TUTORIAL ON FUNCTIONAL DATA ANALYSIS

Ana-Maria Staicu

Department of Statistics, North Carolina State University
astaicu@ncsu.edu

SAMSI

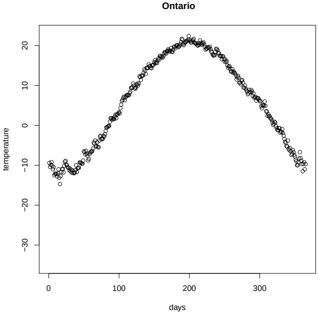
April 5, 2017

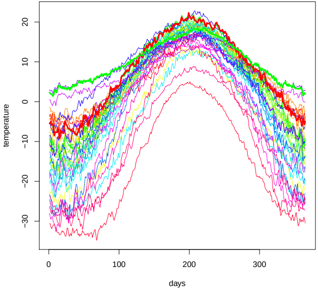
Broad overview of course topics:

1. Introduction to Functional Data
2. Modeling Functional Data with Preset Basis Expansions
3. Modeling Functional Data using Functional Principal Component Analysis
4. Beyond Independent and Identically Distributed Functional Data

Example of functional data

Canadian weather data. Daily (and monthly) temperature and precipitation at 35 different locations in Canada averaged over 35 years from 1960 to 1994.

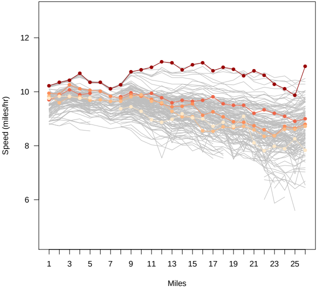




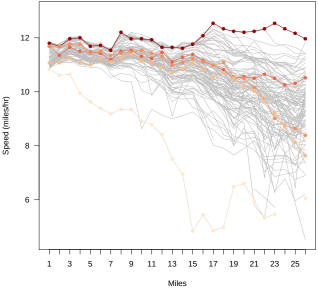
Avg Daily Temp in Ottawa *Avg Daily Temp across Canada*

EXAMPLE OF FUNCTIONAL DATA (CONT'D)

Marathon data. Running time performance of 149 females and 105 male athletes competing in the US Olympic Team Trials Marathon 02/16/2016.



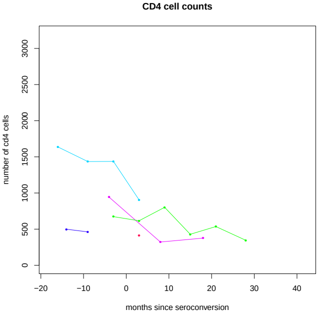
Speed for females



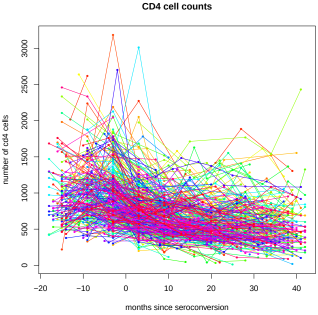
Speed for males

EXAMPLE OF FUNCTIONAL DATA (CONT'D)

CD4 data. CD4 cell count per mm of blood is a useful surrogate of the progression of HIV. Below are CD4 cell counts for 366 subjects between months -18 and 42 months since seroconversion (diagnosis of HIV).



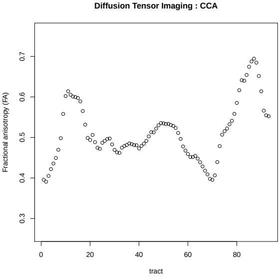
CD4 for few subjects



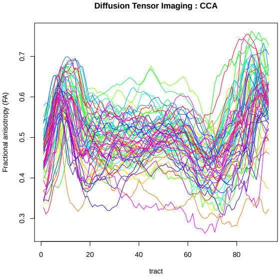
CD4 counts for all the subjects

Example of functional data (cont'd)

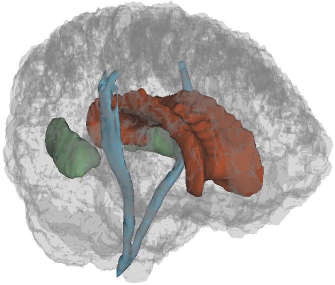
Diffusion tensor imaging (DTI) data. Fractional anisotropy (FA) - measure of the tissue integrity that is useful in diagnosis/progression of multiple sclerosis (MS) - along the main direction of the corpus callosum (CCA) for many subj.



FA obs for one MS subj



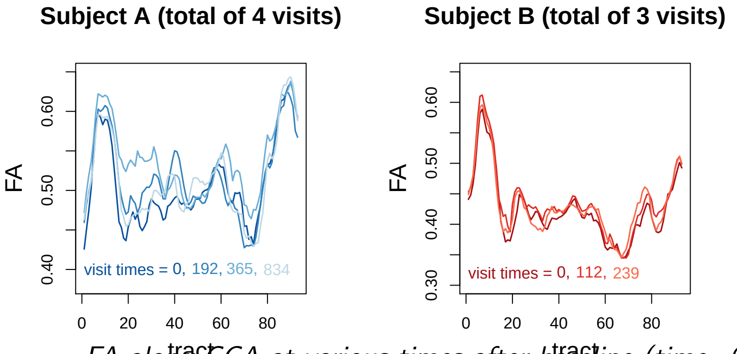
FA for all subj



Brain CCA (red)

Example of functional data (cont'd)

Diffusion tensor imaging (DTI) data. Each subject is observed at several hospital visits and at each time a FA profile is measured. Below: FA for two subjects.
Of interest: what is the dynamic behavior over time ?



FA along CCA at various times after baseline (time=0)

Some characteristics of functional data:

- Often high dimensional
- Typically involves multiple measurements of the same "process"
- Interpretability across subject domains
- Parametric assumptions on the underlying "process" are often *not* made

FDA vs multivariate data analysis

- Topology: well defined topology for FDA / one can permute the elements in MDA
- Model covariance assn: smooth for FDA / unstructured or sparse for MDA

FDA vs longitudinal data analysis

- Model covariance: no assn for FDA / parametric assn for LDA
- Mechanisms of missingness: not very important in FDA / very important for LDA
- Interest more in subject-specific trajectories for FDA / inference for LDA
- Sampling design: typically high frequency for FDA / sparse and irregular for LDA

Some references

- Ramsay & Silverman, 2005, "Functional Data Analysis"
- Ramsay, Hooker & Graves, 2009, "Functional Data Analysis in R and Matlab"
- Horvath & Kokoszka, 2012, "Inference for Functional Data with Applications"
- Bosq, 2002, "Linear Processes on Function Spaces"
- Ferraty & Vieux, 2006, "Nonparametric Functional Data Analysis"
- Zhang, 2013, "Analysis of Variance for Functional Data"
- Hsing & Eubank, 2015, "Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators"

From discrete to functional data. Intuition

The term *functional* in reference to observed data refers to the intrinsic structure of the data being functional; i.e. there is an underlying function that gives rise to the observed data.

From discrete to functional data. Intuition

The term *functional* in reference to observed data refers to the intrinsic structure of the data being functional; i.e. there is an underlying function that gives rise to the observed data.

Advantages of representing the data as a smooth function:

- allows evaluation at any time point
- allows evaluation of rates of change of the underlying curve
- allows registration to a common time-scale

From discrete to functional data. Intuition

The term *functional* in reference to observed data refers to the intrinsic structure of the data being functional; i.e. there is an underlying function that gives rise to the observed data.

