Statistical Inference for Functional Data

May 14, 2018

Outline

Linear operators - a generalization of matrices

- Let H be a linear (sub)space of square integrable functions. A linear and continuous function $\Psi : H \mapsto H$ is called a (linear) **operator**.
- Let $\mathcal{L}$ be the space of continuous linear operators on H.
- We note that if x is a square integrable function also $\Psi(x)$ (Ψx for shorthness) is a function.

Linear operators - a generalization of matrices

- Let H be a linear (sub)space of square integrable functions. A linear and continuous function $\Psi : H \mapsto H$ is called a (linear) **operator**.
- Let $\mathcal{L}$ be the space of continuous linear operators on H.
- We note that if x is a square integrable function also $\Psi(x)$ (Ψx for shorthness) is a function.
- Note and discuss the notational subtleties: $(\Psi x)(t)$, $\Psi x(t)$, and $\Psi(x(t))$ which does not make sense, why?

Examples

- Examples of operators: Consider functions on the interval $[0,1]$. Each of the following is a linear operator. Their continuity has to be addressed separately and often requires to restrict the domain of the operator.
  - $\Psi x = x$ - the identity operator.
  - $\Psi x(t) = \int_0^t x(u)$ – integral operator.
  - $\Psi x(t) = x'(t)$ – derivative operator
  - $\Psi x(t) = x(1-t)$ – symmetric vertical reflection
  - $\Psi x(t) = -x(t)$ – symmetric horizontal reflection
  - $\Psi x(t) = \int x(u)y(t-u)\, du$ – convolution operator
- Discuss vector analogs of these operators, represent them as matrices.

Simplest infinite dimensional operators

- An operator $\Psi \in \mathcal{L}$ is said to be **compact** if there exist two orthonormal bases $\{v_j\}$ and $\{f_j\}$ and a real sequence $\{\lambda_j\}$ converging to zero, such that:

$$\Psi(x) = \sum_{i=1}^{\infty} \lambda_j \langle x, v_j \rangle f_j, \qquad x \in H. \tag{1}$$

  This representation is called the **singular value decomposition (SVD)**.
- The special case is when $v_j = f_j$.
- A compact operator admitting such a representation is said to be a Hilbert-Schmidt operator if $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$.

Spectral decomposition

- An operator $\Psi \in \mathcal{L}$ is said to be, respectively, symmetric and positive definite if:

$$\langle \Psi(x), y \rangle = \langle x, \Psi(y) \rangle, \langle \Psi(x), x \rangle \geq 0 \qquad x, y \in H \tag{2}$$

- A symmetric positive-definite Hilbert-Schmidt operator Ψ admits the decomposition:

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j \qquad x \in H, \tag{3}$$

with orthonormal v_j which are the eigenfunctions of Ψ, i.e. $\Psi(v_j) = \lambda_j v_j$.

Eigenfunctions and eigenvalues

- Recall the concept of eigenvalues and eigenvectors of a matrix.
- Propose a natural extension of this concept to the case of space of functions and an operator acting on them.
- Check that the functions appearing in the spectral decomposition are actually eigenfunctions of the operator. What are the corresponding eigenvalues?

Kernel operator

- Recall that by H (or by L^2) we denote the set of measurable real-valued functions defined on [0,1] satisfying $\int_0^1 x^2(t)dt < \infty$. It is a Hilbert space with the inner product:

$$\langle x, y \rangle = \int x(t)y(t)dt. \tag{4}$$

Kernel operator

- Recall that by H (or by L^2) we denote the set of measurable real-valued functions defined on [0,1] satisfying $\int_0^1 x^2(t)dt < \infty$. It is a Hilbert space with the inner product:

$$\langle x, y \rangle = \int x(t)y(t)dt. \tag{4}$$

- An important class of operators in L^2 are the integral operators defined by:

$$\Psi(x)(t) = \int \psi(t,s)x(s)ds, \qquad x \in L^2, \tag{5}$$

with the real kernel $\psi(\cdot,\cdot)$.

Kernel operator

- Recall that by H (or by L^2) we denote the set of measurable real-valued functions defined on [0,1] satisfying $\int_0^1 x^2(t)dt < \infty$. It is a Hilbert space with the inner product:

$$\langle x, y \rangle = \int x(t)y(t)dt. \tag{4}$$

- An important class of operators in L^2 are the integral operators defined by:

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \qquad x \in L^2, \tag{5}$$

with the real kernel $\psi(\cdot, \cdot)$.

- If $\psi(s, t) = \psi(t, s)$ and $\int\int \Psi(t, s)x(t)x(s) \geq 0$, the integral operator Ψ is symmetric and positive-definite, and it follows that

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j v_j(t)v_j(s), \qquad \in L^2([0, 1] \times [0, 1]) \tag{6}$$

The model of random functional data

- Let X be a H-valued random function.
- If there is $\mu \in H$ such that $E\langle X, y\rangle = \langle \mu, y\rangle$ for all $y \in H$, then μ is called the expectation of X and denoted by EX.
- The expectation commutes with continuous operators, i.e. if $\Psi \in L$ and X is integrable, then $E\Psi(X) = \Psi(EX)$.
- For a random function X, the covariance operator of X is defined by:

$$C(y) = E[\langle X - EX, y\rangle (X - EX)], \qquad y \in H. \tag{7}$$

(We assume implicitly that the right hand side is well defined and also often, as before, we write simply Cy instead of $C(y)$.)

More explicit view on mean

- Let assume that we do not question the change of order of linear operations in all the following derivations.
- The mean:

$$E\langle X, y\rangle = E\int_0^1 X(u)y(u)\,du$$
$$= \int_0^1 E(X(u))y(u)\,du.$$

- Thus if we define $\mu(u) = E(X(u))$, and $\mu \in H$, then the expectation can be viewed as simply pointwise expectation of a random function.

