Modeling Multi-Way Functional Data With Weak Separability

Kehui Chen

Department of Statistics
University of Pittsburgh, USA

@CMStatistics, Seville, Spain

December 09, 2016

Outline

- Introduction.
- Multi-way FPCA, marginal FPCA and product FPCA.
- Weak separability.
- Data examples.

- Based on joint work with Brian Lynch (Pittsburgh), Pedro Delicado (Barcelona) and Hans-Georg Müller (Davis).
- K. Chen and H.G. Müller (2012), *Modeling repeated functional observations, JASA*.
- K. Chen , P. Delicado and H.G. Müller (2016), *Modeling functional-valued stochastic processes, with applications to fertility dynamics, JRSSB*.
- B. Lynch and K. Chen (2016+), *Weak Separability: Concepts and Inference. In Preparation*.
- Partially supported by NSF1612458.

Functional Data

- Functional data consist of fully or partially observed samples of random functions $[0, T] \to \mathbb{R}^q$, usually $q = 1$.
- As infinite-dimensional objects, functional data require dimension reduction.
- Such data are ubiquitous – longitudinal studies; tracking and monitoring.
- Any data observed over a continuum for many individuals.

Function-Valued Stochastic Processes

- Traditional:

$$\mathcal{T} \to X(t) \in \mathcal{R},\ X \in L^2(\mathcal{T}),\ \ X \text{ is smooth}$$

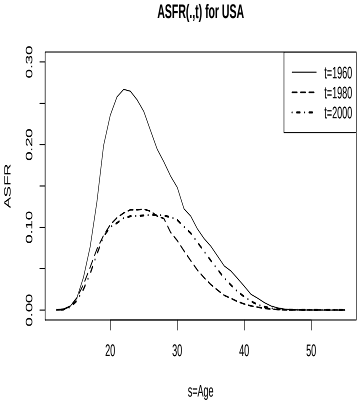
- Now the value of the process at each $t \in \mathcal{T}$ is a random function $X(\cdot, t)$ on a domain $\mathcal{S}$:

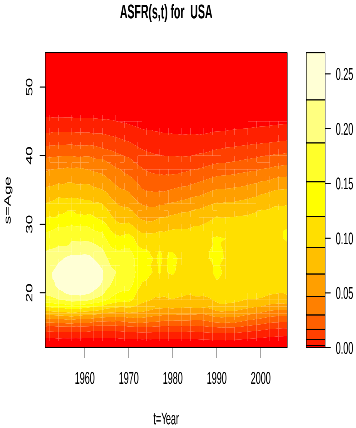
$$\mathcal{T} \to X(\cdot, t) \in L^2(\mathcal{S}),\ X \in L^2(\mathcal{S} \times \mathcal{T}),$$

with

$$E(X(s,t)) = \mu(s,t),\ C(s,t;u,v) = \text{cov}(X(s,t), X(u,v)).$$

Fertility Data





Other Examples of Multi-Way Functional Data

- Multi-way functional data: the time index t of the stochastic process and the argument s of the observed functions; or the other way around. $X(s,t) \in L^2(\mathcal{S} \times \mathcal{T})$.
- Data obtained from tracking apps where animal's/people's 24-hour profiles of activities are recorded every day.
- EEG/fMRI data with spatial and/or temporal index, or even with longitudinal follow-ups over years.
- Different from traditional spatial-temporal data, as we consider a sample of n realizations.

Karhunen-Loève (KL) Representation

$$X(s,t) = \mu(s,t) + \sum_{r=1}^{\infty} Z_r \gamma_r(s,t), \quad s \in \mathcal{S},\, t \in \mathcal{T}. \tag{1}$$

- Here $\{\gamma_r : r \geq 1\}$ are the eigenfunctions of the linear operator with kernel $C(s,t;u,v) = \mathrm{cov}(X(s,t), X(u,v))$.
- $\{Z_r = \int \gamma_r(s,t) X^c(s,t) ds dt : r \geq 1\}$ are the (uncorrelated) functional principal components (FPCs) with $E(Z_r) = 0$.
- Multi-Way FPCA: Modeling the variability in the random process and dimension reduction.
- Serve as building blocks for further modeling and regularization.