Advantages of representing the data as a smooth function:

- ▷ allows evaluation at any time point
- ▷ allows evaluation of rates of change of the underlying curve
- ▷ allows registration to a common time-scale

Main idea in FDA: treat the observed data functions as *single entities*, rather than sequence of individual observations.

Notation

Observed data: $\{(Y_{i1}, t_{i1}), \ldots, (Y_{im_i}, t_{im_i})\}_i$;

Notation

Observed data: $\{(Y_{i1}, t_{i1}), \ldots, (Y_{im_i}, t_{im_i})\}_i$;

- Y_{ij}: the "snapshot" an underlying ith signal/curve -latent- $X_i(\cdot)$, at timepoint t_{ij}, possibly blurred by error
- $X_i(\cdot)$ smooth latent curve on $\mathcal{T}$; $X_i(\cdot)$'s are independent realizations of a stochastic process $X(\cdot)$

Notation

Observed data: $\{(Y_{i1}, t_{i1}), \ldots, (Y_{im_i}, t_{im_i})\}_i$;

- Y_{ij}: the "snapshot" an underlying ith signal/curve -latent- $X_i(\cdot)$, at timepoint t_{ij}, possibly blurred by error
- $X_i(\cdot)$ smooth latent curve on $\mathcal{T}$; $X_i(\cdot)$'s are independent realizations of a stochastic process $X(\cdot)$
- $t_{ij} \in \mathcal{T}$; t_{ij} may vary across i and j

Notation

Observed data: $\{(Y_{i1}, t_{i1}), \ldots, (Y_{im_i}, t_{im_i})\}_i$;

- Y_{ij}: the "snapshot" an underlying ith signal/curve -latent- $X_i(\cdot)$, at timepoint t_{ij}, possibly blurred by error
- $X_i(\cdot)$ smooth latent curve on $\mathcal{T}$; $X_i(\cdot)$'s are independent realizations of a stochastic process $X(\cdot)$
- $t_{ij} \in \mathcal{T}$; t_{ij} may vary across i and j
- If no noise, then $Y_{ij} = X_i(t_{ij})$; otherwise $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$.
  ϵ_{ij} - white noise; Typically ϵ_{ij} are IID

NOTATION

Observed data: $\{(Y_{i1}, t_{i1}), \ldots, (Y_{im_i}, t_{im_i})\}_i$;

- Y_{ij}: the "snapshot" an underlying ith signal/curve -latent- $X_i(\cdot)$, at timepoint t_{ij}, possibly blurred by error
- $X_i(\cdot)$ smooth latent curve on $\mathcal{T}$; $X_i(\cdot)$'s are independent realizations of a stochastic process $X(\cdot)$
- $t_{ij} \in \mathcal{T}$; t_{ij} may vary across i and j
- If no noise, then $Y_{ij} = X_i(t_{ij})$; otherwise $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$. ϵ_{ij} - white noise; Typically ϵ_{ij} are IID

Common objectives :

- Characterize the pattern of variability among curves
- Recover the subject specific trajectories $X_i(\cdot)$'s

Smoothing

Why do we need smoothing?

- Data are often observed with error
- There's a need to "interpolate" to a common grid

Smoothing

Why do we need smoothing?

- Data are often observed with error
- There's a need to "interpolate" to a common grid

How are we going to do smoothing?

- Use prespecified basis functions (e.g. splines, wavelets, Fourier etc.)
- Use data-driven basis functions (e.g. functional principal components)

Prespecified basis expansion

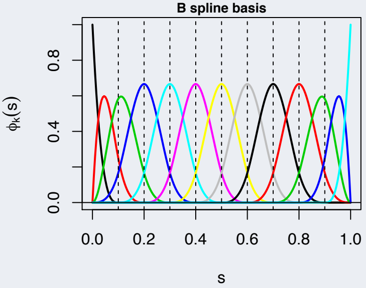
Let $\{\psi_1(\cdot), \psi_2(\cdot), \dots, \psi_K(\cdot)\}$ be a user specified basis. Assume:

$$Y_{ij} = \sum_{k=1}^{K} c_{ik}\psi_k(t_{ij}) + \epsilon_{ij}.$$

- ▷ We only need to estimate the subject-specific scores c_{ik}'s.

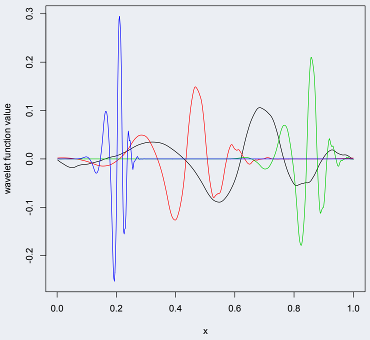
What kind of pre-specified basis to use? B-splines, wavelets etc

Some common basis functions: B-splines



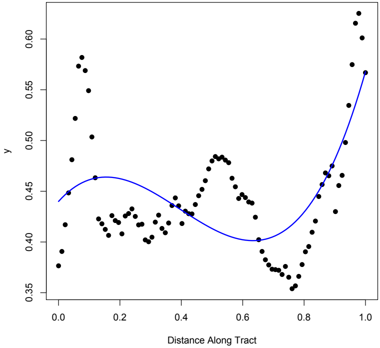
- Continuous
- Easily defined derivatives
- Good for smooth data

Some common basis functions: Wavelets

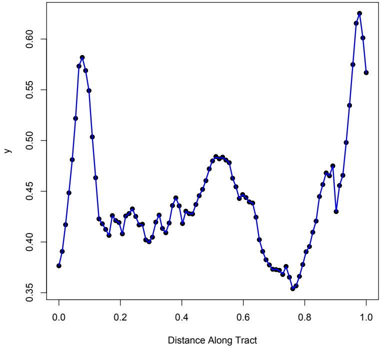


- Formed from a single "mother wavelet" function: $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$
- Orthonormal basis
- Particularly good when there are jumps, spikes, peaks, etc.
- Wavelet representation is *sparse*

Example



Few B-splines fns expansion



Many B-spline fns expansion

Tuning

For any curve, many possible smooths are available

- Depends on the type of basis
- Depends on the number of basis functions
- Depends on the estimation procedure

"Tuning" is the process of adjusting the smoother to the data at hand. This is often implicit (eg. kernel smoothing).

Prespecified basis expansion: estimation

Explicit penalization: use a large number of basis functions and penalize the "roughness" of the fit.

Leads to a penalized SSE:

$$PenSSE_i = \sum_{j=1}^{m_i} \left(Y_{ij} - \sum_{k=1}^{K} \psi_k(t_{ij}) c_{ik} \right)^2 + \lambda \text{Pen} \left(\sum_{k=1}^{K} \psi_k(\cdot) c_{ik} \right)$$

- Measure the roughness using derivatives of the fit, e.g. $\text{Pen}\left(\sum_{k=1}^K \psi_k(\cdot) c_{ik}\right) = \int \{\sum_{k=1}^K \psi_k''(t) c_{ik}\}^2 dt = c_i^T D c_i$, D is $K \times K$ matrix with (k, k') equal to $\int \psi_k''(t) \psi_{k'}''(t)\, ds$.
- Choose c_{ik}'s that minimize $PenSSE_i$; obtain $\widehat{X}_i(\cdot)$
- Need to select *tuning parameter* λ. Common ways: CV and REML

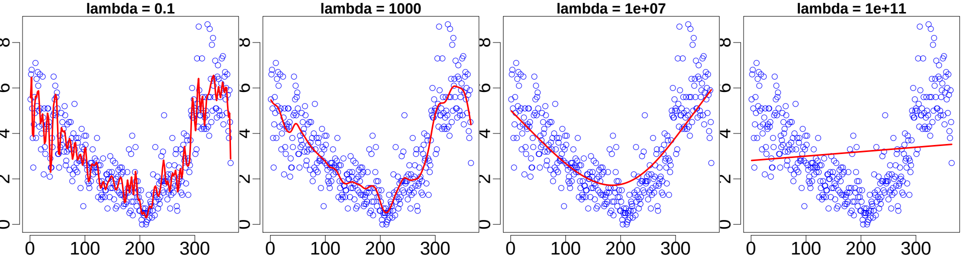
Smoothing parameter λ

- When $\lambda = 0$ the criterion reduces to the SSE (emphasis on fit)
- When $\lambda >> 0$ (is very large) more emphasis on smoothness

Smoothing parameter λ

- When $\lambda = 0$ the criterion reduces to the SSE (emphasis on fit)
- When $\lambda >> 0$ (is very large) more emphasis on smoothness

Illustration : Vancouver mean temperature. Fits using 100 cubic splines (with equally spaced knots) and various λ .



