DDA Meeting
3rd June 2024

Functional data analysis of density data in Bayes spaces

Alessandra Menafoglio

MOX, Department of Mathematics, Politecnico di Milano

alessandra.menafoglio@polimi.it

Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
    - Simplicial Functional Principal Component Analysis
    - Linear Models in B^2
- Conclusions

Functional data & Functional data analysis

- **Functional data** are entities that can be described through a function, e.g., a curve, a surface, an image
- **Large *p* small *n* problems**: classical multivariate methods fail when the number of variable is higher than the number of data (in this case, p=31, n=10)
- **Functional Data Analysis** is concerned with statistical analysis of (virtually) infinite-dimensional objects

Example: Berkeley Growth study
Observation of the height of 10 girls measured along 31 ages

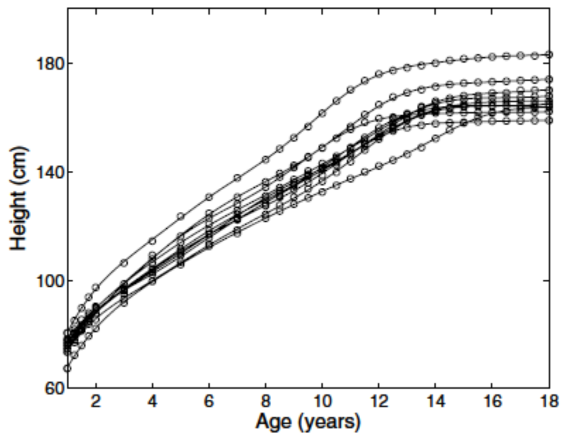


Figure 1.1. The heights of 10 girls measured at 31 ages. The circles indicate the unequally spaced ages of measurement.

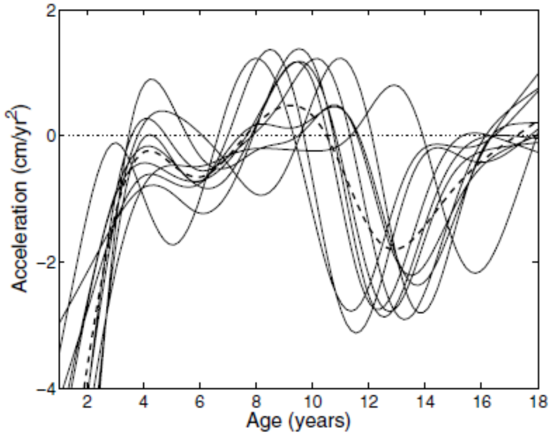


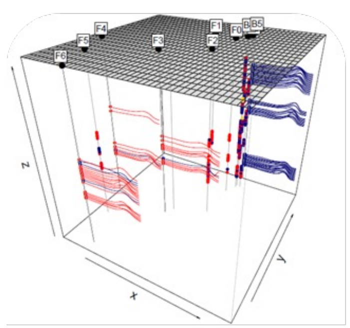
Figure 1.2. The estimated accelerations of height for 10 girls, measured in centimeters per year. The heavy dashed line is the cross-sectional mean, and is a rather poor summary of the curves.

Taken from Ramsay & Silverman (2002)

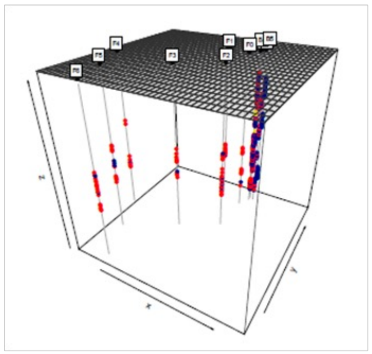
The limit of classical approaches

- Classical approaches would advocate the reduction of the data to simple indicators, which can be analyzed with classical multivariate methods.
- This inevitably yields an information loss.

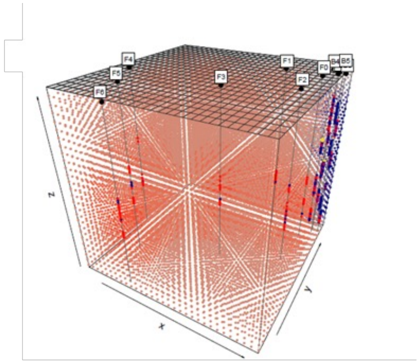
Conceptual example: spatial prediction of functional data through data reduction



Dataset of
complex objects



Reduced
dataset

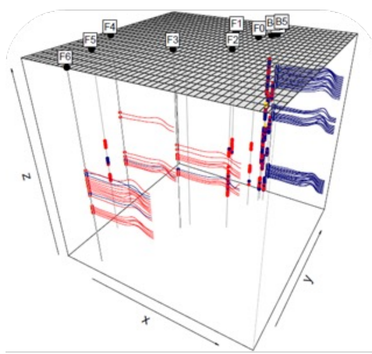


Prediction

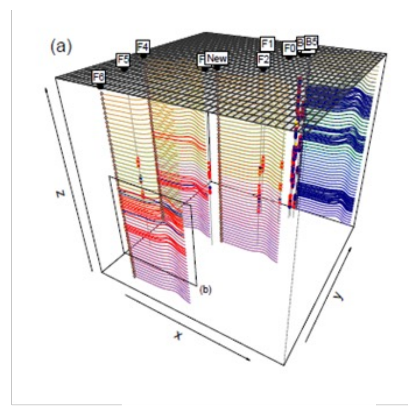
The limit of classical approaches

- Classical approaches would advocate the reduction of the data to simple indicators, which can be analyzed with classical multivariate methods.
- This inevitably yields an information loss.
- In Functional Data Analysis, the **“atom”** of the statistical analysis is the entire function, rather than a limited number of selected features of the data. This potentially allows to **exploit the entire information content embedded within the data**

Conceptual example: spatial prediction of functional data through FDA



Dataset of
complex objects

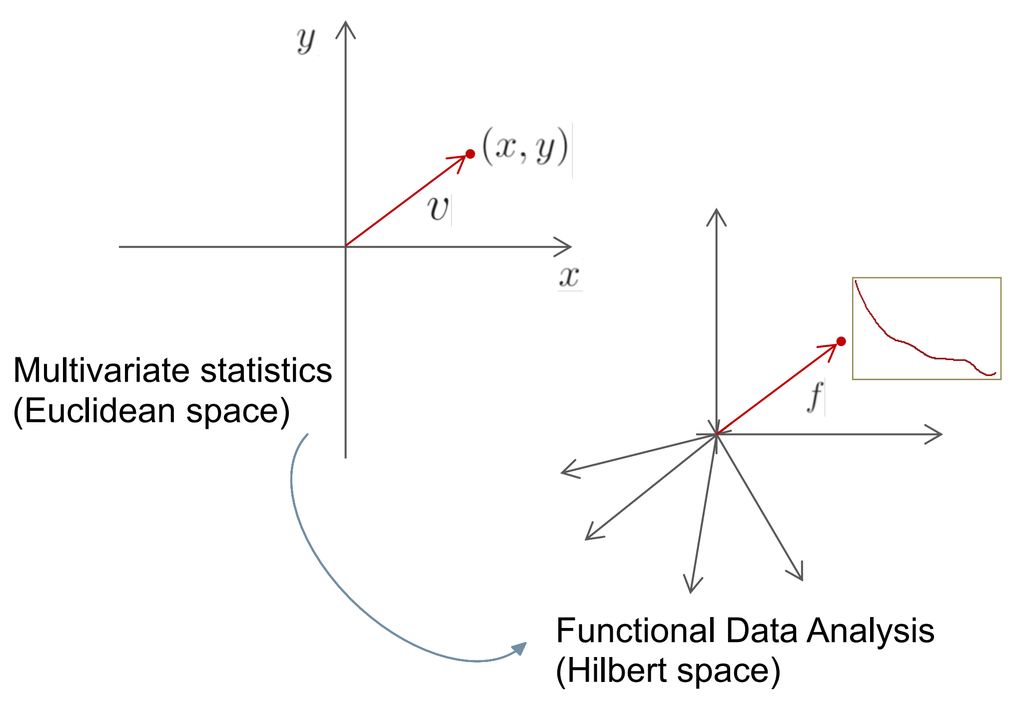


FDA Prediction

From multivariate to functional statistics

- When data are multivariate, we understand them as points in a Euclidean space
- We understand functional data as **points of a space of functions** (typically a Hilbert space)

Conceptual example: from Euclidean to Hilbert spaces



Why Hilbert spaces?

The notion of **Hilbert space** generalizes the concept of Euclidean space to spaces of any dimension

Many techniques in **multivariate statistics can be generalized to data embedded in a Hilbert space**, through the notions of inner product and norm

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

Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

Analysing density data with FDA

- Most methods in Functional Data Analysis implicitly assume that the data objects can be embedded in L^2

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

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

Operations $(+, \cdot)$

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

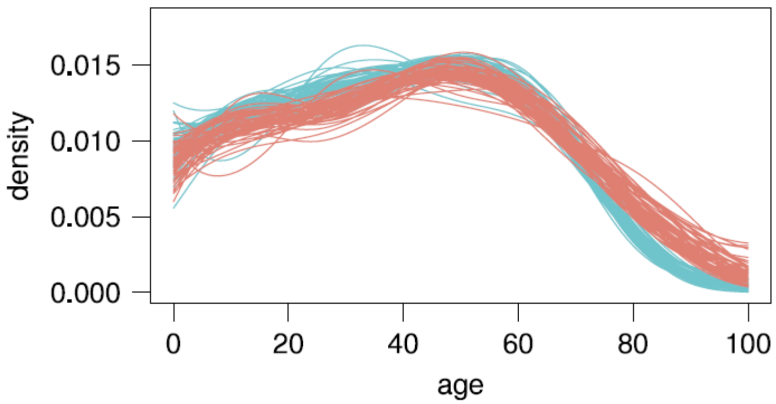
Inner product

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

Analysing density data with FDA

- Most methods in Functional Data Analysis implicitly assume that the data objects can be embedded in L^2
- However, **Density Data (DD)** are constrained functional data

Example: Density functions of Age Distribution in Austria



- Positive
- Integral to unity

Analysing density data with FDA

- Most methods in Functional Data Analysis implicitly assume that the data objects can be embedded in L^2
- However, **Density Data (DD)** are constrained functional data

