Statistical methods of data science

An introduction to Functional Data Analysis

Alessandra Menafoglio[1*]

[1]MOX, Department of Mathematics, Politecnico di Milano

*alessandra.menafoglio@polimi.it

1. Hilbert space model for functional data

Agenda

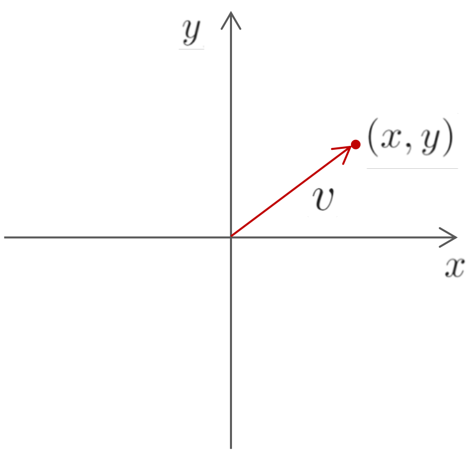
1.1. Basics notions on Hilbert spaces

A Hilbert Space approach to the analysis of Functional Data

The notion of **Hilbert space** generalizes the concept of Euclidean space to spaces of any (even infinite) dimension

- Vectorial structure (linear combinations)
- Distance, angles, projections (measure of dependence, best approximations)

Euclidean space $\mathbb{R}^2$



- Sum: $v_1 + v_2 = (x_1 + x_2, y_1 + y_2)$
- Product by a constant: $c \cdot v = (c \cdot x, c \cdot y)$

$\Big\}$ Operations (+, ·)

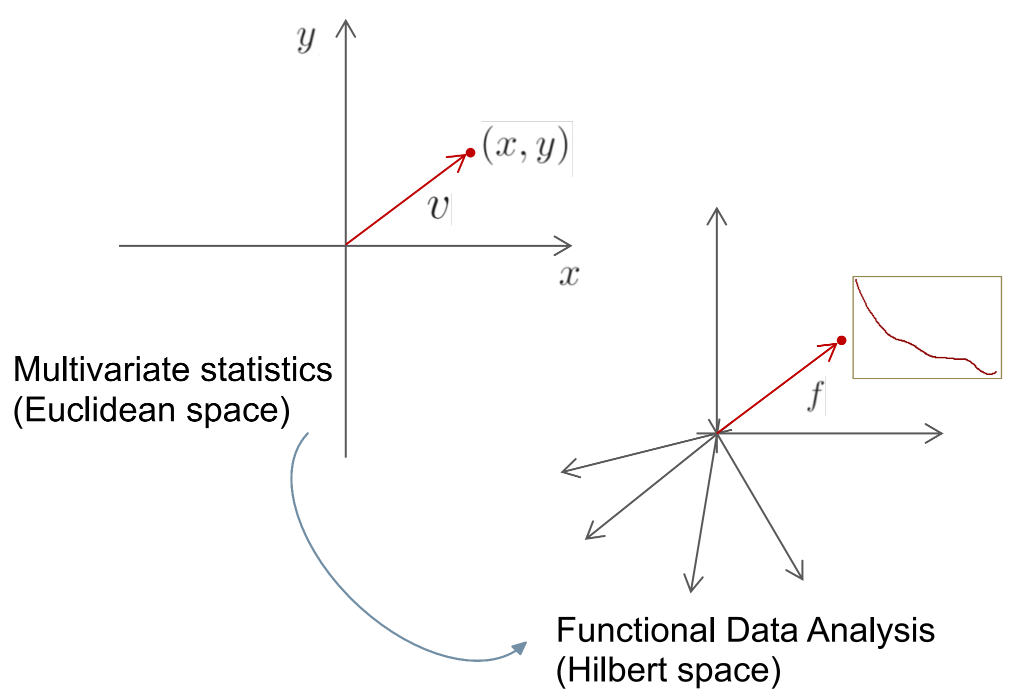
- Norm (lenght of a vector): $\|v\| = (x^2 + y^2)^{1/2}$
- Distance: $\|v_1 - v_2\| = (x_1 - x_2)^2 + (y_1 - y_2)^2$
- Angle: $\vartheta = \arccos \dfrac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|}$

$\Big\}$ Inner product
$\langle v_1, v_2 \rangle = (x_1 \cdot x_2) + (y_1 \cdot y_2)$

The notion of **Hilbert space** generalizes the concept of Euclidean space to spaces of any (even infinite) dimension

- Vectorial structure (linear combinations)
- Distance, angles, projections (measure of dependence, best approximations)

Conceptual example: from Euclidean to Hilbert spaces



Why Hilbert spaces?