DATA-DRIVEN BASIS

- Previous bases don't depend on the data; only the loadings do
- Functional principal component analysis (FPCA) gives a "data-driven" basis (constructed from the observed data)
- Similar representation of the observed data $Y'_{ij}s$:

$$Y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} c_{ik}\phi_k(s_{ij}) + \epsilon_{i\ell}.$$

- Difference is that the $\phi_k(\cdot)$'s are not known.
- $\{\phi_1(\cdot), \phi_2(\cdot), \ldots\}$ describe the main directions of variability in the observed data

fPCA: main idea

Idea of fPCA: find projections of maximum variance.
Let $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID zero-mean curves in $L^2[\mathcal{T}]$.

fPCA: main idea

Idea of fPCA: find projections of maximum variance.
Let $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID zero-mean curves in $L^2[\mathcal{T}]$.

- The 1st fPC $= \phi_1(t)$ for which

$$\xi_{1i} =< \phi_1, X_i >= \int_{\mathcal{T}} \phi_1(t) X_i(t)\, dt$$

has maximum variance subject to the constraint $\|\phi_1\|^2 = 1$.

fPCA: main idea

Idea of fPCA: find projections of maximum variance.
Let $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID zero-mean curves in $L^2[\mathcal{T}]$.

- The 1st fPC $= \phi_1(t)$ for which

$$\xi_{1i} =< \phi_1, X_i >= \int_{\mathcal{T}} \phi_1(t) X_i(t)\, dt$$

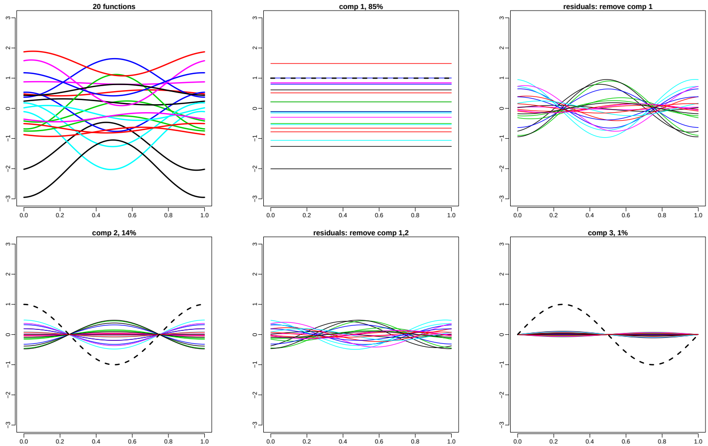
  has maximum variance subject to the constraint $\|\phi_1\|^2 = 1$.
- The 2nd fPC $= \phi_2(t)$ for which

$$\xi_{2i} =< \phi_2, X_i >= \int_{\mathcal{T}} \phi_2(t) X_i(t)\, dt$$

  has maximum variance, subject to $< \phi_2, \phi_1 >= 0$, $\|\phi_2\|^2 = 1$.
- ...

Data-driven basis: Illustration

FPCA for a sample of 20 curves. Displayed in dashed black line are 3 FPC: top middle panel, bottom left and right panels.



Model: $X_i(t) = \xi_{i1}\phi_1(t) + \xi_{i2}\phi_2(t) + \xi_{i3}\phi_3(t)$

How to obtain the fPC ?

Notation:

- $\Sigma(s,t) := E[\{X_i(s) - \mu(s)\}\{X_i(t) - \mu(t)\}]$ is the covar fn
- $\Sigma(s,t)$ induces an integral operator Σ

$$\Sigma f\ (t) = \int_{\mathcal{T}} \Sigma(t,s) f(s) ds, \qquad f(t) \in L^2[\mathcal{T}]$$

How to obtain the fPC ?

Notation:

- $\Sigma(s,t) := E[\{X_i(s) - \mu(s)\}\{X_i(t) - \mu(t)\}]$ is the covar fn
- $\Sigma(s,t)$ induces an integral operator Σ

$$\Sigma f\ (t) = \int_{\mathcal{T}} \Sigma(t,s) f(s) ds, \qquad f(t) \in L^2[\mathcal{T}]$$

Re-write the steps of the algorithm that defines fPCs:

- 1st fPC: $\phi_1 = \arg\max_f < \Sigma f, f >, \|f\|^2 = 1$
- 2nd fPC:
  $\phi_2 = \arg\max_f < \Sigma f, f >, < f, \phi_1 >= 0$ and $\|f\|^2 = 1$
- ...

fPCs ϕ_k's are the *eigenvectors* of the cov operator Σ: $\Sigma\phi_k = \lambda_k \phi_k$.

MERCER'S THM

- Assume $\Sigma(s,t)$ is continuous over $\mathcal{T} \times \mathcal{T}$. Then there exists: an orthonormal basis $\{\phi_k\}_k$ of continuous fns in $L^2[\mathcal{T}]$ and $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ such that
  - $\Sigma\phi_k = \lambda_k\phi_k$ where Σ is the operator induced by the cov fn
  - $\Sigma(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t), \quad t,s \in \mathcal{T}$, where the series converges uniformly on $\mathcal{T}^2$

KARHUNEN-LOÈVE EXPANSION

- $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID *zero-mean* curves in $L^2[\mathcal{T}]$ and $\{\phi_k\}_k$ - eigenfns of the cov fn $\Sigma(\cdot, \cdot)$.

$$X_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) \text{ (series converges in } L^2 \text{ norm)}$$

where

KARHUNEN-LOÈVE EXPANSION

- $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID *zero-mean* curves in $L^2[\mathcal{T}]$ and $\{\phi_k\}_k$ - eigenfns of the cov fn $\Sigma(\cdot,\cdot)$.

$$X_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) \text{ (series converges in } L^2 \text{ norm)}$$

where

- $\xi_{ik} := \int X_i(t)\phi_k(t)dt$ are random variables
- $E[\xi_{ik}] = 0$, $Var[\xi_{ik}] = \lambda_k$ and $\{\xi_{ik} : k \geq 1\}$ are mutually uncorrelated
- ξ_{ik}'s are called fPC scores for X_i

Implications KL expansion

- One-to-one mapping $X_i(\cdot) \to (\xi_{i1}, \xi_{i2}, \ldots)^T$

Implications KL expansion

- One-to-one mapping $X_i(\cdot) \to (\xi_{i1}, \xi_{i2}, \ldots)^T$
- λ_k is the average variance of the kth fPC ϕ_k, i.e.:

$$\lambda_k = \lambda_k \int \phi_k^2(t)dt = \int Var\{\xi_{ik}\phi_k(t)\}dt$$

Implications KL expansion

- One-to-one mapping $X_i(\cdot) \to (\xi_{i1}, \xi_{i2}, \ldots)^T$
- λ_k is the average variance of the kth fPC ϕ_k, i.e.:

$$\lambda_k = \lambda_k \int \phi_k^2(t) dt = \int Var\{\xi_{ik}\phi_k(t)\} dt$$

- $\int \{VarX_i(t)\} dt = \sum_{k=1}^{\infty} \lambda_k$

Implications KL expansion

- One-to-one mapping $X_i(\cdot) \rightarrow (\xi_{i1}, \xi_{i2}, \ldots)^T$
- λ_k is the average variance of the kth fPC ϕ_k, i.e.:

$$\lambda_k = \lambda_k \int \phi_k^2(t)dt = \int Var\{\xi_{ik}\phi_k(t)\}dt$$

- $\int \{VarX_i(t)\}dt = \sum_{k=1}^{\infty} \lambda_k$
- Percentage variance explained (PVE) by the first K fPCs:

$$\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{l=1}^{\infty} \lambda_l}$$

fPCA for sample of IID true curves

Recall: $\{X_i(t) : t \in \mathcal{T}\}_i$ be IID curves in $L^2[\mathcal{T}]$.