The space of PDFs is not a linear (Hilbert) space when the L^2 geometry is used

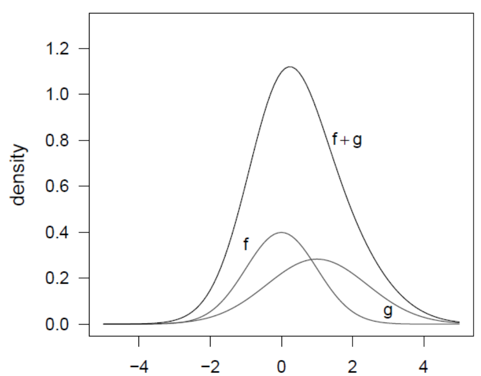
Analysing density data with FDA

- Most methods in Functional Data Analysis implicitly assume that the data objects can be embedded in L^2
- However, **Density Data (DD)** are constrained functional data

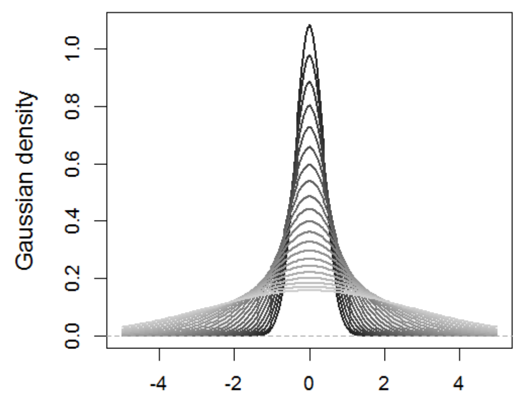
The space of PDFs is not a linear (Hilbert) space when the L^2 geometry is used

Examples of DD analyses in L^2

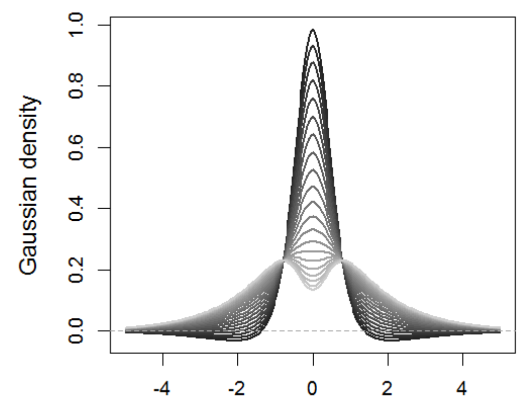
L^2 sum of two Gaussian PDFs



Principal Component Analysis in L^2 for a dataset of Gaussian PDFs



Original densities



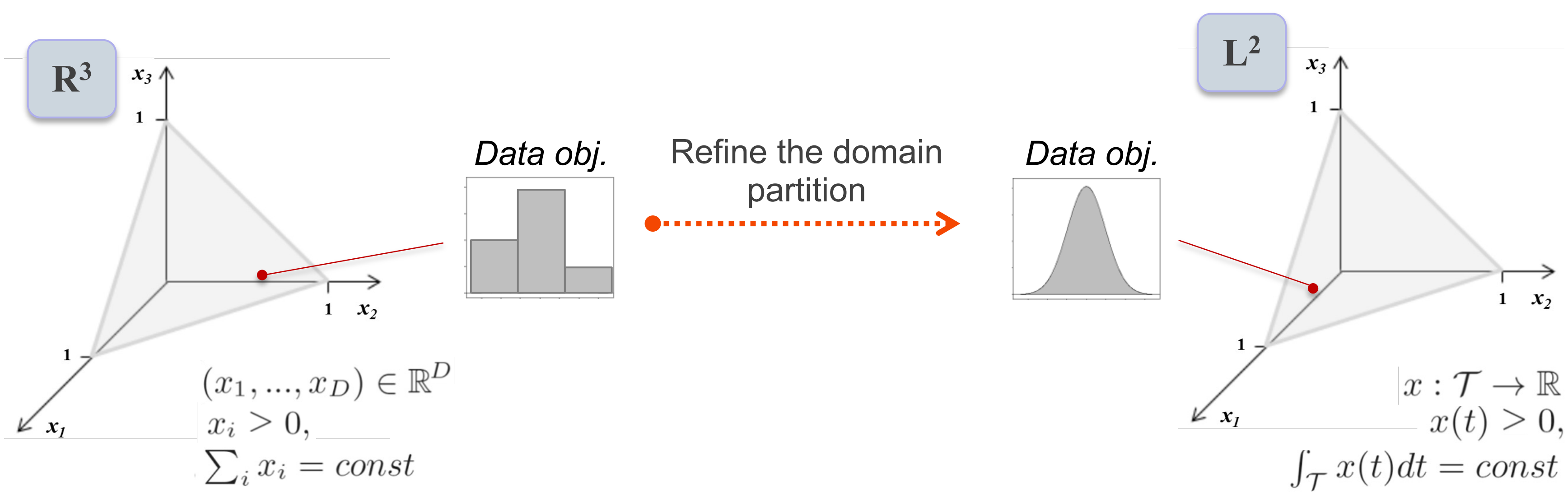
Projection on FPC_1
(residual variance: 9.44%)

Analysing density data with FDA

- Most methods in Functional Data Analysis implicitly assume that the data objects can be embedded in L^2
- However, **Density Data (DD)** are constrained functional data

The space of PDFs is not a linear (Hilbert) space when the L^2 geometry is used

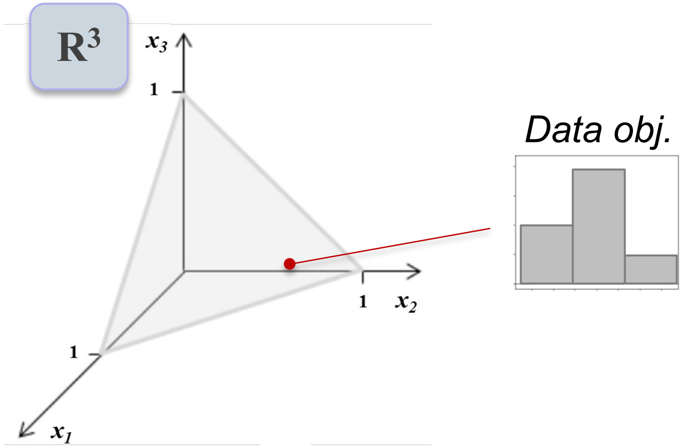
From multivariate compositions to DD



Historic Background: Compositions & Aitchison simplex

Three equivalent views on compositions

- **Equivalence classes of vectors** of amounts of each component
  $\{\boldsymbol{x} \in \Re^D : x_i > 0\}, \qquad \boldsymbol{x} =_A \boldsymbol{y} :\Leftrightarrow \exists c \in \Re^+ : \boldsymbol{x} = c\boldsymbol{y}$
- **The Simplex:** positive vectors adding to 1 aka portion of the total
  $$S^D = \left\{\boldsymbol{x} \in \Re^D : x_i > 0, \sum_i x_i = 1\right\}$$
- By a **discrete probability distribution**
  $\{P : P \text{ is a probablity measure with support } \{1, \dots, D\}\},$
  $x_i = P(\{i\})$



Historic Background: Compositions & Aitchison simplex

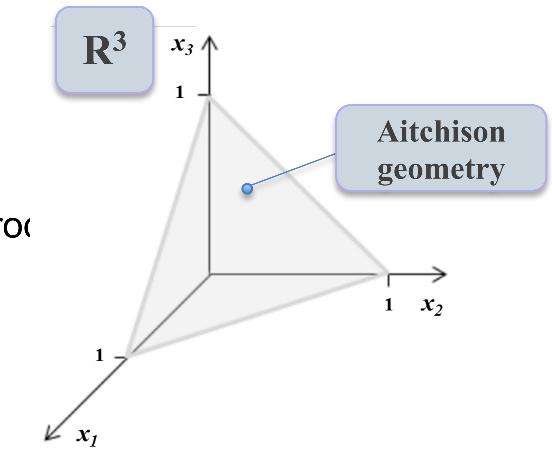
Aitchison geometry for the simplex

With this equality $(x_i) =_A (y_i) \Leftrightarrow \exists c \in \Re^+ : \forall i \colon x_i = c y_i$ and the two operations $\oplus$ and $\odot$ (Aitchison, 1984, Pawlowsky-Glahn, Egozcue, 2001)

- Perturbation $\underline{x}, \underline{y} \in S^D$: $\underline{x} \oplus \underline{y} = \mathcal{C}[(x_1 y_1, \ldots, x_D y_D)']$
- Powering $\underline{x} \in S^D, \alpha \in \mathbb{R}$: $\alpha \odot \underline{x} = \mathcal{C}[(x_1^\alpha, \ldots, x_D^\alpha)']$

the simplex is a vector space (Barcelò-Vidal et al. 2001), with scalar proc
defined as for $\underline{x}, \underline{y} \in S^D$,

$$\langle \underline{x}, \underline{y} \rangle = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \ln\left(\frac{x_i}{x_j}\right) \ln\left(\frac{y_i}{y_j}\right)$$

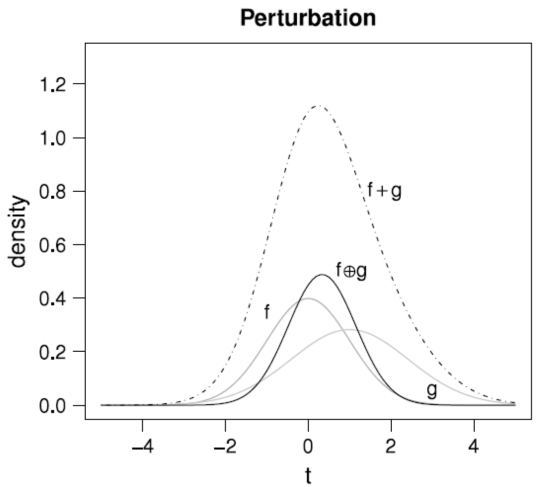


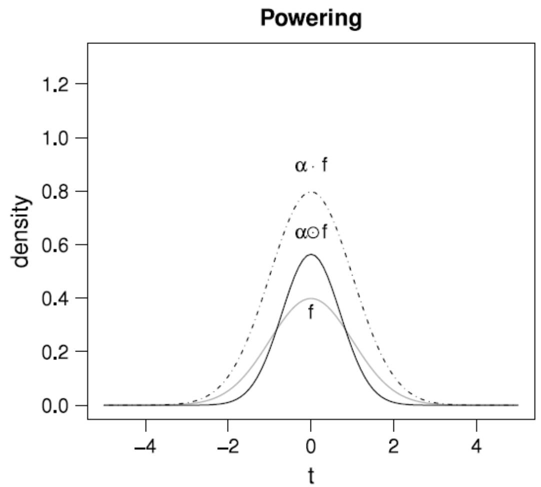
Note. Neutral element of perturbation: uniform composition, i.e., $\underline{e} = (1, \ldots, 1)' \in S^D$