- We understand functional data as **points of a space of functions**
- Many techniques in **multivariate statistics can be generalized to data embedded in a Hilbert space**, through the notions of inner product and norm

Inner product spaces

Let *H* be a linear space. An inner product on *H* is a bilinear, symmetric, positive definite form

$$\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$$

that satisfies

$$\text{(i)} \quad \langle \lambda x + y, z \rangle = \lambda \langle x, z \rangle + \langle y, z \rangle \qquad \forall \lambda \in \mathbb{R}, \quad \forall x, y, z \in H$$

$$\text{(ii)} \quad \langle x, y \rangle = \langle y, x \rangle \qquad \forall x, y \in H$$

$$\text{(iii)} \quad \langle x, x \rangle \geq 0 \qquad \forall x \in H$$

$$\text{(iv)} \quad \langle x, x \rangle = 0 \quad \iff \quad x = 0$$

In particular:

- The inner product allows to measure lengths and angles
- It allows defining orthogonality: two vectors in *H* are orthogonal if $\langle x, y \rangle = 0$
- The inner product induces a norm and a metric
- The inner product allows generalizing the Pythagoras' Theorem:

$$||x + y||^2 = ||x||^2 + ||y||^2 \quad \text{if and only if} \quad \langle x, y \rangle = 0$$

A (real) Hilbert space *H* is an inner product space that is complete, in the norm induced by the inner product.

- A Hilbert space is complete in the sense that it contains all the limit points of its Cauchy sequences
- Useful properties:
  - In a Hilbert space one has the notion of orthogonal projection and of best approximations
  - A Hilbert space *H* always admits an orthonormal basis $\{u_n\}_{n\in\mathbb{N}}$
  - If *H* is separable Hilbert space, $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal basis and $x \in H$. Then

$$x = \sum_{n=1}^{\infty} \langle x, u_n \rangle u_n.$$

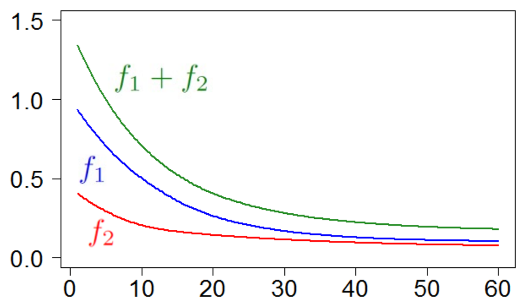
Basis expansion

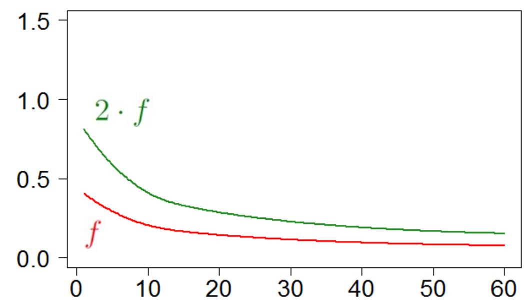
An Example: the Hilbert space L^2

L^2: space of real-valued square-integrable functions

Operations $(+, \cdot)$

- Sum: $(f_1 + f_2)(t) = f_1(t) + f_2(t)$
- Product by a constant: $(c \cdot f)(t) = c \cdot f(t)$





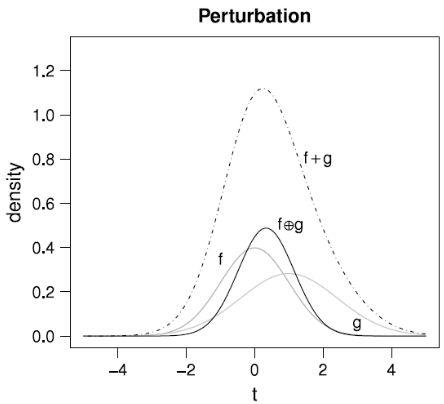
- Norm: $\|f\|^2 = \int (f(t))^2 dt$
- Distance: $\|f_1 - f_2\|^2 = \int (f_1(t) - f_2(t))^2 dt$
- Angle: $\vartheta = \arccos \dfrac{\langle f_1, f_2 \rangle}{\|f_1\| \|f_2\|}$

