Adaptive optimal estimation of irregular mean and covariance functions

Steven Golovkine[1,2], Nicolas Klutchnikoff[3], Valentin Patilea[2]

[1]Groupe Renault
[2]CREST, Ensai
[3]IRMAR, Université Rennes 2

CFE-CMStatistics
Virtual Conference, 18-20 December 2021

FDA challenges and opportunities

- Ideally, one observes several (sometimes many) curves/images/signals
- Ideally, data represent the continuous time measurements of sample paths from a same stochastic process
- A challenge is then to find the suitable parsimonious representations of the data
  - the 'optimal' representations are usually determined by the purpose of the analysis

FDA challenges and opportunities

- Real data are
  - discretely observed, possibly at random points, which may be sparsely distributed
  - noisy measurements
- Another challenge is then to recover the curves in a suitable way
  - the quality of the recovery will influence the quality of the subsequent inference/prediction methods
  - the meaning of 'optimal recovery' depends on the application!
- Solutions come from the replication structure of the functional data
  - one observes several (sometimes many) curves/images/signals
  - data are generated by sample paths of a same stochastic process

Sample path regularity – a key concept

- Data: (randomly) discrete time noisy measurements of the curves
- Recover the curves
  - A nonparametric regression problem where optimality depends on the curve regularity; see, e.g., Tsybakov (2009)
- Estimate the mean and the covariance
  - The optimal rates depend on the curve regularity; see, e.g., Cai and Yuan (2010, 2011, 2016)
- Fit usual predictive models (linear...)
  - The optimal convergence rates depend on the predictor curve regularity; see, e.g., Hall and Horowitz (2007)

Mean and covariance functions

- Let X be a random function taking values in $L^2(\mathcal{T})$
- Assume $\mathbb{E}(\|X\|) < \infty$. The mean function (curve) is

$$\forall t \in \mathcal{T}, \qquad \mu(t) = \mathbb{E}[X(t)]$$

- Assume $\mathbb{E}(\|X\|^2) < \infty$. The covariance function: $\forall t, s \in \mathcal{T}$,

$$\Gamma(t,s) = \mathbb{E}\left[\{X(t) - \mu(t)\}\{X(s) - \mu(s)\}\right]$$

- The covariance operator is the continuous linear operator $C : L^2(\mathcal{T}) \to L^2(\mathcal{T})$ defined by

$$C(y)(t) = \int_{\mathcal{T}} \Gamma(t,s) y(s) ds$$

Eigenvalues decrease rate

- In FDA literature, the sample paths regularity is usually hidden in the decrease rate for the eigenvalues of the covariance operator
- Usual condition

$$\lambda_j \sim j^{-\nu}, \qquad j \geq 1,$$

  for some $\nu > 1$
- The value of ν is usually supposed given!

Observed data

- Let $X^{(1)}, \ldots, X^{(N)}$ be an independent sample of a random process $X = (X(t) : t \in \mathcal{T})$ with continuous trajectories
- For each $1 \leq n \leq N$
  - M_n is a random positive integer
  - $T_m^{(n)} \in \mathcal{T}, 1 \leq m \leq M_n$ be the (random) observation times, design points, for the curve $X^{(n)}$
- The observations are $(Y_m^{(n)}, T_m^{(n)})$, $1 \leq m \leq M_n, 1 \leq n \leq N$, where
  - $$Y_m^{(n)} = X^{(n)}(T_m^{(n)}) + \sigma(T_m^{(n)}, X^{(n)}(T_m^{(n)}))e_m^{(n)}$$
  - $e_m^{(n)}$ are independent copies of a standardized error term

Aims

- We aim to
  - estimate $X^{(n)}(t)$ for an arbitrary point $t \in \mathcal{T}$ and for each n;
  - calculate mean curve estimate $\widehat{\mu}(t)$, $t \in \mathcal{T}$;
  - calculate covariance function estimate $\widehat{\Gamma}(\cdot,\cdot)$, $s, t \in \mathcal{T}$;

Local regularity

- Let $\mathcal{O}_\star$ be a neighborhood of t
- For $H_t \in (0,1)$, and $L_t > 0$, assume that the stochastic process X satisfies the condition:

$$\mathbb{E}(X_u - X_v)^2 \asymp L_t^2 |v-u|^{2H_t}, \quad \text{whenever } u \text{ and } v \text{ belong to } \mathcal{O}_\star$$

- H_t is called *the local regularity of the process X* on $\mathcal{O}_\star$
- Our parameter H_t is related to Hurst exponent
- The definition extends to smoother sample paths using the derivatives of X_u and X_v instead