1. Sample mean $\bar{X}(t)$
2. Sample covariance fn:
   $C_X(s,t) = (n-1)^{-1} \sum_{i=1}^{n} \{X_i(s) - \bar{X}(s)\}\{X_i(t) - \bar{X}(t)\}$
3. Spectral decomposition (or eigenanalysis) of C_X gives pairs eigenfn/eigenval $\{\widehat{\phi}_k(\cdot), \lambda_k\}$

fPCA for sample of IID true curves

4. Select finite truncation K using PVE or information criteria (AIC, BIC etc.)

fPCA for sample of IID true curves

4. Select finite truncation K using PVE or information criteria (AIC, BIC etc.)
5. Estimate fPC scores: $\widehat{\xi}_{ik} = \int_{\mathcal{T}} \{X_i(t) - \bar{X}(t)\} \widehat{\phi}_k(t) dt$

fPCA for sample of IID true curves

4. Select finite truncation K using PVE or information criteria (AIC, BIC etc.)
5. Estimate fPC scores: $\widehat{\xi}_{ik} = \int_{\mathcal{T}} \{X_i(t) - \bar{X}(t)\} \widehat{\phi}_k(t) dt$
6. KL expansion $X_i(t) = \bar{X}(t) + \sum_{k=1}^{K} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$

Theoretical properties fPCA

Consistency results for the sample mean, sample covariance functions, and the corresponding eigenvalues/eigenfunctions.

- $E\bar{X} = \mu$ and $E\|\bar{X} - \mu\|^2 = O(n^{-1})$.

Theoretical properties fPCA

Consistency results for the sample mean, sample covariance functions, and the corresponding eigenvalues/eigenfunctions.

- $E\bar{X} = \mu$ and $E\|\bar{X} - \mu\|^2 = O(n^{-1})$.
- The sample covariance is unbiased and is mean squared consistent estimator of the covariance function

  Assume $\lambda_1 > \lambda_2 > \ldots > \lambda_K > \lambda_{K+1} \geq 0$

  Let $\{\widehat{\lambda}_k, \widehat{\phi}_k\}_k$ be the eigenvalues/eigenfunctions and $\widehat{c}_k = sign \int \widehat{\phi}_k(t)\phi_k(t)dt$. Then

Theoretical properties fPCA

Consistency results for the sample mean, sample covariance functions, and the corresponding eigenvalues/eigenfunctions.

- $E\bar{X} = \mu$ and $E\|\bar{X} - \mu\|^2 = O(n^{-1})$.
- The sample covariance is unbiased and is mean squared consistent estimator of the covariance function

  Assume $\lambda_1 > \lambda_2 > \ldots > \lambda_K > \lambda_{K+1} \geq 0$

  Let $\{\widehat{\lambda}_k, \widehat{\phi}_k\}_k$ be the eigenvalues/eigenfunctions and $\widehat{c}_k = sign \int \widehat{\phi}_k(t)\phi_k(t)dt$. Then

  $$\limsup_{n\to\infty} nE[\|\widehat{c}_k\widehat{\phi}_k - \phi_k\|^2] < \infty \qquad \limsup_{n\to\infty} nE[|\widehat{\lambda}_k - \lambda_k|^2] < \infty$$

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i is large $\forall i$

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i is large $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i is large $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Common approach: first smooth each curve, then do fPCA

- For each i smooth data $\{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_i\} \rightarrow \widehat{X}_i(\cdot)$

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i is large $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Common approach: first smooth each curve, then do fPCA

- For each i smooth data $\{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_i\} \rightarrow \widehat{X}_i(\cdot)$
- $\widehat{\mu}$ sample mean of $\widehat{X}_i$'s

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i is large $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Common approach: first smooth each curve, then do fPCA

- For each i smooth data $\{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_i\} \rightarrow \widehat{X}_i(\cdot)$
- $\widehat{\mu}$ sample mean of $\widehat{X}_i$'s
- $\widehat{\Sigma}$ sample cov of $\widehat{X}_i$'s

fPCA for dense sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \dots, m_i\}_i$; m_i is large $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Common approach: first smooth each curve, then do fPCA

- For each i smooth data $\{(Y_{ij}, t_{ij}) : j = 1, \dots, m_i\} \rightarrow \widehat{X}_i(\cdot)$
- $\widehat{\mu}$ sample mean of $\widehat{X}_i$'s
- $\widehat{\Sigma}$ sample cov of $\widehat{X}_i$'s
- Eigenanalysis of $\widehat{\Sigma}$ to get $\{\widehat{\lambda}_k, \widehat{\phi}_k\}_k$
- Use PVE or other criteria to select truncation K

Trajectories reconstruction

- Estimate fPC scores

$$\widehat{\xi}_{ik} = \sum_{j=1}^{m_i} \{Y_{ij} - \widehat{\mu}(t_{ij})\}\widehat{\phi}_k(t_{ij})(t_{ij} - t_{ij-1})$$

(accounting for possibly unequal grids of points)

Trajectories reconstruction

- Estimate fPC scores

$$\widehat{\xi}_{ik} = \sum_{j=1}^{m_i} \{Y_{ij} - \widehat{\mu}(t_{ij})\}\widehat{\phi}_k(t_{ij})(t_{ij} - t_{ij-1})$$

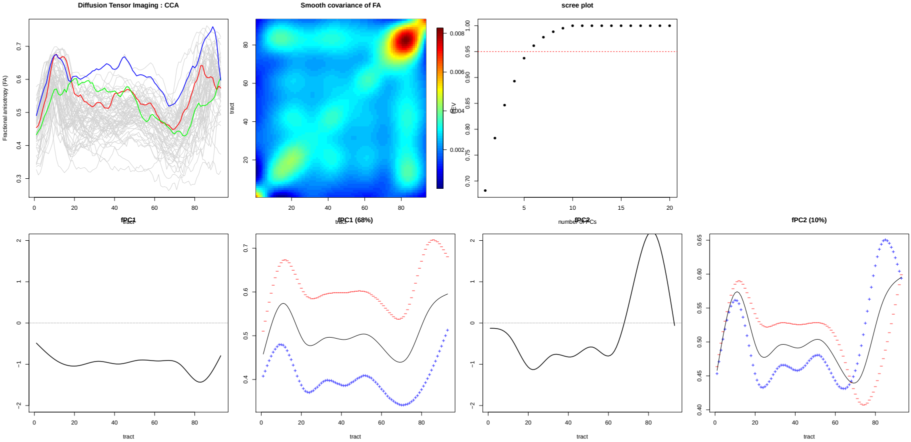
(accounting for possibly unequal grids of points)

- Finite dimension approx:

$$\widehat{X}_i^K(t) = \widehat{\mu}(t) + \sum_{k=1}^{K} \widehat{\phi}_k(t)\widehat{\xi}_{ik}$$

Illustration: DTI

Top: FA along corpus callosum for MS subjects (left); covariance estimate (middle); scree plot (right). Bottom: The two leading fPCs and their effect relative to the population mean



3.4. fPCA for sparse sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i small $\forall i$

3.4. fPCA for sparse sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i small $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

3.4. fPCA FOR SPARSE SAMPLING DESIGN PLUS NOISE

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i small $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Challenges:

- Number of measurements per curve is very small;
- Reconstruction of the curves is very important; however smoothing each curve individually is not realistic !

