Bayesian Smoothing and Classification of Sparse Functional Data Using Gaussian Process

Tahmidul Islam*, Paramita Chakraborty, James Lynch, John Grego
Department of Statistics, University of South Carolina

Goal

Functional data (FD) refers to trajectories (curves) observed from study subjects over a domain such as time and/or space. In many cases, trajectories are observed at domain points that vary widely across subjects giving rise to "sparse" functional data. Analyzing sparse FD is more challenging than regular FD (traditionally spline and functional principal component-based models are used).

Gaussian stochastic processes provide a more flexible tool for studying sparse FD's. The goal of this project regarding sparse FD is as follows

- Propose a general (sparse/non-sparse) FD model accounting for within and between subject variations in the curves within a Bayesian framework using the Gaussian process (GP) prior over the space of functions.
- Estimate the model parameters and derive the posterior distribution.
- Use of the posterior distribution for estimating the mean trajectory, and perform supervised classification.

GP Regression

A Gaussian process (GP) is a stochastic process $\{f(x) : x \in \mathcal{X}\}$. In a Bayesian nonparametric regression setting, GP with kernel $K(\cdot, \cdot)$ is used as a prior on a function as:

$$Y = f(x) + \epsilon$$
$$f(\cdot) \sim \mathcal{GP}(\mu(\cdot), K(\cdot, \cdot))$$
$$\epsilon \sim N(0, \sigma^2).$$

The posterior distribution can be made tractable by assuming normally distributed error term. The posterior predictive distribution for test points X^* is a multivariate normal distribution with following mean and covariance matrix.

$$\boldsymbol{\mu}^* = K(X^*, X)(K(X, X) + \sigma^2 I)^{-1}\boldsymbol{y}$$
$$\Sigma^* = K^*(X^*, X^*) + \sigma^2 I -$$
$$K(X^*, X)(K(X, X) + \sigma^2 I)^{-1} K(X, X^*).$$

Bayesian Model for Sparse Functional Data

We assume that the functional curve (response over time t) produced by the j^{th} subject is given by a stochastic process $Y_j(t)$ of the form:

$$Y_j(t) = \mu(t) + g_j(t) + \varepsilon_j$$
$$g_j(\cdot) \sim \mathcal{GP}(0, \Omega(\cdot, \cdot))$$
$$\varepsilon_j \sim N(0, \sigma^2),$$

where, $\mu(t)$ is the fixed mean function for the population, $g(t)$ is a random component that accounts for the subject's variation from the overall mean function (think mixed effects) and finally ε is the i.i.d normally distributed measurement error. We impose a GP prior on the mean function $\mu(t)$ with covariance function $\Sigma(\cdot, \cdot)$.

$$\mu \sim \mathcal{GP}(0, \Sigma(\cdot, \cdot)).$$

A vector-matrix representation of the whole observed data leads to compact and easier derivation of further results.

$$\begin{bmatrix} \boldsymbol{Y_1(t)} \\ \boldsymbol{Y_2(t)} \\ \vdots \\ \boldsymbol{Y_n(t)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu(t)} \\ \boldsymbol{\mu(t)} \\ \vdots \\ \boldsymbol{\mu(t)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{g_1(t)} \\ \boldsymbol{g_2(t)} \\ \vdots \\ \boldsymbol{g_n(t)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_n} \end{bmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{g} + \boldsymbol{\varepsilon}$$

Estimation: Hyperparameters θ_Σ, θ_Ω and error variance σ^2 are estimated by maximizing the marginal likelihood function where the marginal likelihood is $\boldsymbol{Y} \sim MN(\boldsymbol{0}, K = Cov(\boldsymbol{\mu}) + Cov(\boldsymbol{g}) + Cov(\boldsymbol{\varepsilon}))$.

Posterior: Due to the Gaussian assumption, we can derive the exact tractable posterior distribution for some test time points $\boldsymbol{t}^*$.

$$\boldsymbol{Y}^* | \boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{t}^* \sim MN(\mu_p = K^* K^{-1} \boldsymbol{y}, \Sigma_p = K^{**} - K^* K^{-1} K'^*).$$

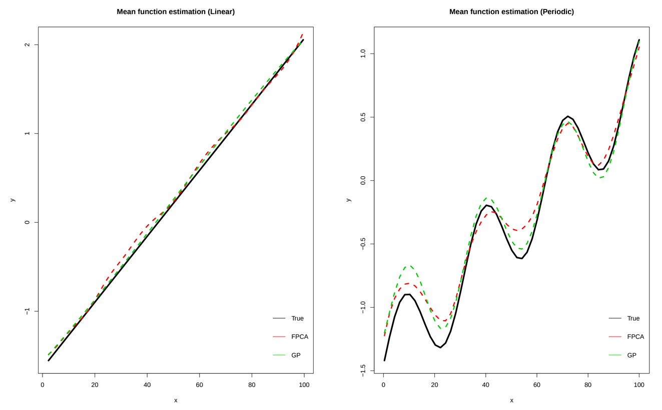
Where $K^* = \Sigma'(\boldsymbol{t}^*, \boldsymbol{t}),\ K^{**} = \Sigma(\boldsymbol{t}^*, \boldsymbol{t}^*) + \Omega(\boldsymbol{t}^*, \boldsymbol{t}^*) + \sigma^2 I.$

Bayesian Smoothing

Suppose observations from the j^{th} subject are available on time points $(t_{j_1}, t_{j_2}, \ldots, t_{j_{n_j}})$ (discrete and varying). Parameters are estimated using all available observations on a pooled grid. The mean function on a desirable time grid can be estimated by the posterior mean, while simply plugging-in the grid as test time points on the posterior distribution. The average squared error (ASE) of our method (with different kernels) compared to FPCA on simulated datasets is presented below.

	Bayesian Smoothing				FPCA
ASE	Gaussian	Laplace	Matern 5/2	Matern 3/2	
Linear	0.0060	0.0080	0.0045	0.0049	0.0099
Periodic	0.0071	0.0100	0.0068	0.0074	0.0191

Table 1: Average Squared Error (ASE) by fitting our method with different kernels and FPCA.



Classification

Further, our method can perform supervised classification of sparse/non-sparse FD using a discriminant analysis approach. The steps of classification are:

- Estimate separate models for each of the class.
- For a new test curve (observed at any time points), compute the posterior distribution (mean and covariance).
- Compute posterior densities for each class and apply Bayes' classifier.

We apply our Bayesian method on spinal bone mineral density data and classify subjects to 4 ethnic groups. The following table shows the number of correct classifications when only 153 female subjects were considered. Our model outperforms functional linear discriminant analysis (FLDA) and functional robust support vector machine (FSVM) (for certain kernels).

	Bayesian Method				FLDA	RSVM
Kernel	Gaussian	Laplace	Matern 5/2	Matern 3/2		
Correct	68	81	70	72	63	71

Reference

[1] G. M. James and T. J. Hastie, *Functional linear discriminant analysis for irregularly sampled curves*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63 (2001), pp. 533–550.

[2] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2, MIT press Cambridge, MA, 2006.

[3] Y. Wu and Y. Liu, *Functional robust support vector machines for sparse and irregular longitudinal data*, Journal of computational and Graphical Statistics, 22 (2013), pp. 379–395.

[4] F. Yao, H.-G. Müller, and J.-L. Wang, *Functional data analysis for sparse longitudinal data*, Journal of the American Statistical Association, 100 (2005), pp. 577–590.