Robust functional principal components for sparse longitudinal data

Matías Salibián-Barrera
Graciela Boente

Department of Statistics, University of British Columbia
Department of Mathematics, Universidad de Buenos Aires

ICSA, Banff, July 2022

While I still have your attention

- FPCA for sparsely observed functional data
- Compares favourably with current proposals
- Naturally robustified

“Robust functional principal components for sparse longitudinal data” (2021) Boente & SB, *Metron*

`https://github.com/msalibian/sparseFPCA`

Functional Data Analysis and PCA

- Observations are functions

$$X : (\Omega, \mathcal{A}, P) \longrightarrow L^2(\mathcal{I})$$

- Best lower dimensional representation

  Describe the process

  Reconstruct curves

Functional PCA

- Covariance function

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

- Covariance operator – self adjoint

$$(\mathbf{\Gamma} f)(a) = \int G(a,s) f(s)\, ds\,, \; \langle f\,, \mathbf{\Gamma} g\,\rangle = \langle \mathbf{\Gamma} f\,, g\,\rangle$$

- Eigenfunctions & eigenvalues

$$(\mathbf{\Gamma}\phi_k)(s) = \lambda_k\, \phi_k(s)$$

Functional PCA

- Orthonormal basis (Mercer theorem)

$$G(s,t) = \sum_{j\geq 1} \lambda_j \, \phi_j(s) \, \phi_j(t) \qquad \lambda_\ell \geq \lambda_{\ell+1}$$

$$\langle \phi_j, \phi_\ell \rangle = \int \phi_j(t) \phi_\ell(t) \, dt = \delta_{j\ell}$$

Functional PCA

- Karhunen-Loève expansion

$$X_i(s) \;=\; \mu(s) + \sum_{j \geq 1} \lambda_j^{1/2}\, \xi_{ij}\, \phi_j(s)$$

$$\xi_{ij} \;=\; \langle X_i - \mu\,,\phi_j\,\rangle\,, \quad E\,(\xi_{ij}) = 0$$

$$\mathrm{cov}\,(\xi_{ij}\,,\xi_{i\ell}) = \delta_{j\ell}$$

Functional PCA

- With estimated eigenfunctions, eigenvalues and scores:

$$\widehat{X}_i = \widehat{\mu} + \sum_{j=1}^{q} \hat{\lambda}_j^{1/2} \, \widehat{\xi}_{ij} \, \widehat{\phi}_j \,, \quad 1 \leq i \leq n$$

  "predictions" / reconstructed curves

The "sparse" case

- Observations:

$$X_i(t_{ij})\,, \quad j = 1, \ldots, n_i\,, \quad i = 1, \ldots, n$$

- t_{ij}'s and n_i's are iid
- n_i's can be small (e.g. 2, 3, 4, etc.)

The “sparse” case

Two main approaches:

- Assume a finite rank process

  James, Hastie, Sugar (2000); Gervini (2008)

- Estimate covariance function

  Staniswallis & Lee (1998); Yao, Müller & Wang (2005)

PACE

- Since

$$\begin{aligned} G(s,t) &= \operatorname{cov}(X(s), X(t)) \\ &= E\Big[(X(s)-\mu(s))\,(X(t)-\mu(t))\Big] \end{aligned}$$

- $\hat{G} \longleftarrow$ smoothed raw "covariances"

$$G_{ij\ell} = (X_i(t_{ij}) - \hat{\mu}(t_{ij}))\,(X_i(t_{i\ell}) - \hat{\mu}(t_{i\ell}))$$

Staniswallis & Lee (1998); Yao, Müller & Wang, (2005)

PACE

- Compute $\hat{\lambda}_k$, $\hat{\phi}_k$ from $\hat{G}(s, t)$
- Best (linear) scores predictor:

$$\hat{\xi}_{ik} = \hat{\lambda}_k^{1/2} \, \hat{\phi}_k' \, \hat{\Sigma}_i^{-1} \, (\mathbf{X}_i - \hat{\boldsymbol{\mu}}(\mathbf{t}_i))$$

Yao, Müller & Wang, (2005)

$$\widehat{X}_i = \widehat{\mu} + \sum_{j=1}^{q} \hat{\lambda}_j^{1/2} \, \widehat{\xi}_{ij} \, \widehat{\phi}_j \, , \quad 1 \le i \le n$$

Robust PACE?