Challenges with Multi-Way FPCA

- A non-parametric modeling of $C(s,t;u,v) = \text{cov}(X(s,t), X(u,v))$ is very difficult.
- For dense regular design with p_1 grids in $\mathcal{T}$ and p_2 grids in $\mathcal{S}$, a nonparametric estimation of $C(s,t;u,v)$ based on sample covariance has a huge dimension $p_1^2 p_2^2$.
- Covariance estimation for sparse designs requires 4-dimensional smoothing.
- The effects of s and t are hard to separate and to visualize.

Marginal Kernel

- Define the the marginal covariance functions

$$C_{\mathcal{S}}(s,u) = \frac{\int_t C(s,t;u,t)dt}{\int_{s,t} C(s,t;s,t)dsdt}, \quad C_{\mathcal{T}}(t,v) = \frac{\int_s C(s,t;s,v)ds}{\int_{s,t} C(s,t;s,t)dsdt}.$$

- The marginal kernels are normalized to have $\int_s C_{\mathcal{S}}(s,s)ds = 1$ and $\int_t C_{\mathcal{T}}(t,t)dt = 1$.
- The eigen decompositions are $C_{\mathcal{S}}(s,u) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s)\psi_j(u)$ and $C_{\mathcal{T}}(t,v) = \sum_{k=1}^{\infty} \gamma_k \phi_k(t)\phi_k(v)$,
  where $\lambda_1 \geq \lambda_2 \geq \ldots$, and $\gamma_1 \geq \gamma_2 \geq \ldots$, are the marginal eigenvalues.

A Tensor Product Representation

- The product of marginal eigen functions $\phi_k \psi_j$ forms an orthogonal basis on $L^2(\mathcal{S} \times \mathcal{T})$.
- We have the representation of $X(s,t)$,

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \chi_{jk} \phi_k(t) \psi_j(s).$$

- Product FPCA

$$X(s,t) \approx \mu(s,t) + \sum_{j=1}^{P} \sum_{k=1}^{K} \chi_{jk} \phi_k(t) \psi_j(s).$$

Loss bound on product FPCA

For $P \geq 1$ and $K \geq 1$, consider the following loss minimization

$$E\left(\int_{\mathcal{S},\mathcal{T}}\{X(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K}\langle X, \phi_k\psi_j\rangle\phi_k(t)\psi_j(s)\}^2 dsdt\right).$$

Let Q^* be the minimum unexplained variance that can be achieved using any basis $f_k(t)g_j(s)$, we have $Q < Q * + aE\|X\|^2$ where $a = \min(a_{\mathcal{T}}, a_{\mathcal{S}})$, with $(1 - a_{\mathcal{T}})$ denoting the fraction of variance explained by K terms for $G_{\mathcal{T}}(t,v)$ and analogously for $a_{\mathcal{S}}$.

The product FPCA under strong separability

- Strong separability:

$$C(s,t;u,v) = aC_{\mathcal{S}}(s,u)C_{\mathcal{T}}(t,v),$$

  with $\int_s C_{\mathcal{S}}(s,s)ds = 1$ and $\int_t C_{\mathcal{T}}(t,t)dt = 1$.
- We can show that $C_{\mathcal{S}}$ and $C_{\mathcal{T}}$ are the same as the marginal kernels.
- Under strong separability, the eigenfunctions of $C(s,t;u,v)$ is in the form of the product of marginal eigenfunctions.

Strong separability

- Under strong separability:

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \chi_{jk} \phi_k(t) \psi_j(s),$$

  is the same as the Multi-Way KL expansion.
- The scores χ_{jk} are mutually uncorrelated.
- However, the strong separability assumption of a covariance is often too restrictive.

Concept of weak separability

Brian & Chen (2016+)

- For any orthogonal bases $\{f_j, j \geq 1\}$ in $L^2(\mathcal{S})$ and $\{g_k, k \geq 1\}$ in $L^2(\mathcal{T})$, we can have

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \tilde{\chi}_{jk} f_j(s) g_k(t).$$

- $X(s,t)$ is weakly separable if there exist orthogonal bases $\{f_j, j \geq 1\}$ and $\{g_k, k \geq 1\}$ such that $\mathrm{cov}(\tilde{\chi}_{jk}, \tilde{\chi}_{j'k'}) = 0$ for $j \neq j'$ or $k \neq k'$., i.e., the scores $\{\tilde{\chi}_{jk}, j \geq 1, k \geq 1\}$ are uncorrelated to each other.