... and on covariance

- For]the covariance operator of X

$$\begin{aligned} Cy(t) &= E[\langle X - EX, y\rangle] \\ &= E\left(\int_0^1 (X(u) - EX(u))\, y(u)\, du (X(t) - EX(t))\right) \\ &= \int_0^1 E\left((X(u) - EX(u))\, (X(t) - EX(t))\right) y(u)\, du \\ &= \int_0^1 Cov(X(u), X(t)) y(u)\, du. \end{aligned}$$

- Thus the covariance operator can be viewed as a kernel operator with the kernel $c(t, u)$ being pointwise covariance of random functions.
- There are some technical aspects, why the mean and the covariance are not introduced this way but for all practical understanding, it is good to think about them this way.

Further properties

- The covariance operator C is symmetric and positive-definite, with eigenvalues λ_i satisfying:

$$\sum_{i=1}^{\infty} \lambda_i = E\|X - EX\|^2 < \infty. \tag{8}$$

- Thus it admits the SVD (or, in other terminology, the spectral decomposition).

Outline

A sample of functions and their sample charectiristics

- Let $X_1, X_2, \ldots, X_N$ be observed H-valued random functions.
- The mean can be estimated naturally by

$$\hat{\mu}_N(t) = \frac{1}{N}\sum_{k=1}^{N} X_k(t).$$

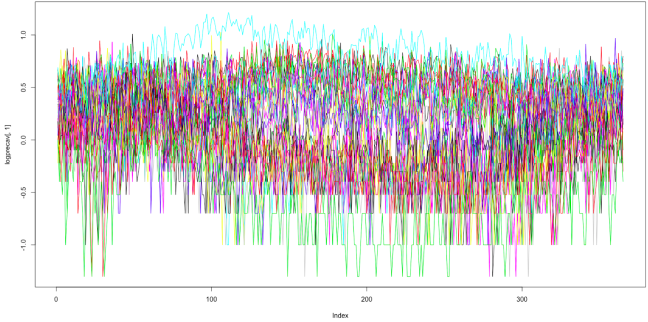
- Similarily, the covariance kernel c(t,s) can be estimated by:

$$\hat{c}(t,s) = \frac{1}{N}\sum_{k=1}^{N}(X_k(t) - \hat{\mu}_N(t))(X_k(s) - \hat{\mu}_N(s)).$$

- Alternativly, one can estimate eigenvalues of the spectral decomposition for the covariance operator using eigenvectors and their eigenvalues for sample covariance.

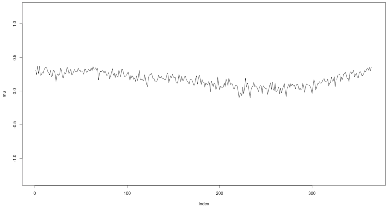
Log-precipitation data example

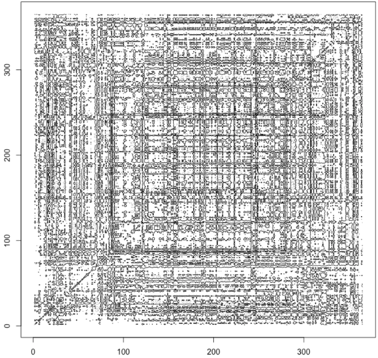
- Raw data



Raw mean and covariance

- Raw mean (left) and covariance (right)





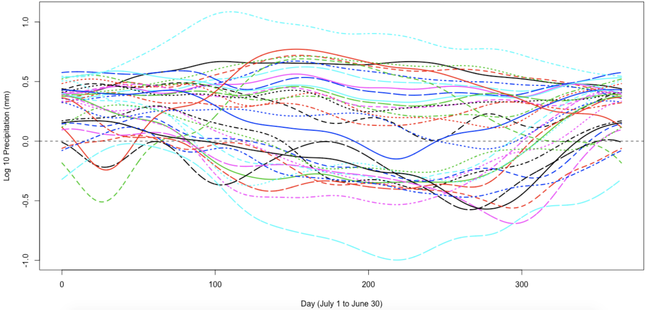
- Some data smoothing is needed.

Smoothing using `fda` package - Section 5.4

- We follow the steps of the textbook using the script.

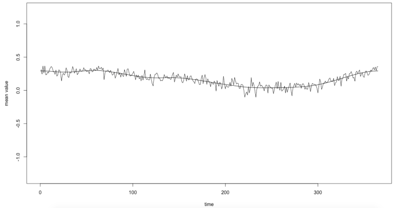
Smoothed log-precipitation data

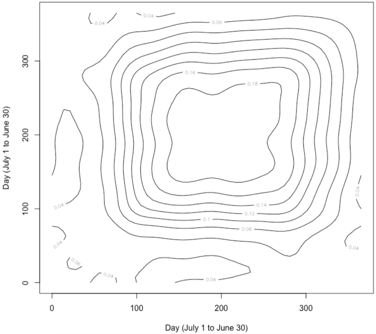
- Smoothed data using the fda package and optimal filter



Mean and covariance for smoothed data

- Raw mean (left) and covariance (right)





- Some data smoothing is needed.

Rationale behind smoothing

- When we smooth, we indirectly assume that the data of interest come distorted by a certain noise:

$$y_{obs}(t) = y(t) + \epsilon(t), \; t \in [t_0, t_1].$$

- This noise is not of interest and needs to be filtered.
- The validity of such assumptions is the case dependent and often is frequently decide on an ad hoc basis.