An Introduction to Functional Data Analysis for Density Functions in Bayes spaces

Alessandra Menafoglio[1*], Karel Hron[2], Jitka Machalova[2]

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

[2]Department of Mathematical Analysis and Application of Mathematics, Faculty of Science, Palacky University Olomouc

*alessandra.menafoglio@polimi.it

INTRODUCTION

About us

Alessandra Menafoglio
Politecnico di Milano

Karel Hron
Palacky University Olomouc

Jitka Machalova
Palacky University Olomouc

An Introduction to Functional Data Analysis for Density Functions in Bayes spaces

Functional data & Functional data analysis

- **Functional data** are entities that can be described through a function, e.g., a curve, a surface, an image
- The observed values reflect a **smooth variation of the phenomenon**.
- **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

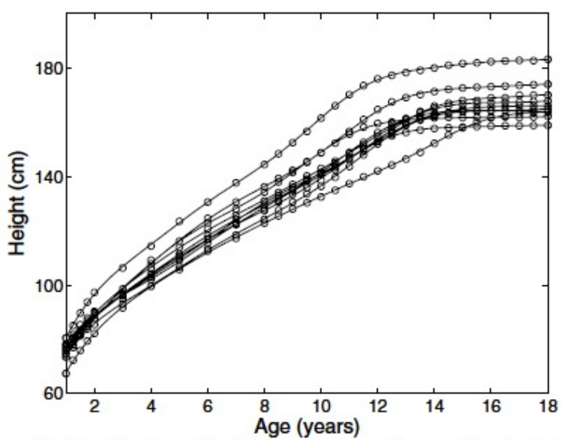


Figure 1.1. The heights of 10 girls measured at 31 ages. The circles indicate the unequally spaced ages of measurement.

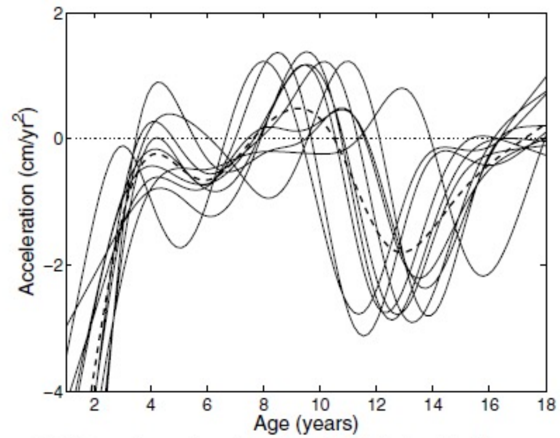
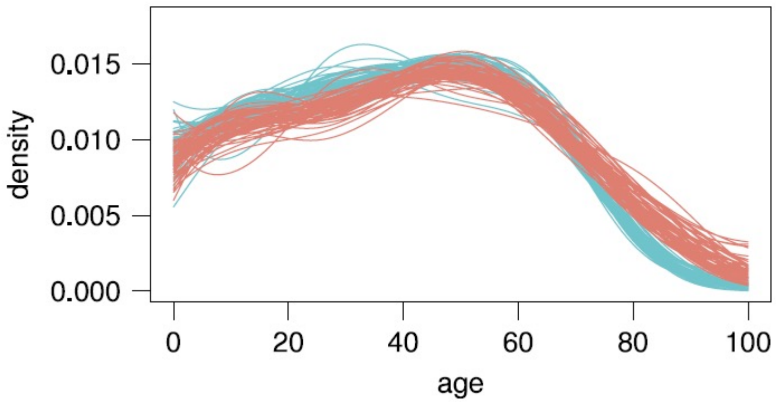


Figure 1.2. The estimated accelerations of height for 10 girls, measured in centimeters per year. The heavy dashed line is the cross-sectional mean, and is a rather poor summary of the curves.

Not only unconstrained functions

Example: Density functions of Age Distribution in Austria (constrained data)

