Description of Functional Data and Mathematical Foundations

School of Statistics and Data Science

Overview

Diffusion Tensor Imaging Example

- Fractional anisotropy (FA): a measure of water diffusion in the brain.
- FA tract profiles for the corpus callosum (CCA) in Multiple Sclerosis patients.

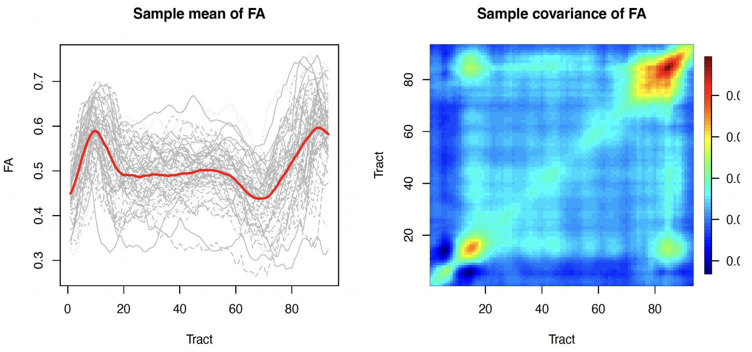


Figure 1: Mean and Covariance of FA traits.

Formal Characterization

- Samples: $X_1(t), \ldots, X_n(t), t \in T$, n realizations of a random function $X(t)$.
- We can think of $X(t)$ in two complementary ways.
- A random element in a Hilbert space
  - a random variable that takes values in a Hilbert space.
- A stochastic process, with sample paths
  - a collection of random variables $X = \{X(t) : t \in T\}$ defined on a common probability space Ω, and sample paths $t \to X(t)$.

Square Integrable Functions

- Let $X(t)$ be a random function with mean $\mu(t)$ and covariance $C(s,t) = Cov(X(t), X(s))$.
- $X(t)$ has realizations in a function space.
- Work with the space of square integrable functions L^2, as it insures the existence of the first two moments.

Definition

A function $f(t) : T \to \mathbb{R}$ is said to be square integrable, $f \in L^2(T)$ if

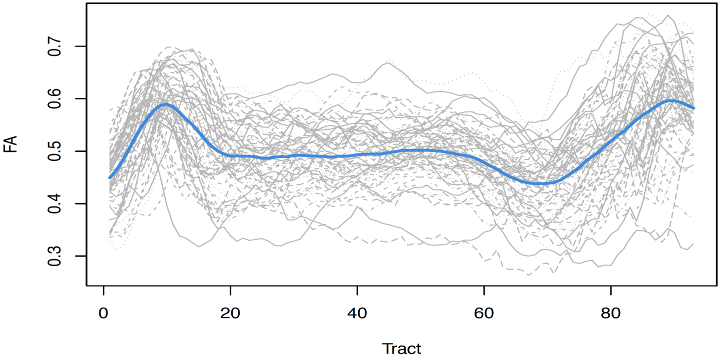
$$\int_T f(t)^2 dt < \infty.$$

- Square integrable functions form a vector space. Therefore, if $f, g \in L^2(T)$, $\forall a, b \in \mathbb{R}$,

$$(af + bg)(t) = af(t) + bg(t) \in L^2(T).$$

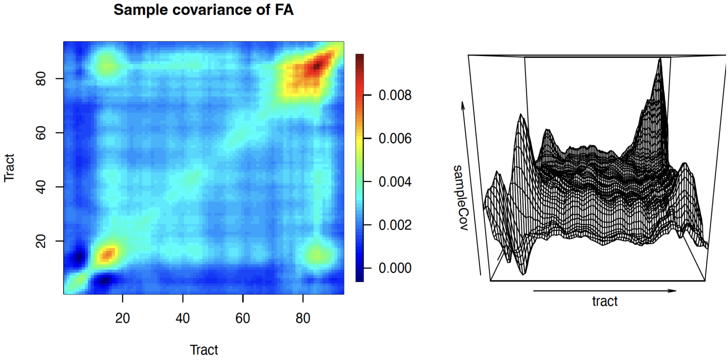
Sample Mean

$$\widehat{\mu}(t) = \bar{X}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$$



