Spatio-temporal Functional Mixed Effects Model for Longitudinal Functional Responses

Ricardo Batista

Introduction

The aim of "FMEM: Functional Mixed Effects Models for Longitudinal Functional Responses" (Zhu et al, 2019) is to conduct a systematic and theoretical analysis of estimation and inference for a class of functional mixed effects models (FMEM). Such FMEMs consist of fixed effects that characterize the association between longitudinal functional responses and covariates of interest and random effects that capture the spatio-temporal correlations of longitudinal functional responses. Focus is on sparse longitudinal data.

Motivation

Suppose we observe longitudinal functional data and covariates for n independent subjects. Let

- T_i be the total number of longitudinal measurements for the i-th subject, $i = 1, ..., n$, and t_{ij} be the j-th measurement time point for the i-th subject, so $j = 1, ..., T_i$.
- s_m represent a specific grid point of the functional template space $\mathcal{S}$ for $m = 1, ..., M$.

  For instance, M grid points along a nerve tract in the brain. For ease of notation, it is assumed that $S = [0, 1]$ and $0 = s_1 \leq \cdots \leq s_M = 1$, but the results can be easily extended to higher dimensions when S is a compact subset of a Euclidean space.

Specifically, for the i-th subject at time t_{ij}, we observe functional data, denoted by $y_{ij}(s_m) = y_i(t_{ij}, s_m)$ for $1 \leq m \leq M$, and a p_x dimensional covariate vector x_i of interest, denoted by $x_{ij} = x_i(t_{ij})$, at time t_{ij}. The x_i may include time-independent as well as time-dependent covariates, such as age, gender, and genetic markers.

FMEM: Functional Mixed Effects Model

The FMEM consists of a measurement model and a hierarchical factor model:

$$y_{ij}(s) = \mu(x_{ij}, \beta(s)) + z_{ij}^T b_i(s) + e_{ij} \qquad (1)$$

- $\mu(\cdot,\cdot)$ is a known function

  In many applications, $\mu(x_{ij}, \beta(s)) = x_{ij}^T\beta(s)$ is a linear function of x_{ij}. Since, marginally, for a fixed s, the model above with $\mu(x_{ij}, \beta(s)) = x_{ij}^T\beta(s)$ is a standard linear mixed effects model, this motivates us to adopt standard notation for linear mixed effects models. Extensions to nonlinear cases is not covered in this poster but are trivial.
- $\beta(s) = (\beta_1(s), ..., \beta_{p_\beta}(s))^T$ is a $p_\beta \times 1$ vector of the fixed-effect functions of s
- $b_i(s) = (b_{i1}(s), ..., b_{ip_z}(s))^T$ is a vector of the random effects that characterize the spatial temporal correlation structures across the functional domain space
- $z_{ij} = z_i(t_{ij}) = (z_{ij1}, ..., z_{ijp_z})^T$ is a $p_z \times 1$ vector of the random-effect covariates associated with the random effects $b_i(s)$

  Since z_{ij} may include time-independent as well as time-dependent covariates, the inclusion of $z_{ij}^T b_i(s)$ allows us to capture a large portion of the variation in the spatial and temporal correlation structures.
- $e_{ij}(s)$ is a spatial random process delineated from the $b_i(s)$, i.e., after filtering out $z_{ij}^T b_i(s)$

  The spatial random process e_{ij} is further decomposed into two parts,

  $$e_{ij}(s) = e_{ij,G}(s) + e_{ij,L}(s), \qquad (2)$$

  where $e_{ij,G}(s)$ is a smooth stochastic process representing the global dependency that depicts the medium-to-long-range spatial dependence, $e_{ij,L}(s)$ is a measurement error representing local variability, and $e_{ij_1,L}(\cdot)$ and $e_{ij_2,L}(\cdot)$ are independent for any j_1 and j_2.

