Theory of functional principal components analysis for noisy and discretely observed data

Hang Zhou [1]

Department of Statistics
University of California, Davis

[1] jointly with Fang Yao and Dongyi Wei

Functional Data: Smoothness and Infinite-dimensionality

- A sample of subjects or experimental unit: one or more functions $X(t), t \in \mathcal{T}$, for each subject. WLOG, let $\mathcal{T} = [0, 1]$.
- Infinite dimensionality and smoothness.
- **Discretely observed (realistic):** *measurements are taken at discrete time points with noise:* $X_{ij} = X_i(t_{ij}) + \varepsilon_{ij}, i = 1, \ldots, n; j = 1 \ldots, N$.
- Q: How the discretized observations affect the estimation?
- Phase transition for mean and covariance : $N \gtrsim n^{1/4}$ (Pooling)

Q: How the discretization and noise contamination affect the dimension reduction via FPCA? (eigenfunction and eigenvalue with diverging index)

$$C(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t), \quad s, t \in \mathcal{T}$$

Theory for Eigen-Component with Diverging Index

- The compact covariance operator $C(f) = \int_0^1 C(s,t)f(s)ds$ is non-invertible.
- Regularization is needed in models involving inverse problem (FLM, gFLM, fCOX). Eigen-component used in constructing the regression estimator should tend to infinity as sample size grows.
- Highly depend on the convergence rate of a diverging number (with n) of eigenfunction estimates.
- **Contribution:** $\mathcal{L}^2$, uniform convergence of eigenfunction/asymptotic normality of eigenvalues with diverging index.

$$\mathbb{E}\|\hat{\phi}_k - \phi_k\|^2 \lesssim \frac{k^2}{n}\left\{1 + \left(\frac{k^a}{N}\right)^2\right\} + \frac{k^a}{nNh}\left(1 + \frac{k^a}{N}\right) + h^4 k^{2c+2}.$$