A Bayesian Model for Sparse Functional Data

Wes Thompson, PhD
Assistant Professor
Department of Statistics and Psychiatry
University of Pittsburgh

Ori Rosen, PhD
Associate Professor
Department of Mathematics
University of Texas, El Paso
Outline of Talk

- Introduction
- Brief Review of Models for Functional Data
- Our Bayesian Model for Sparse Functional Data
  - Specification of model (with priors)
  - MCMC Sampling Scheme
  - Bayesian Inference
- Applications to Some Real and Simulated Data
- Conclusions
FDA is concerned with estimation and inference involving data which consist of curves across individuals.

Often curves are sampled longitudinally (over time).

A common application of FDA is to the nonparametric analysis of longitudinal data.

Research in the area has been quite active in the last 10 years or so (Ramsay and Silverman, 1997).
In theory, functional data could be sampled infinitely often.

In practice, of course, functions are sampled discretely at a finite number of time points, and with noise.

The usual first step in an analysis of functional data is to obtain smoothed estimates of the individual functions.
Introduction: Functional Data Analysis

- Potential issues:
  - Functional form may be complex.
  - Number and timing of measurements may differ between individuals.
  - Measurements on any one curve may be “sparse”.

- Desirable model properties:
  - Make few unwarranted assumptions about functional form.
  - Fitting procedure locally adaptive.
  - Borrow strength across individuals.
Examples:

- Growth of children born to HIV infected mothers (Hoover, et al., 1998).
- Progesterone curves during menstrual cycles (Brumback and Rice, 1998).
- Angles of the hip over walking cycles (Rice and Wu, 2001).
- Follicle-stimulating hormone levels over time (Staniswalis and Lee, 1998).
Example: Body Fat Data

- Subset of MIT Growth and Development study.
  

- Longitudinal data from a prospective study on body fat accretion.

- 162 Girls measured annually over 6 years pre- to 4 years post-menarche.

- There are 1049 total body fat percentage measurements.
Body Fat Data

Plot of Percent Body Fat over Time for 162 Girls
Body Fat Data

- Unequal number of measurements per subject. (range: 3-10; 6.4 per subject).

- Change point at or around menarche.

- Data contain a lot of variation both between and within subjects over time.
Potential Goals:

- Estimate a smooth functional form for overall mean and individual curves which makes few a priori assumptions about shape.
- Obtain some idea of the precision for overall mean and individual curve estimates.
- Use these smoothed functional data in further analyses. E.g., FPCA or functional linear models. (Ramsay and Silverman, 2005)
Potential Goals:

- Estimate a smooth functional form for overall mean and individual curves which makes few \textit{a priori} assumptions about shape.

(Ramsay and Silverman, 2005)
Potential Goals:

- Estimate a smooth functional form for overall mean and individual curves which makes few *a priori* assumptions about shape.

- Obtain some idea of the precision for overall mean and individual curve estimates.
Functional Data Analyses

Potential Goals:

- Estimate a smooth functional form for overall mean and individual curves which makes few *a priori* assumptions about shape.
- Obtain some idea of the precision for overall mean and individual curve estimates.
- Use these smoothed functional data in further analyses.
Potential Goals:

- Estimate a smooth functional form for overall mean and individual curves which makes few *a priori* assumptions about shape.
- Obtain some idea of the precision for overall mean and individual curve estimates.
- Use these smoothed functional data in further analyses. E.g., FPCA or functional linear models.

*(Ramsay and Silverman, 2005)*
The response for individual $i$ as a function of time $t$ is assumed to arise from the model:

$$Y_i(t) = \mu(t) + g_i(t) + \epsilon_i(t), \quad 0 \leq t \leq T$$  \hspace{1cm} (1)

- $\mu(t)$ is the overall mean curve over time.
- $g_i(t)$ is the $i$th subject’s smooth variation from the mean.
- $\epsilon_i(t)$ is the unexplained short-term variation.

$f_i(t) = \mu(t) + g_i(t)$ is the $i$ subject’s (latent) curve.

$\epsilon_i(t)$ are independent white-noise processes.
Curves are sampled at a finite number of discrete time points.