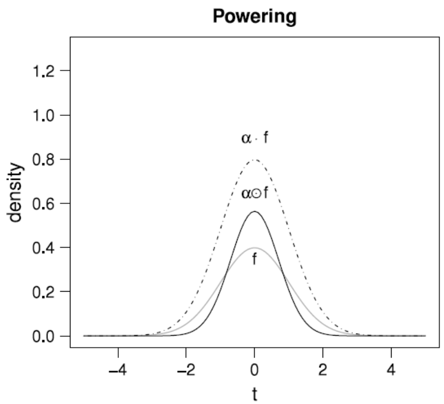
Inner product
$\langle f_1, f_2 \rangle = \int (f_1(t) \cdot f_2(t)) dt$

More precisely, L^2 is a quotient space with respect to the equivalence relation: $x = y$ *if* $\int [x(t) - y(t)]^2 dt = 0$)

B^2: space of density functions on a close interval I, with log in L^2

- Equivalence relation: f, g are equivalent if they are proportional (*scale invariance*)
- Sum (perturbation): $(f \oplus g)(t) = \dfrac{f(t)g(t)}{\int_I f(s)g(s)\,ds}$,
- Product by a constant (powering): $(\alpha \odot f)(t) = \dfrac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$
- Inner product: $\langle f, g\rangle_{\mathcal{B}} = \dfrac{1}{2\eta}\displaystyle\int_I\int_I \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)}\,dt\,ds$
- Norm: $\|f\|_{\mathcal{B}} = \left[\dfrac{1}{2\eta}\displaystyle\int_I\int_I \ln^2\frac{f(t)}{f(s)}\,dt\,ds\right]^{1/2}$





Note: the geometry of L^2 wouldn't make sense for density functions

B^2: space of density functions on a close interval I, with log in L^2

- B^2 is isomorphic to L^2 (in fact, all the Hilbert spaces are isomorphic). An isometric isomorphism is provided, e.g., by the **centred log-ratio transformation**

$$\mathrm{clr}(f)(t) = f_c(t) = \ln f(t) - \frac{1}{\eta}\int_I \ln f(s)\, ds.$$

Exercise: prove that

$$\mathrm{clr}(f \oplus g)(t) = f_c(t) + g_c(t), \qquad \mathrm{clr}(\alpha \odot f)(t) = \alpha \cdot f_c(t), \quad \langle f, g\rangle_{\mathcal{B}} = \langle f_c, g_c\rangle_2 = \int_I f_c(t) g_c(t)\, dt.$$

- Hilbert space structure for functional compositional data (e.g., probability density functions)
- Account for the key properties of compositional data: scale invariance, relative scale, sub-compositional coherence
- Meaningful interpretations in mathematical statistics, e.g.,
  - Exponential families as affine finite-dimensional subspaces
  - Perturbation $\oplus$ as a Bayes update of information

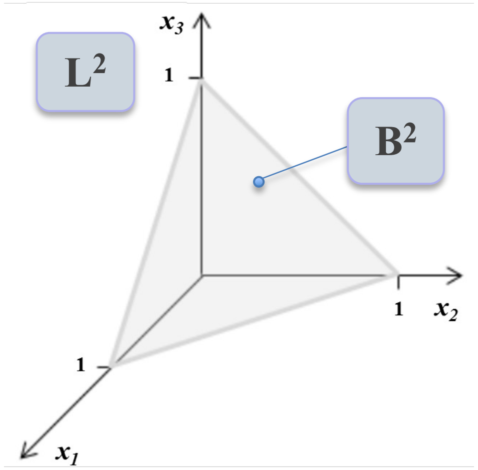
B^2: space of density functions on a close interval I, with log in L^2

- B^2 is isomorphic to L^2 (in fact, all the Hilbert spaces are isomorphic). An isometric isomorphism is provided, e.g., by the **centred log-ratio transformation**

$$\mathrm{clr}(f)(t) = f_c(t) = \ln f(t) - \frac{1}{\eta}\int_I \ln f(s)\, ds.$$

Exercise: prove that

$$\mathrm{clr}(f \oplus g)(t) = f_c(t) + g_c(t), \qquad \mathrm{clr}(\alpha \odot f)(t) = \alpha \cdot f_c(t), \quad \langle f, g\rangle_{\mathcal{B}} = \langle f_c, g_c\rangle_2 = \int_I f_c(t) g_c(t)\, dt.$$