Unique bases for weak separability

Brian & Chen (2016+)

- If X is weakly separable, the pair of bases $\{f_j, j \geq 1\}$ and $\{g_k, k \geq 1\}$ that satisfies weak separability is unique:

$$f_j(s) \equiv \psi_j(s), \;\; g_k(t) \equiv \phi_k(t),$$

  where $\psi_j(s)$ and $\phi_k(t)$ are the eigenfunctions of the marginal kernels.
- Weak separability is testable.

Weak separabile covariance structure

- Under weak separability, the covariance becomes

$$C(s,t;u,v) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \eta_{jk}\psi_j(s)\psi_j(u)\phi_k(t)\phi_k(v), \tag{2}$$

  where $\eta_{jk} = \mathrm{var}(\chi_{jk})$.
- Strong separability is a special case of weak separability:
  Strong separability is equivalent to the condition that
  $\eta_{jk} = a\lambda_j\gamma_k$.

Weak separable covariance structure cont.

- Define the two-way array $V = (\eta_{jk}, j \geq 1, k \geq 1)$, where $\eta_{jk} = \mathrm{var}(\chi_{jk})$.

  $$\mathrm{rank}_+(V) = \min\{\ell : V = V_1 + ... + V_\ell,\ V_i \geq 0,\ \mathrm{rank}(V_i) = 1,\ \forall\, i\},$$

  where $V_i \geq 0$ means that V_i is entrywise nonnegative.

- If V has $\mathrm{rank}_+(V) = L$, then V has nonnegative decomposition $V = \sum_{l=1}^{L} a^l \Lambda^l (\Gamma^l)^T$, where $\Lambda^l = (\lambda_j^l, j \geq 1)$ and $\Gamma^l = (\gamma_k^l, k \geq 1)$ are all nonnegative for $l = 1, \ldots, L$.

Weak separable covariance structure cont.

- **Brian & Chen (2016+)**: a notion of L-separability, and the strong separability corresponds to 1-separability.
- We have

$$C(s,t;u,v) = \sum_{l=1}^{L} a^l C_{\mathcal{S}}^l(s,u) C_{\mathcal{T}}^l(t,v),$$

where $C_{\mathcal{S}}^l(s,u) = \sum_j \lambda_j^l \psi_j(s)\psi_j(u)$ and $C_{\mathcal{T}}^l(t,v) = \sum_k \gamma_k^l \phi_k(t)\phi_k(v)$.

