Presentation for the paper “Functional Data Analysis for Sparse Longitudinal Data”

Boyi Hu

SFU

February 25, 2019

Introduction

Target data set: **sparse** functional data as **noisy** sampled points from a collection of trajectories that are assumed to be independent realizations of a smooth random function, with unknown mean function

$$\mathrm{E}X(t) = \mu(t)$$

and covariance function

$$\mathrm{cov}(X(s), X(t)) = G(s, t).$$

The **pooled** time points are sufficiently **dense** in the domain T of $X(\cdot)$.

Notations

- Let Y_{ij} be the jth observation of the random function $X_i(\cdot)$, at a time T_{ij}.
- Let ϵ_{ij} be the IID measurement errors with

$$\mathrm{E}\epsilon_{ij} = 0, \ \mathrm{var}(\epsilon_{ij}) = \sigma^2$$

- The model being considered is given by

$$\begin{aligned} Y_{ij} &= X_i(T_{ij}) + \epsilon_{ij} \\ &= \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(T_{ij}) + \epsilon_{ij}. \end{aligned}$$

- The number of measurements N_i made on the ith subject is considered random, reflecting sparse.
- $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, $k = 1, 2, \ldots$.

Estimation of model components

- For a first step, $\hat{\mu}(t)$ is estimated by using a local linear smoother on the pooled data.
- Given $\hat{\mu}(t)$, let $G(s,t) = \text{cov}[X(s), X(t)]$, the "raw" covariance can be obtained

  $$G_Y(s,t) = (Y_i - \hat{\mu}(s))(Y_i - \hat{\mu}(t)),$$

  and we can observe that

  $$\mathrm{E}[G_Y(s,t)|s,t] \approx G(s,t) + \sigma^2\delta_{st}.$$

- Then $\hat{G}(s,t)$ and $\hat{\sigma}^2$ are estimated based on $G_Y(s,t)$.
- Given $\hat{G}(s,t)$, $\hat{\phi}_k$ and $\hat{\lambda}_k$ are estimated.
- For the last step, $\hat{\xi}_{ik}$ is estimated by conditional expectation.

Estimate $\hat{\mu}(t)$

The estimate of $\hat{\mu}(t)$ is straightforward. A local linear smoother is used on the pooled data. $\hat{\mu}(t)$ is estimated by minimizing

$$\sum_i \sum_j k_1 \left(\frac{T_{ij} - t}{h_\mu}\right) \{Y_{ij} - \beta_0 - \beta_1 (t - T_{ij})\}^2$$

with respect to β_0 and β_1. It is a point-wise minimization. The estimate of $\mu(t)$ is $\hat{\mu}(t) = \hat{\beta}_0(t)$.

Estimate $\hat{G}(s,t)$

Then we focus on the estimation of $\hat{G}(s,t)$ and $\hat{\sigma}^2$.

- Since

$$\mathrm{E}[G_Y(s,t)|s,t] \approx G(s,t)(= \mathrm{cov}[X(s),X(t)]) + \sigma^2\delta_{st},$$

$\hat{G}(s,t)$ is estimated by using a local linear surface smoother on the "raw" covariance $G_Y(s,t)$ after the removal of its diagonals,

$$\sum_i \sum_{j \neq l} k_2\left(\frac{T_{ij}-s}{h_G}, \frac{T_{il}-t}{h_G}\right) \times$$
$$\{G_Y(T_{ij},T_{il}) - \beta_0 - \beta_{11}(s-T_{ij}) - \beta_{12}(t-T_{il})\}^2.$$

The estimate of $G(s,t)$ is $\hat{G}(s,t) = \hat{\beta}_0(s,t)$.

Estimate $\hat{\sigma}^2$

In the paper, they claim that the covariance of X(t) is maximal along the diagonal, a local quadratic rather than a local linear fit is expected to better approximate the shape of the surface in the direction orthogonal to the diagonal.

- An estimator $\hat{V}(t)$ focusing on diagonal values $\{G(t,t)+\sigma^2\}$ is the diagonal of $\bar{G}(s,t)$, which minimizes the following function without removing the diagonal entries of $G_Y(s,t)$,

$$\sum_i \sum_{j \neq l} k_2\left(\frac{T^*_{ij}-s}{h_G}, \frac{T^*_{il}-t}{h_G}\right) \times$$
$$\{G_Y(T^*_{ij}, T^*_{il}) - \gamma_0 - \gamma_1(s - T^*_{ij}) - \gamma_2(t - T^*_{il})^2\}^2.$$

  where (T^*_{ij}, T^*_{il}) is the coordinates after we rotate the x-axis and y-axis by 45 degrees.