Remark: $b_i(s)$, $e_{ij,L}(s)$, and $e_{ij,G}(s)$ are mutually independent and are iid $\text{SP}(0, \Sigma_{e,L})$, $\text{SP}(0, \Sigma_b)$, and $\text{SP}(0, \Sigma_{e,G})$, respectively, where $\text{SP}(\mu, \Sigma)$ denotes a stochastic process vector with mean function (or function vector) $\mu(s)$ and covariance function (or function matrix) $\Sigma(s, s')$. So the covariance structure of $y_i(s) = (y_{i1}(s), ..., y_{iT_i}(s))^T$, denoted by $\Sigma_{y_i}(s, s')$, is

$$\Sigma_{y_{ij_1j_2}}(s, s') = z_{ij_1}^T \Sigma_b(s, s') z_{ij_2} + \Sigma_{e,G}(s, s') I(j_1 = j_2) + \Sigma_{e,L}(s, s') I(j_1 = j_2, s = s').$$

Estimation Procedure

Step (I): Calculate an initial estimator $\hat{\beta}(s)$ of $\beta(s)$ for each $s \in \mathcal{S}$.

First we simplify the problem by assuming there is no structure to the e_{ij}, i.e., there is no $e_{ij,G}$, and perform local linear regression of Y on x: recall that, for s near s_m,

$$\beta(s_m) \approx \beta(s) + \dot{\beta}(s)(s_m - s) = A(s) s_{h_1}(s_m - s)$$

where

- $s_{h_1}(s_m - s) = (1, (s_m - s)/h_1)^T$
- $A(s) = \left[\beta(s)h_1 \;\; \dot{\beta}(s)\right]$ is a $p_x \times 2$ matrix
- $\dot{\beta}(s) = (\dot{\beta}_1(s), ..., \dot{\beta}_{p_x}(s))^T$ is a $p_x \times 1$ vector
- $\dot{\beta}_l(s) = d\beta_l(s)/ds$ for $l = 1, ..., p_x$

Then we define $\hat{\beta}(s)$ to minimize the least squares function

$$\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}\{y_{ij}(s_m) - x_{ij}^T A(s) s_{h_1}(s_m - s)\}^2 K_{h_1}(s_m - s),$$

where $K(s)$ and $K_h(s) = h_1^{-1}K(s/h_1)$ are the kernel function and the rescaled kernel function with bandwidth h_1, respectively. In practice, we may select h_1 via LOOCV.

Step (II): Estimate the covariance operators $\Sigma_b(s, s')$, $\Sigma_{e,G}(s, s')$, and $\Sigma_{e,L}(s, s')$, which, note, fully specify the distributions of the $b_i(s)$, $e_{ij,G}(s)$, and $e_{ij,L}(s)$, respectively.

(S1) We estimate the covariance surfaces $\Sigma_{bkk'}(s, s')$ and $\Sigma_e(s, s')$ on a *point-by-point basis*. (Note (k, k') is the entry index.) That is, for each $(s_m, s_{m'}) \in \mathcal{S} \times \mathcal{S}$ we regress the residuals from Step (I) $\hat{u}_{ij}(s) = y_{ij}(s) - x_{ij}^T\hat{\beta}(s)$ on the random effect covariates z_{ij} to find the minimizers $\hat{\Sigma}_b^{LS}(s_m, s_m')$ and $\hat{\Sigma}_e^{LS}(s_m, s_m')$ of the least squares function

$$\sum_{i=1}^{n}\sum_{j_1 \neq j_2}\left\{\hat{u}_{ij_1}(s_m)\hat{u}_{ij_2}(s_{m'}) - z_{ij_1}^T\Sigma_b(s_m, s_{m'})z_{ij_2}\right\}^2 + \sum_{i=1}^{n}\sum_{j=1}^{T_i}\left\{\hat{u}_{ij}(s_m)\hat{u}_{ij}(s_{m'}) - z_{ij}^T\Sigma_b(s_m, s_{m'})z_{ij} - \Sigma_e(s_m, s_{m'})\right\}^2,$$

where $\Sigma_e(s, s')$ is the covariance function of $e_{ij}(s)$.