Bayes Hilbert spaces

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): $(\alpha \odot f)(t) = \dfrac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$





Bayes Hilbert spaces

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): $(\alpha \odot f)(t) = \dfrac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$

Meaningful interpretations in mathematical statistics, e.g.,
- Perturbation $\oplus$ as a Bayes update of information
- Exponential families as affine finite-dimensional subspaces

Bayes spaces contain
- probability measures,
- improper priors
- and likelihoods

and behave like loglikelihoods

Bayes Hilbert spaces

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): $(\alpha \odot f)(t) = \dfrac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$

Meaningful interpretations in mathematical statistics, e.g.,

- Perturbation $\oplus$ as a Bayes update of information
- Exponential families as affine finite-dimensional subspaces

$$f(t,\boldsymbol{\alpha}) =_{\mathcal{B}} g(t) \cdot \exp\left\{\sum_{j=1}^{k} \vartheta_j(\boldsymbol{\alpha}) T_j(t)\right\}, \quad t \in I,$$

$$f(t,\boldsymbol{\alpha}) =_{\mathcal{B}} g(t) \oplus \bigoplus_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \odot \exp\{T_j(t)\}\right], \quad t \in I,$$

Extended exponential families

Bayes Hilbert spaces

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): $(\alpha \odot f)(t) = \dfrac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$

Bayes Hilbert spaces

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: $\langle f, g\rangle_{\mathcal{B}} = \dfrac{1}{2\eta}\displaystyle\int_I\int_I \ln\frac{f(t)}{f(s)}\ln\frac{g(t)}{g(s)}\,dt\,ds$

References

- *Egozcue, J.J., Dìaz-Barrero, J.L., Pawlowsky-Glahn, V., 2006. Hilbert space of probability density functions based on Aitchison geometry. Acta Math. Sin. (Engl. Ser.) 22 (4), 1175–1182.*
- *Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. Aust. N. Z. J. Stat. 56 (2), 171–194.*

Bayes Hilbert spaces

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: $\langle f, g\rangle_{\mathcal{B}} = \dfrac{1}{2\eta}\displaystyle\int_I \int_I \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)}\,dt\,ds$

- 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
- Could be defined for reference measures different from the Lebesgue on *I*

DDA in Bayes spaces

Bayes Hilbert space $B^2(I)$

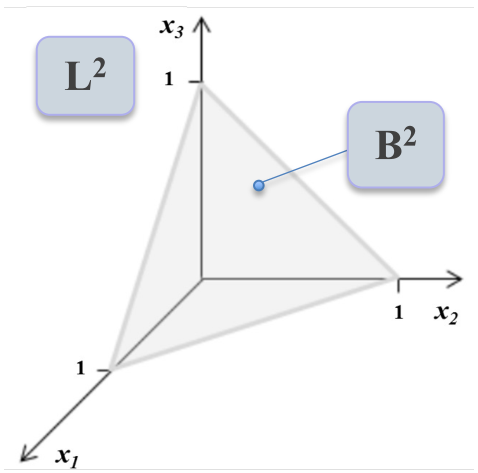
(Egozcue et al., 2006; van den Boogaart et al., 2014)

Space (of equivalence classes of) positive functions on I with square-integrable log

$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds},$$

$$(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \qquad t \in I.$$

$$\langle f, g\rangle_{\mathcal{B}} = \frac{1}{2\eta}\int_I\int_I \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)}\,dt\,ds$$



Strategy for the analysis: embedd the data in the Bayes space and here perform the statistical analysis

DDA in Bayes spaces

Bayes Hilbert space $B^2(I)$

(Egozcue et al., 2006; van den Boogaart et al., 2014)

Space (of equivalence classes of) positive functions on I with square-integrable log

$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds},$$

$$(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \qquad t \in I.$$

$$\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)}\,dt\,ds$$

Remarks:

- **Problems with zeros**: the B^2 space is suitable for strictly positive densities (zeros create problems with logs!)
- For simplicity, we focus on the space B^2 for densities with **support over a closed interval**
- Extensions exist to deal with infinite supports (the entire real line), which use different reference measures – not covered in this presentation

Strategy for the analysis: embedd the data in the Bayes space and here perform the statistical analysis

Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

Computational tricks

- B^2 is a Hilbert space. As such, it is isomorphic to L^2 and to ℓ^2
- Possible isometric isomorphisms:
  - Centered log-ratio (clr) transformation

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

    where η stands for the length of the interval PDF support *I.*
    Note: by construction clr-transformed data have zero integral

  - Projection over an orthonormal functional basis of B^2 and consideration of (a truncation of) the basis coefficients

$$x =_{\mathcal{B}} \bigoplus_{k=1}^{\infty} \langle x, u_k \rangle_{\mathcal{B}} \odot u_k$$

- Computations of operations and inner products in B^2 can be performed by relying on routines for L^2 based on clr-transformations, or multivariate routines based on basis coefficients

Computational tricks

- B^2 is a Hilbert space. As such, it is isomorphic to L^2 and to ℓ^2
- Possible isometric isomorphisms:

  - Centered log-ratio (clr) transformation

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

    where η stands for the length of the interval PDF support *I.*
    Note: by construction clr-transformed data have zero integral

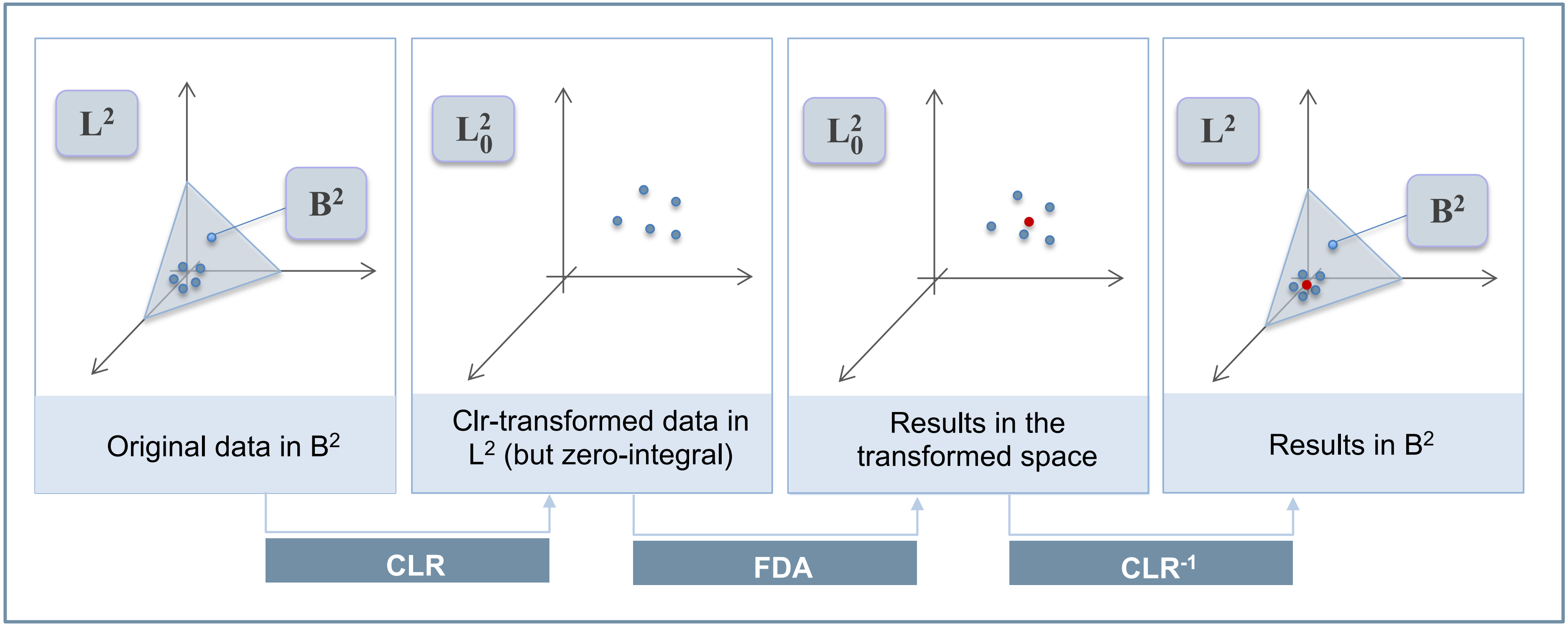
  - Projection over an orthonormal functional basis of B^2 and consideration of (a truncation of) the basis coefficients

$$x =_{\mathcal{B}} \bigoplus_{k=1}^{\infty} \langle x, u_k \rangle_{\mathcal{B}} \odot u_k$$

- Computations of operations and inner products in B^2 can be performed by relying on routines for L^2 based on clr-transformations, or multivariate routines based on basis coefficients

Using the clr transformation

- We can use the clr transformation to perform computations in the Bayes space using convenient formulations in L^2 .



