Discussion of "Functional Models for Time-varying Random Objects"

Shahin Tavakoli
(Proposer)

Warwick Statistics

October 23, 2019

Random Objects — Functional Data

Datapoints are

1. real, vector-valued, or Hilbert space valued curves
2. Covariance matrices/operators
3. measures/density functions
4. Trees

Abstract View

Fall into (roughly) 3 categories

1. Hilbert space (distance, directions, geometry)
2. Riemannian Manifold (distance, local Euclidean approximations $\Longrightarrow$ local directions, local geometry)
3. Metric space (only distance)

(local) geometry $\Longrightarrow$ we can do (local) moves in directions and (local) PCA

only distance $\Longrightarrow$ no notion of “moving into a direction”, **no PCA (?)**

Main contribution of Dubey & Müller

Data

Time-varying random objects, i.e. $X(t) \in \Omega, t \in [0, 1]$, where (Ω, d) is a *bounded* metric space.

Main Question

> *"We aim here at identifying dominant directions of variation [. . .] where random objects are indexed by time and in a general metric space"*

How to define functional (object) PCA (fPCA); i.e.

- $\rightarrow$ main modes of variations?
- $\rightarrow$ (real-valued) scores for pair plots?
- $\rightarrow$ estimation, consistency, examples

Key ideas

fPCA for $X(t) \in \mathbb{R}, t \in [0,1]$ is based on **covariance kernel** $c(t,s) = \mathrm{Cov}(X(t), X(s))$, so **covariance between random variables is key!**

Key observation (U, V random, U', V' iid copies)
If $U, V \in \mathbb{R}$ random, $d_E(U,V) = |U - V|$,

$$\begin{aligned}\mathbb{R} \ni \mathrm{Cov}(U,V) &= \mathbb{E}\left[(U - \mathbb{E}\,U)(V - \mathbb{E}\,V)\right] \\ &= \frac{1}{4}\mathbb{E}\left[d_E^2(U,V') + d_E^2(U',V) - 2d_E^2(U,V)\right].\end{aligned}$$

If $U, V \in \mathbb{R}^p$ random, $d_E(U,V) = \sqrt{(U-V)^\mathsf{T}(U-V)}$,

$$\begin{aligned}\mathbb{R} \ni \mathrm{Cov}_\Omega(U,V) &:= \frac{1}{4}\mathbb{E}\left[d_E^2(U,V') + d_E^2(U',V) - 2d_E^2(U,V)\right] \\ &= \mathbb{E}\left[(U - \mathbb{E}\,U)^\mathsf{T}(V - \mathbb{E}\,V)\right] \\ &\neq \mathrm{Cov}(U,V) \in \mathbb{R}^{p \times p}\end{aligned}$$

Key ideas

If $U, V \in H$ random element of Hilbert space, $d_H(U,V) = \|U - V\|$,

$$\mathbb{R} \ni \mathrm{Cov}^*(U,V) := \frac{1}{4}\mathbb{E}\left[d_H^2(U,V') + d_H^2(U',V) - 2d_H^2(U,V)\right].$$
$$= \mathbb{E}\left[\langle U - \mathbb{E}\,U, V - \mathbb{E}\,V\rangle\right]$$
$$\neq \mathrm{Cov}(U,V) \in S_1(H) \quad \text{(operators on } H\text{)}$$

If $U, V \in H$ random object in metric space (Ω, d)

$$\mathbb{R} \ni \mathrm{Cov}_\Omega(U,V) := \frac{1}{4}\mathbb{E}\left[d^2(U,V') + d^2(U',V) - 2d^2(U,V)\right].$$
$$\overset{?!}{=} \mathbb{E}\left[\langle U - \mathbb{E}\,U, V - \mathbb{E}\,V\rangle\right]$$

Great! Now we can define $c(s,t) := \mathrm{Cov}_\Omega(X(s), X(t))\ldots$ but positive definiteness is **not** guaranteed (hence no PCA. . .)