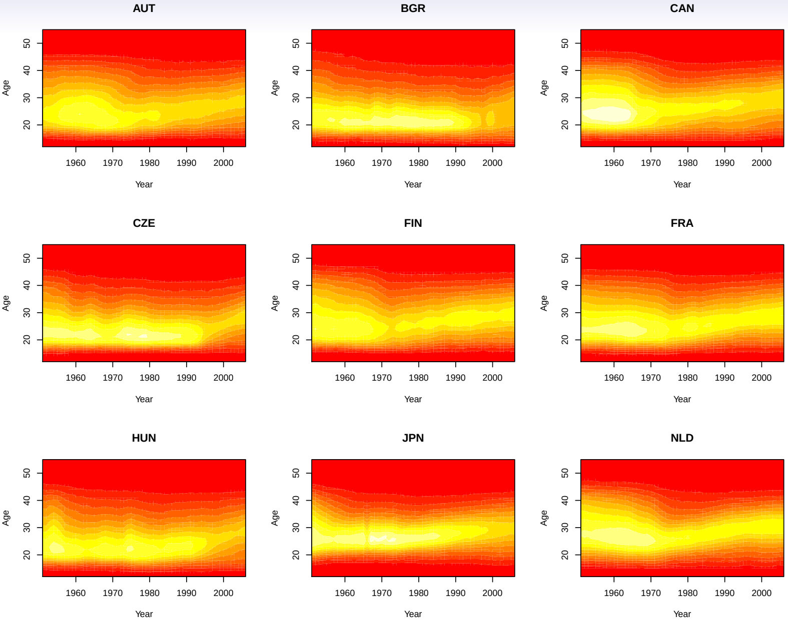
A variation: Marginal FPCA

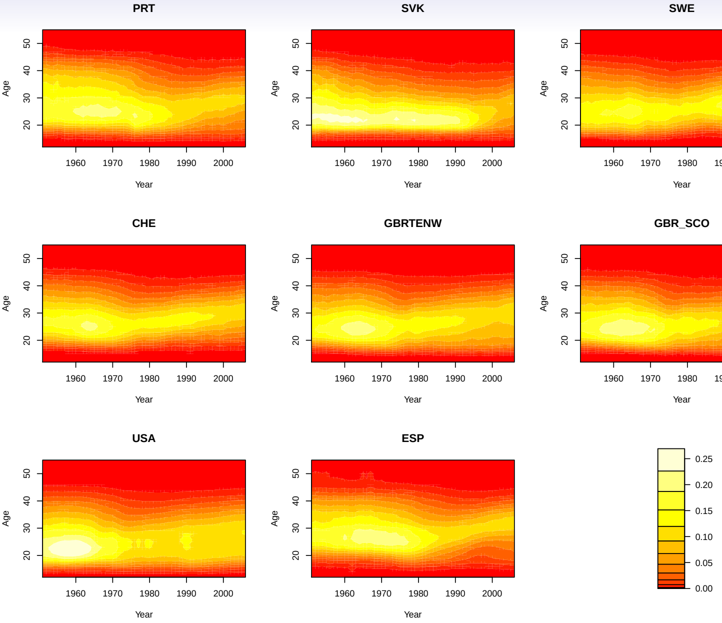
- Using the marginal eigenfunctions $\{\psi_j : j \geq 1\}$ of $L^2(\mathcal{S})$,
$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(t)\psi_j(s)$$
with random coefficient functions $\{\xi_j : j \geq 1\}$.
- Applying the KL representations of $\xi_j(t)$, $\xi_j(t) = \sum_{k=1}^{\infty} \chi_{jk}\phi_{jk}(t)$, with eigenfunctions ϕ_{jk} and FPCs χ_{jk}, leads to
$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \chi_{jk}\phi_{jk}(t)\psi_j(s).$$

Fertility Data

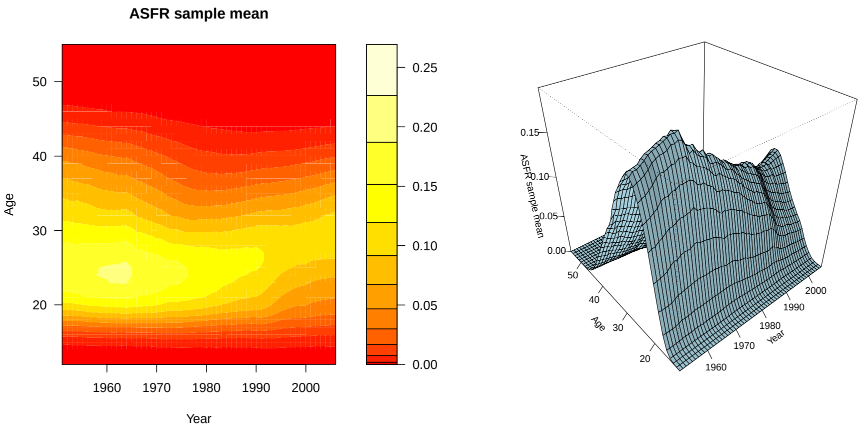
- Age-Specific Fertility Rate (AFSR) for 17 countries, 1951 to 2006. Ages of mothers s range from 12 to 55 years old. (Human Fertility Database 2013 (HFD-2013))
- Data: 17 independent units (countries), corresponding to a realization of the function valued stochastic process $\text{ASFR}(\cdot, t)$ at each year t. Observation grid $(\text{age}, \text{calendar-year})$ has 44×56 equidistant points.
- AFSR for age s (expressed in years) and calendar year t:

$$\frac{\text{Births during the year } t \text{ to women of age } s}{\text{Person-years lived for the year } t \text{ by women of age } s}.$$


AUT
BGR
CAN
CZE
FIN
FRA
HUN
JPN
NLD
Age
Year
20
30
40
50
1960
1970
1980
1990
2000


PRT
SVK
SWE
CHE
GBRTENW
GBR_SCO
USA
ESP
Age
Year
1960
1970
1980
1990
2000
20
30
40
50
0.00
0.05
0.10
0.15
0.20
0.25

Mean function $\mu(s, t)$



Comparing 2-d FPCA with Marginal FPCA and Product FPCA

(1) As expected, standard FPCA based on the two-dimensional Karhunen-Loève expansion needs less components to explain a certain amount of variance, as 4 eigenfunctions lead to a FVE of 89.74%, while marginal FPCA representation achieves a FVE of 87.51% with 6 terms, and product FPCA needs 7 terms to explain 87.42%.

Table: Fraction of Variance Explained (FVE) of ASFR(s,t) for the leading terms in the proposed marginal FPCA, product FPCA and 2d FPCA. Number of terms are selected to have accumulative FVE more than 85%.