- Given $\hat{G}(s,t)$ and $\hat{V}(t)$, σ^2 is then estimated by their average difference

$$\hat{\sigma}^2 = \frac{1}{|T|}\int_T \{\hat{V}(t) - \mathrm{diag}[\hat{G}(s,t)]\} dt,$$

One-curve-leave-out cross validation

Let $\hat{\mu}^{(-i)}$ and $\hat{\phi}_k^{(-i)}$ be the estimated mean and eigenfunctions after removing the data for the ith subject. We choose h_μ, h_G and K so as to minimize the cross-validation score based on the squared prediction error,

$$\mathrm{CV(K)} = \sum_i \sum_j \{Y_{ij} - \hat{Y}_{ij}^{(-i)}(T_{ij})\}^2.$$

Estimate eigenvalues and eigenfunctions

The estimate of eigenfunctions and eigenvalues correspond to the solutions $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigenequations

$$\int_T \hat{G}(s,t)\hat{\phi}_k(s) = \hat{\lambda}_k \hat{\phi}_k(t),$$

under the constraints $\left\|\hat{\phi}_k(t)\right\| = 1$ and $< \hat{\phi}_k(t), \hat{\phi}_m(t) >= 0$.

Estimate FPC scores

Traditionally FPC scores are estimated by numerical integration

$$\hat{\xi}_{ik} = \int (X_i(t) - \hat{\mu}(t))\hat{\phi}_k(t)dt.$$

However, for sparse functional data, the numerical integration can not provide a reasonable approximation.

Estimate FPC scores

- A Fact: Under the assumption that ξ_{ik} and ϵ_{ij} are jointly Gaussian, the best prediction is given by

$$\tilde{\xi}_{ik} = \mathrm{E}(\xi_{ik}|Y_i) = \lambda_k \phi_{ik} \Sigma_{Y_i}^{-1}(Y_i - \mu_i),$$

  where $Y_i = (Y_{i1}, \ldots, Y_{iN_i})$, $\Sigma_{Y_i} = \mathrm{cov}(Y_i, Y_i)$, $\phi_{ik} = (\phi_k(T_{i1}), \ldots, \phi_k(T_{iN_i}))$, $\mu_i = (\mu(T_{i1}), \ldots, \mu(T_{iN_i}))$.
- But the true values of λ_k, ϕ_{ik}, Σ_{Y_i} and μ_i are unknown.
- In this paper, they still use the conditional expectation to estimate scores but substitute estimates of these true values, their score estimator is given by

$$\hat{\xi}_{ik} = \hat{\lambda}_k \hat{\phi}_{ik} \hat{\Sigma}_{Y_i}^{-1}(Y_i - \hat{\mu}_i).$$

Prediction of trajectory $X_i(t)$

- Assume that the infinite-dimensional process under consideration are well approximated by the projection on the function space spanned by the first K eigenfunctions.
- The choice of K is decided by cross validation based on one-curve-leave-out prediction error.
- Then the prediction for the trajectory $X_i(t)$ is given by

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

Theoretical results for mean curve and covariance function

- Under some regularity conditions,

$$\sup_T |\hat{\mu}(t) - \mu(t)| = O_p(\frac{1}{\sqrt{n}h_\mu});$$

and

$$\sup_{T\times T} |\hat{G}(s,t) - G(s,t)| = O_p(\frac{1}{\sqrt{n}h_G^2}).$$

- Under some regularity conditions, for any k

$$|\hat{\lambda}_k - \lambda_k| = O_p(\frac{1}{\sqrt{n}h_G^2});$$

and

$$\sup_T |\hat{\phi}_k(t) - \phi_k(t)| = O_p(\frac{1}{\sqrt{n}h_G^2}).$$

Theoretical results for FPC scores

- Under some regularity conditions,

$$\lim_{n\to\infty} \hat{\xi}_{ik} = \xi_{ik}, \text{ in probability}$$

and

$$\lim_{K\to\infty} \lim_{n\to\infty} \hat{X}_i^K(t) = X_i(t), \text{ in probability}$$

Theoretical results for asymptotic inference

Define $\phi_{k,t} = (\phi_1(t), \dots, \phi_K(t))$, $H = \text{cov}(\xi_{k,i}, Y_i)$,
$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$, $\Omega_K = \Lambda - H\Sigma_{Y_i}^{-1}H^T$, $\omega(s,t) = \phi_{K,s}^T \Omega_K \phi_{K,t}$.

- Under some regularity conditions,

$$\lim_{K\to\infty} \lim_{n\to\infty} P\{\frac{\hat{X}_i^K(t) - X_i(t)}{\sqrt{\hat{\omega}(t,t)}} \le x\} = \Phi(x);$$

and

$$\lim_{n\to\infty} P\{\sup_{t\in T} \frac{|\hat{X}_i^K(t) - X_i(t)|}{\sqrt{\hat{\omega}(t,t)}} \le \sqrt{\chi^2_{K,1-\alpha}}\} \ge 1-\alpha.$$

where $\Phi(x)$ is the standard Gaussian cdf and $\chi^2_{K,1-\alpha}$ is the $1-\alpha$ percentile of the chi-squared distritbuion with K degrees of freedom.

Asymptotic confidence bands for individual trajectory

The theoretical results from the previous slide lead to constructions of point-wise and simultaneous confidence intervals for trajectory $X_i(t)$.

- An asymptotic point-wise $(1-\alpha)$ confidence interval for $X_i(t)$ is given by
$$\hat{X}_i^K(t) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\omega}(t,t)}.$$
- An asymptotic simultaneous $(1-\alpha)$ band for $X_i(t)$ is given by
$$\hat{X}_i^K(t) \pm \sqrt{\chi^2_{K,1-\alpha}\hat{\omega}(t,t)}.$$