A covariance operator estimator with dense functional data

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

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

Journées de Statistique
Nancy, June 2019

Functional Data Analysis

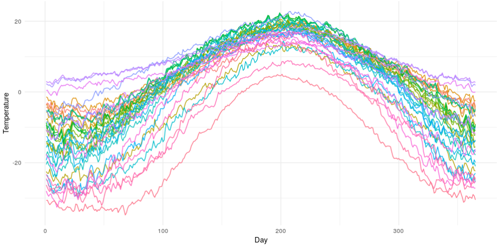

Figure 1: Mean daily temperature for the Canadian weather stations (Ramsay and Silverman (2005)).

Examples

- Spectroscopy;
- Sounds recognition;
- Electroencephalography comparison;
- Various sensors.

Model

- We are interested by independent realizations of the stochastic process

$$X = \{X(t) : t \in [0, 1]\}$$

taking values in $L^2[0, 1]$.

- Usually the process is decomposed

$$X(t) = \mu(t) + U(t), \quad t \in [0, 1].$$

where

  - $X(t) \in \mathbb{R}$, $\forall t$;
  - $\mu(t) = \mathbb{E}(X(t))$, $\forall t$;
  - $U(\cdot)$ represents the stochastic part of $X(\cdot)$ and $\mathbb{E}(U_n(\cdot)) = 0$.

Mean and covariance

- For any $(s, t) \in [0, 1]^2$, the covariance function is defined by
$$\phi(s, t) = \mathbb{E}\left(\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}\right).$$
- We aim to estimate the mean $\mu(\cdot)$ and the covariance $\phi(\cdot, \cdot)$ functions for the different regimes of functional data.

The data

- Let $X_n, n \in \{1, \ldots, N\}$ be independent trajectories of X.
- In practice, such trajectories cannot be observed at any t.
- Moreover, only noisy data are available;
  - the observed values on the trajectory $X_n(\cdot)$ are contaminated with additive errors.
- For any $1 \leq n \leq N$, we observe $M_n \geq 2$ random pairs (T_{ni}, Y_{ni}) which are defined as:

$$Y_{ni} = X_n(T_{ni}) + \sigma(X_n(T_{ni}))\epsilon_{ni}, \quad i = 1, \ldots, M_n$$

  where
  - $(T_{n1}, \ldots, T_{nM_n})$ are i.i.d. random sampling points in $[0, 1]$;
  - ϵ_{ni} are i.i.d. random errors;
  - $\sigma^2(\cdot)$ is an unknown conditional variance function.

The different sampling regimes and our aims

- If the realizations of X are observed without error, the mean $\mu(\cdot)$ and the covariance $\phi(\cdot, \cdot)$ could be estimated at rate $N^{-1/2}$.
- When the trajectories $X_n(\cdot)$ are noisy, different rates of convergence are expected, depending on the relative orders of M_n and N.
- Question: When it will still be possible to achieve the rate $N^{-1/2}$ for estimating $\mu(\cdot)$ and $\phi(\cdot, \cdot)$?

- Zhang and Wang (2016) proposed several regimes of functional data, depending on the answer to the question.
- The $N^{-1/2}-$convergence rate cannot be achieved for the estimation of $\mu(\cdot)$ and $\phi(\cdot,\cdot)$:
  - sparse – typically the M_n are bounded;
  - non-dense – M_n tends to infinity but not fast enough.
- The $N^{-1/2}-$convergence rate can be achieved:
  - semi-dense – a suitable choice of the smoothing parameter is needed to make the asymptotic bias negligible;
  - ultra-dense – the asymptotic bias is negligible.
- We mainly focus on the semi and ultra-dense situations.