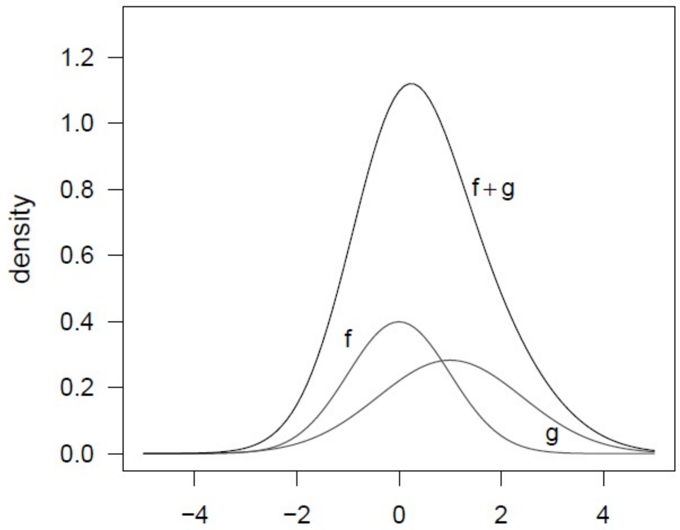


- Positive
- Integral to unity

A closer look into density data

- Density data are constrained functional data
- Most methods in Functional Data Analysis implicitly assumes that the data object can be embedded in L^2
- The L^2 geometry becomes meaningless in the presence of density data

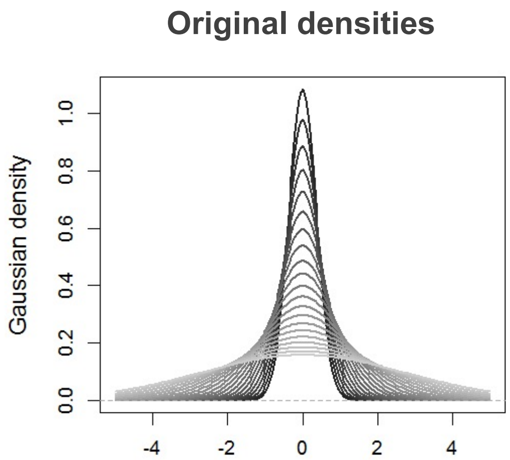
Example:
L^2 sum of two
Gaussian PDFs

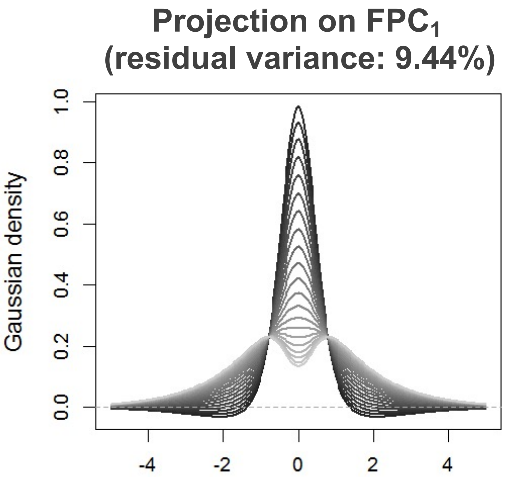


A closer look into density data

- Density data are constrained functional data
- Most methods in Functional Data Analysis implicitly assumes that the data object can be embedded in L^2
- The L^2 geometry becomes meaningless in the presence of density data
- FDA methods may provide uninterpretable results

Example: Functional Principal Component Analysis of a dataset of Gaussian densities





A closer look into density data

- Density data are constrained functional data
- Most methods in Functional Data Analysis implicitly assumes that the data object can be embedded in L^2
- The L^2 geometry becomes meaningless in the presence of density data
- FDA methods may provide uninterpretable results

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

In this short course: FDA for density data

- When data are constrained, **FDA methods can be extended to deal with data constraints**, by changing the usual working space (L^2) into different types of spaces (this is also done with scalar data, when data transformations are used – e.g., logs!)

- Special embedding for density functions: **Bayes Hilbert spaces** (generalization of the Aitchison geometry for functional data)

In this **short course**, we will be particularly concerned with

- The **geometry of the Bayes Hilbert space**: how to sensibly interpret the data as points within an infinite-dimensional simplex, defining operations and inner product
- **Representing the data**: given raw/discrete observations, represent the data through a continuous functional form (density function)
- **Reduce the dimensionality** of the dataset and highlights the main sources of its variability (as in Principal Component Analysis)
- **Identify anomalies** through functional control charts
- **Estimate linear relations** between (scalar or functional) inputs and (scalar or functional) outputs
- Deal with possible **spatial dependence**

Course Agenda

The course is organized in 6 modules

Day 1 (14:00-17:00 GMT+2)

- The geometry of Bayes spaces
- Smoothing splines for densities
- Dimensionality reduction in Bayes spaces: Simplicial Functional Principal Component Analysis

Day 2 (14:00-17:00 GMT+2)

- Anomaly detection for density data based on control charts
- Density-on-scalar, scalar-on-density and density-on-density functional regression
- Spatial statistics for distributional observations (spatial modeling & kriging)