(S2) Using the least squares (LS) estimates from (S1), we use a local constant smoother to produce the covariance surface estimates $\hat{\Sigma}_{bkk'}(s, s')$ and $\hat{\Sigma}_{e,G}(s, s')$ by minimizing

$$\min_{\Sigma_{bkk'}(s,s')}\sum_{m,m'=1}^{M}\left\{\hat{\Sigma}_{bkk'}^{LS}(s_m, s_{m'}) - \Sigma_{bkk'}(s, s')\right\}^2 K_{h_2}(s_m - s)K_{h_2}(s_{m'} - s')$$

$$\min_{\Sigma_{e,G}(s,s')}\sum_{m \neq m'}\left\{\hat{\Sigma}_{e}^{LS}(s_m, s_{m'}) - \Sigma_{e,G}(s, s')\right\}^2 K_{h_3}(s_m - s)K_{h_3}(s_{m'} - s')$$

Finally, we perform the spectral decomposition of $\hat{\Sigma}_{bkk'}(s, s')$ and $\hat{\Sigma}_{e,G}(s, s')$ to calculate $\hat{\Sigma}_{e,L}(s_m, s_m')$.

Step (III): We proceed as in Step (I): we run local linear regression of Y on x, but this time we include the estimated covariance function $\hat{\Sigma}_{y_i,G}(s, s')$ to get the refined estimator $\tilde{\beta}(s)$.

We use the estimated covariance operators obtained from Step (II) to improve the estimate in Step (I) with a refined estimator of $\beta(s)$, denoted by $\tilde{\beta}(s)$, the minimizer of

$$\sum_{i=1}^{n}\sum_{m=1}^{M}\left[\left\{y_i(s_m) - X_i^T A(s) s_{h_3}(s_m - s)\right\}^T \hat{\Sigma}_{y_i,G}(s_m, s_m)^{-1/2}\right]^{\otimes 2} K_{h_3}(s_m - s)$$

where $a^{\otimes 2} = aa^T$ for any vector a, and $\hat{\Sigma}_{y_i,G}(s, s')$ is the estimator of the covariance function of $u_{i,G}(s) = (u_{i1,G}(s), ..., (u_{iT_i,G}(s))^T$ where $u_{ij,G}(s) = z_{ij}^T b_i(s) + e_{ij,G}(s)$. We obtained $\hat{\Sigma}_{y_i,G}(s, s')$ based on $\hat{\Sigma}_b(s, s')$ and $\hat{\Sigma}_{e,G}(s, s')$.

Step (IV): Obtain individual random effect functions $u_{ij,G}(s) = z_{ij}^T b_i(s) + e_{ij,G}(s)$.

We produce estimates for $\mu_{ij,G}(s)$ using local linear regression on the residuals based on the new $\beta(s)$ estimate, i.e., $\{\tilde{u}_{ij}(s_m) = y_{ij}(s_m) - x_{ij}^T\tilde{\beta}(s_m)\}_{m=1}^{M}$. Furthermore, if there is an interest in recovering the subject-specfic random effect $b_i(s)$, one could use the BLUPs.