- Can we use a robust smoother?
- Robust estimator for expected values?
- Will not work

Robust FPCA

- Eigenfunctions of the covariance function (plug-in, spherical PCA)

  Locantore et al 1999, Gervini (2008); Kraus & Panaretos (2012); Sawant, Billor, Shin (2012)

  (cannot accommodate sparsely observed curves)

Robust FPCA

- “Directions” of maximal variability (dispersion)

  Hyndman & Ullah (2007); Projection-pursuit (Bali, Boente, Tyler and Wang, 2011)

- “Best” lower-dimensional approximations

  Boente, SB (2015); Lee, Shin, Billor (2013)

  (cannot accommodate sparsely observed curves)

Robust FPCA

Conceptual framework

- Elliptical random elements
  - Heavy tails, no second moments
  - Karhunen-Loève type
  - Best lower dim. approximation

Boente, SB, Tyler (2014)

Elliptical random elements

For a random element on $\mathcal{H}$

$$X : (\Omega, \mathcal{A}, P) \longrightarrow \mathcal{H}$$

if $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$ is a self-adjoint, positive semi-definite and compact operator, we say:

$$X \sim \mathcal{E}(\boldsymbol{\mu}, \mathbf{\Gamma}, \varphi)$$

if for all $\mathbf{A} : \mathcal{H} \to \mathbb{R}^d$ bounded & linear

$$\mathbf{A}X \sim \mathcal{E}_d(\mathbf{A}\,\boldsymbol{\mu}, \mathbf{A\Gamma A}^*, \varphi)$$

Elliptical random elements

Important result: if X is an elliptical element on $L^2(\mathbf{I})$, then

$$\mathbf{X} = \begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_k) \end{pmatrix}$$

has an ellipitical distribution on $\mathbb{R}^k$

Elliptical random elements

If X has an elliptical distribution on $L^2(\mathbf{I})$

$$X(t_1) \,\big|\, X(t_2) \;\sim\; \mathcal{E}_1\left(\mu_{1|2},\, \sigma_{1|2}\right)$$

$$\mu_{1|2} = \boldsymbol{\mu}(t_1) + \mathbf{G}(t_1, t_2)\, \left(X(t_2) - \boldsymbol{\mu}(t_2)\right) / \mathbf{G}(t_2, t_2)$$

$$X(t_1)|X(t_2) = \boldsymbol{\mu}(t_1) + \beta_{1|2}\, \left(X(t_2) - \boldsymbol{\mu}(t_2)\right) + \varepsilon_{1|2}$$

Robust sparse FPCA

A local M-estimator for μ:

$$\hat{\boldsymbol{\mu}}(t) = \min_{a \in \mathbb{R}}{}^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \rho\left(\frac{X_i(t_{ij}) - a}{\hat{\sigma}(t)}\right) \mathcal{K}\left(\frac{t_{ij} - t}{h}\right)$$

Robust sparse FPCA

Estimate $G(t, t)$: local M-scale

$$\sum_{i=1}^{N} \sum_{j=1}^{n_i} \chi\left(\frac{X_i(t_{ij}) - \hat{\mu}(t_{ij})}{\hat{G}(t, t)} \right) w_{ij}(t) = b$$

where

$$w_{ij}(t) = \mathcal{K}\left(\frac{t_{ij} - t}{h} \right) \Big/ \sum_k \sum_\ell \mathcal{K}\left(\frac{t_{k\ell} - t}{h} \right)$$

Robust sparse FPCA

Estimate the off-diagonal elements:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\overset{-1}{\min}} \sum_{i=1}^{N} \sum_{j \neq \ell} \rho \left(\frac{\tilde{X}_i(t_{ij}) - \boldsymbol{\beta} \tilde{X}_i(t_{i\ell})}{\hat{\sigma}(t_0, s_0)} \right) \mathcal{K}_{ij}(t_0) \, \mathcal{K}_{i\ell}(s_0)$$

$$\tilde{G}(t_0, s_0) = \hat{\boldsymbol{\beta}} \, \hat{G}(s_0, s_0)$$

and then

$$\hat{G}(t_0, s_0) = \left(\tilde{G}(t_0, s_0) + \tilde{G}(s_0, t_0) \right) \Big/ 2$$