The geometry of Bayes spaces

Alessandra Menafoglio[1*]

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

*alessandra.menafoglio@polimi.it

Table of contents

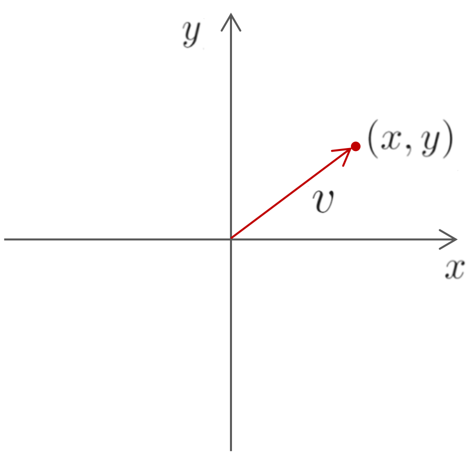
Table of contents

A Hilbert space model for 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: $c \cdot v = (c \cdot x, c \cdot y)$

Operations (+, ·)

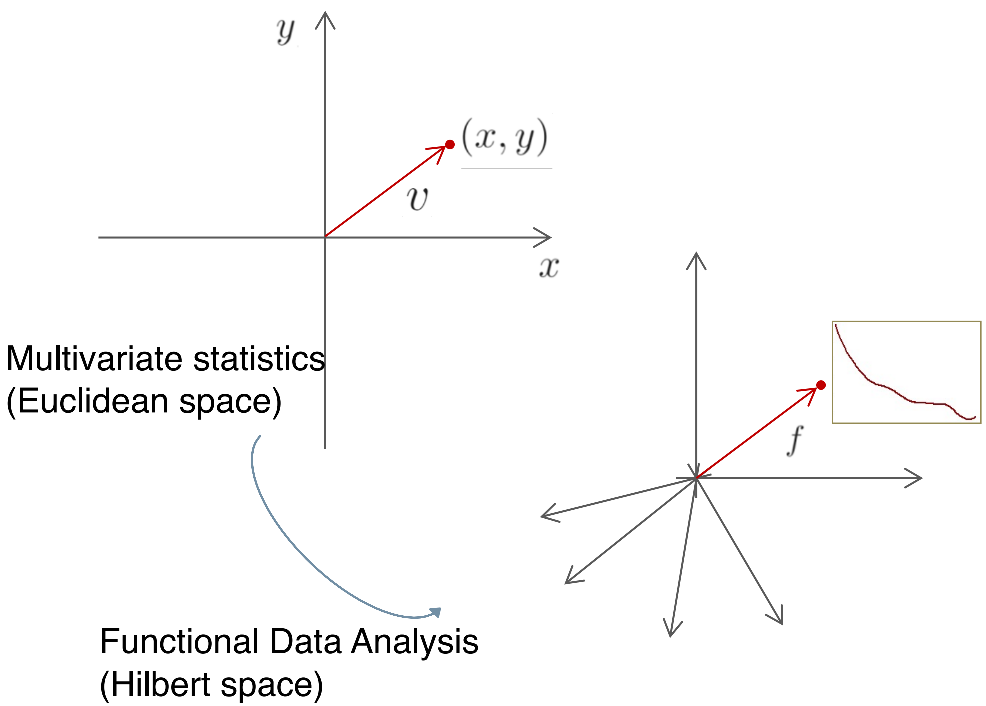
- Norm (lenght of a vector): $\|v\| = (x^2 + y^2)^{1/2}$
- Distance: $\|v_1 - v_2\| = (x_1 - x_2)^2 + (y_1 - y_2)^2$
- Angle: $\vartheta = \arccos \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|}$

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

A Hilbert space model for 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)



Why Hilbert spaces?

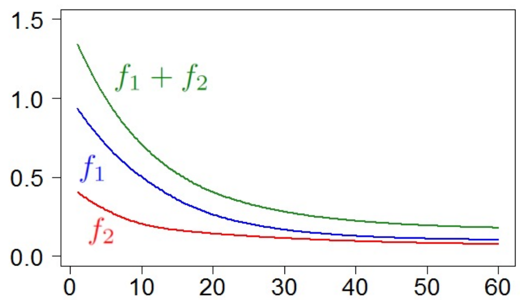
- We understand functional data as **points of a space of functions**
- Many techniques in **multivariate statistics can be generalized to data embedded in a Hilbert space**, through the notions of inner product and norm