3.4. fPCA for sparse sampling design plus noise

Observed data: $\{Y_{ij}, t_{ij} : j = 1, \ldots, m_i\}_i$; m_i small $\forall i$

Model assn: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$; $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$.

Challenges:

- Number of measurements per curve is very small;
- Reconstruction of the curves is very important; however smoothing each curve individually is not realistic !

Common approach: pool subjects data to do fPCA

Reasoning behind estimation procedure

Recall model: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$ $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$

- Mean estimator $\widehat{\mu}(t)$: by smoothing the data $\{(t_{ij}, Y_{ij}) : i, j\}$

Reasoning behind estimation procedure

Recall model: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$ $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$

- Mean estimator $\widehat{\mu}(t)$: by smoothing the data $\{(t_{ij}, Y_{ij}) : i, j\}$
- To estimate the covariance, note the following:

$$\begin{aligned} Cov(Y_{ij}, Y_{ij'}) &= \Sigma(t_{ij}, t_{ij'}) \qquad \text{if } j \neq j' \\ Var(Y_{ij}) &= \Sigma(t_{ij}, t_{ij}) + \sigma^2 \end{aligned}$$

Reasoning behind estimation procedure

Recall model: $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$ $X_i \sim$ IID (μ, Σ), $\epsilon_{ij} \sim WN(0, \sigma^2)$

- Mean estimator $\widehat{\mu}(t)$: by smoothing the data $\{(t_{ij}, Y_{ij}) : i, j\}$
- To estimate the covariance, note the following:

$$\begin{aligned} Cov(Y_{ij}, Y_{ij'}) &= \Sigma(t_{ij}, t_{ij'}) \qquad \text{if } j \neq j' \\ Var(Y_{ij}) &= \Sigma(t_{ij}, t_{ij}) + \sigma^2 \end{aligned}$$

Treat differently diagonal (t_{ij}, t_{ij}) & off-diagonal $(t_{ij}, t_{ij'})$ terms

Sparse fPCA

- 'Off-diagonal' smoothing: Define "raw -covariances"

$$G_{ijj'} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}\{Y_{ij'} - \widehat{\mu}(t_{ij'})\}$$

Sparse fPCA

▷ 'Off-diagonal' smoothing: Define "raw -covariances"

$$G_{ijj'} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}\{Y_{ij'} - \widehat{\mu}(t_{ij'})\}$$

- two-dimensional smoothing of $\{G_{ijj'},\ t_{ij}, t_{ij'}) : j \neq j'\}_i$
  working model: $G_{ijj'} = \Sigma(t_{ij}, t_{ij'}) + e_{ijj'}$, $e_{ijj'} \sim$ IID,
  where $\Sigma(\cdot, \cdot)$ symmetric + positive-definite fn
  Model Σ using bivariate basis fns or tensor product of two univ bases fns. Different penalization (Wood, 2005)

Sparse fPCA

- ▷ 'Off-diagonal' smoothing: Define "raw -covariances"

$$G_{ijj'} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}\{Y_{ij'} - \widehat{\mu}(t_{ij'})\}$$

- two-dimensional smoothing of $\{G_{ijj'},\ t_{ij}, t_{ij'}) : j \neq j'\}_i$
  working model: $G_{ijj'} = \Sigma(t_{ij}, t_{ij'}) + e_{ijj'}$, $e_{ijj'} \sim$ IID,
  where $\Sigma(\cdot,\cdot)$ symmetric + positive-definite fn
  Model Σ using bivariate basis fns or tensor product of two univ bases fns. Different penalization (Wood, 2005)

- ▷ Adjust estimate to be symmetric; zero out negative eigenvals !

Sparse fPCA

▷ 'Diagonal' smoothing: Define raw -variances

$$G_{ij} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}^2$$

- one-dimensional smoothing of $\{(G_{ij},\ t_{ij}) : i, j\} \to \widehat{\sigma}^2_Y(t)$

Sparse fPCA

- ‘Diagonal’ smoothing: Define raw -variances

$$G_{ij} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}^2$$

- one-dimensional smoothing of $\{(G_{ij},\ t_{ij}) : i, j\} \to \widehat{\sigma}^2_Y(t)$
- Estimate σ^2: $\widehat{\sigma}^2 = \int_{\mathcal{T}} \{\widehat{\sigma}^2_Y(t) - \widehat{\Sigma}(t,t)\}\, dt$

Sparse fPCA (cont'd)

- Eigenanalysis of $\widehat{\Sigma}(t,s)$ gives eigenvals/fns, $\{\widehat{\lambda}_k, \widehat{\phi}_k(t)\}_k$,
- Choose K using PVE or other approaches (e.g. 95%)
- Estimate true signal $X_i(\cdot)$ by the truncated KL expansion:

$$\widehat{X}_i^K(t) = \widehat{\mu}(t) + \sum_{k=1}^{K} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$$

Sparse fPCA (cont'd)

- Eigenanalysis of $\widehat{\Sigma}(t,s)$ gives eigenvals/fns, $\{\widehat{\lambda}_k, \widehat{\phi}_k(t)\}_k$,
- Choose K using PVE or other approaches (e.g. 95%)
- Estimate true signal $X_i(\cdot)$ by the truncated KL expansion:

$$\widehat{X}_i^K(t) = \widehat{\mu}(t) + \sum_{k=1}^{K} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$$

- How to estimate the fPC $\widehat{\xi}_{ik}$?
  $\sum_{j=1}^{m_i} \{Y_{ij} - \widehat{\mu}(t_{ij})\} \widehat{\phi}_k(t_{ij})(t_{ij} - t_{ij-1})$ is no longer feasible !?

Sparse fPCA (cont'd)

- Eigenanalysis of $\widehat{\Sigma}(t,s)$ gives eigenvals/fns, $\{\widehat{\lambda}_k, \widehat{\phi}_k(t)\}_k$,
- Choose K using PVE or other approaches (e.g. 95%)
- Estimate true signal $X_i(\cdot)$ by the truncated KL expansion:

$$\widehat{X}_i^K(t) = \widehat{\mu}(t) + \sum_{k=1}^{K} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$$

- How to estimate the fPC $\widehat{\xi}_{ik}$?
  $\sum_{j=1}^{m_i} \{Y_{ij} - \widehat{\mu}(t_{ij})\} \widehat{\phi}_k(t_{ij})(t_{ij} - t_{ij-1})$ is no longer feasible !?
  Instead estimate fPC using *conditional expectation* !

