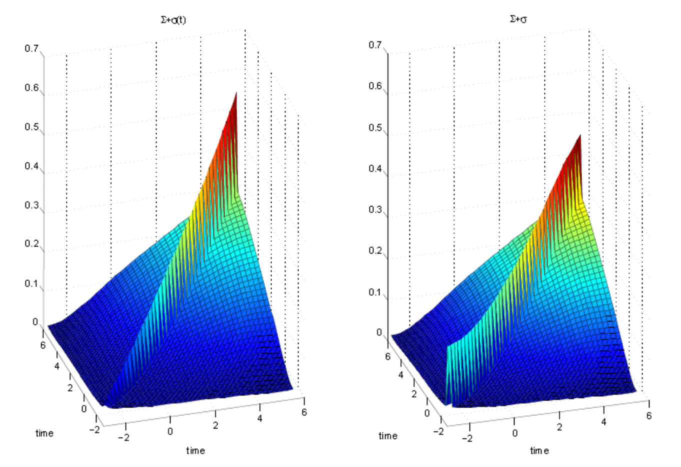
$$Y(t) = \mu(t, \beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon(t).$$

- Jiang and W. (2010) extended the "MAVE" approach by Xia et al (2002) to functional/longitudinal data and established $\sqrt{n}$-consistency of the estimates for β_j.

AIDS CD4: Estimated Mean



AIDS: Estimated Covariance + measurement error



End of Multidimensional Covariates

Outline

1. Introduction to FPCA
   - What Does FPCA Offer?

2. Covariate adjusted FPCA
   - FPCA with Multidimensional Covariates
   - What's Next After FPCA?

What's Next After FPCA?

- FPCA can be the end product, or it can be further used

  - to explore the covariate effects,
  - to recover the trajectories of each subject,
  - to explore the modes of variation
  - etc.

What's Next After FPCA?

- FPCA can be the end product, or it can be further used

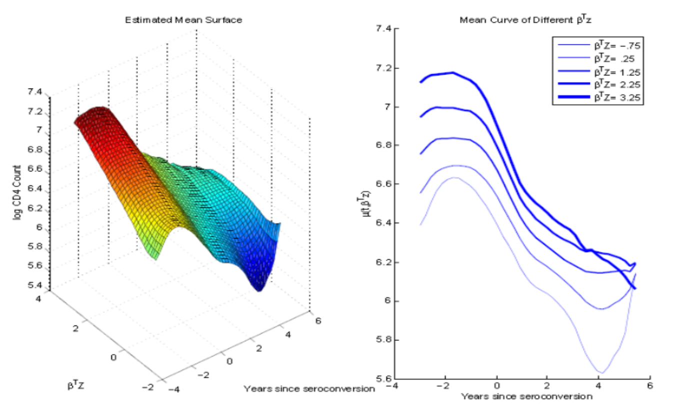
  - to explore the covariate effects,
  - to recover the trajectories of each subject,
  - to explore the modes of variation
  - etc.

- FPCA can help to find a more parsimonious model.

  - We have illustrated this already when no covariates are involved.

  - Next we explore model building when covariates are present.

AIDS CD4: Estimated Mean with Covariates adjusted



AIDS CD4 Data

- This suggests the possibility of a more parsimonious model with additive or multiplicative covariate effects.

  $Y(t) = \mu(t) + \psi(\beta^T z) + e(t) \rightarrow \mu(t)$ could be parametric, eg. a polynomial.

AIDS CD4 Data

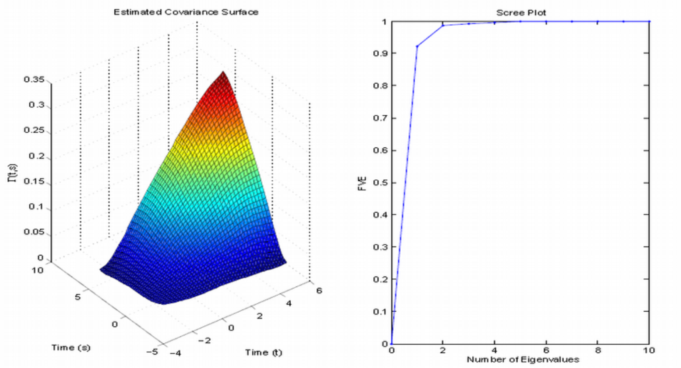
- This suggests the possibility of a more parsimonious model with additive or multiplicative covariate effects.