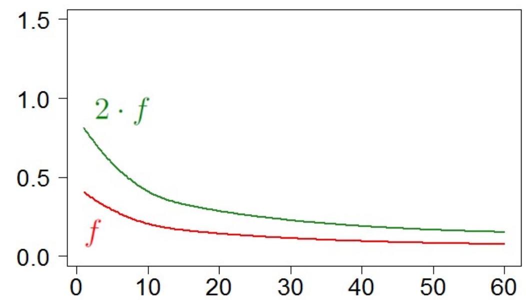
An Example: the Hilbert space L^2

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 (+, ·)





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

Inner product

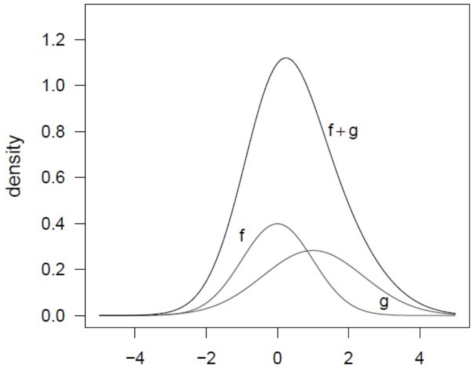
$$\langle f_1, f_2 \rangle = \int (f_1(t) \cdot f_2(t)) dt$$

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

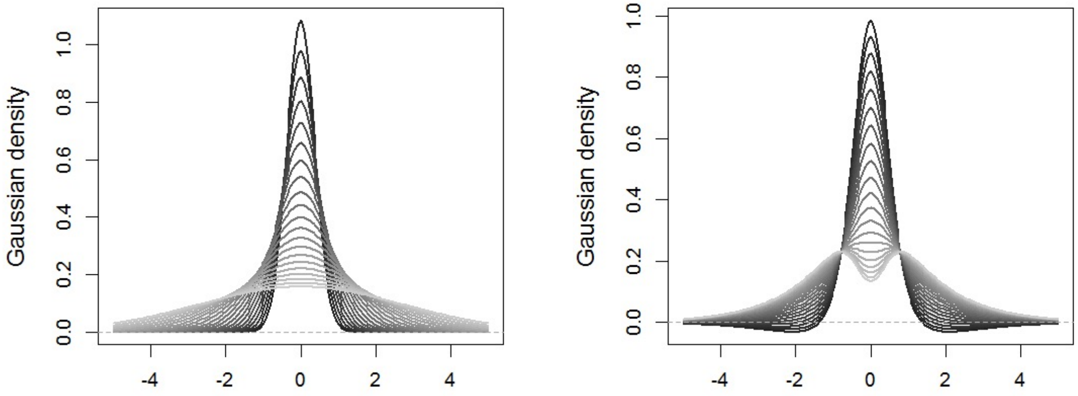
The L^2 space for density data

- The space L^2 can be used as embedding for *unconstrained data*
- In practice, all the observed data are actually finite, so the square-integrability is not a strict constraint
- However, embedding density data in L^2 presents limitations

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%)

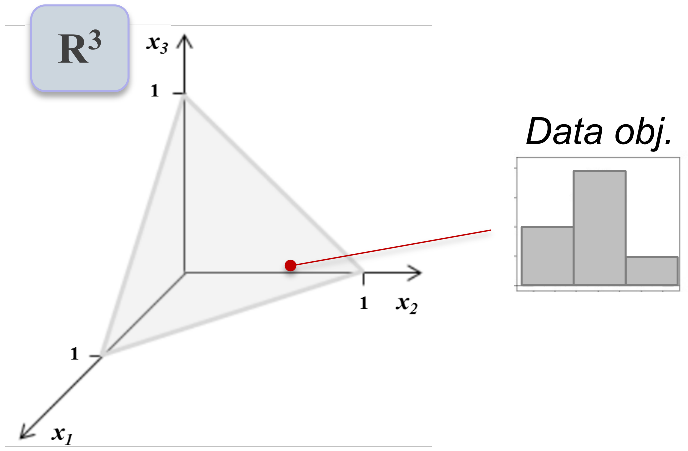
Table of contents

PDFs as functional compositional data

Discrete PDFs as multivariate compositions

$$(x_1, ..., x_D) \in \mathbb{R}^D$$
$$x_i > 0, \sum_i x_i = const$$

- x_i represents a *parts of a whole* according to a partition of the domain
- Convey only relative information: (log) ratios between parts provide the meaningful info.