Robust sparse FPCA

- Bandwidth selection
  - K-fold robust CV
    - “Robust prediction error”

- Number of FPC’s K

  - Proportion of total eigenvalues

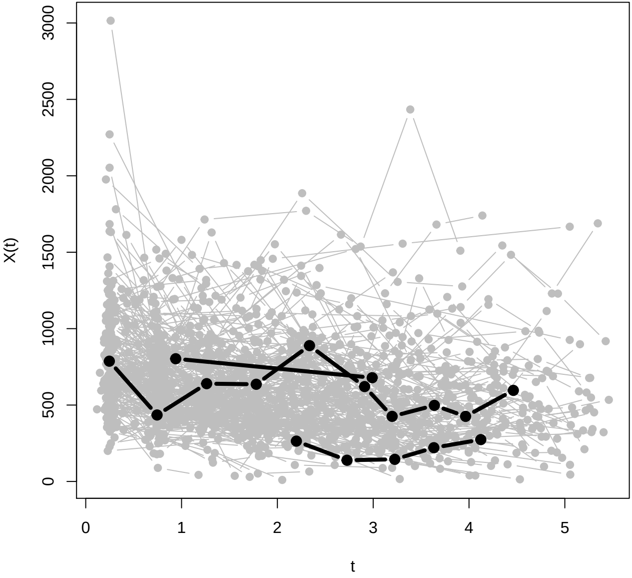
Need not be robust

- This approach can be used with any $\hat{\mu}$, $\hat{G}(t, t)$, and local regression estimator $\hat{\beta}$
- When we use the "least squares" version, and no atypical observations are present, this approach works well

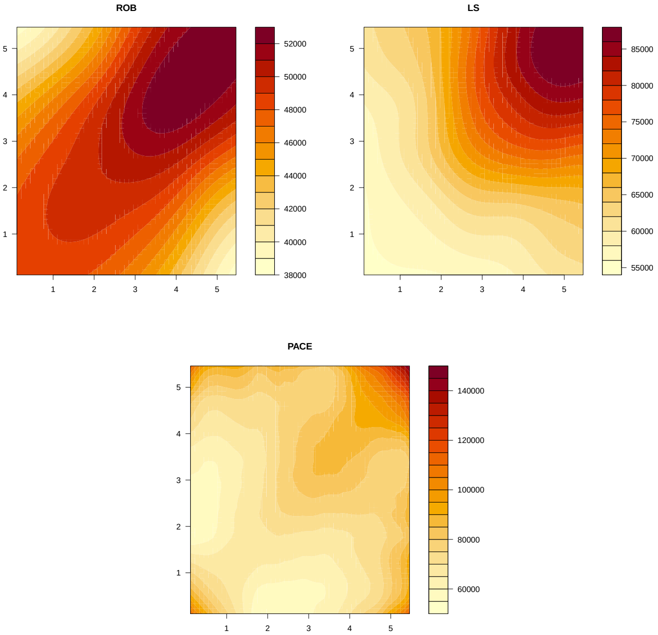
Example - CD4 counts

- Multicentre AIDS Cohort Study
- 2376 measurements
- 292 curves, observations after seroconversion
- Between 2 and 11 (median: 5) observations per curve