The functional data model (1) then becomes:

$$Y_i = \mu_i + g_i + \epsilon_i$$  \hspace{1cm} (2)

- $\mu_i = \mu(t_i)$
- $g_i = g(t_i)$
- $\epsilon_i \sim \text{iid } N(0, \sigma^2_\epsilon)$
- $t_i = (t_{i1}, \ldots, t_{in_i})$
The “direct method” smooths each curve individually.

(e.g. Rice and Silverman, 1991)

E.g., each curve can be individually fit with a truncated basis expansion or with a penalized smoothing spline.

- Direct method ignores information from other individuals when estimating each curve.
- Individuals with few measurements may be estimated poorly.
- Subsequent analyses using these estimated curves, treat each curve equally.
- Inferences conditional on selected model.
Review of FDA Smoothing Models

- “Penalized smoothing spline mixed-model” approach.  
  (Brumback and Rice, 1998)

- Generalization of Kimmeldorf and Wahba (1970) to multiple curves.

- This approach handles missing or “sparse” functional data.
  - Estimation procedure (REML with EM) very computationally intensive.
  - Inference following estimation uses partially parametric bootstrap
  - Not locally adaptive.
  - Assumes a particular form for covariance kernel.
“Mixed-effects regression spline” approach.

(Rice and Wu, 2000)

Uses B-spline truncated basis expansion.

This approach also handles missing or “sparse” functional data.

- Need a separate procedure (GCV, AIC) to choose number and placement of knots.
- This may involve a huge number of models to select from.
- Inference following estimation conditional on “best” model selected.
Our Approach

- We use a Bayesian mixed-model regression spline.
  - This extends the approach of Rice and Wu (2000).

- This model also incorporates automatic selection of number and placement of knots for B-splines.
  - Handles missing or “sparse” functional data.
  - Computationally feasible using MCMC methods.
  - Posterior inferences simple.
  - Posterior inferences also include uncertainty about model (knot) selection.
  - Model selection/fitting/inference are unified.
Mixed-Effects Model

We assume we can well-approximate the model (2) by:

\[ y_{ij} = \phi_{ij}^T \beta + \phi_{ij}^T b_i + \epsilon_{ij} \]  \hspace{1cm} (3)

- \( \phi_{ij} = (\phi_1(t_{ij}), \ldots, \phi_K(t_{ij}))^T \) are B-spline basis functions evaluated at time \( t_{ij} \).
- \( \mu_{ij} = \phi_{ij}^T \beta \) is the overall mean at time \( t_{ij} \).
- \( g_{ij} = \phi_{ij}^T b_i \) is the smooth deviation of the \( i \)th curve.
- \( b_i \sim iid \ N(0, \Sigma_b) \)
- \( \epsilon_{ij} \sim iid \ N(0, \sigma^2_\epsilon) \)
Conditional on $K$ and the placement of the breakpoints, this is a standard linear mixed-effects model.

In practice, this needs to be determined somehow.

We take a Bayesian approach:

- Start with a large number of basis functions (breakpoints).
- Include a latent indicator variable for each basis function.
- Place a prior on each with nonzero probability that it equals zero.
Model (3) becomes:

\[ y_{ij} = \phi_{\gamma ij}^T \beta_\gamma + \phi_{\gamma ij}^T b_{\gamma i} + \epsilon_{ij} \]  \hspace{1cm} (4)

\[ \gamma = (\gamma_1, \ldots, \gamma_{K-p}) \] are indicators for the interior breakpoints.

\[ \phi_\gamma = (\phi_{\gamma 1}, \ldots, \phi_{\gamma q_\gamma}) \] are the $q_\gamma$ basis functions selected by $\gamma$.

\[ \beta_\gamma \] is the $q_\gamma$ vector of fixed coefficients.

\[ b_{\gamma i} \] is the $q_\gamma$ vector of random coefficients.
Prior specification for (4) is hierarchical:

- $b_i$,
- $b_j$, $N(0, c_1)$,
- $b_j$, $IW(S, 1)$,
- $Q_{Kk} = 1$, $K_{1-k}(1-k)$,
- Beta($c, d$).
Prior specification for (4) is hierarchical:

\[
\cdot \ b_i \mid \beta, \Sigma_b, \gamma \sim N(\beta, \Sigma_b, \gamma)
\]
Prior specification for (4) is hierarchical:

\[
\begin{align*}
\cdot b_i & \mid \beta, \Sigma_b, \gamma \sim N( \beta_\gamma, \Sigma_{b,\gamma} ) \\
\cdot \beta & \mid \Sigma_b, \gamma \sim N( 0, cI )
\end{align*}
\]
Prior specification for (4) is hierarchical:

- $b_i \mid \beta, \Sigma_b, \gamma \sim N(\beta_\gamma, \Sigma_{b,\gamma})$
- $\beta \mid \Sigma_b, \gamma \sim N(0, cI)$
- $\Sigma_b \mid \gamma \sim IW(S_\gamma; \eta)$

$c$ and $d$ can be adjusted to give a priori probability of including a pre-specified number of breakpoints.
Prior specification for (4) is hierarchical:

\[ b_i \mid \beta, \Sigma_b, \gamma \sim N( \beta_\gamma, \Sigma_{b,\gamma} ) \]
\[ \beta \mid \Sigma_b, \gamma \sim N(0, cI) \]
\[ \Sigma_b \mid \gamma \sim IW( S_\gamma; \eta ) \]
\[ \sigma^2_\epsilon \sim IG( c_\epsilon; , d_\epsilon ) \]
Prior specification for (4) is hierarchical:

\[ b_i \mid \beta, \Sigma_b, \gamma \sim N(\beta, \Sigma_b, \gamma) \]
\[ \beta \mid \Sigma_b, \gamma \sim N(0, cI) \]
\[ \Sigma_b \mid \gamma \sim IW(S_{\gamma}, \eta) \]
\[ \sigma^2_\epsilon \sim IG(c_\epsilon, d_\epsilon) \]
\[ \gamma \mid \pi \sim \prod_{k=1}^{K} \pi^{\gamma_k} (1 - \pi)^{1-\gamma_k} \]
Prior specification for (4) is hierarchical:

\[
\begin{align*}
&\cdot b_i \mid \beta, \Sigma_b, \gamma \sim N(\beta, \Sigma_{b,\gamma}) \\
&\cdot \beta \mid \Sigma_b, \gamma \sim N(0, cI) \\
&\cdot \Sigma_b \mid \gamma \sim IW(S_\gamma, \eta) \\
&\cdot \sigma^2_\epsilon \sim IG(c_\epsilon; d_\epsilon) \\
&\cdot \gamma \mid \pi \sim \prod_{k=1}^{K} \pi^{\gamma_k} (1 - \pi)^{1-\gamma_k} \\
&\cdot \pi \sim Beta(c_\pi, d_\pi)
\end{align*}
\]
Priors for Model with Breakpoint Selection

- Prior specification for (4) is hierarchical:

\[
\begin{align*}
\cdot b_i & \mid \beta, \Sigma_b, \gamma \sim N(\beta, \Sigma_b, \gamma) \\
\cdot \beta & \mid \Sigma_b, \gamma \sim N(0, cI) \\
\cdot \Sigma_b & \mid \gamma \sim IW(S_\gamma, \eta) \\
\cdot \sigma^2 & \sim IG(c_\epsilon, d_\epsilon) \\
\cdot \gamma & \mid \pi \sim \prod_{k=1}^{K} \pi^{\gamma_k}(1 - \pi)^{1-\gamma_k} \\
\cdot \pi & \sim Beta(c_\pi, d_\pi)
\end{align*}
\]

- \(c_\pi\) and \(d_\pi\) can be adjusted to give a priori probability of including a pre-specified number of breakpoints. (Kohn, Smith, and Chan, 2001)
Model is estimated via an MCMC sampling algorithm:
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

**Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) \mid \gamma_{(k)}, \sum_b, \sigma^2_e, \text{data}$
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

**Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) \mid \gamma(k), \Sigma_b, \sigma_e^2, \text{data}$

**Step 2(a):** Sample $\gamma_k \mid \gamma(k), \Sigma_b, \sigma_e^2, \text{data}$
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

**Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) | \gamma, \Sigma_b, \sigma^2_\epsilon, \text{data}$

**Step 2(a):** Sample $\gamma_k | \gamma, \Sigma_b, \sigma^2_\epsilon, \text{data}$

**Step 2(b):** Sample $\beta | \gamma, \Sigma_b, \sigma^2_\epsilon, \text{data}$
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

**Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) \mid \gamma_{(k)}, \Sigma_b, \sigma^2_\epsilon, \text{data}$

- **Step 2(a):** Sample $\gamma_k \mid \gamma_{(k)}, \Sigma_b, \sigma^2_\epsilon, \text{data}$
- **Step 2(b):** Sample $\beta \mid \gamma, \Sigma_b, \sigma^2_\epsilon, \text{data}$
- **Step 2(c):** Sample $\{b_i\} \mid \gamma, \beta, \Sigma_b, \sigma^2_\epsilon, \text{data}$
Model is estimated via an MCMC sampling algorithm:

**Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

**Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) \mid \gamma_{(k)}, \Sigma_b, \sigma^2_\varepsilon, \text{data}$

- **Step 2(a):** Sample $\gamma_k \mid \gamma_{(k)}, \Sigma_b, \sigma^2_\varepsilon, \text{data}$
- **Step 2(b):** Sample $\beta \mid \gamma, \Sigma_b, \sigma^2_\varepsilon, \text{data}$
- **Step 2(c):** Sample $\{b_i\} \mid \gamma, \beta, \Sigma_b, \sigma^2_\varepsilon, \text{data}$

**Step 3:** Sample $\sigma^2_\varepsilon \mid \gamma, \beta, \{b_i\}, \text{data}$
Estimation of Model with Breakpoint Selection

- Model is estimated via an MCMC sampling algorithm:

  **Step 1:** Sample $k$ from $p(k) = \frac{1}{K-p}$

  **Step 2:** Sample $(\gamma_k, \beta, \{b_i\}) \mid \gamma(k), \Sigma_b, \sigma^2_\epsilon, \text{data}$
  
  **Step 2(a):** Sample $\gamma_k \mid \gamma(k), \Sigma_b, \sigma^2_\epsilon, \text{data}$
  
  **Step 2(b):** Sample $\beta \mid \gamma, \Sigma_b, \sigma^2_\epsilon, \text{data}$
  
  **Step 2(c):** Sample $\{b_i\} \mid \gamma, \beta, \Sigma_b, \sigma^2_\epsilon, \text{data}$

  **Step 3:** Sample $\sigma^2_\epsilon \mid \gamma, \beta, \{b_i\}, \text{data}$

  **Step 4:** Sample $\Sigma_b \mid \gamma, \beta, \{b_i\}, \text{data}$
The posterior conditional distribution for $\gamma_k$ is:

$$p(\gamma_k = 1) = \frac{\theta_1 L_1}{\theta_0 L_0 + \theta_1 L_1}$$
The posterior conditional distribution for $\gamma_k$ is:

$$p(\gamma_k = 1) = \frac{\theta_1 L_1}{\theta_0 L_0 + \theta_1 L_1}$$

$$\theta_0 = \frac{K - q_\gamma_1 + d_\pi}{K + c_\pi + d_\pi - 1} \quad \theta_1 = \frac{q_\gamma_1 + c_\pi - 1}{K + c_\pi + d_\pi - 1}$$
The posterior conditional distribution for $\gamma_k$ is:

$$p(\gamma_k = 1) = \frac{\theta_1 L_1}{\theta_0 L_0 + \theta_1 L_1}$$

$$\theta_0 = \frac{K - q\gamma_1 + d_\pi}{K + c_\pi + d_\pi - 1} \quad \theta_1 = \frac{q\gamma_1 + c_\pi - 1}{K + c_\pi + d_\pi - 1}$$

$$L_s \approx (nc + 1)^{-q_{\gamma_s}} \left\{ \prod_{i=1}^{n} \det \left[ \Phi_{\gamma_s,i}^T \Phi_{\gamma_s,i} \Sigma_{b,\gamma_s} / \sigma^2_\epsilon + (1 - \frac{c}{nc+1})I \right]^{-\frac{1}{2}} \right\} \times \exp \left\{ \frac{1}{\sigma^4} \sum_{i=1}^{n} \mathbf{y}_i^T \Phi_{\gamma_s,i} \left[ \Phi_{\gamma_s,i}^T \Phi_{\gamma_s,i} / \sigma^2_\epsilon + (1 - \frac{c}{nc+1})\Sigma_{b,\gamma}^{-1} \right]^{-1} \Phi_{\gamma_s,i} \mathbf{y}_i \right\}$$

$\Phi_{\gamma_s,i}$ is the $n_i \times q_\gamma$ matrix of basis functions evaluated at $t_i$
Estimate of Posterior Mean Curves

- Suppose we have $L$ draws of the parameters.