Application: Modeling white matter across the corpus callosum over time

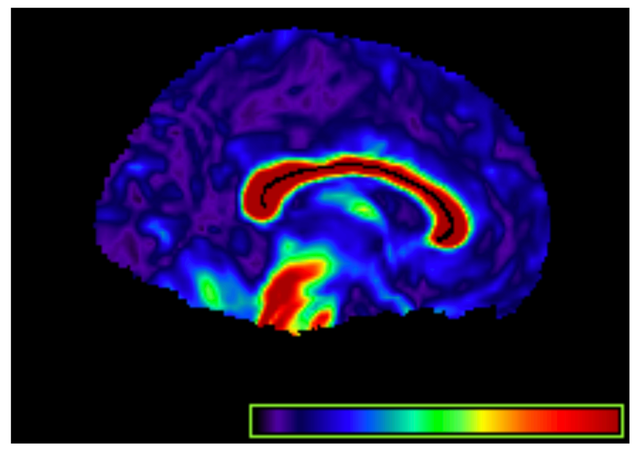

Figure 1. 3D visualization of the corpus callosum in the sagittal view

The researchers fit FMEM on MRI data collected on 253 infants over several visits. More specifically, the response $y_{ij}(s) \in [0, 1]$ is fractional anisotropy (FA) – a useful measure of connectivity in the brain – that can be derived from a diffusion tensor imaging (DTI) – an MRI method – dataset. The functional template space $\mathcal{S}$ is the corpus callosum, the black arc with red boarder highlighted in Figure 1. In this manner $y_{ij}(s)$ is the FA value for the ith patient at position $s \in [0, 45]$ (i.e., arc length) on day $j = [0, 8000]$. The goal of the analysis is to assess the development of brain connectivity over time.

They fit FMEM (1) and (2) with $x_i = (1, \text{Gender}, \log(\text{Age}), \log(\text{Age})^2)^T$ and $z_i = (1, \log(\text{Age}))^T$ to the selected FA tracts obtained from all 253 subjects. The coefficient functions associated with $\log(\text{Age})$ and $\log(\text{Age})^2$ were included to detect age effect in FA changes. Additionally, as shown in Figure 2, there are random subject-to-subject variations in FA measures at each grid point along this tract as well as those in the age effect on FA measures. The researchers included random intercept and age effects in the model in order to account for the inter-subject variations.

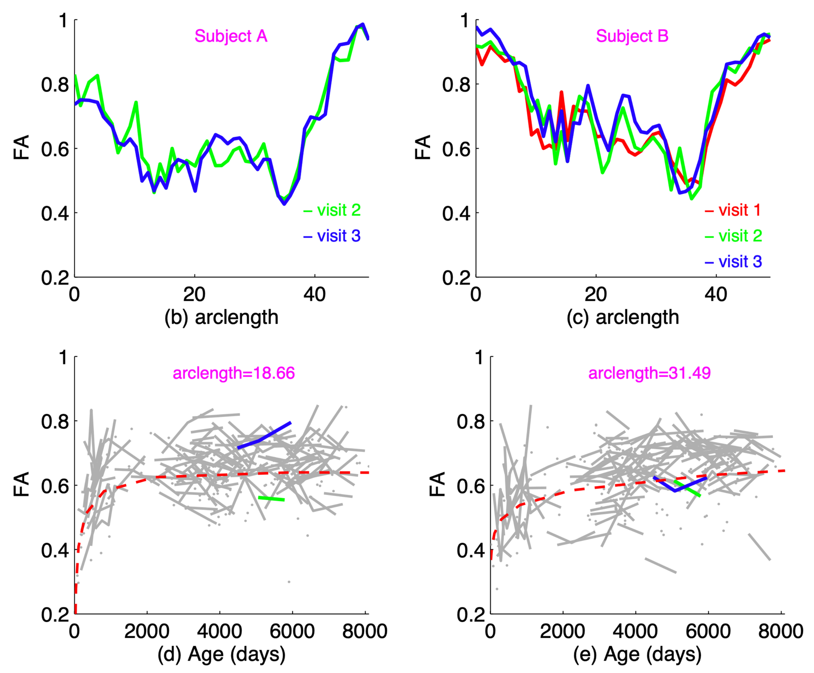

Figure 2. (b) and (c) FA's along the corpus callosum obtained from 2 selected subjects A (b) and B (c) with 2 or 3 visits. Different visits for the same subjects are indicated by color. (d) and (e) FA values varying over age at selected locations: arclength=18.66 (d) and arclength=31.49 (e) along the corpus callosum for all 253 subjects, with green and blue lines corresponding to subjects A and B, respectively. Red dashed lines represent the fitted lines for the male group.