Metric spaces of Negative Types

Not all hope is lost: if (Ω, d) is of **negative type**, then there is a Hilbert space $(H, \langle . \rangle)$ and $f : \Omega \to H$ injective such that

$$d(U,V) = \langle f(U) - f(V), f(U) - f(V) \rangle \,.$$

(Sejdinovic et al. 2013). Now we have

$$\begin{aligned} \mathbb{R} \ni \mathrm{Cov}_\Omega(U,V) &:= \frac{1}{4}\mathbb{E}\left[d^2(U,V') + d^2(U',V) - 2d^2(U,V)\right]. \\ &= \mathbb{E}\left[\langle f(U) - \mathbb{E}\, f(U), f(V) - \mathbb{E}\, f(V)\rangle\right] \end{aligned}$$

The blue bit implies that

$$c(s,t) = \mathrm{Cov}_\Omega(X(s), X(t)),$$

the **metric auto-covariance operator** is now symmetric positive semi-definite, so we are in business!

Functional (Object) PCA

Now usual fPCA realm!

Assume metric auto-covariance kernel $c(s,t) = \mathrm{Cov}_\Omega(X(s), X(t))$ is **continuous** (typo in paper).

Mercer's Theorem:

$$c(s,t) = \sum_{k \geq 1} \lambda_k \phi_k(t) \phi_k(s),$$

uniform convergence (in t, s) and ϕ_ks are continuous.

Usually we then use the Karhunen–Loeve expansion

$$X(t) = \sum_{i \geq 1} \xi_k \phi_k(t); \quad \xi_k = \int X(t) \phi_k(t) dt$$

but not possible here!

Enters Fréchet

For usual functions $h(t) \in \mathbb{R}$, if $\int \phi(t)dt = 1$, we have

$$\int h(t)\phi(t)dt = \arg\inf_{\omega \in \mathbb{R}} \int d_E^2(\omega, h(t))\phi(t)dt.$$

Use this to define, for $S(t) \in \Omega$, the **generalized Fréchet integral**

$$\int_{\oplus} S(t)\phi(t)dt := \arg\inf_{\omega \in \Omega} \int d^2(\omega, S(t))\phi(t)dt$$

for $\int \phi = 1$.

$\rightarrow$ need uniqueness and existence of infimum, hence **completeness** of the metric space (typo in paper).

Object FPC ("scores" in metric space)

Define **object FPC**

$$\psi_{\oplus}^{k} := \int_{\oplus} X\phi_k^* \in \Omega,$$

where $\phi_k^* = \phi_k / \int \phi_k$.

$\implies$ summarise $t \mapsto X(t) \in \Omega$ by a few $\psi_{\oplus}^{k}, k = 1, 2, 3$ (say).
$\implies$ We can plot these! (each one is in metric space)

How about scalar scores?

Letting $\mu_\oplus(t) := \arg\min_{\omega\in\Omega} \mathbb{E}\left[d^2(\omega, X(t))\right]$, define the k-th (scalar) **Fréchet score** by

$$\beta_k := \int d(X(t), \mu_\oplus(t))\phi_k(t)dt \in \mathbb{R},$$

Now we can do pairs plots!

Remarks

We aim here at identifying dominant directions of variation. . . random objects are indexed by time and in a general metric space

So we wanted

1. General Metric Space
2. Dominant modes of variations (measuring co-variations and maximum variations)

Remarks (General Metric Space)

All examples considered have a structure richer than metric space:

1. densities
2. networks
3. covariances

If **only a metric** is available
$\implies \mathrm{Cov}_\Omega(s,t)$ and its empirical eigenfunctions can be estimated,
but Object PC or Fréchet scores will not be computable easily because

$$\mu_\oplus(t) := \arg\min_{\omega\in\Omega} \mathbb{E}\left[d^2(\omega, X(t))\right]$$

and

$$\psi_\oplus^k := \arg\inf_{\omega\in\Omega} \int d^2(\omega, X(t))\phi_k^*(t)dt$$

involve minimizations, but **no gradient descent available!**

Remarks (measuring co-variations)

A lot of information is lost with metric covariance

If $\Omega = H$ a separable Hilbert space,

$$\begin{aligned} \mathbb{R} \ni \mathrm{Cov}^*(U,V) &:= \frac{1}{4}\mathbb{E}\left[d_H^2(U,V') + d_H^2(U',V) - 2d_H^2(U,V)\right]. \\ &= \mathbb{E}\left[\langle U - \mathbb{E}\,U, V - \mathbb{E}\,V\rangle\right] \\ &= \mathrm{Tr}(\mathrm{Cov}(U,V)) \\ &\neq \mathrm{Cov}(U,V) \in S_1(H) \quad \text{(operators on } H\text{)} \end{aligned}$$

Trace as measure of magnitude is

good for **self-adjoint positive semi-definite operator**,

but **not for general operators**.