Local regularity vs. eigenvalue decrease rate

- The (local) regularity is related to the decrease rate of the eigenvalues of the covariance operator of the process
- Under some conditions, if

$$\lambda_j \sim j^{-\nu}, \qquad j \geq 1,$$

for some $\nu > 1$, then

$$2(H + \delta) = \nu - 1$$

when the sample paths admit derivative up to order δ and H is the regularity of the derivatives of order δ

Estimation

- For $s, t \in \mathcal{O}_\star$, let

$$\theta(s,t) = \mathbb{E}\left[(X_t - X_s)^2\right] \approx L^2 |t-s|^{2H_{t_0}}.$$

- Let t_1 and t_3 be such that $[t_1, t_3] \subset \mathcal{O}_\star$, and denote t_2 the middle point of $[t_1, t_3]$.
- A natural proxy of H_{t_0} is given by

$$\frac{\log(\theta(t_1,t_3)) - \log(\theta(t_1,t_2))}{2\log 2}, \quad \text{if } t_3 - t_1 \text{ is small.}$$

Estimation

- For $s, t \in \mathcal{O}_\star$, let

$$\theta(s,t) = \mathbb{E}\left[(X_t - X_s)^2\right] \approx L^2 |t-s|^{2H_{t_0}}.$$

- Let t_1 and t_3 be such that $[t_1, t_3] \subset \mathcal{O}_\star$, and denote t_2 the middle point of $[t_1, t_3]$.
- An estimator of H_{t_0} is given by

$$\frac{\log(\widehat{\theta}(t_1,t_3)) - \log(\widehat{\theta}(t_1,t_2))}{2\log 2}, \quad \text{if } t_3 - t_1 \text{ is small.}$$

where, given a nonparametric estimator $\widetilde{X_t}$ of X_t,

$$\widehat{\theta}(s,t) = \frac{1}{N}\sum_{n=1}^{N}\left(\widetilde{X_t^{(n)}} - \widetilde{X_s^{(n)}}\right)^2.$$

Unfeasible estimators

- If the realizations of the process were observed, the estimators of the mean and covariance functions would be

$$\widetilde{\mu}_N(t) = \frac{1}{N}\sum_{n=1}^{N} X^{(n)}(t), \quad t \in \mathcal{T},$$

$$\widetilde{\Gamma}_N(s,t) = \frac{1}{N-1}\sum_{n=1}^{N}\left(X^{(n)}(s) - \widetilde{\mu}_N(s)\right)(X^{(n)}(t) - \widetilde{\mu}_N(t)), \quad s,t \in \mathcal{T}.$$

Estimator of the mean function

- For any $t \in \mathcal{T}$, let
$$W_N(t, h) = \sum_{n=1}^{N} w_n(t, h).$$
- The estimator of the mean function is
$$\widehat{\mu}_N(t, h) = \frac{1}{W_N(t, h)} \sum_{n=1}^{N} w_n(t, h) \widehat{X}^{(n)}(t), \quad t \in \mathcal{T}.$$
- An adaptive optimal bandwidth is
$$\widehat{h}_\mu^\star = C_\mu (N\mathfrak{m})^{-1/(1+2\widehat{H}_t)}.$$
- Note $\widehat{\mu}^\star$ the estimation of the mean using the bandwidth $\widehat{h}_\mu^\star$.