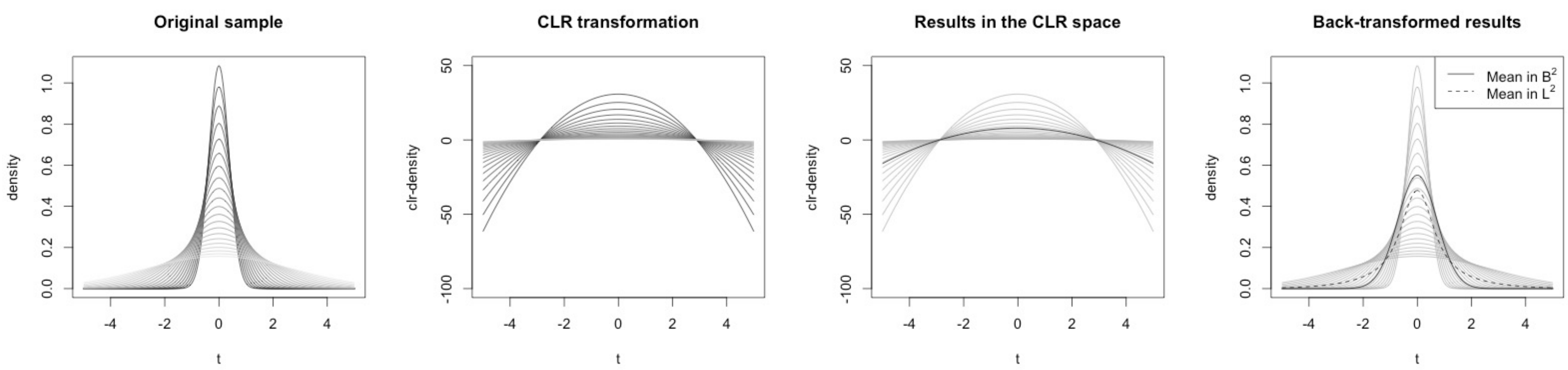
Using the clr transformation

- We can use the clr transformation to perform computations in the Bayes space using convenient formulations in L^2 .

Example: Computing the Frechét mean in Bayes spaces

The sample mean function in L^2 can be computed point-wise, in B^2 it has to be computed as Fréchet sample mean. We can thus map the data in L^2, make computations on clr-transformed data and map back the results in B^2. Let $X_1,\dots,X_N$ be i.i.d. random element in B^2

$$\overline{X} = clr^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} clr(X_i)\right)$$



Using the clr transformation

- We can use the clr transformation to perform computations in the Bayes space using convenient formulations in L^2 .

Example: Computing the sample covariance operator in Bayes spaces

The covariance operator in L^2 can be computed through a kernel, which consists of point-wise covariances among evaluations of the functions, i.e. (with zero-mean r.f.) $c(t,s) = \mathbb{E}[X(t)X(s)]$:

$$Cx(t) = \mathbb{E}[\langle X, x\rangle X] = \int c(t,s)x(s)ds \quad \text{for } x \in L^2$$

Similarly, the sample covariance operator can be obtained from the sample estimate of the kernel:

$$Sx(t) = \frac{1}{N}\sum_{i=1}^{N}\langle X_i, x\rangle X_i = \int \hat{c}(t,s)x(s)ds \quad \text{for } x \in L^2$$

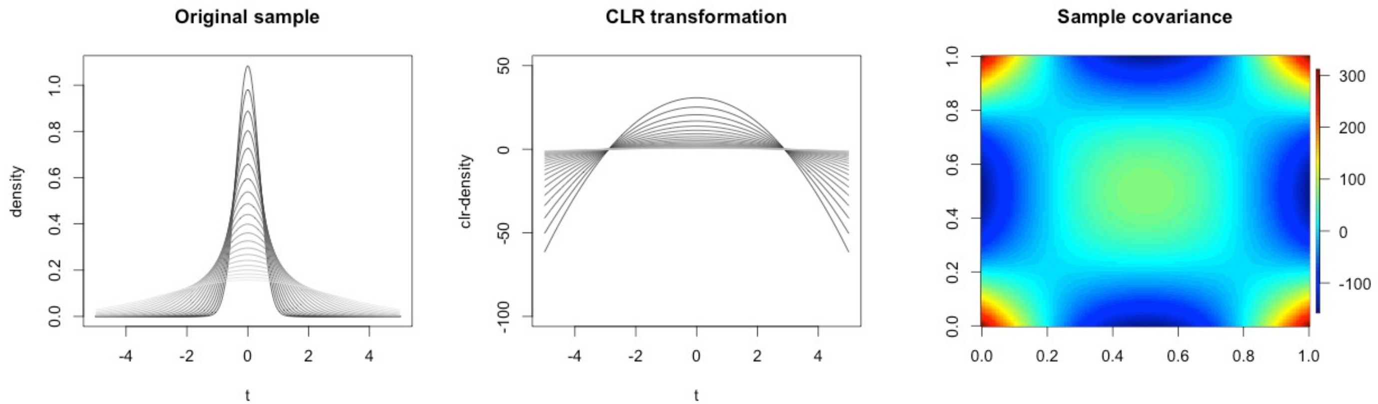
Using the clr transformation

- We can use the clr transformation to perform computations in the Bayes space using convenient formulations in L^2 .

Example: Computing the sample covariance operator in Bayes spaces

Using the clr transformation, the kernel representation can be used in B^2 too. Let $X_1,\dots,X_N$ be zero-mean i.i.d. random element in B^2 , S be their sample covariance operator, and S_{clr} the sample covariance operator of the clr-transformed data. Then:

$$\langle Sx, y\rangle_{B^2} = \langle S_{clr}\tilde{x}, \tilde{y}\rangle_{L^2} = \int \hat{c}_{clr}(t,s)\tilde{x}(s)\tilde{y}(t)dsdt$$



Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

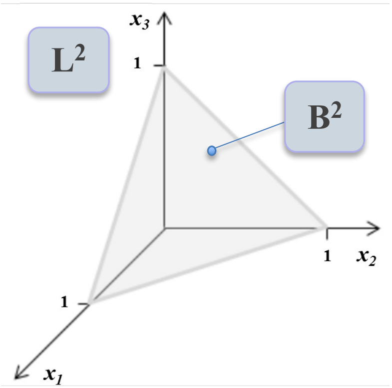
Simplicial Functional Principal Components

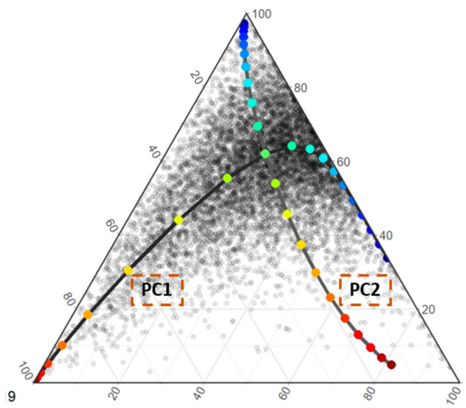
Problem: Functional Principal Component Analysis in B^2

Given a dataset of N zero-mean functional observations in H, $X_1, \ldots, X_N$, find the directions of maximum variability (in B^2) of the dataset, i.e., those maximizing

$$\frac{1}{N}\sum_{i=1}^{N}\langle \xi, X_i\rangle_{\mathcal{B}}^2 \quad \text{subject to} \quad \|\xi\|_{\mathcal{B}} = 1$$
$$\langle \xi_k, \xi\rangle_{\mathcal{B}} = 0, \quad j < k$$

Maximum sample variance **Orthonormality in B^2**





Simplicial Functional Principal Components

Problem: Functional Principal Component Analysis in B^2

Given a dataset of N zero-mean functional observations in H, $X_1, \ldots, X_N$, find the directions of maximum variability (in B^2) of the dataset, i.e., those maximizing

$$\frac{1}{N}\sum_{i=1}^{N}\langle \xi, X_i\rangle^2_{\mathcal{B}} \text{ subject to } \|\xi\|_{\mathcal{B}} = 1$$

$$\langle \xi_k, \xi\rangle_{\mathcal{B}} = 0, \quad j < k$$

Maximum sample variance **Orthonormality in B^2**

Solution

Call S the sample covariance operator of $X_1, \ldots, X_N$. Then, the functional principal components $\xi_1, \ldots, \xi_{N-1}$ are found as the eigenfunctions of the operator S, i.e., they solve the eigen-equations

$$S\xi_k = \lambda_k \xi_k$$

The eigenvalue λ_k associated with the eigenfunction ξ_k represents the variability along the direction ξ_k .

We call *functional score* x_{ik} the projection of the observation X_i along the direction ξ_k , i.e.,

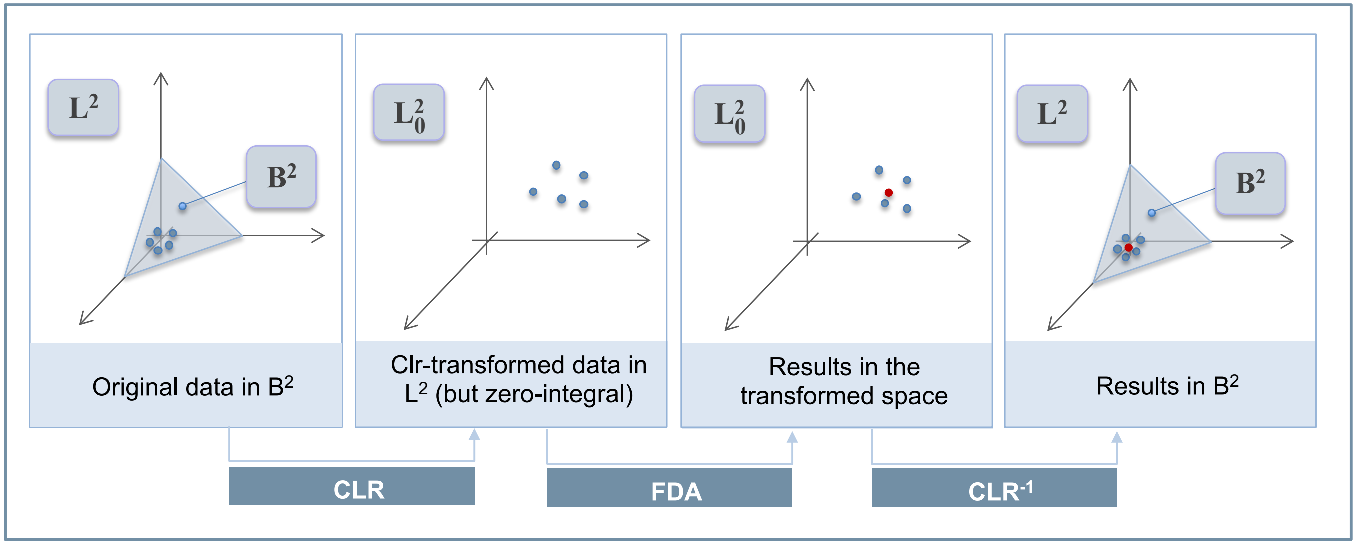
$$x_{ik} = \langle X_i, \xi_k\rangle$$

Simplicial Functional Principal Components

In practice, to perform SFPCA we can:

1. Transform the data to clr
2. Estimate the covariance operator from clr transformed data (in L^2) via the sample estimator
3. Compute the eigenfunctions of the sample covariance operator
4. Back-transform the results to get the eigenfunctions in B^2

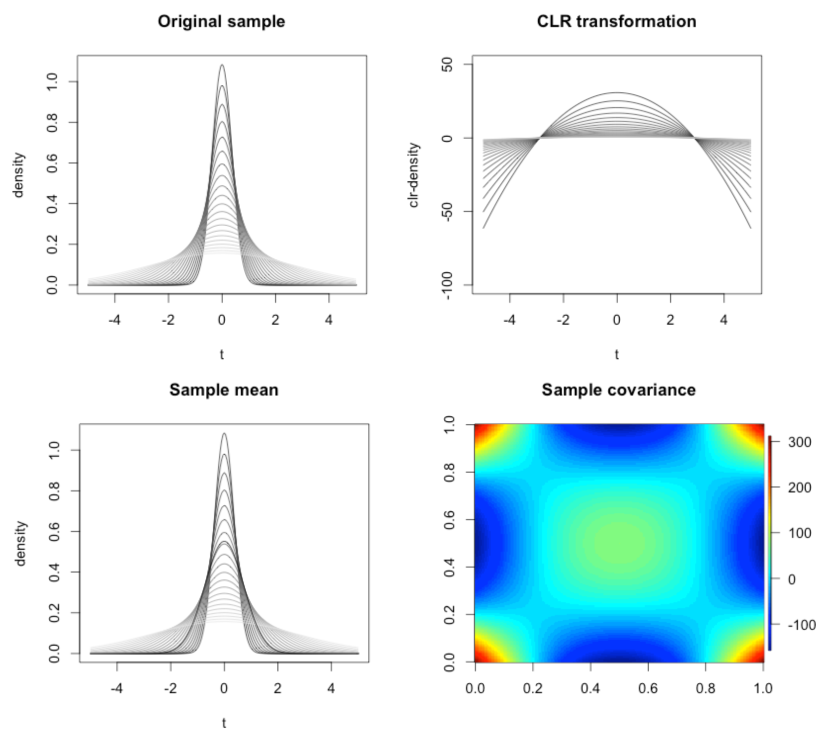
Note. The functional scores in B^2 and L^2 coincide, as well as the eigenvalues



An example with Gaussian densities

Example 1: Dataset of Gaussian distributions

Consider a set of Gaussian distributions with zero-mean and variance $\sigma_i = \exp(-1 + (i-1)/10)$ for $i = 1, \ldots, 21$. Note that we already computed the mean and covariance functions for this sample.

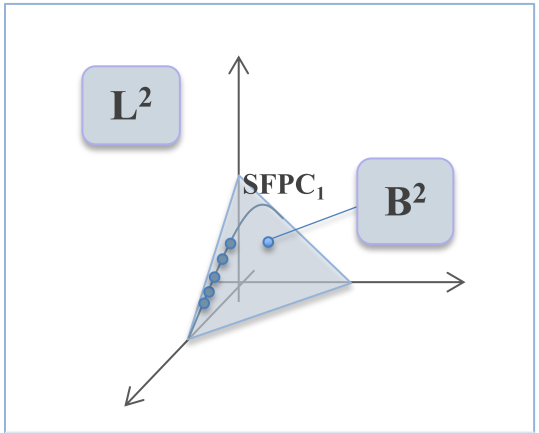


An example with Gaussian densities

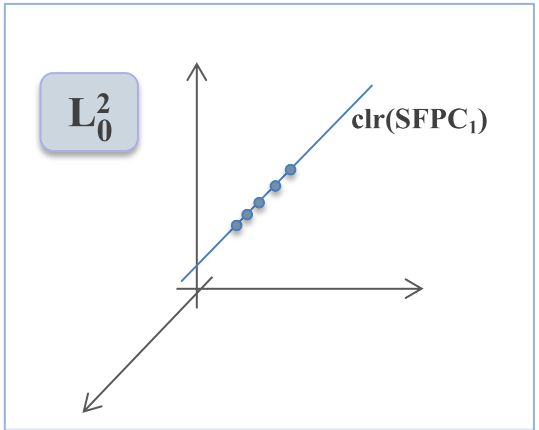
Example 1: Dataset of Gaussian distributions

The Gaussian distribution with zero-mean belongs to a **1-parametric** exponential family

➤ We expect to single out a single direction of variability (related with the varying parameter σ_i)



Original data in B^2



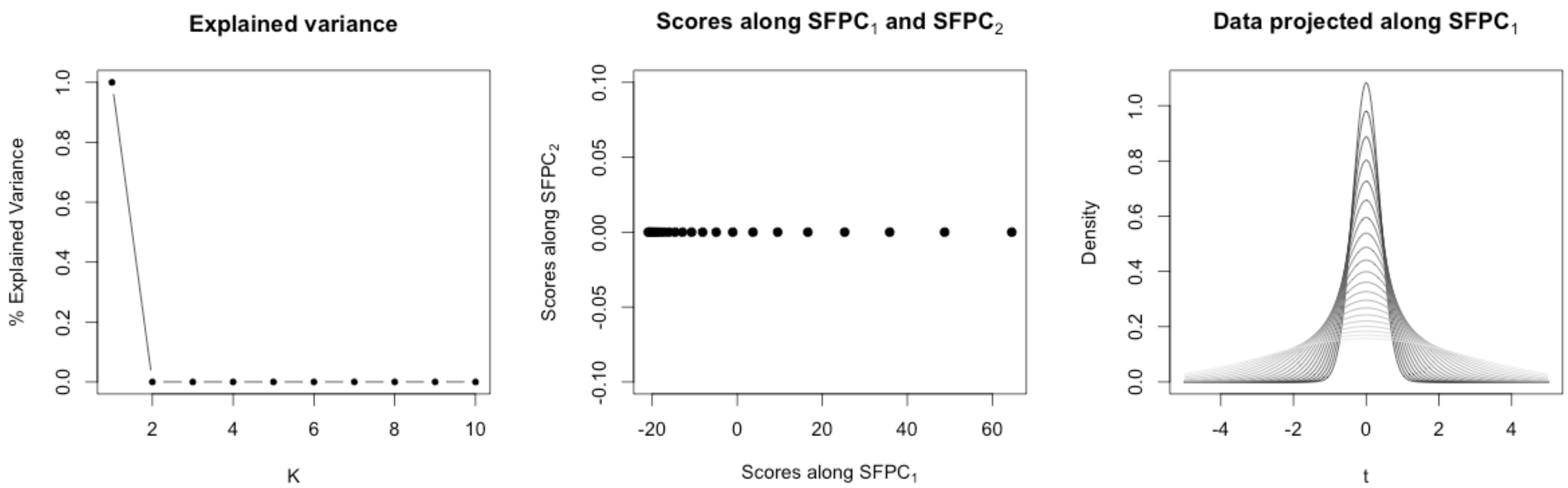
Clr-transformed data in L^2

An example with Gaussian densities

Example 1: Dataset of Gaussian distributions

The Gaussian distribution with zero-mean belongs to a **1-parametric** exponential family

➤ We expect to single out a single direction of variability (related with the varying parameter σ_i)



Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

Functional Linear Models for DD

- In FDA, linear models can be described by three families of models, depending on the type of predictors and response involved

From Classical to Functional Linear Models

	Scalar Covariates	Functional Covariates
Scalar Response	$x_i = \sum_{l=1}^{L} z_{il}\beta_l + \varepsilon_i$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in \mathbb{R}$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=0}^{L} \langle z_{il}, \beta_l \rangle + \varepsilon_i(t)$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in L^2$ $z_{il} \in L^2$
Functional Response	$x_i = \sum_{l=1}^{L} z_{il}\beta_l + \varepsilon_i$ $x_i, \varepsilon_i \in L^2$ $\beta_l \in L^2$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=1}^{L} \mathcal{B}_l z_{il} + \varepsilon_i$ $x_i, \varepsilon_i \in L^2$ $(\mathcal{B}_l z)(t) = \beta_l(t)z(t)dt$ or $(\mathcal{B}_l z)(t) = \int_I \beta_l(t,s)z(s)ds$ $z_{il} \in L^2$

Functional Linear Models for DD

- In FDA, linear models can be described by three families of models, depending on the type of predictors and response involved
- In Bayes spaces, most of these models can be developed, but operations and inner products need to be interpreted consistently
- Models available in the literature include: density-on-scalar, scalar-on-density, density-on-density regressions (*Talska et al., CSDA 2018; Talska et al., CSDA 2021, Scimone et al., SPASTA 2022*)

Functional Linear Models in B²

	Scalar Covariates	**Density Covariates**
Scalar Response	$x_i = \sum_{l=1}^{L} z_{il}\beta_l + \varepsilon_i$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in \mathbb{R}$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=0}^{L} \langle z_{il}, \beta_l \rangle + \varepsilon_i(t)$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in B^2$ $z_{il} \in B^2$
Density Response	$x_i = \bigoplus_{l=1}^{L} z_{il} \odot \beta_l \oplus \varepsilon_i$ $x_i, \varepsilon_i \in B^2$ $\beta_l \in B^2$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=1}^{L} \mathcal{B}_l z_{il} \oplus \varepsilon_i$ $x_i, \varepsilon_i \in B^2$ $\mathcal{B}_l \in \mathcal{L}(B^2, B^2)$ $z_{il} \in B^2$

Functional Linear Models for DD

- In FDA, linear models can be described by three families of models, depending on the type of predictors and response involved
- In Bayes spaces, most of these models can be developed, but operations and inner products need to be interpreted consistently
- Models available in the literature include: density-on-scalar, scalar-on-density, density-on-density regressions (*Talska et al., CSDA 2018; Talska et al., CSDA 2021, Scimone et al., SPASTA 2022*)