Sample Covariance

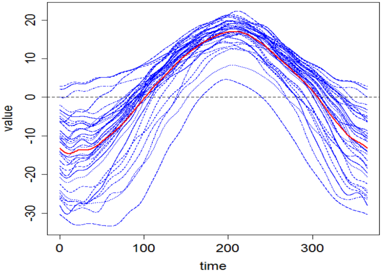
$$\widehat{\rho}(s,t) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i(s)-\bar{X}(s)\right)\left(X_i(t)-\bar{X}(t)\right).$$

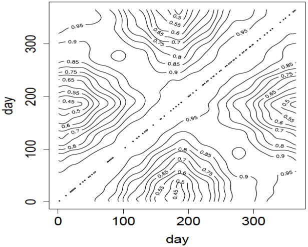


- Covariance surfaces provide insight but do not describe the major directions of variation

Sample Correlation

$$\widehat{C}(s,t) = \frac{\sum_{i=1}^{n}\left(X_i(s)-\bar{X}(s)\right)\left(X_i(t)-\bar{X}(t)\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i(s)-\bar{X}(s)\right)^2\sum_{i=1}^{n}\left(X_i(t)-\bar{X}(t)\right)^2}}$$





Inner Product

- if $f, g \in L^2(T)$, the inner product is

$$\langle f, g \rangle = \int_T f(t)g(t)dt.$$

- Orthogonal: $\langle f, g \rangle = 0$.
- The inner product norm:

$$\|f\| = \sqrt{\langle f, f \rangle} = \left\{ \int_T f^2(t)dt \right\}^{1/2}.$$

- The inner product norm:

$$d(f, g) = \|f - g\| = \left\{ \int_T [f(t) - g(t)]^2 dt \right\}^{1/2}.$$

Inner Product

Properties of the inner product norm: $\forall f, g \in L^2(T)$, $\forall a \in \mathbb{R}$,

- $\|af\| = a\|f\|$
- Cauchy-Schwartz inequality:

$$|\langle f, g\rangle| \leq \|f\|\|g\|$$

- Triangle Inequality:

$$\|f+g\| \leq \|f\| + \|g\|$$

-

$$d(f,g) = d(g,f) \geq 0, d(f,f) = 0$$

$$d(f,g) \leq d(f,z) + d(z,g)$$

- Metric Space with the metric $\|\cdot\|$.

Basis of L^2

- A set of functions $\{e_1, e_2, \dots\}$ is a basis of $L^2(T)$ if, any $f \in L^2(T)$ admits a unique expansion

$$f(t) = \sum_{j=1}^{\infty} a_j e_j(t), \quad a_j \in \mathbb{R}$$

- A Hilbert space: a complete metric space, i.e. an inner product space where every Cauchy sequence has a limit.

Linear Operators

- Let V_1, V_2 be two vector spaces. A function $L : V_1 \to V_2$ is called a linear transformation or linear operator if

$$L(ax + by) = aL(x) + bL(y),\ \forall a, b \in \mathbb{R},\ x, y \in V_1.$$

Example (Integral Operator)

1. $\Psi_1(x) : L^2(T) \to \mathbb{R}$,

$$\Psi_1(x) = \int_T \psi(t)x(t)dt$$

   with $\Psi_1(x) < \infty$ if $\psi(t) \in L^2(T)$.

2. A Hilbert-Schmidt operator:$\Psi_2(x) : L^2(T) \to L^2(T)$,

$$\Psi_2(x) = \int_T \psi(s,t)x(t)ds$$

   with $\int_T \int_T \psi(s,t)^2 dsdt < \infty$ with the kernel $\psi(s,t)$.

Spectral Theory

Suppose $L : V \to V$ is a linear operator on a vector space V.

- If there exist a non-zero vector $x \in V$ such that

$$L(x) = \lambda x$$

- An Eigenvalue of L: λ
- An Eigenvector of L: x
- Non-negative definite operator: $\forall x \in V, \langle L(x), x \rangle \geq 0$, all eigenvalues are non-negative.
- Positive definite operator: $\forall x \in V, \langle L(x), x \rangle = 0$ if and only if $x = 0$, all eigenvalues are positive.
- L is symmetric if $\forall x, y \in V, \langle L(x), y \rangle = \langle x, L(y) \rangle$.
- Eigenvectors for distinct eigenvalues of a symmetric operator are orthogonal.