- Estimated posterior overall mean curve:
  
  $$
  \hat{\mu}(t) = \frac{1}{L} \sum_{l=1}^{L} \phi_{\gamma_l}^T(t) \beta_{\gamma_l}^{[l]} 
  $$

- Estimated posterior mean of the $i$th curve:
  
  $$
  \hat{f}_i(t) = \frac{1}{L} \sum_{l=1}^{L} \phi_{\gamma_l}^T(t) b_{\gamma_l,i}^{[l]} 
  $$

- Pointwise posterior credible intervals easy to obtain.
Two Small Simulation Studies

- 100 datasets generated from each of two settings of:

\[ y_{ij} = b_i^T \phi (t_{ij}) + \epsilon_{ij} \]

- \( n = 50 \) curves each with \( n_i \) observations each
- \( n_i \sim \text{Pois}(5) + 2 \)
- \( t_i | n_i \sim \text{iid U}(0,1) \)
- \( b_i \sim \text{iid N}(\beta, \Sigma) \)
- \( \epsilon_i \sim \text{iid N}(0, 1) \)
Simulated Data (left) and Estimated Means (right)
Two Small Simulation Studies: Sampling Procedure

- Select from 40 equally-spaced breakpoints.

- Specified noninformative priors.

- 5000 iterates obtained after a burn-in of 5000.

- For each simulated dataset we:

  1. Determine if true mean falls within pointwise PCIs.

  2. Compute posterior probabilities of breakpoint inclusion.
Coverage Levels of PCIs (left) and Post. Prob. of BP Inclusion (right)
The “smooth” within-individual covariance function can be estimated by:

\[
\hat{\Sigma}_\tau = \frac{1}{S} \sum_{s=1}^{S} X_{\tau,\gamma^s} \Sigma_{b,\gamma}^s X'_{\tau,\gamma^s}.
\]

\(X_{\tau,\gamma}\) is the \(\tau \times q_\gamma\) matrix of B-splines selected by \(\gamma\) and evaluated at the time points in \(t\).

\(t\) is a fine grid of \(\tau\) time values on the interval \([0, T]\).
True Covariance/Correlation Surfaces
Randomly-Selected Dataset: Estimated Covariance/Correlation Surfaces
Body Fat Data: Mean and 95% PCIs

Plot of Percent Body Fat over Time for 162 Girls
Body Fat Data: Four Individual Trajectories

Subject 1

Subject 2

Subject 3

Subject 4
Body Fat Data: Covariance (left) and Correlation (right)
First Two Eigenfunctions (left) with Individual Trajectories (right).
The B-spline mixed-effects model handles missing or “sparse” functional data appropriately.
Conclusions

- The B-spline mixed-effects model handles missing or “sparse” functional data appropriately.

- Latent indicator variables provide an “automatic” locally-adaptive method for fitting the B-spline mixed-effects model.
Conclusions

- The B-spline mixed-effects model handles missing or “sparse” functional data appropriately.

- Latent indicator variables provide an “automatic” locally-adaptive method for fitting the B-spline mixed-effects model.

- Inferences on curves including posterior probabilities of breakpoint inclusion are straightforward and should be more realistic than conditioning on the best model.
Conclusions

- The B-spline mixed-effects model handles missing or "sparse" functional data appropriately.

- Latent indicator variables provide an "automatic" locally-adaptive method for fitting the B-spline mixed-effects model.

- Inferences on curves including posterior probabilities of breakpoint inclusion are straightforward and should be more realistic than conditioning on the best model.

- The model and the MCMC sampling scheme are fairly easy to extend to more complicated models.