PDFs as functional compositional data

Discrete PDFs as multivariate compositions

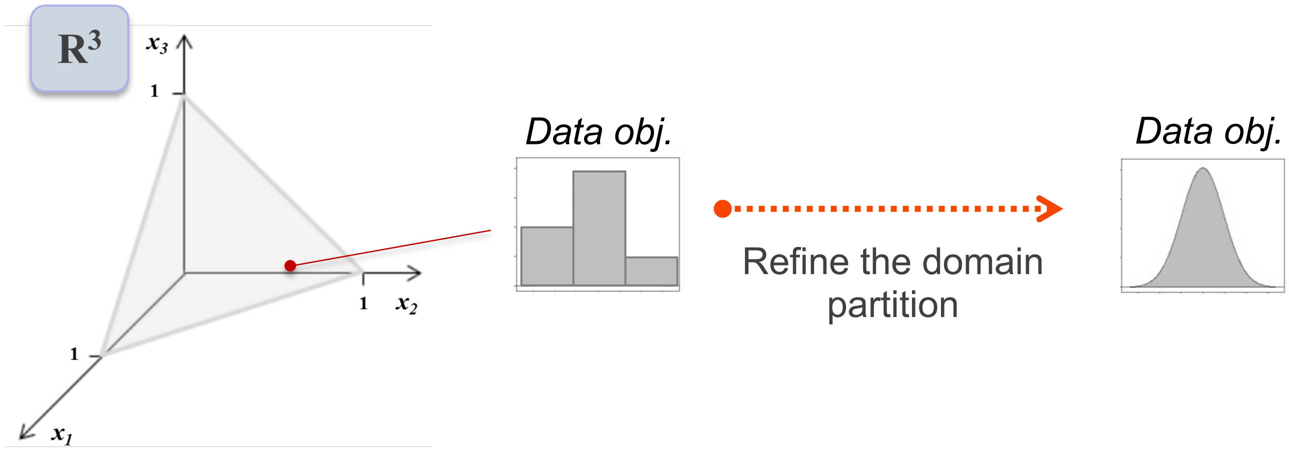
$$(x_1, ..., x_D) \in \mathbb{R}^D$$
$$x_i > 0, \sum_i x_i = const$$

- x_i represents a *parts of a whole* according to a partition of the domain
- Convey only relative information: (log) ratios between parts provide the meaningful info.

Continuous PDFs as functional compositions:

$$x : \mathcal{T} \to \mathbb{R}$$
$$x(t) > 0, \int_{\mathcal{T}} x(t)dt = const$$

- Infinite-dimensional object (a function)
- Point-wise values represent infinitesimal parts of a whole (e.g., unity).



PDFs as functional compositional data

Discrete PDFs as multivariate compositions

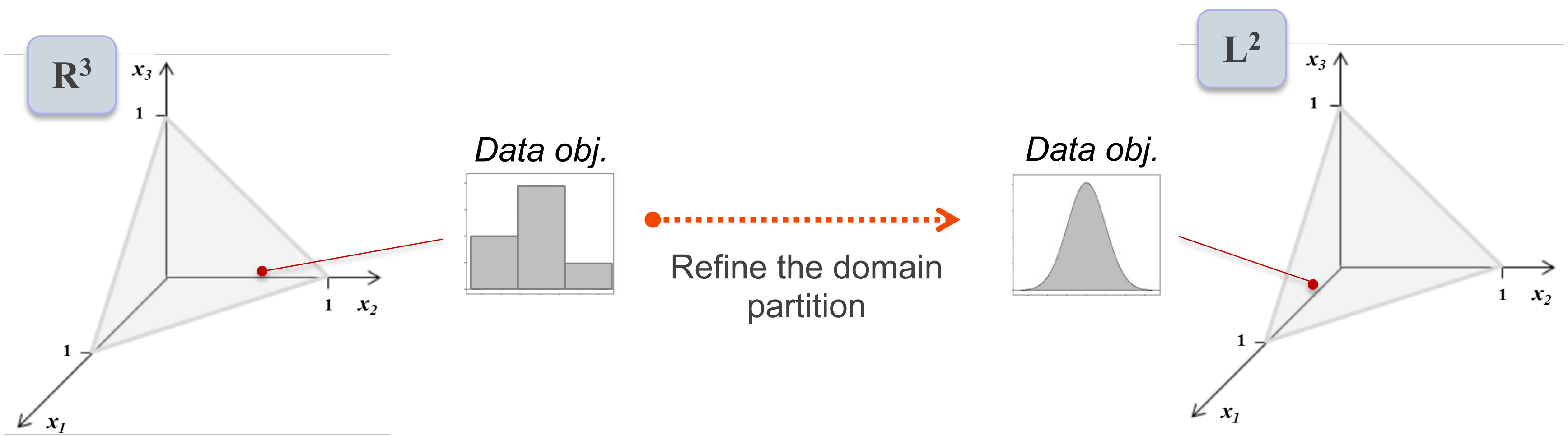
$$(x_1, \ldots, x_D) \in \mathbb{R}^D$$
$$x_i > 0, \sum_i x_i = const$$

- x_i represents a *parts of a whole* according to a partition of the domain
- Convey only relative information: (log) ratios between parts provide the meaningful info.