The estimated functional coefficients of $\beta(s)$ and their 95% simultaneous confidence bands were constructed. Figure 3 presents the estimated coefficient functions along with their 95% simultaneous confidence bands. The intercept function describes the overall trend of FA along the corpus callosum. In general, the central regions of the corpus callosum show smaller FA values, whereas the peripheral regions show larger FA values. Note that, for the gender effect, the simultaneous confidence band contains the zero horizontal line, whereas the zero line is out of the 95% simultaneous confidence band for the age effect, indicating a significant age effect.

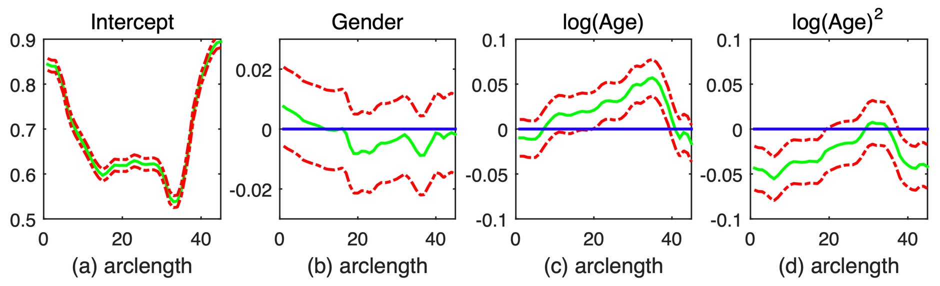

Figure 3. 95% simultaneous confidence bands for coefficient functions.

Figure 4 shows the top eigenfunctions for b and e_G, respectively. It shows that 31.4% of the variability is explained by the first principal component (PC) for b and 18.2% by the first PC for e_G. Overall, the first 8 PCs for b explain 62.5% of the total variability, whereas the first 8 PCs for e_G explain 32.18% of the total variability. This indicates that the random effects b capture most of the variation in the data. Within b, 53.6% and 8.9% of the total variation are explained by the random functional intercept and the subject-specific random slope, respectively. The within-curve measurement error explains only 5.4% of the total variation.

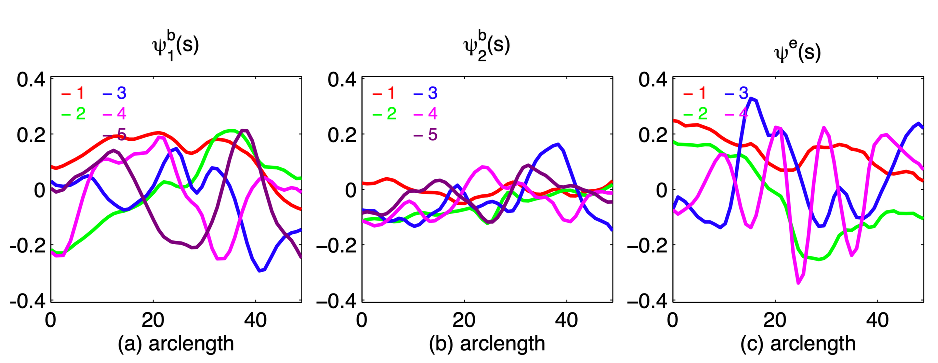

Figure 4. (a) (b) The first five estimated eigenfunctions $\psi_{l,k}^b(s)$, $l = 1, 2$ for the random intercept and slope processes. $\psi_{1,k}^b(s)$ and $\psi_{2,k}^b(s)$ correspond to the random functional intercept and random functional slope, respectively. (c) The first four estimated eigenfunctions $\psi_k^e(s)$ for the visit specific deviation process.

rbatista@iastate.edu