Example (weakness of metric covariance)

Let $U(t) \in \mathbb{R}, t \in [0, 1/2]$ with mean zero and $\mathrm{Cov}(U(t), U(t)) = 1$.
Define $X(t) \in \mathbb{R}^2, t \in [0, 1]$ by

$$X(t) = \begin{cases} \begin{pmatrix} U(t) \\ 0 \end{pmatrix}, & t \in [0, 1/2], \\ \\ \begin{pmatrix} 0 \\ U(t - 1/2) \end{pmatrix}, & t \in (1/2, 1]. \end{cases}$$

We have $\mathrm{Cov}_\Omega(X(t), X(t + 1/2)) = 0$, but the true covariance matrix is

$$\mathrm{Cov}(X(t), X(t + 1/2)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and is non-zero. Hence metric covariance fails to measure association here...

Remarks (Modes of Variation)

fPCA is based on decomposition of covariance kernel/operator, but strong motivation comes from **variance maximization**:

In Hilbert space, find φ such that $\mathrm{Var}(\langle \varphi, X\rangle)$ is maximized. If $X = X(t) \in H$ for each $t \in [0,1]$, $\varphi(t) \in H$ and

$$\lambda_1 = \max_{\|\varphi\|=1} \mathrm{Var}(\langle \varphi, X\rangle) = \int \left\langle \underbrace{\mathrm{Cov}(X(s), X(t))}_{\text{operator on } H} \varphi(t), \varphi(s) \right\rangle dsdt$$

In the paper,

$$\lambda_1 = \max_{\int \phi^2 = 1} \int c(s,t)\phi(s)\phi(t)dsdt = \int \phi(s)\phi(t)\,\mathrm{Tr}[\mathrm{Cov}(X(s), X(t))]dsdt$$

fPCA Interpretation is lost if the space is actually a Hilbert space...

Fréchet scores

In the real-valued case ($\Omega = \mathbb{R}$),

$$\begin{aligned} \beta_k &:= \int d(X(t), \mu_\oplus(t))\phi_k(t)dt \in \mathbb{R} \\ &= \int |X(t) - \mu(t)|\phi_k(t)dt \end{aligned}$$

This is **not the usual PC scores**, so interpretation is not straightforward...

Possible extensions

1. Instead of looking at $X(t), t \in [0,1]$, we could have $t \in E$. For instance $E =$ Great-Britain and $X(t)$ is a covariance matrix at location $t \in E$. (Tavakoli et al. 2019)
2. Somehow use the implicit injection into Hilbert space to work with full covariance:

$$\begin{aligned}\mathbb{R} \ni \mathrm{Cov}_\Omega(U,V) &:= \frac{1}{4}\mathbb{E}\left[d^2(U,V') + d^2(U',V) - 2d^2(U,V)\right].\\ &= \mathbb{E}\left[\langle f(U) - \mathbb{E}\, f(U), f(V) - \mathbb{E}\, f(V)\rangle\right]\\ &= \mathrm{Tr}\,\underbrace{\mathbb{E}\left[(f(U) - \mathbb{E}\, f(U)) \otimes (f(V) - \mathbb{E}\, f(V))\right]}_{\text{full covariance}}\end{aligned}$$

but not clear how to do since f is implicit. . .

Conclusions

Great thought provoking paper:

$\rightarrow$ moving beyond usual covariance (detects only linear dependencies),
$\rightarrow$ hence away from Gaussian
(Jim Ramsay's comments in SAMSI 2010, Diablerets 2016)

I propose the vote of thanks!