Continuous PDFs as functional compositions:

$$x : \mathcal{T} \to \mathbb{R}$$
$$x(t) > 0, \int_{\mathcal{T}} x(t)dt = const$$

- Infinite-dimensional object (a function)
- Point-wise values represent infinitesimal parts of a whole (e.g., unity).



PDFs as functional compositional data

Discrete PDFs as multivariate compositions

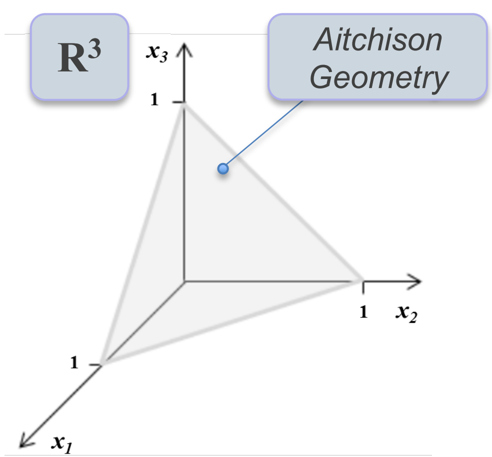
$$(x_1, ..., x_D) \in \mathbb{R}^D$$
$$x_i > 0, \sum_i x_i = const$$

- x_i represents a *parts of a whole* according to a partition of the domain
- Convey only relative information: (log) ratios between parts provide the meaningful info.

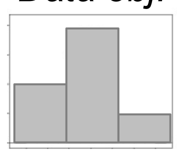
Continuous PDFs as functional compositions:

$$x : \mathcal{T} \to \mathbb{R}$$
$$x(t) > 0, \int_{\mathcal{T}} x(t)dt = const$$

- Infinite-dimensional object (a function)
- Point-wise values represent infinitesimal parts of a whole (e.g., unity).

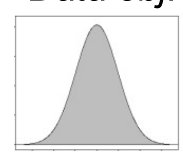


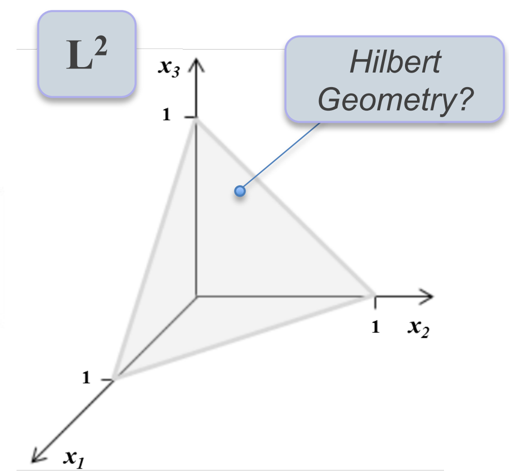
Data obj.



Refine the domain partition

Data obj.



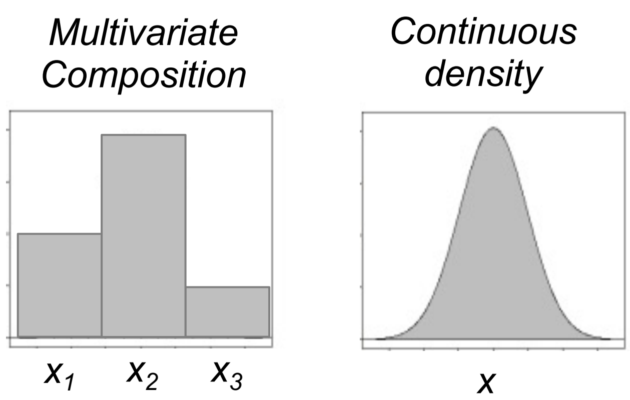


The Bayes Hilbert space B^2

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*)

Easier understood with multivariate compositions: changing the unit of measure (e.g., percentages to proportions) does not change the information content within the composition