Functional Linear Models in B^2

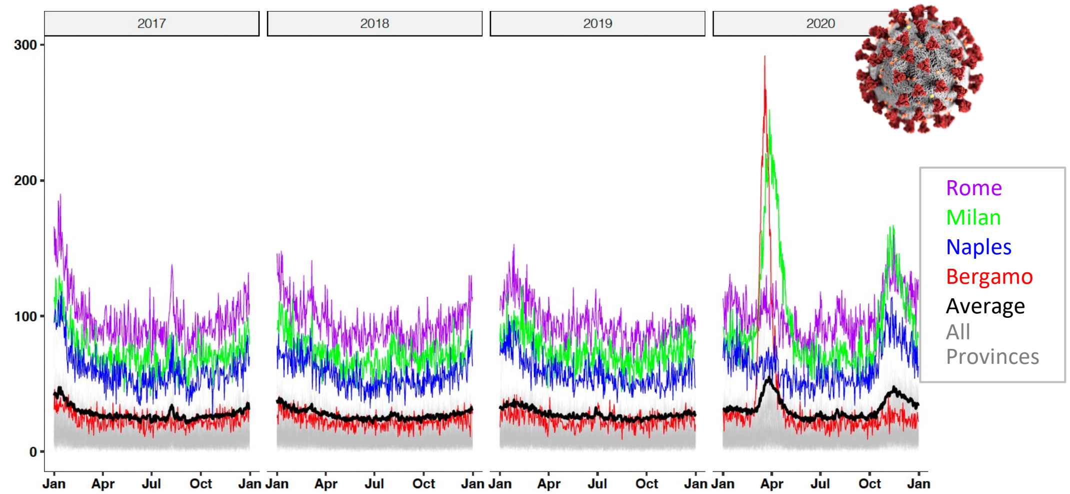
	Scalar Covariates	Density Covariates
Scalar Response	$x_i = \sum_{l=1}^{L} z_{il}\beta_l + \varepsilon_i$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in \mathbb{R}$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=0}^{L} \langle z_{il}, \beta_l \rangle + \varepsilon_i(t)$ $x_i, \varepsilon_i \in \mathbb{R}$ $\beta_l \in B^2$ $z_{il} \in B^2$
Density Response	$x_i = \bigoplus_{l=1}^{L} z_{il} \odot \beta_l \oplus \varepsilon_i$ $x_i, \varepsilon_i \in B^2$ $\beta_l \in B^2$ $z_{il} \in \mathbb{R}$	$x_i = \sum_{l=1}^{L} \mathcal{B}_l z_{il} \oplus \varepsilon_i$ $x_i, \varepsilon_i \in B^2$ $\mathcal{B}_l \in \mathcal{L}(B^2, B^2)$ $z_{il} \in B^2$

A case study on mortality densities

Example: Analysis of Mortality data in Italian Provinces

- Counts of daily deaths *from all causes* in Italian Provinces

109 Italian Provinces



Daily death counts over 2017-2020, **70+ years**

Scimone, Menafoglio, Sangalli, Secchi (SPASTA, 2022)

A case study on mortality densities

Example: Analysis of Mortality data in Italian Provinces

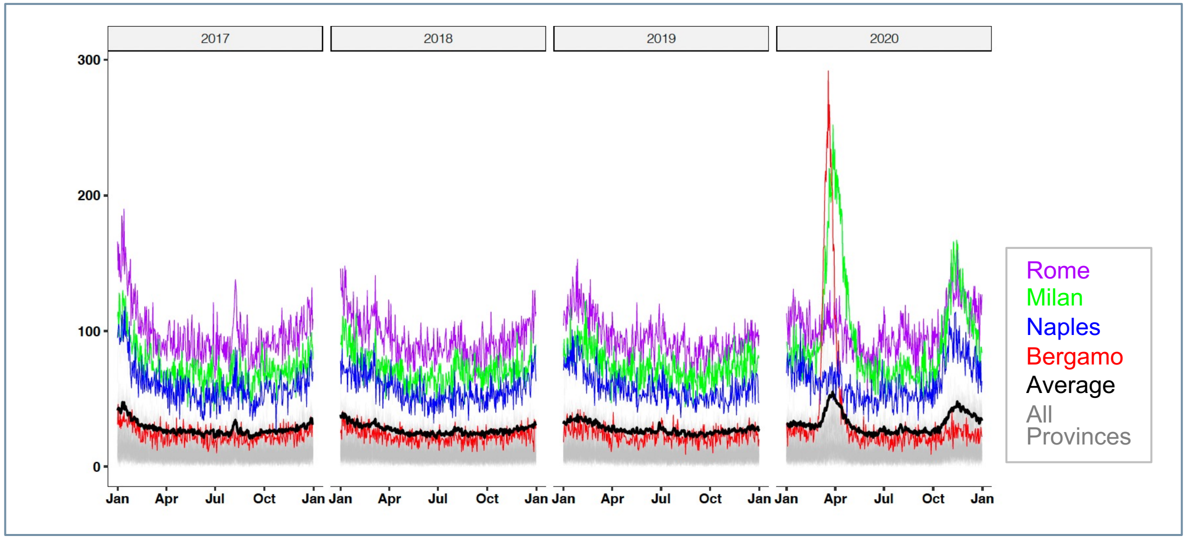
- Counts of daily deaths *from all causes* in Italian Provinces

- ➤ **A proxy** of immediate impact of the pandemic on Italian communities
- ➤ **High-quality** data (source: ISTAT) at a very fine space-time scale
- ➤ **Not affected by varying definitions or testing capabilities**
- ➤ **Integrate the direct and indirect effects of the pandemic shock,** registering the consequences of the containment policies, as well as the disruption at the local level of the health and welfare system overwhelmed by the struggle against COVID-19.

Scimone, Menafoglio, Sangalli, Secchi (SPASTA, 2022)

A case study on mortality densities

Example: Analysis of Mortality data in Italian Provinces

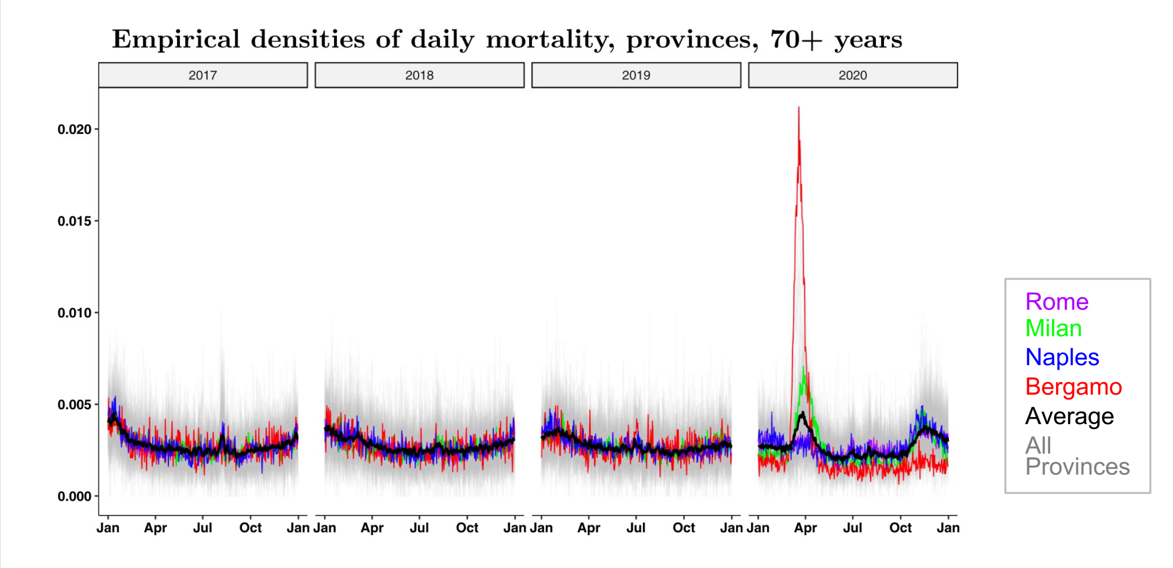


- If data are just considered on an **absolute scale**, highly populated provinces will mask the entire variability of the data
- If just considering the total over the year, the dynamic along year is lost

A case study on mortality densities

Example: Analysis of Mortality data in Italian Provinces

- **Data object**: the distribution (probability density function - PDF) of the variable *time of death* along the year, considered as an element of B^2
- **Relative scale**: we do not want to look at the absolute counts, but on the relative proportion of deaths within each Province



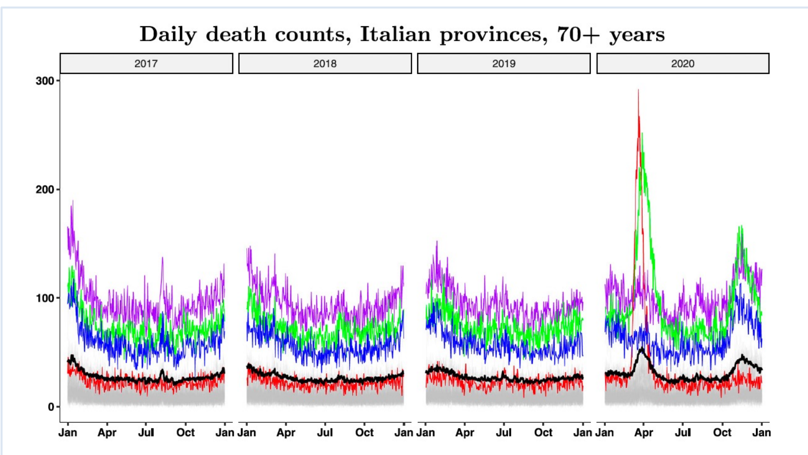
A case study on mortality densities

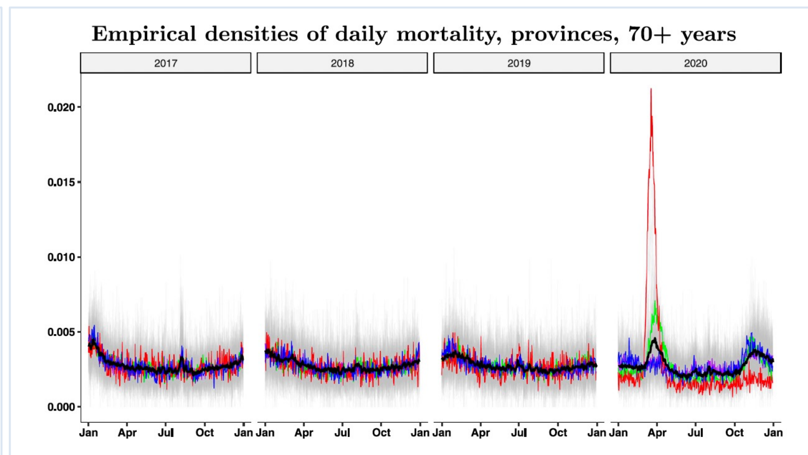
Example: Analysis of Mortality data in Italian Provinces

- **Data object**: the distribution (probability density function - PDF) of the variable *time of death* along the year, considered as an element of B^2
- **Relative scale**: we do not want to look at the absolute counts, but on the relative proportion of deaths within each Province

Equivalence relation: f and g are equivalent if they are proportional (scale invariance)

Once embedded in B^2, the scale of the phenomenon is of no relevance!