Conceptually, the B2 geometry equip the infinite dimensional simplex of PDFs with a Hilbert Geometry (as the Atichison geometry does in finite dimension)

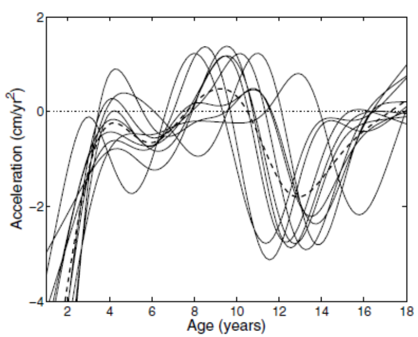
Agenda

1. **Hilbert space model for functional data**
    1.1. Basics notions on Hilbert spaces
    1.2. Hilbert space embedding for functional data
    1.3. Formal definition of functional data

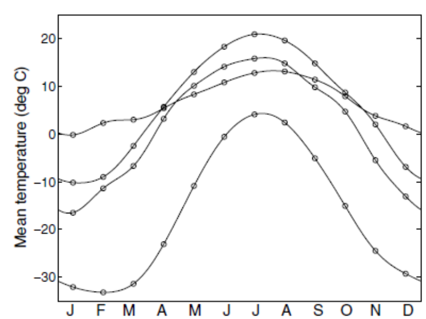
1.2. Hilbert space embedding for functional data

- As a first step of any functional data analysis, one need to **choose the embedding** for the data
- **Separable Hilbert spaces are a convenient choice** (projections, best approximations).
  Note: Not all the interesting spaces are Hilbert: e.g., the space of continuous functions is not a Hilbert space. Other interesting spaces: Riemannian manifolds (OODA)
- Examples of Hilbert spaces for FDA:
  - L^2, space of square integrable functions: OK for most data analyses (especially if data are unconstrained)
  - B^2, space of functional compositions: useful for density functions

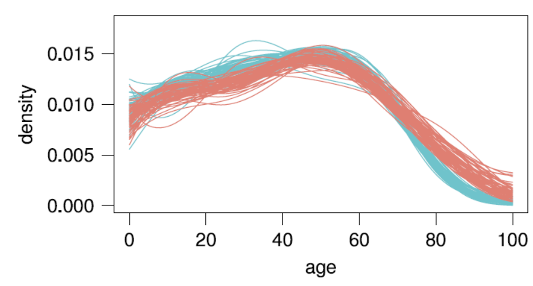
Examples:



Acceleration in Berkeley Growth data



Temperatures in Canada



Population Age densities

Agenda

Functional random variables and functional data

- Let H be a Hilbert space, whose points are functions defined on a closed interval $T = [t_{min}, t_{max}]$ (e.g., range of time during which the data are collected)
- Hereafter, we will always consider functional data in Hilbert spaces

Definition 1

A **functional random variable** is a random element on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ in the space $X : \Omega \to H$

Definition 2

A **functional datum** x is a realization of a functional random variable, i.e., for $\omega \in \Omega$,

$$x = X(\omega) : T = [t_{min}, t_{max}] \to \mathbb{R}$$

Definition 3

A **functional dataset** is a collection of functional data.

Let $X : \Omega \rightarrow H$ be a functional random variable in *H.* Hereafter, we always assume that $\mathbb{E}[\|X\|_H^4] < \infty$

Definition 4

We call Fréchet mean of X the (unique) element μ of *H* that solves

$$\underset{x \in H}{\operatorname{arginf}}\ \mathbb{E}[\|X - x\|_H^2].$$

- If *H=L*^{2} (space of square-integrable functions), the Fréchet mean coincides a.e. with the point-wise mean
$$\mathbb{E}[X(t)] = \mu(t), \quad t \in T$$
- If *H=B*^{2} (Bayes space of PDFs), the Fréchet mean can be computed as
$$\mu = clr^{-1}(\mathbb{E}[clr(X)])$$
(in particular, one can define the mean of the clr-transformed variable point-wise)
- In any *H*, one can **estimate the mean via the sample estimator**