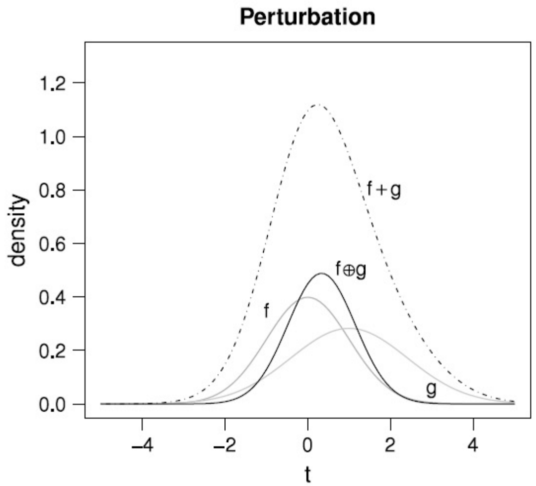


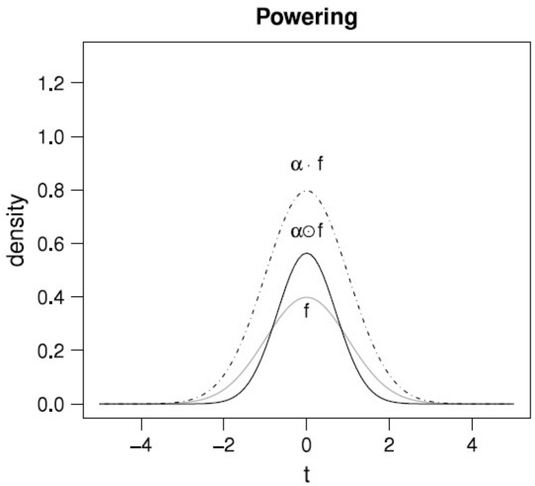
The y-axis per se is not particularly meaningful!

The Bayes Hilbert space B^2

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





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

Note: subtraction is defined accordingly as

$$(f \ominus g)(t) = (f \oplus [(-1) \odot g])(t), \; t \in I$$

The Bayes Hilbert space B^2

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*

The Bayes Hilbert space B^2

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

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

- Norm: $\|f\|_{\mathcal{B}} = \left[\frac{1}{2\eta}\int_I\int_I \ln^2\frac{f(t)}{f(s)}\, dt\, ds\right]^{1/2}$

Result: *B^2 equipped with the above operations and inner product is a Hilbert space*

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.*

The Bayes Hilbert space B²

Bayes Hilbert space $B^2(I)$

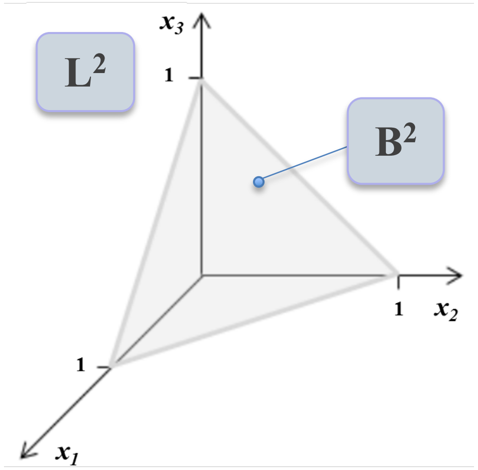
(Egozcue et al., 2006; van den Boogaart et al., 2014)

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

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

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

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



The Bayes Hilbert space B^2

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

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

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

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

Remarks:

- The B^2 geometry is based on the key observation that only (log-)ratios between probabilities are meaningful (~ odds-ratio) as data represent the distribution of a *total* mass (=1) over a domain
- This is a key principle of all the compositional data analysis
- More in general, the B^2 geometry is built upon the principles of compositional data analysis (scale invariance, relative scale, subcompositional coherence)

The Bayes Hilbert space B^2

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

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

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

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

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 course.