Smoothing first, then estimation

- Zhang and Chen (2007) consider dense functional data. They smooth the individual curves first, and then estimates the mean and covariance.
- Curves' smoothing, for example, by kernel smoothing:

$$\widehat{X}_n(t) = \frac{1}{M_n} \sum_{i=1}^{M_n} Y_{ni} K_h(T_{ni} - t).$$

  where
  - $h > 0$ is the bandwidth;
  - $K_h(\cdot) = K(\cdot/h)/h$ with $K : \mathbb{R} \to \mathbb{R}$ the kernel.
- Then, the estimation of the mean function is

$$\widehat{\mu}(t) = \frac{1}{N} \sum_{n=1}^{N} \widehat{X}_n(t).$$

- The estimator of the covariance function is

$$\widehat{\phi}(s,t) = \frac{1}{N-1} \sum_{n=1}^{N} \left(\widehat{X}_n(s) - \widehat{\mu}(s)\right) \left(\widehat{X}_n(t) - \widehat{\mu}(t)\right).$$

- Zhang and Wang (2016) consider weighted local linear estimators.
- The weights are defined according to the sampling regime (sparse or dense).
- The local constant version of their estimators are

$$\widehat{\mu}(t) = \sum_{n=1}^{N} w_n \sum_{i=1}^{M_n} Y_{ni} K_h(T_{ni} - t).$$

and

$$\widehat{\phi}(s,t) = \sum_{n=1}^{N} v_n \sum_{1 \le k \ne l \le M_n} \left(Y_{nk} - \widehat{\mu}(T_{nk})\right)\left(Y_{nl} - \widehat{\mu}(T_{nl})\right) \times K_h(T_{nk} - s) K_h(T_{nl} - t).$$

- The different weighting scheme are:
  - $w_n = 1/\sum_{n=1}^{N} M_n$ and $v_n = 1/\sum_{n=1}^{N} M_n(M_n - 1)$ (OBS);
  - $w_n = 1/NM_n$ and $v_n = 1/NM_n(M_n - 1)$ (SUBJ, Li and Hsing (2010)).

- The estimator of Zhang and Wang (2016) bridges the gap between the different sampling regimes using suitable weights w_n and v_n.
- Zhang and Wang (2016) provide asymptotic theory conditional of the sequence M_n; they assume a given regularity for the mean and covariance functions.
- When the sampling regime is such that M_n are drawn from a same law, the OBS et SUBJ estimators are essentially asymptotically equivalent.

Bridging the gap: a new approach

- We aim proposing an estimator of the mean $\mu(\cdot)$ and the covariance functions $\phi(\cdot, \cdot)$ that adapts for
  - the type of sampling regime;
  - the regularity of the target functions.
- For simplicity, here we only consider the case where
  - $\sigma(\cdot) \equiv \sigma^2$;
  - M_n are i.i.d.
- Here we focus on the covariance function.
- Starting from the proposal of Zhang and Wang (2016), we consider the following extensions:
  - for each n separately, smooth the value $Y_{n1}, \ldots, Y_{nM_n}$;
  - consider a leave-one-out version of the mean in the definition of the covariance estimator.

The new estimator of the covariance

- Define the Leave-One-Out Kernel Smoothing curve by

$$\widehat{X}_n^{(k,l)}(t) = \frac{1}{M_n - 2} \sum_{1 \le i \le M_n, i \notin \{k,l\}} Y_{ni} K_b(T_{ni} - t).$$

- Define the Leave-One-Out Kernel Smoothing mean curve

$$\overline{X}_N^{(n)}(t) = \frac{1}{N-1} \sum_{1 \le m \le N, m \neq n} \frac{1}{M_m} \sum_{1 \le i \le M_m} Y_{mi} K_b(T_{mi} - t).$$

- Then, we follow the construction of Zhang and Wang (2016) and replace
  - Y_{nk} (resp. Y_{nl}) by $\widehat{X}_n^{(k,l)}(T_{nk})$ (resp. $\widehat{X}_n^{(k,l)}(T_{nk})$);
  - $\widehat{\mu}(t)$ (resp. $\widehat{\mu}(s)$) by $\overline{X}_N^{(n)}(T_{nk})$ (resp. $\overline{X}_N^{(n)}(T_{nl})$).