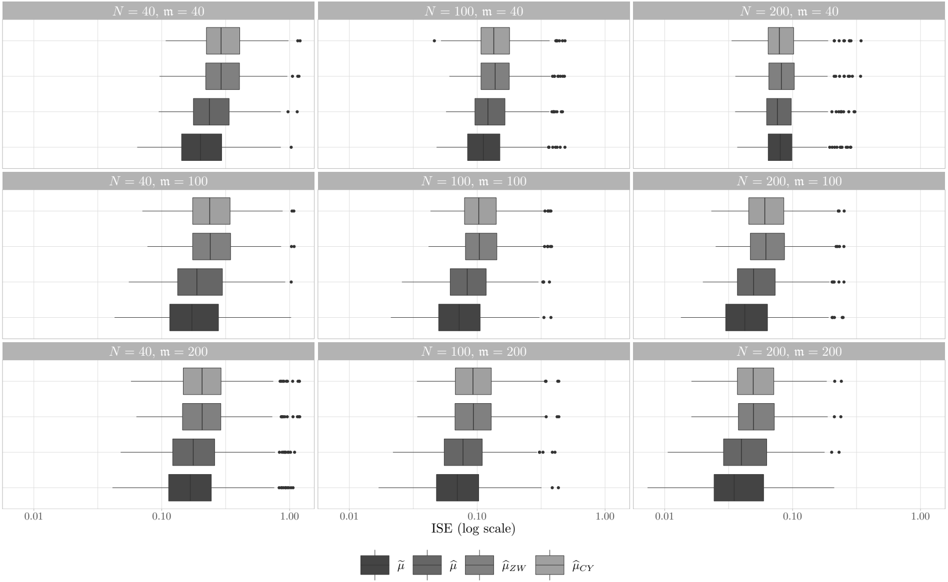

Figure 1: ISE with respect to the true mean function μ

Estimator of the covariance function

- For any $s \neq t$, let

$$W_N(s,t,h) = \sum_{n=1}^{N} w_n(s,h) w_n(t,h).$$

- The estimator of the covariance function is, for $|s-t| > \delta$,

$$\widehat{\Gamma}_N(s,t,h) = \frac{1}{W_N(s,t,h)} \sum_{n=1}^{N} w_n(s,h)\widehat{X}^{(n)}(s) w_n(t,h)\widehat{X}^{(n)}(t) - \widehat{\mu}^\star(s)\widehat{\mu}^\star(t).$$

- An adaptive optimal bandwidth is

$$\widehat{h}_\Gamma^\star = C_\Gamma (N\mathfrak{m})^{-1/(1+2\min(\widehat{H}_s, \widehat{H}_t))}.$$

Estimator of the diagonal band of the covariance function

- The previous estimator can be used only outside the diagonal set.
- The covariance function estimator of the diagonal band, for $|s-t| \leq \delta$ is defined as

$$\widehat{\Gamma}_N(s,t,h) = \frac{1}{W_N(u,h)} \sum_{n=1}^{N} w_n(u,h)\widehat{(X^{(n)})^2}(u) - \widehat{\mu}_N^{\star 2}(u,h), \quad u = (s+t)/2$$

$$= \frac{1}{W_N(u,h)} \sum_{n=1}^{N} w_n(u,h)\{Y_m^{(n)}\}^2 W_m^{(n)}(u,h) - \widehat{\mathbb{E}}\left[\sigma^2(u,X_u)\right] - \widehat{\mu}_N^{\star 2}(u)$$

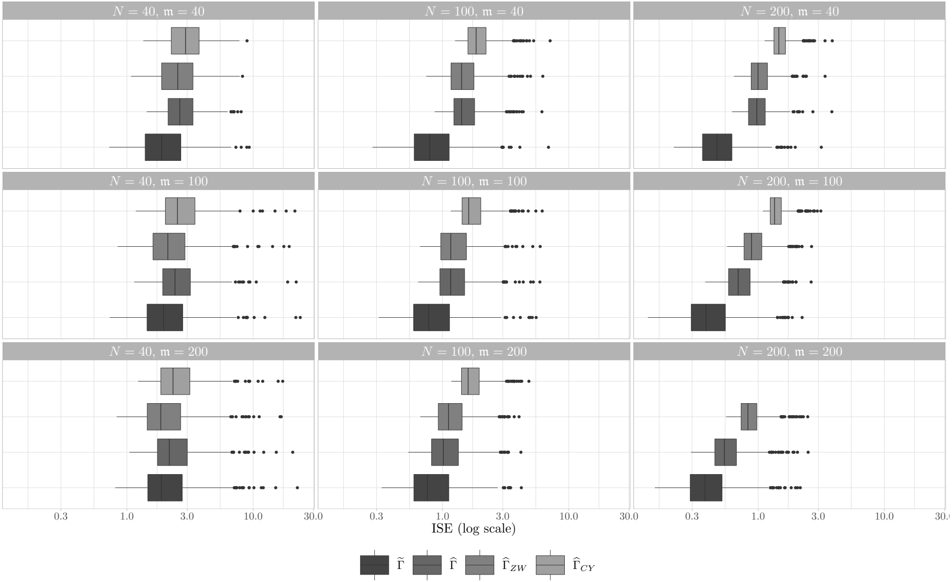

Figure 2: ISE with respect to the true covariance function Γ

Take away ideas

- The available data in FDA are usually noisy measurements at a discrete, possibly random design points
- The usual FDA methods require the reconstruction of the curves
- The optimal curve recovery depends on the purpose, but in most cases depends on the regularity of the sample paths
- We formalize the concept of local regularity of the process, propose a first simple estimator for it and mean and covariance function.
- A preprint of the paper for the estimation of the regularity is available at

  `https://arxiv.org/abs/2009.03652`
- A preprint of the paper for the estimation of the mean and covariance functions is available at

  `https://arxiv.org/abs/2108.06507`

References I

Cai, T. T. and Yuan, M. (2010). Nonparametric Covariance Function Estimation for Functional and Longitudinal Data. *University of Pennsylvania and Georgia institute of technology*, page 36.

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Annals of Statistics*, 39(5):2330–2355.

Cai, T. T. and Yuan, M. (2016). Minimax and Adaptive Estimation of Covariance Operator for Random Variables Observed on a Lattice Graph. *Journal of the American Statistical Association*, 111(513):253–265.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, New York, NY.