Prediction of fPC

- Consider the reduced rank model (mixed effects model)

$$Y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} \phi_k(t_{ij}) \xi_{ik} + \epsilon_{ij} \qquad j = 1, 2, \ldots, m_i$$

Prediction of fPC

- Consider the reduced rank model (mixed effects model)

$$Y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} \phi_k(t_{ij})\xi_{ik} + \epsilon_{ij} \qquad j = 1, 2, \ldots, m_i$$

$\xi_{ik} \sim_i \text{IID}(0, \lambda_k)$, uncorrelated over k, and $\epsilon_{ij} \sim \text{IID}(0, \sigma^2)$
Assume for now that $\mu(\cdot), \phi_k(\cdot)$, λ_k and σ^2 are known

Prediction of fPC

- Consider the reduced rank model (mixed effects model)

$$Y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} \phi_k(t_{ij}) \xi_{ik} + \epsilon_{ij} \qquad j = 1, 2, \dots, m_i$$

$\xi_{ik} \sim_i \text{IID}(0, \lambda_k)$, uncorrelated over k, and $\epsilon_{ij} \sim \text{IID}(0, \sigma^2)$
Assume for now that $\mu(\cdot), \phi_k(\cdot)$, λ_k and σ^2 are known

- Prediction of ξ_{ik} by $\widetilde{\xi}_{ik} = E[\xi_{ik} | Y_{i1}, \dots, Y_{im_i}]$

Prediction of fPC

- Consider the reduced rank model (mixed effects model)

$$Y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} \phi_k(t_{ij})\xi_{ik} + \epsilon_{ij} \qquad j = 1, 2, \ldots, m_i$$

  $\xi_{ik} \sim_i \mathrm{IID}(0, \lambda_k)$, uncorrelated over k, and $\epsilon_{ij} \sim \mathrm{IID}(0, \sigma^2)$
  Assume for now that $\mu(\cdot), \phi_k(\cdot)$, λ_k and σ^2 are known
- Prediction of ξ_{ik} by $\widetilde{\xi}_{ik} = E[\xi_{ik}|Y_{i1}, \ldots, Y_{im_i}]$
- Assume ξ_{ik}'s and Y_{ij}'s jointly Gaussian.

  Then $\widetilde{\xi}_{ik}$ is the best linear unbiased predictor (BLUP) of ξ_{ik}.
  In matrix notation; subscript $i \rightarrow$ to allow for t_{ij}'s per subject:

$$\widetilde{\xi}_{ik} = \lambda_k \Phi_{ik}^T (\Sigma_i + \sigma^2 I_{m_i})^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i).$$

Prediction of fPC in practice

- Use the approximated reduced rank model as working model

$$Y_{ij} = \widehat{\mu}(t_{ij}) + \sum_{k=1}^{K} \widehat{\phi}_k(t_{ij})\xi_{ik} + \epsilon_{ij}$$

where $\xi_{ik} \sim_i \text{IID}(0, \widehat{\lambda}_k)$ and $\epsilon_{ij} \sim \text{IID}\ (0, \widehat{\sigma}^2)$

Prediction of fPC in practice

- Use the approximated reduced rank model as working model

$$Y_{ij} = \widehat{\mu}(t_{ij}) + \sum_{k=1}^{K} \widehat{\phi}_k(t_{ij})\xi_{ik} + \epsilon_{ij}$$

  where $\xi_{ik} \sim_i \text{IID}(0, \widehat{\lambda}_k)$ and $\epsilon_{ij} \sim \text{IID}\ (0, \widehat{\sigma}^2)$

- Predict ξ_{ik} by the empirical BLUP:

$$\widetilde{\xi}_{ik} = \widehat{\lambda}_k \widehat{\Phi}_{ik}^T (\widehat{\Sigma}_i + \widehat{\sigma}^2 I_{m_i})^{-1} (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i).$$

Theoretical properties of sparse fPCA

Under suitable regularity conditions

- Let $\widehat{\mu}(t)$ be the local linear mean estimator of $\mu(t)$. Then

$$\sup_t |\widehat{\mu}(t) - \mu(t)| = O_p(n^{-3/10})$$

THEORETICAL PROPERTIES OF SPARSE FPCA

Under suitable regularity conditions

- Let $\widehat{\mu}(t)$ be the local linear mean estimator of $\mu(t)$. Then

$$\sup_t |\widehat{\mu}(t) - \mu(t)| = O_p(n^{-3/10})$$

- Let $\widehat{\Sigma}(s,t)$ be the local linear estimator of $\Sigma(s,t)$. Then

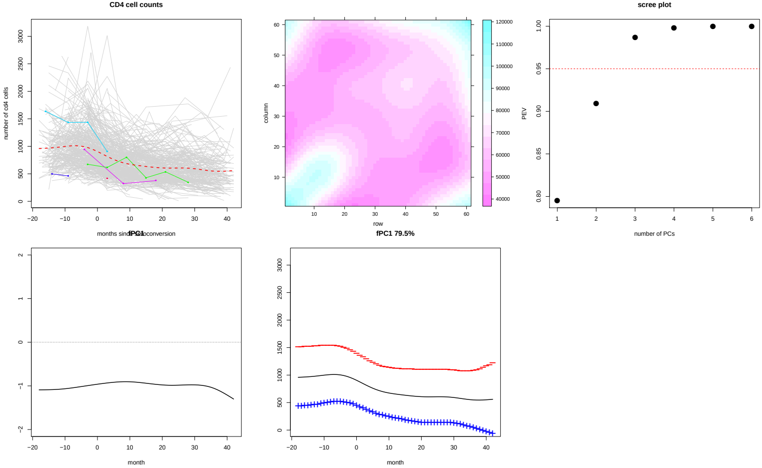
$$\sup_{t,s} |\widehat{\Sigma}(s,t) - \Sigma(s,t)| = O_p(n^{-1/10})$$

Assume $\lambda_1 > \lambda_2 > \ldots > \lambda_K > \lambda_{K+1} \geq 0$
Let $\{\widehat{\lambda}_k, \widehat{\phi}_k\}_k$ be the eigenvalues/eigenfunctions of $\widehat{\Sigma}(\cdot,\cdot)$ and $\widehat{c}_k = sign \int \widehat{\phi}_k(t)\phi_k(t)dt$. Then for all $k = 1, \ldots, K$

$$\|\widehat{c}_k\widehat{\phi}_k - \phi_k\|^2 = O_p(n^{-1/10}), \quad |\widehat{\lambda}_k - \lambda_k| = O_p(n^{-1/10})$$

Illustration: CD4

Top: estimated mean; estimated covariance; scree plot
Bottom: Top fPC and the effect of changes along this direction.



Software implementation

- R
  - for estimation
    - `face` for fast covariance estimation for sparse functional data
    - `fda` for functional data analysis in R
    - `fpca` for functional principal component analysis
    - `mgcv` for generalized additive (mixed) models; semi-parametric smoothing
    - `refund` for regression with functional data;
  - for visualization
    - `fields` for 2d image plots
    - `lattice` for various plots
    - `refund.shiny` for interacting plots for functional data analyses
    - `rgl` for 3d plots
- MATLAB
  - PACE

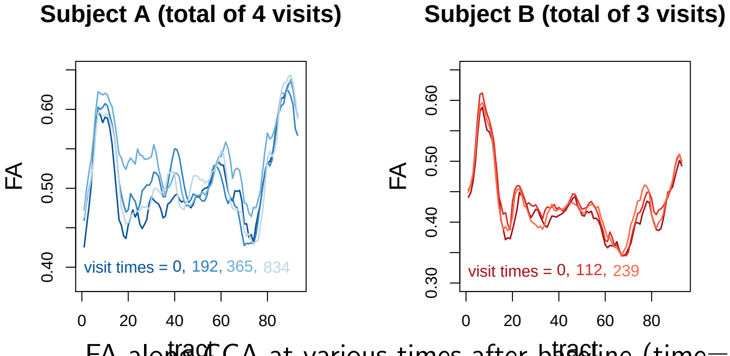
Beyond iid functional data

Although the iid case is quite common, here are other situations:

- Multilevel functional data:
  - $\{Y_{ij}(s),\, s \in \mathcal{S},\, i = 1, \ldots, n,\, j = 1, \ldots, m_i\}$
  - Eg: crypt-level biomarker data in colon-carcinogenesis studies
- Longitudinal functional data:
  - $[\{Y_{ij}(s), T_{ij}\},\, s \in \mathcal{S},\, i = 1, \ldots, n,\, j = 1, \ldots, m_i\}$
  - Eg: DTI data (multiple clinical visits)

DTI data (revisit)

Each subject is observed at several hospital visits and at each time a FA profile is measured. Below: FA for two subjects. Of interest: what is the dynamic behavior over time ?