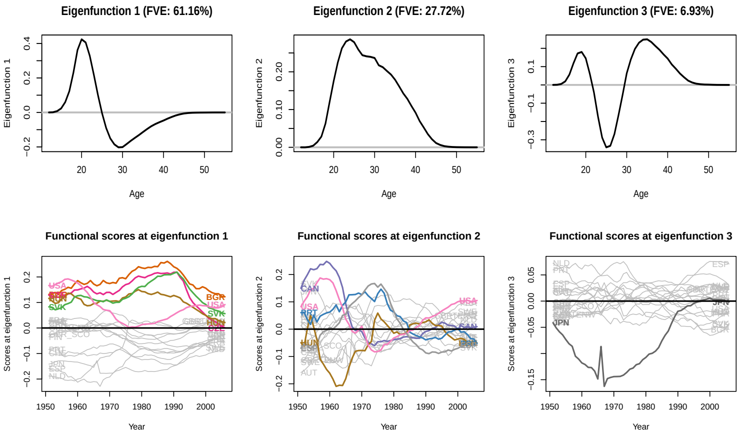
marginalFPCA	FVE	productFPCA	FVE	2d FPCA	FVE
Six	87.51	Seven	87.42	Four	89.74
$\hat{\phi}_{11}(t)\hat{\psi}_1(s)$	54.33	$\hat{\phi}_1(t)\hat{\psi}_1(s)$	53.7	$\hat{\gamma}_1(s,t)$	53.94
$\hat{\phi}_{21}(t)\hat{\psi}_2(s)$	13.04	$\hat{\phi}_2(t)\hat{\psi}_2(s)$	8.18	$\hat{\gamma}_2(s,t)$	13.71
$\hat{\phi}_{22}(t)\hat{\psi}_2(s)$	6.88	$\hat{\phi}_1(t)\hat{\psi}_2(s)$	8.06	$\hat{\gamma}_3(s,t)$	11.04
$\hat{\phi}_{12}(t)\hat{\psi}_1(s)$	4.63	$\hat{\phi}_3(t)\hat{\psi}_2(s)$	5.54	$\hat{\gamma}_4(s,t)$	6.05
$\hat{\phi}_{23}(t)\hat{\psi}_2(s)$	4.41	$\hat{\phi}_2(t)\hat{\psi}_1(s)$	4.4		
$\hat{\phi}_{31}(t)\hat{\psi}_3(s)$	4.22	$\hat{\phi}_4(t)\hat{\psi}_2(s)$	3.86		
		$\hat{\phi}_1(t)\hat{\psi}_3(s)$	3.68		

Comparison Cont.

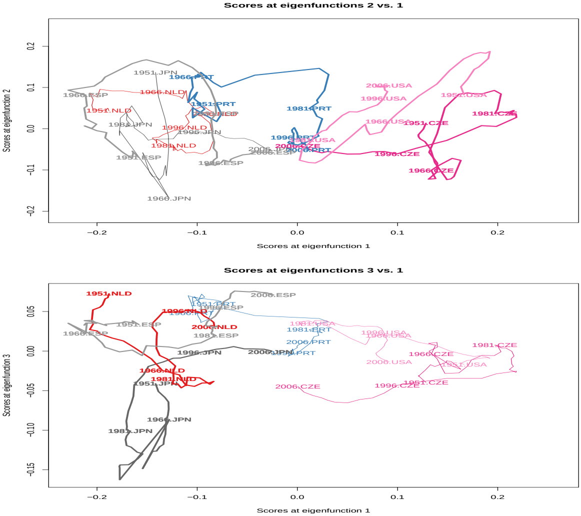
(2) Product FPCA and Marginal FPCA represent the functional data as a sum of terms that are products of two functions, each depending on only one argument. This provides for much better interpretability and feature discovery.

(3) Marginal FPCA makes it much easier than standard FPCA to analyze the time dynamics of the fertility process.

Eigen Functions and Score Functions



Track Plots For Selected Countries



Summary

- Observe a sample of $X_i(s,t)$, we want to model the variability in the random process and perform dimension reduction.
- Tensor product representation: allows for analyzing the separate (possibly asymmetric) effects of s and t.
- Optimal representation based on marginal kernels: Marginal FPCA and product FPCA.
- The new concept of weak separability: much more flexible than strong separable covariance assumption.

Thank You!

Questions?