Predictable and unpredictable components for mortality densities

Predictable component of mortality

Idea:

We fit a ***linear model*** *for the mortality density at year y (response variable, a density in B^2), based on the average mortality density in the previous 4 years (predictors, a density in B^2)*

Predictable and unpredictable components for mortality densities

Predictable component of mortality

Linear prediction model targeting the mortality density at year y given the average mortality densities from years *y-1, …, y-4*

$$f_{iy} = \beta_{0,y} \oplus \mathcal{B}_y \bar{f}_{iy} \oplus \varepsilon_i$$

$$\bar{f}_{iy} = \frac{1}{n} \odot \bigoplus_{r=y-4}^{y-1} f_{ir}$$

Predictable and unpredictable components for mortality densities

Predictable component of mortality

Linear prediction model targeting the mortality density at year y given the average mortality densities from years *y-1, …, y-4*

$$f_{iy} = \beta_{0,y} \oplus \mathcal{B}_y \bar{f}_{iy} \oplus \varepsilon_i$$

$$\bar{f}_{iy} = \frac{1}{n} \odot \bigoplus_{r=y-4}^{y-1} f_{ir}$$

Predictable component from previous years

Forecast for year y+1 with model calibrated at year y

$$\hat{f}_{i,y+1} = \hat{\beta}_{0,y} \oplus \hat{\mathcal{B}}_y \bar{f}_{i,y+1}$$

Predictable and unpredictable components for mortality densities

Predictable component of mortality

Linear prediction model targeting the mortality density at year y given the average mortality densities from years *y-1, …, y-4*

$$f_{iy} = \beta_{0,y} \oplus \mathcal{B}_y \bar{f}_{iy} \oplus \varepsilon_i$$
$$\bar{f}_{iy} = \frac{1}{n} \odot \bigoplus_{r=y-4}^{y-1} f_{ir}$$

Predictable component from previous years

Forecast for year y+1 with model calibrated at year y

$$\hat{f}_{i,y+1} = \hat{\beta}_{0,y} \oplus \hat{\mathcal{B}}_y \bar{f}_{i,y+1}$$

Regression analysis in practice

- Data smoothed in CLR space on a cubic splines basis in B^2 (Machalova et al (2015, 2020)), with a penalization on the second derivative (52 evenly spaced internal knots).
- Linear model fitted by ordinary least squares, using the cubic spline coefficients (on CLRs)

$$\boxed{f_{iy}^c(t)} = \boxed{\beta_{0,y}^c(t)} + \int_I \boxed{\beta_y(t,s)}\boxed{\bar{f}_{iy}^c(s)}ds + \boxed{\varepsilon_i^c(t)}$$
$$\boxed{\bar{f}_{iy}^c} = \frac{1}{n} \sum_{r=y-4}^{y-1} \boxed{f_{ir}^c}$$

Expressed on a B-spline basis

Note: penalized regression is also possible

Predictable and unpredictable components for mortality densities

Predictable component of mortality

Linear prediction model targeting the mortality density at year y given the average mortality densities from years *y-1, …, y-4*

$$f_{iy} = \beta_{0,y} \oplus \mathcal{B}_y \bar{f}_{iy} \oplus \varepsilon_i$$

$$\bar{f}_{iy} = \frac{1}{n} \odot \bigoplus_{r=y-4}^{y-1} f_{ir}$$

Predictable component from previous years

Forecast for year y+1 with model calibrated at year y

$$\hat{f}_{i,y+1} = \hat{\beta}_{0,y} \oplus \hat{\mathcal{B}}_y \bar{f}_{i,y+1}$$

Model error

We consider the prediction error of the model above, i.e., the difference between the observed mortality density, and its forecast from the model calibrated at year y

Unpredictable component from previous years

Prediction error (not the residuals!)

$$\delta_{i,y+1} = f_{i,y+1} \ominus \hat{f}_{i,y+1}$$

Predictable and unpredictable components for mortality densities

Predictable component of mortality

Linear prediction model targeting the mortality density at year y given the average mortality densities from years *y-1, …, y-4*

$$f_{iy} = \beta_{0,y} \oplus \mathcal{B}_y \bar{f}_{iy} \oplus \varepsilon_i$$

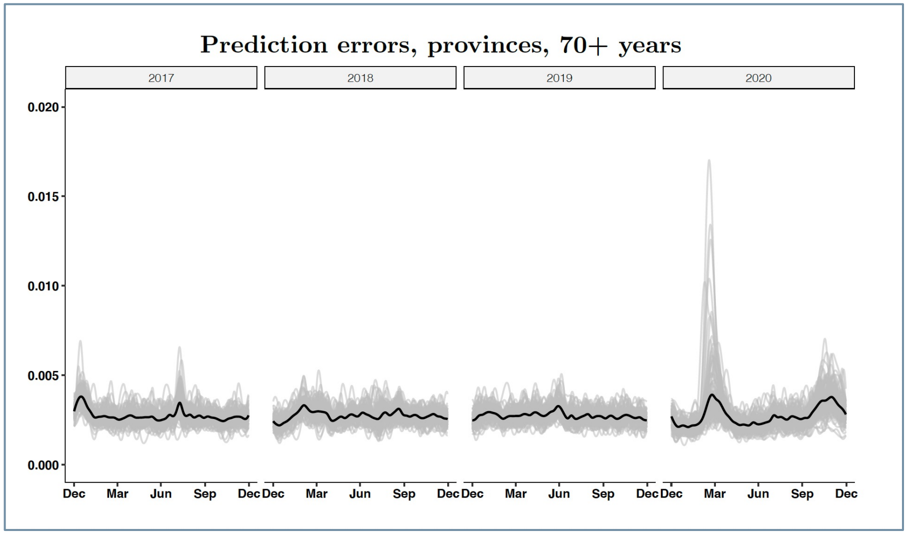
$$\bar{f}_{iy} = \frac{1}{n} \odot \bigoplus_{r=y-4}^{y-1} f_{ir}$$

Predictable component from previous years

Forecast for year y+1 with model calibrated at year y

$$\hat{f}_{i,y+1} = \hat{\beta}_{0,y} \oplus \hat{\mathcal{B}}_y \bar{f}_{i,y+1}$$

Model error

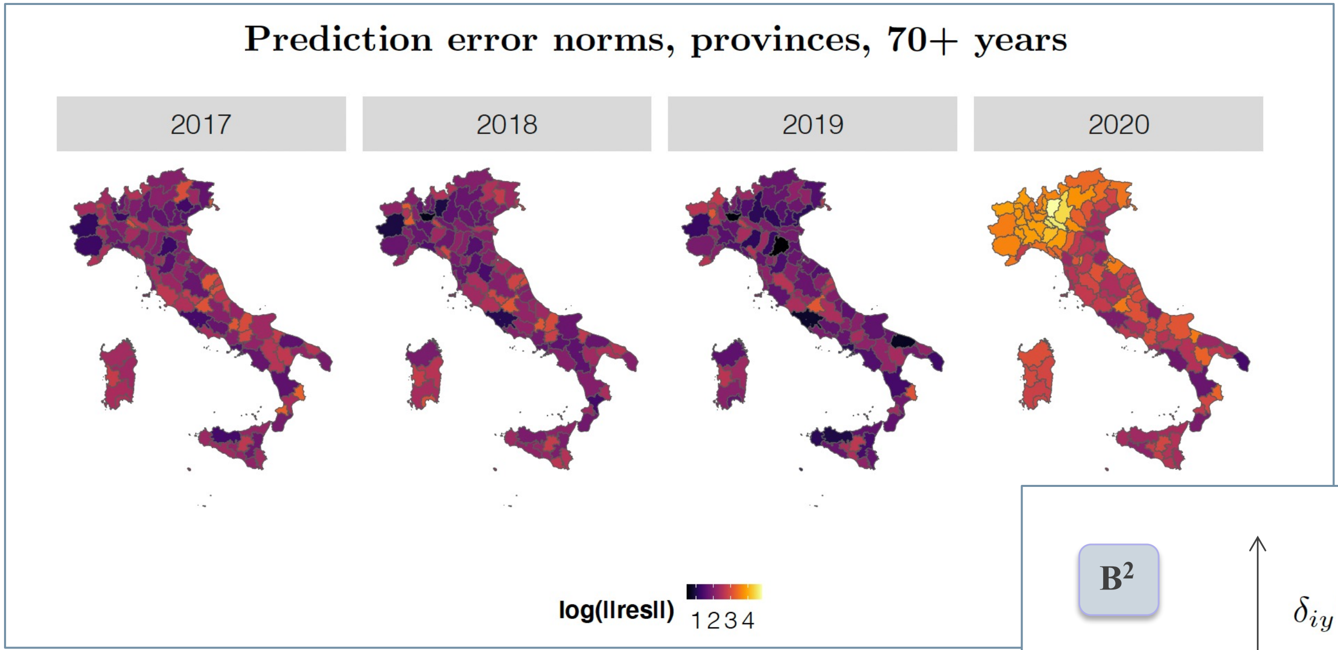


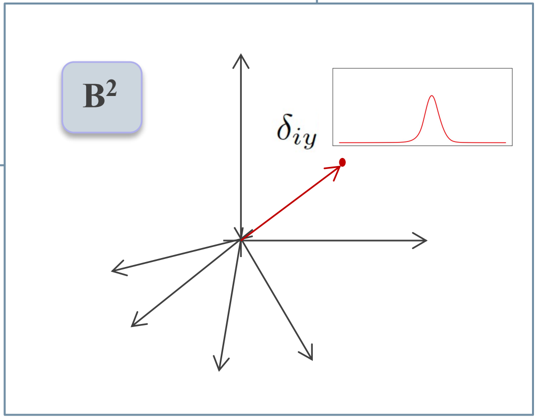
Note. The "0" in B^2 is the uniform distribution