FA along CCA at various times after baseline (time=0)

Longitudinal functional data analysis (LFDA)

Data Structure

$$[\{Y_{ij}(\cdot),\ T_{ij}\} : i = 1, \ldots, n,\ j = 1, \ldots, m_i]$$

- $Y_{ij}(\cdot)$ is jth response of the ith subject observed on *fine grid*
- T_{ij} is the time corresponding to the $Y_{ij}(\cdot)$

  T_{ij} is observed in a longitudinal design

Objective:

- Dynamic behavior of the process over time
- Predict full curve using subjects' past history

Related literature

- Morris & Carroll(JRSSB06); Baladandayuthapani et al(Bcs08); Di et al(AoAS09); Aston et al(JRSSC10); S. et al (Biost10); Scheipl et al(JCGS14); Morris (AnRev15)
- Greven et al(EJS10); Chen & Müller(JASA12);

  ...

Limitation: no methodology that captures the process dynamics over time and provides prediction of full trajectory at future time

Park & S. (Stat15)

Proposed Model

$$Y_{ij}(s) = \mu(s, T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}(T_{ij})\phi_k(s) + \epsilon_{ij}(s)$$

- $\mu(\cdot, T_{ij})$ mean response profile (FA along CCA) at time T_{ij}
- $\phi_k(\cdot)$'s are orthogonal functions in $L^2(\mathcal{S})$
- $\xi_{ik}(T)$ zero-mean time-varying random coefficients; they quantify the process dynamics
- $\epsilon_{ij}(\cdot)$ zero-mean IID residual process

Flexibility of the proposed framework

PAR : $\mathrm{cov}\{\xi_{ik}(T), \xi_{ik}(T')\} = \lambda_k \rho_k(|T - T'|)$, $\lambda_k > 0$;

relates to Gromenko et al(AnnAS12) and Gromenko and Kokoszka (CSDA13) for $n = 1$

REM : $\xi_{ik}(T) = \zeta_{0,ik} + \zeta_{1,ik} T$;

relates to Greven et al(EJS10) for $\phi_k^{X_1}(s) = \phi_k^{X_2}(s) := \phi_k(s)$

NP : $\xi_{ik}(T) = \sum_{\ell \geq 1} \zeta_{ik\ell} \psi_{k\ell}(T)$;

relates to Chen & Müller(JASA12) for $\phi_k(s|T_{ij}) := \phi_k(s)$

Selecting the time-invariant basis functions

Recall model : $Y_{ij}(s) = \mu(s, T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}(T_{ij})\, \phi_k(s) + \epsilon_{ij}(s)$

- Pre-specified basis functions (e.g. Fourier, B-splines, wavelets)
- Data-driven basis functions like the eigenfunctions of some covariance function. Which covariance function to use ?

Time-invariant basis functions, $\phi_k(s)$'s (cont'd)

Recall model : $Y_{ij}(s) = \mu(s, T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}(T_{ij})\, \phi_k(s) + \epsilon_{ij}(s)$

- Let $W_i(s, T) = \sum_{k=1}^{\infty} \xi_{ik}(T)\phi_k(s)$ and denote its covariance by
$$\mathrm{cov}\{W_i(s, T),\ W_i(s', T')\} = c\{(s, T), (s', T')\}$$
- Define $\Sigma(s, s') := \int c\{(s, T), (s', T)\}\, g(T)\, dT,$

  $g(T)$ the sampling density of T.
- $\Sigma(s, s') \rightarrow$ *proper* cov fn. Call Σ *marginal cov induced by W_i*.
  Similar concept: Chen et al.(JRSSB16+), Aston et al.(srXiv15)

TIME-INVARIANT BASIS FUNCTIONS, $\phi_k(s)$'S (CONT'D)

- Choose $\{\phi_k(\cdot)\}_k$'s as the eigenfunctions of $\Sigma(s, s')$
- Proxy time-varying basis coefficients

$$\xi_{W,ijk} = \int \{ Y_{ij}(s) - \mu(s, T_{ij}) \} \, \phi_k(s) ds$$

  - For each k, $\{(\xi_{W,ijk}, T_{ij})_{j=1}^{m_i}\}_i$ describe the process dynamics.

ESTIMATION ROADMAP

Step 1 : Mean function $\mu(s, T)$

Step 2 : Marginal covariance function $\Sigma(s, s')$ and the orthogonal basis $\phi_k(s)$'s

Step 3 : Time-varying coefficients $\xi_{ik}(T)$ for every T

Step 4 : Full trajectories $Y_i(\cdot, T)$ for every T

ESTIMATION

CASE: FULLY OBSERVED CURVES $L^2[\mathcal{S}]$ ERROR

Step 1 : Mean function

- Estimate $\mu(s, T)$ using bivariate (tensor product) splines smoothing Wood (Bcs2006) + working independence $\Rightarrow \widehat{\mu}(s, T)$

Estimation (cont'd)

Step 2 : Marginal Covariance Function

- Demean data: $\widetilde{Y}_{ij}(s) := Y_{ij}(s) - \widehat{\mu}(s, T_{ij})$
- Estimate $\Sigma(s, s')$ by sample covariance of the demeaned data

$$\widehat{\Sigma}(s, s') = \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \widetilde{Y}_{ij}(s) \widetilde{Y}_{ij}(s')$$

- Spectral decomposition of $\widehat{\Sigma}(s, s')$ gives $\{\widehat{\lambda}_k, \widehat{\phi}_k(s)\}_k$

Estimation (cont'd)

Step 3 : Time-varying coefficients $\xi_{ik}(T)$ for every T

For each k estimate proxy $\xi_{W,ijk} = \xi_{ik}(T_{ij}) + e_{ijk}$ by

$$\widetilde{\xi}_{W,ijk} = \int \widetilde{Y}_{ij}(s)\, \widehat{\phi}_k(s) ds \text{ (numerical integ.)}$$

- Use standard longitudinal/sparse functional data methods to analyse $\{(\widetilde{\xi}_{W,ijk}, T_{ij})_{j=1}^{m_i}\}_i$ and estimate temporal covaraince
- Predict $\xi_{ik}(T)$: $\widehat{\xi}_{ik}(T)$ for every T using this estimated covariance

Estimation (cont'd)

Step 4 : Predict the full trajectory $Y_i(\cdot, T)$ for every T

$$\widehat{Y}_i(\cdot, T) = \widehat{\mu}(\cdot, T) + \sum_{k=1}^{K} \widehat{\xi}_{ik}(T)\widehat{\phi}_k(\cdot)$$

Theoretical properties

Lemma 1: Marginal covariance

Under regularity assumptions, $\|\widehat{\Sigma}(\cdot,\cdot) - \Sigma(\cdot,\cdot)\| \rightarrow_p 0$

- Consistency results for $\widehat{\lambda}_k$ and $\widehat{\phi}_k(\cdot)$.