Hilbert-Schmidt

Theorem (Hilbert-Schmidt)

Let $\Psi(x) : H \to H$ be a symmetric HS operator. There exists an orthonormal system $\{e_j, j \geq 1\}$ consisting of eigenvectors of Ψ corresponding to non-zero eigenvalues λ_j, such that $\forall x \in H$ has a unique representation

$$x = \sum_{j=1}^{\infty} a_j e_j + v,$$

with v satisfying $\Psi(v) = 0$.

Corollary (Mercer's Theorems)

Let $\Psi(x) : H \to H$ *be a symmetric HS operator. There exists an orthonormal system* $\{e_j, j \geq 1\}$ *such that* $\forall x \in H$

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, e_j \rangle e_j,$$

where λ_j *is an eigenvalue corresponding to the eigenfunctions of* e_j *of* Ψ*.*

Corollary

If Ψ *is with eigenvalues* $\lambda_1 > \lambda_2 > \cdots$, *and corresponding eigenfunctions* $e_1, e_2, \cdots$,

$$\sup\{\langle \Psi(x), x \rangle : \|x\| = 1, \langle x, e_j \rangle = 0, 1 \leq j \leq i-1\} = \lambda_i,$$

and the sup is achieved at if $x = e_i$, *unique up to a sign.*

- An equivalent re-statement of Mercer's theorem is based on the spectral decomposition of HS kernels.

Theorem

Let $\psi(\cdot,\cdot)$ be the kernel of Ψ, a symmetric HS operator, then

$$\psi(s,t) = \sum_{j=1}^{\infty} \lambda_j e_j(s) e_j(t), \ \ \forall s,t \in T.$$

Covariance Operators

- Let $y \in L^2$, with integral operator $C : L^2 \to L^2$ defined as

$$C_y(t) := \int c(t,s)\, y(s)\, ds, \quad \forall t \in T$$

where $C(t,s) = E[\{X(t) - \mu(t)\}\{X(s) - \mu(s)\}]$, is known as the covariance operator, with kernel function $C(t,s)$.

- A simple application of Mercer's theorem yields

$$C(t,s) = \sum_{j=1}^{\infty} \lambda_j e_j(t) e_j(s)$$

$$C_y(t) = \sum_{j=1}^{\infty} \lambda_j \langle y, e_j \rangle\, e_j.$$

Reproducing Kernel Hilbert Space (RKHS)

Let $k(s,t)$ be the kernel of a symmetric, pd, HS operator. Equivalently we assume:

- $k(s,t) : T \times T \to \mathbb{R}$.
- For any $t_1, \ldots, t_m \in T$, the matrix $\{k(t_i, t_j) : i = 1, \ldots, m, j = 1, \ldots, m\}$ is pd.
- $k(s,t) = k(t,s)$.
- $\int k(t,s)^2 \, dt \, ds < \infty$.

The kernel $k(\cdot,\cdot)$ is said to be a reproducing kernel for a Hilbert space H if:

1. $k(\cdot,t) \in H$, for all $t \in T$.
2. For any $f \in H$ and $t \in T$, $f(t) = \langle f, k(\cdot,t) \rangle$.
3. $\langle k(\cdot,t), k(\cdot,s) \rangle = k(t,s)$

Reproducing Kernel Hilbert Space (RKHS)

A set of functions obtained as a finite linear combination of a HS kernel $k(\cdot, \cdot)$ is known as a RKHS.
Formally the function space defined as:

$$H_k := \left\{ f : f(t) = \sum_{j=1}^{J} \alpha_j k(t, s_j), \text{ for some } 1 \leq J \leq \infty, \, s_j \in T, \, \alpha_j \in \mathbb{T} \right\}$$

is known as RKHS.

RKHS and Smoothing

- Let $y(x_i) = f(x_i) + \varepsilon_i$, $i = 1, \ldots, n$, and $k(s,t)$ be a reproducing kernel.
- Assuming $f(x) \in H_k$ a RKHS, $f(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i)$.
- Estimation is achieved minimizing

$$\sum_{i=1}^{n} (y(x_i) - f(x_i))^2 + \lambda ||f||^2$$

  or in vector form

$$(y - K\alpha)^T (y - K\alpha) + \lambda \alpha^T D\alpha.$$

- Example: Gaussian Kernel

$$k(s,t) = \exp\{-\sigma |s-t|^2\}.$$

Rationale behind smoothing

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

$$y_{obs}(t) = y(t) + \varepsilon(t), \ t \in T.$$

- 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.

Thank You!

Questions? Comments?