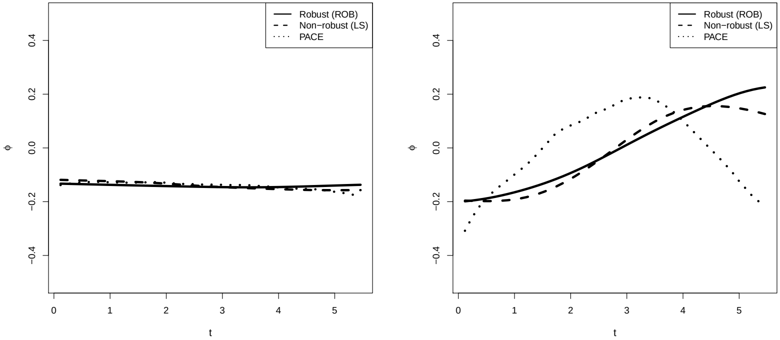
CD4 counts



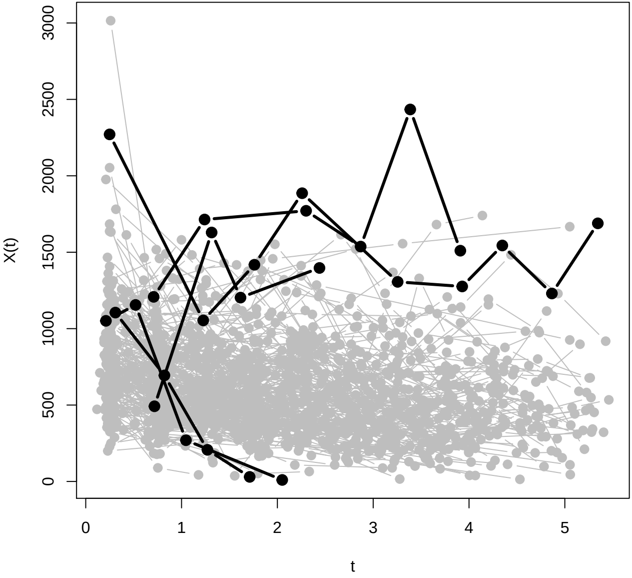
CD4 counts


ROB
LS
PACE
52000
50000
48000
46000
44000
42000
40000
38000
85000
80000
75000
70000
65000
60000
55000
140000
120000
100000
80000
60000
1
2
3
4
5

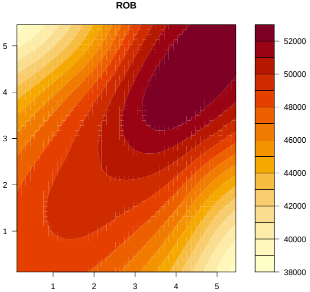
Eigenfunctions

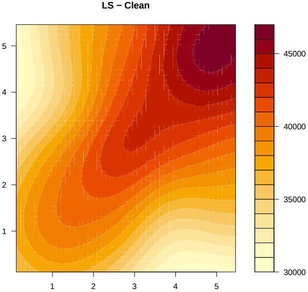


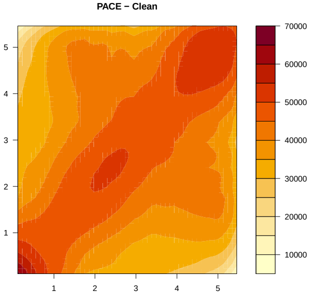
Most outlying



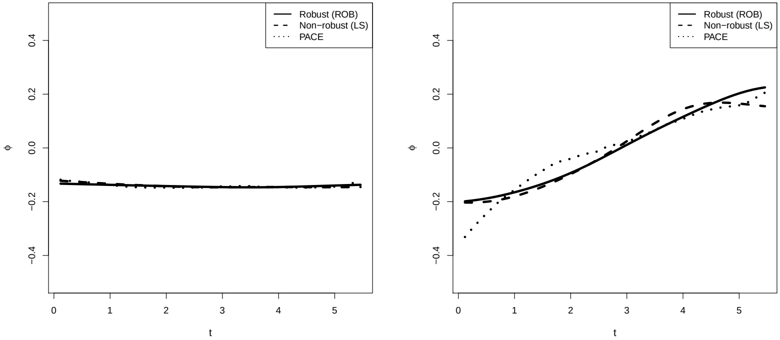
CD4 counts - clean







Eigenfunctions - clean



Conclusion, reproducibility

- FPCA for sparsely observed functional data
- Compares favourably with current proposals
- Naturally robustified

“Robust functional principal components for sparse longitudinal data” (2021) Boente & SB, *Metron*

`https://github.com/msalibian/sparseFPCA`