The expression of the new estimator

- Let $M_n^{\otimes p} = M_n(M_n - 1) \cdots (M_n - p + 1)$.
- When $s \neq t$, we define $\widehat{\phi}(s, t)$ as

$$\widehat{\phi}(s,t) = \frac{1}{N} \sum_{1 \le n \le N} \frac{1}{M_n^{\otimes 4}} \sum_{1 \le k \neq l \neq i \neq j \le M_n} \left\{ Y_{ni} K_b(T_{ni} - T_{nk}) - \overline{X}_N^{(n)}(T_{nk}) \right\}$$
$$\times \left\{ Y_{nj} K_b(T_{nj} - T_{nl}) - \overline{X}_N^{(n)}(T_{nl}) \right\}$$
$$\times K_h(T_{nk} - s) K_h(T_{nl} - t).$$

The bias

- When all $M_n \approx M_N \to \infty$, the bias term has the rate

$$O_{\mathbb{P}}(b^{s_X} + h^{s_\phi}) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right) \times O_{\mathbb{P}}\left(\frac{\log N}{\sqrt{Nb^2}} + \frac{\log M_N}{\sqrt{M_N b^2}}\right).$$

- s_X is related to the regularity of the trajectories $X_n(\cdot)$ and is determined by the regularity of the mean $\mu(\cdot)$ and the covariance $\phi(\cdot,\cdot)$.
- s_ϕ is the regularity of the covariance function.
- To achieve parametric rates, one necessarily needs $M_N b^2/\log^2 M_N \to \infty$ and $Nb^2/\log^2 N \to \infty$.

The variance

- When all $M_n \approx M_N \to \infty$ and $\log(M_N) << N$, the variance term has the rate

$$\frac{1}{N} \times O_{\mathbb{P}}\left(1 + \frac{1}{M_N h}\frac{\log^2 N}{Nb^2} + \frac{1}{M_N^2 h^2}\frac{\log^2 M_N}{Nb^2}\right).$$

- Without smoothing the curves, Zhang and Wang (2016) considered all the regularities equal to 2, and thus they have a bias term of order h^2, and they obtained the variance

$$\frac{1}{N} \times O_{\mathbb{P}}\left(1 + \frac{1}{M_N h} + \frac{1}{M_N^2 h^2}\right).$$

- To improve over the variance of Zhang and Wang (2016) we need

$$Nb^2/\{\log^2 N + \log^2 M_N\} \to \infty.$$

Simulation

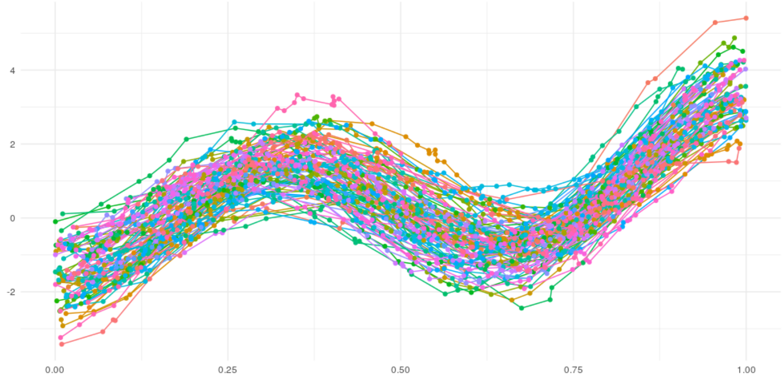

Figure 2: Simulation from Zhang and Wang (2016)

Simulation

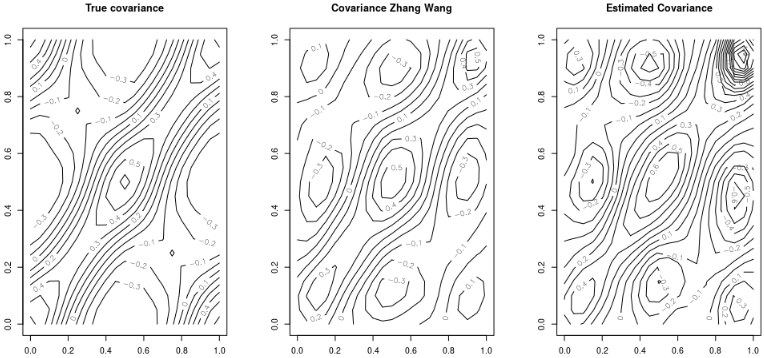

Figure 3: Comparison between true and estimated covariances

Thank you for your attention!
Questions?

References I

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321.