The Bayes Hilbert space B^2

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

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

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

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

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

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

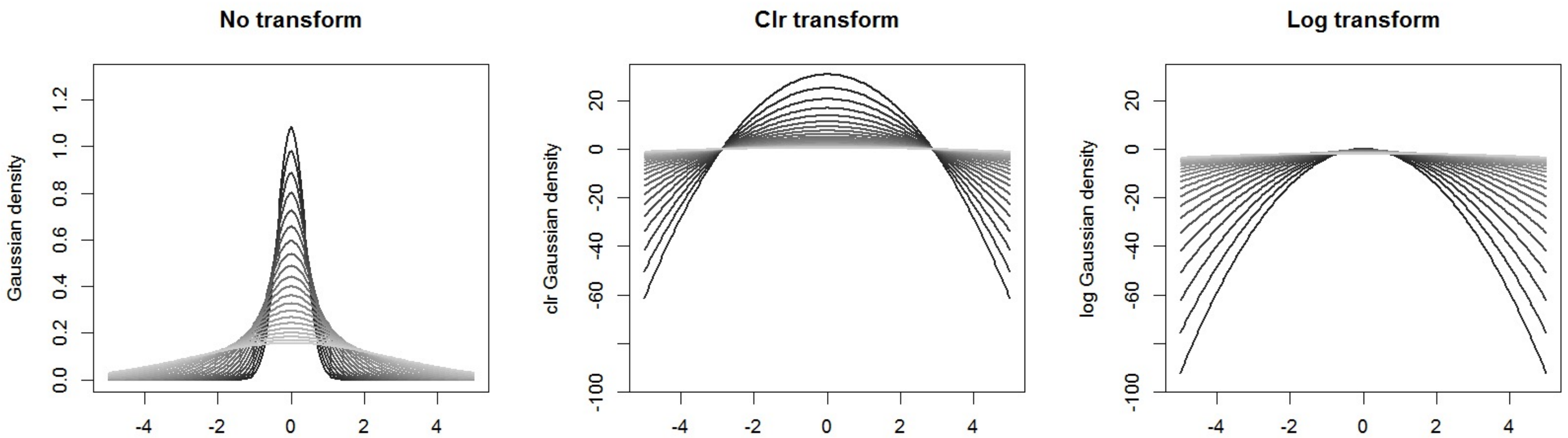
  - ❖ Projection over an orthonormal functional basis of B^2 and consideration of (a truncation of) the basis coefficients

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

- Computations of operations and inner products in B2 can be performed by relying on routines for L^2 based on clr-transformations, or multivariate routines based on basis coefficients

Clr transformation

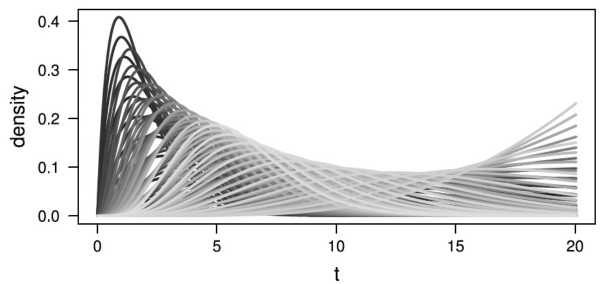
Example: Gaussian densities



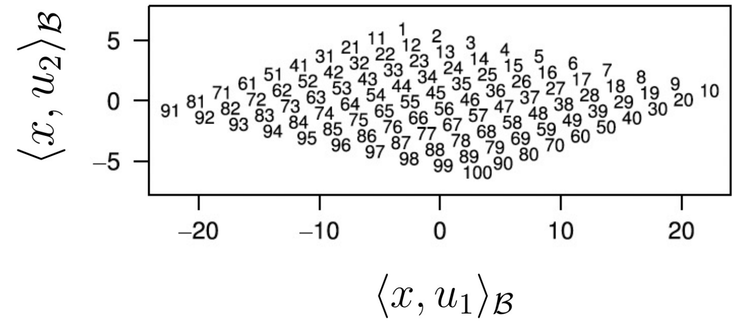
Note: a clr-transform is not just a log-transform!

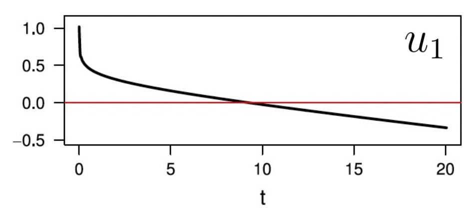
Basis projection

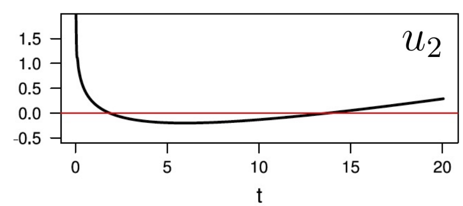
Example: Gamma densities



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







More on this in the module on Simplicial FPCA

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

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.
- Talská, R., Menafoglio, A., Hron, K., Egozcue, J.J., Palarea-Albaladejo, J. (2020) “Weighting the domain of probability densities in functional data analysis”, Stat. DOI: 10.1002/sta4.283.
- 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.

The geometry of Bayes spaces

Alessandra Menafoglio[1*]

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

*alessandra.menafoglio@polimi.it