Interpreting the unpredictable component

Unpredictable component: prediction error from the model

$$\delta_{i,y+1} = f_{i,y+1} \ominus \hat{f}_{i,y+1}$$



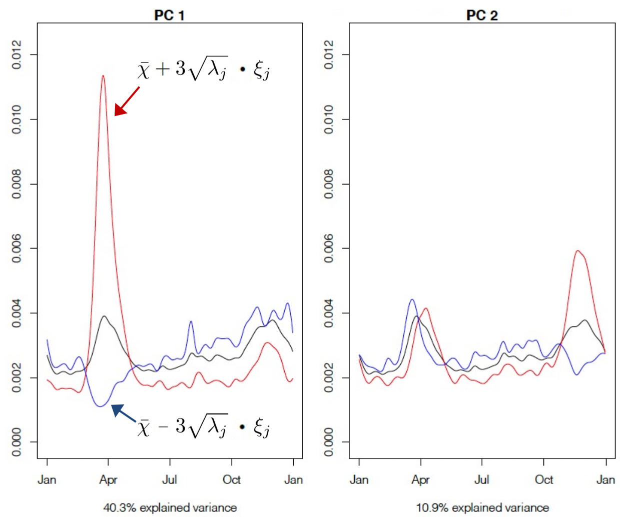


There is indeed a spatial structure in the "size" of the errors

Interpreting the unpredictable component

Interpretation of the first two SFPCs

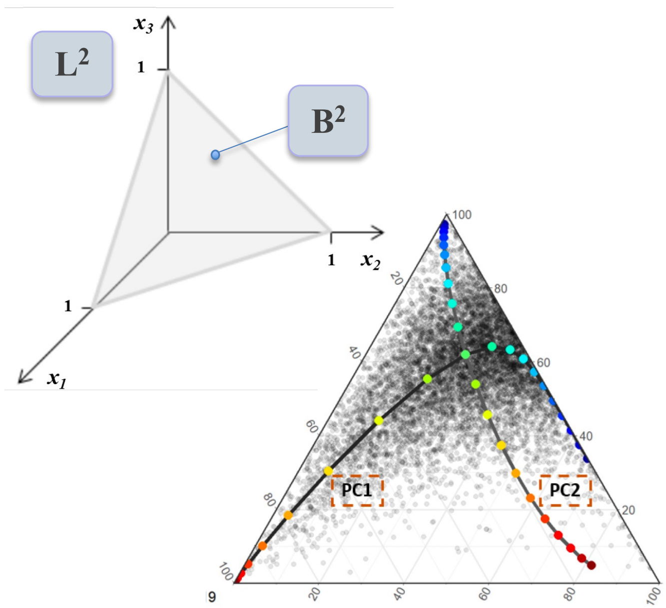
Mean perturbed by the PCs



First wave *Second wave*

SFPCs seems to be associated with the first two waves of Covid pandemic

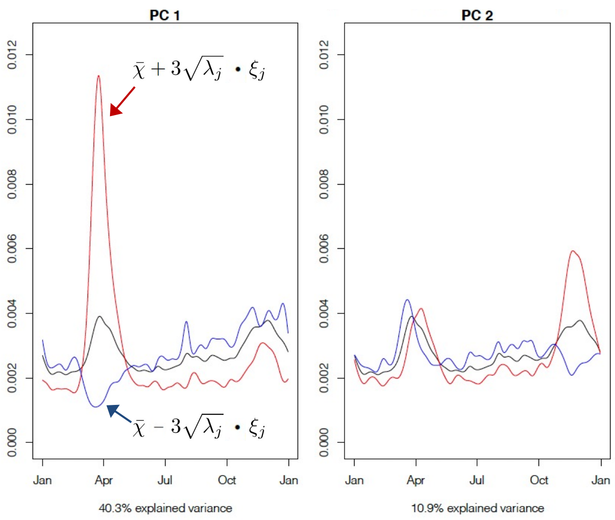
Recall: SFPCs represent direction of max variability in the infinite dimensional simplex B^2



Interpreting the unpredictable component

Interpretation of the first two SFPCs

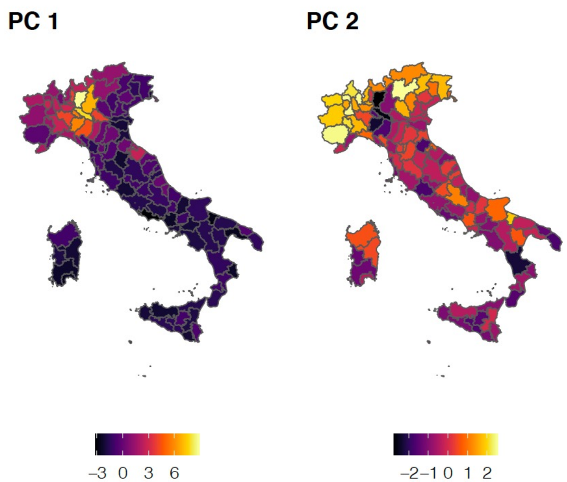
Mean perturbed by the PCs



First wave *Second wave*

SFPCs seems to be associated with the first two waves of Covid pandemic

Maps of the PC scores

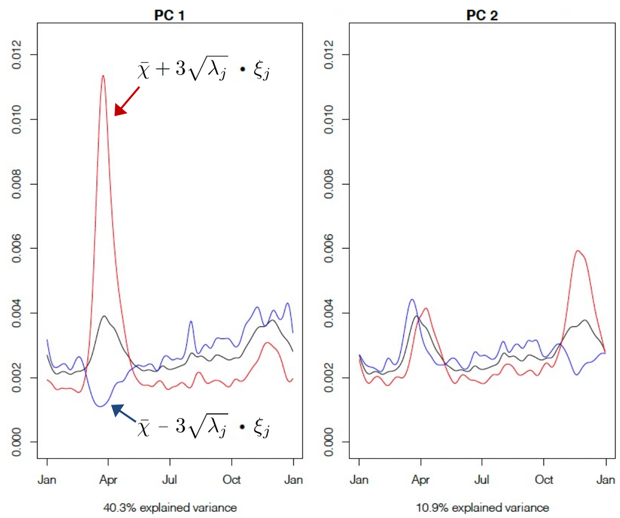


These components have a clear spatial structure

Interpreting the unpredictable component

Interpretation of the first two SFPCs

Mean perturbed by the PCs



First wave *Second wave*

SFPCs seems to be associated with the first two waves of Covid pandemic

Concluding remarks on the case study

- The death process has been critically perturbed in 2020
- The compositional approach provides novel insights, and allow for consistent analyses
- The unpredictable component of the death process is consistently attributed to the pandemic waves via SFPCA decomposition in B^2.

Perspectives and developments

Apply the analysis pipeline to different data, with "early warning" purposes

Codes to replicate the analysis are available at:
https://github.com/RiccardoScimone/Mortality-densities-italy-analysis

Agenda

- Functional Data Analysis in Hilbert spaces
- The Bayes space approach to the analysis of Density Data in FDA
- DDA in practice: the CLR transformation
- A few example of methods:
  - Simplicial Functional Principal Component Analysis
  - Linear Models in B^2
- Conclusions

Conclusions and take home messages

- The B^2 space is a suitable embedding space for density functions, and enjoys of nice properties widely-studied in compositional data analysis
- Functional data analysis of density functions is indeed possible in the B^2 space, by using a geometric approach (Hilbert space model)
- Computations of interesting quantities is eased by transformations (e.g., clr) and bases representations
- Bases representation can also be used to smooth the data and find a functional representation from raw (discrete) data
- Methods and approaches here presented can be extended to the multivariate DD framework

Selected references

- Egozcue, J.J., Pawlowsky-Glahn, V., Tolosana-Delgado, R., Ortego, M.I., van den Boogaart, K.G., 2013. Bayes spaces: use of improper distributions and exponential families. Rev. R. Acad. Cienc. Exactas F.s. Nat. Ser. A Mat. 107, 475–486.
- Egozcue, J.J., Dìaz-Barrero, J.L., Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. Acta Mathematica Sinica (English Series) 22(4), 1175-1182.
- Horvath, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. In: Springer Series in Statistics, Springer.
- Hron, K., Menafoglio, A., Templ, M., Hruzova, C., Filzmoser, P. (2016): "Simplicial principal component analysis for density functions in Bayes spaces", Computational Statistics & Data Analysis, 94, 330–350.
- Menafoglio, A., Guadagnini, A., Secchi, P., 2014. A kriging approach based on Aitchison geometry for the characterization of particle-size curves in Heterogeneous Aquifers", Stochastic Environmental Research and Risk Assessment, 28(7), 1835–1851.
- Menafoglio, A., Grasso, M., Secchi, P., Colosimo, B.M. (2018): "Profile Monitoring of Probability Density Functions via Simplicial Functional PCA with application to Image Data", Technometrics, 60(4), 497-510.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. Stoch. Environ. Res. Risk Assess. 15 (5), 384–398.
- R. Scimone, A. Menafoglio, L. M. Sangalli, P. Secchi (2022): "A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities". Spatial Statistics, 49, 100541.
- Talská, R., Menafoglio, A., Hron, K., Egozcue, J.J., Palarea-Albaladejo, J. (2020) "Weighting the domain of probability densities in functional data analysis", Stat 9 (1), e283, 2020. DOI: 10.1002/sta4.283.
- Talská, R., Menafoglio, A., Machalová, J., Hron, K., Fišerová, E.: Compositional regression with functional response. Computational Statistics and Data Analysis 123, 66-85, 2018.
- Talská, R., Hron, K., Matys Grygar, T.: Compositional scalar-on-function regression with application to sediment particle size distributions. Mathematical Geosciences 53, 1667-1695, 2021.
- Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2010. Bayes linear spaces. Statist. Oper. Res. Trans. 34 (2), 201–222.
- Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. Aust. N. Z. J. Stat. 56 (2), 171–194.