  $Y(t) = \mu(t) + \psi(\beta^T z) + e(t) \rightarrow \mu(t)$ could be parametric, eg. a polynomial.

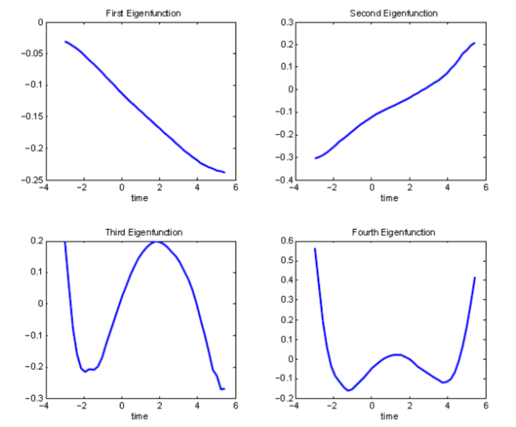
- Common marginal models for longitudinal data take the additive form, and employ parametric models for both the mean and covariance function.

  Both parametric forms are difficult to detect for sparse and noisy longitudinal data.

AIDS CD4: Estimated Covariance



AIDS CD4: Estimated Eigenfunctions

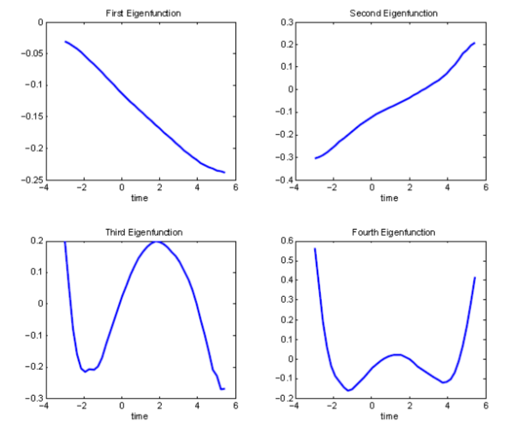


FVE		AIC (BIC)	
MSE	K	MSE	K
0.1154	1	0.0937	3

Adding Random Effects

Help to identify the form of the random effects.

$Y(t) = \mu(t)\psi(\beta^T z)$
$a + bt + e(t)$
$\downarrow \quad \downarrow$
random effects



Semiparametric Product Model

- If the first eigenfunction is proportional to the population mean function$\mu(t, Z)$ and explains almost all the variations of the data,

  we can discards the remaining eigenfunctions and arrive at the following multiplicative random effect model:

$$
\begin{aligned}
Y(t) =&\mu(t,z) + A\mu(t,z) + e(t)\\
&b\mu(t,z) + e(t)\\
&\downarrow
\end{aligned}
$$

Random effects

Semiparametric Product Model

- If the first eigenfunction is proportional to the population mean function$\mu(t, Z)$ and explains almost all the variations of the data,

  we can discards the remaining eigenfunctions and arrive at the following multiplicative random effect model:

$$
\begin{aligned}
Y(t) = &\mu(t,z) + A\mu(t,z) + e(t) \\
&b\mu(t,z) + e(t) \\
&\downarrow
\end{aligned}
$$

Random effects

- Examples of such multiplicative random effects models are bountiful and includes the PET data in Jiang, Aston and W. (2009 and the PBC data (bilirubin) in Ding and W. (2008).

References

- Jiang and Wang (2010, AoS)
  covariate adjusted FPCA
- Jiang and Wang (2011, AoS)
  dimension reduction (Semi-parametric index) model
- Jiang, Aston and Wang (2009, NeuroImage)
  multiplicative random effects model for PET data
- Ding and Wang (2008)
  multiplicative random effects model for PBC data

End of Part II