Lemma 2 : Proxy time-varying basis coefficients

Under regularity assumptions, $\sup_j |\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \rightarrow_p 0$

- Consistency of the predicted trajectories, $\widehat{Y}_i(s, T)$.

Implementation in R

Software implementation in R (Wrobel, Park, S., and Goldsmith, Stat 2016)

- `refund`: `fpca.lfda`
- `refund.shiny`: `plot_shiny` for visualization

Simulation experiment

Generating Model:

$$Y_{ij}(s) = \mu(s, T_{ij}) + \xi_{i1}(T_{ij})\phi_1(s) + \xi_{i2}(T_{ij})\phi_2(s) + \epsilon_{ij}(s)$$

- Covariance structures for the time-varying coef, $\xi_{ik}(T)$'s
  (i) NP ; (ii) REM ; (iii) Exp
- Error structure $\epsilon_{ij}(s)$: smooth + white noise (SNR$=1$)
- Grid points for s: 101 equally spaced points in $[0, 1]$
- For each i, $\{T_{ij} : j = 1, 2, \ldots, m_i\}$ are randomly sampled from 41 equally spaced points in $[0, 1]$

Prediction performance and computation efficiency comparison:

- Proposed method
- Naïve: take average the subject's previous curves to predict the current trajectory
- Chen&Müller (JASA12): using time-dependent orthogonal functions $\phi_k(s|T)$

Results

		$m_i \sim \{8, \ldots, 12\}$			
		IN-IPE	IN-IPE$_{\text{naive}}$	OUT-IPE	OUT-IPE$_{\text{naive}}$
NP	$n = 100$	0.406	7.790	0.988	11.478
	$n = 300$	0.313	7.773	0.559	11.349
	$n = 500$	0.288	7.779	0.455	11.262
REM	$n = 100$	0.328	1.199	1.011	2.160
	$n = 300$	0.265	1.197	0.675	2.160
	$n = 500$	0.247	1.197	0.571	2.150
Exp	$n = 100$	0.554	1.528	1.426	2.520
	$n = 300$	0.508	1.531	1.143	2.498
	$n = 500$	0.494	1.530	1.074	2.492

Results (cont'd)

- Computationally faster compared to available approaches

		Chen&Müller (JASA12)			Proposed method		
		$m_i \sim \{8, \ldots, 12\}$					
		IN-IPE	OUT-IPE	time (sec)	IN-IPE	OUT-IPE	time (sec)
NP	$n = 100$	0.880	2.221	983.872	0.406	0.988	7.369
	$n = 300$	0.622	1.468	1659.611	0.313	0.559	15.892
	$n = 500$	0.556	1.298	2502.462	0.288	0.455	21.418
REM	$n = 100$	0.424	1.359	1084.753	0.328	1.011	9.282
	$n = 300$	0.289	0.729	1955.193	0.265	0.675	11.347
	$n = 500$	0.257	0.614	2947.126	0.247	0.571	22.559
Exp	$n = 100$	0.634	1.642	1556.182	0.554	1.426	7.514
	$n = 300$	0.549	1.251	1959.219	0.508	1.143	16.229
	$n = 500$	0.531	1.155	2865.041	0.494	1.074	17.109

DTI data analysis

- FA along CCA for MS patients
- 162 MS patients observed at between 1 to 8 hospital visits
- T_{ij} - hospital visit time (mean = 2.6 visits/subj)
- $Y_{ij}(\cdot)$ - FA profile (93 locn along CCA) for i subj at T_{ij}
- 421 total curves

Objective:

- Dynamic behavior of FA over time
- Predict FA profile at a subject's future visit

DTI STUDY (CONT'D)

DTI data exploratory analysis

Model assumption

- $Y_{ij}(s) = \mu(s) + \sum_{k=1}^{K} \xi_{ik}(T_{ij})\, \phi_k(s) + \epsilon_{ij}(s)$
- Time-varying coef: $\xi_{ik}(T_{ij}) = b_{0ik} + b_{1ik} T_{ij}$ (REM)

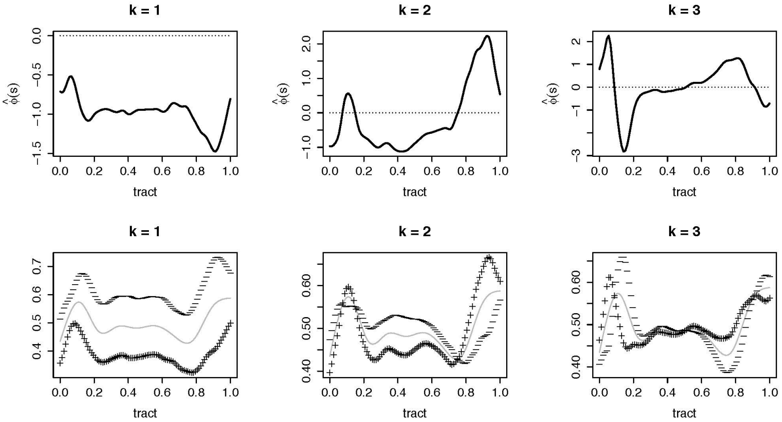
Estimation

Longitudinal functional data analysis for DTI data

- $\widehat{\phi}_k(\cdot)$'s - eigenfns of the estimated marginal covariance $\widehat{\Xi}(\cdot, \cdot)$
- Fix percentage of explained variance to 95% $\rightarrow K = 10$

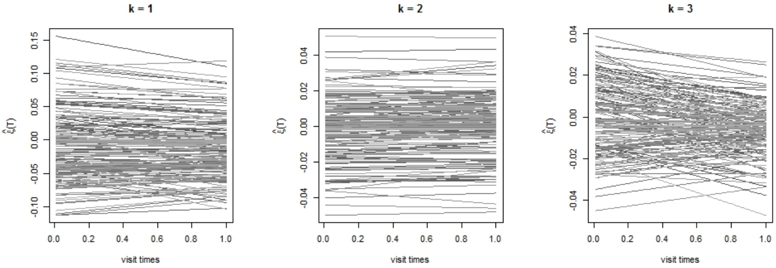
DTI study results (time dynamics)

Figure: Estimated basis functions $\widehat{\phi}_k(s)$ for $k = 1, 2$ and 3



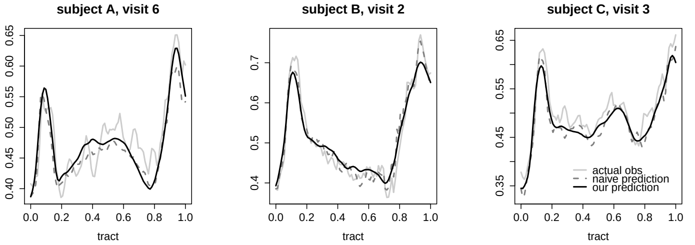
DTI STUDY RESULTS (TIME DYNAMICS)

FIGURE: Time-varying coefficients $\widehat{\xi}_{ik}(T)$ for $k = 1, 2$ and 3 (REM)



DTI study (cont'd)

Predicted values of FA for the last visits of three randomly selected subjects; actual(gray) / proposed (black) / naïve (dashed)



	Naïve	Greven et al. (EJS10)	Chen & Müller (JASA12)	Proposed
IN-IPE$\times 10^2$	-	2.66	3.76	2.31
OUT-IPE$\times 10^2$	3.52	-	8.71	3.48

Final remarks

- Sample of independent curves
  - Smoothing using pre-specified basis
  - FPCA
- Longitudinally observed curves
  - Process dynamics + future curve prediction
- Software implementation/visualization in R

Thank you!

Comments? Questions ? astaicu@ncsu.edu