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

In *H=L*^{2}, this is the point-wise sample mean

Let. $X : \Omega \to H$ be a **zero-mean** functional random variable in *H*, such that $\mathbb{E}[\|X\|_H^4] < \infty$

Definition 5

We call covariance operator of X the operator from *H* to *H* defined as

$$Cx = \mathbb{E}[\langle X, x\rangle X], \quad x \in H$$

- If *H*=*L*^{2} (space of square-integrable functions), the covariance operator can be equivalently defined as the kernel operator

$$[Cx](t) = \int_T c(s,t)x(s)d(s) \ , \quad x \in L^2$$

  where the covariance kernel is precisely the point-wise covariance

$$c(s,t) = \mathbb{E}[X(s)X(t)]$$

- In *H*=$\mathbb{R}^p$, the covariance operator coincides with the linear operator defined by the covariance matrix

Mean and covariance operator

Let. $X : \Omega \rightarrow H$ be a **zero-mean** functional random variable in *H*, such that $\mathbb{E}[||X||_H^4] < \infty$

Definition 5

We call covariance operator of X the operator from *H* to *H* defined as

$$Cx = \mathbb{E}[\langle X, x \rangle X], \quad x \in H$$

- In any *H*, the covariance operator can be estimated through the sample covariance operator

$$Sx = \frac{1}{N} \sum_{i=1}^{N} \langle X_i, x \rangle X_i, \quad x \in H$$

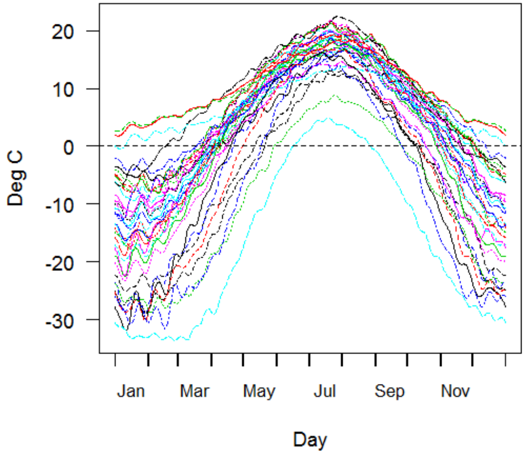
- If $H=L^2$, one can use the alternative definition

$$[Sx](t) = \int_T \widehat{c}(s,t) x(s) d(s)], \quad x \in L^2 \text{ with } \widehat{c}(s,t) = \frac{1}{N} \sum_{i=1}^{N} X(s) X(t)$$

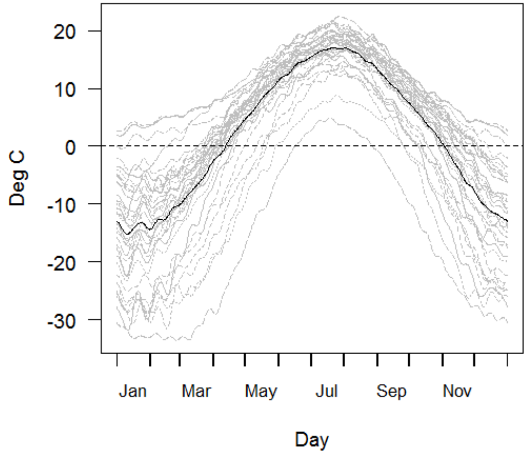
1.3. Formal definition of functional data

An example in L^2

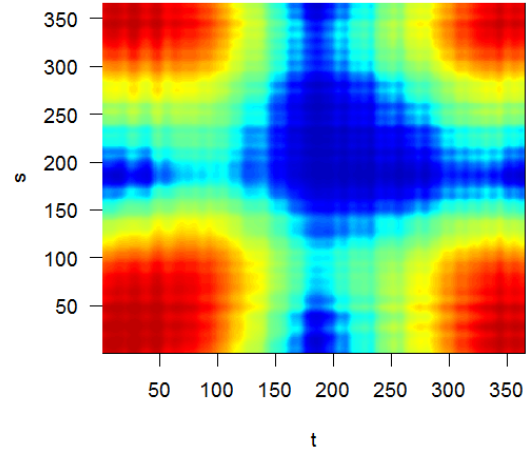
Example: Canadian temperature data (35 observations)



Functional dataset



Sample mean



Sample covariance kernel

1.3. Formal definition of functional data

An example in B^2

Example: Dataset of Age Densities in Upper Austria (114 observations)