Dimensionality reduction in Bayes spaces: Simplicial functional principal component analysis

29 April 2021

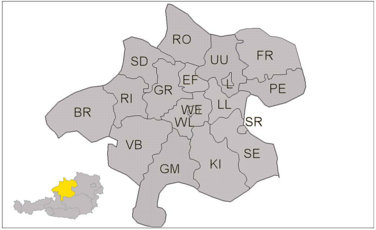
Karel Hron

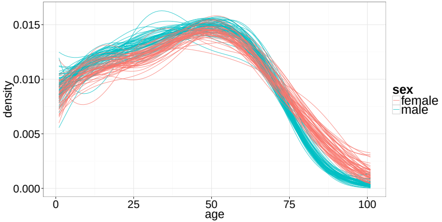
Department of Mathematical Analysis and Applications of Mathematics

Faculty of Science – Palacký University, Olomouc, Czech Republic

Outline

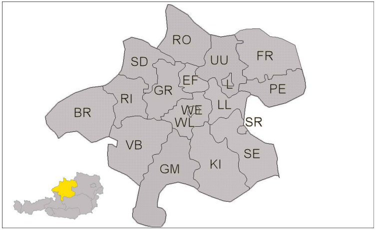
Population age distributions in Upper Austria

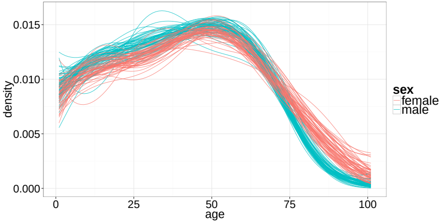




- 15 political districts, age distributions of men and women living in 114 municipalities of Upper Austria (*population pyramids*)

Population age distributions in Upper Austria





- 15 political districts, age distributions of men and women living in 114 municipalities of Upper Austria (*population pyramids*)
- **Aim**: to characterize the available population age densities performing a dimensionality reduction (PCA)

EFDA: sample mean

- ... something any exploratory functional data analysis (EFDA) usually starts with ...
- Given a sample $X_1, ..., X_N$ in $\mathcal{B}^2(I)$, $I = [a, b]$, $a, b \in \mathbb{R}$, $a < b$

EFDA: sample mean

- ...something any exploratory functional data analysis (EFDA) usually starts with ...
- Given a sample $X_1, ..., X_N$ in $\mathcal{B}^2(I)$, $I = [a, b]$, $a, b \in \mathbb{R}$, $a < b$
- **Sample mean:** $\overline{X} = \frac{1}{N} \odot \bigoplus_{i=1}^{N} X_i$
- It can be computed through the back-transform of the sample mean in L_0^2 of the clr-transformed data (the latter being defined point-wise)

$$\overline{X} = \mathrm{clr}^{-1}(\overline{X}^c), \quad \overline{X}^c = \frac{1}{N} \sum_{i=1}^{N} X_i^c$$

EFDA: sample covariance function

- Specifies the *covariance* between density values at $t, s \in \Omega$
- Assigned to one FDA object (here PDF, or a *sample* of PDFs)
- Defined directly in the clr-space (as in the usual L^2 space)

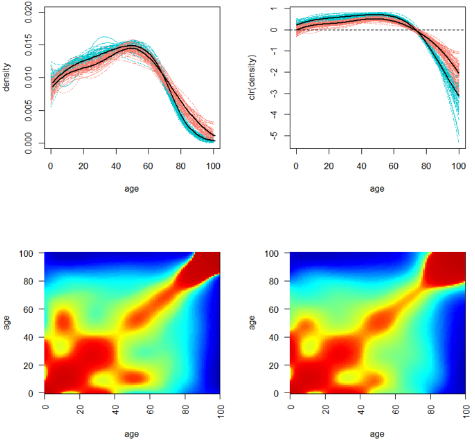
EFDA: sample covariance function

- Specifies the *covariance* between density values at $t, s \in \Omega$
- Assigned to one FDA object (here PDF, or a *sample* of PDFs)
- Defined directly in the clr-space (as in the usual L^2 space)
- **Sample covariance function:**

$$v(s,t) = \frac{1}{N}\sum_{i=1}^{N}(X_i^c(s) - \overline{X}^c(s))(X_i^c(t) - \overline{X}^c(t))$$

- Can be visualized as function of two variables (for smoothed clr-transformed densities)

EFDA: Population age distributions in Upper Austria



Male and female populations: Sample mean and sample covariance function

Functional principal component analysis (FPCA)

- Consider a *centred* functional random sample $X_1, ..., X_N$ in $L^2(I)$, i.e. from all observations $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ is subtracted
- FPCA looks firstly for the main mode of variability, i.e., for the element ξ_1 in $L^2(I)$ – called first functional principal component (FPC)– maximizing over $\xi \in L^2(I)$

$$\frac{1}{N}\sum_{i=1}^{N} \langle X_i, \xi \rangle_2^2 \text{ subject to } \|\xi\|_2 = 1.$$

- **Aim**: to capture the main modes of variability of the data by means of a small number K of linear combinations of the original variables: $X_i \approx \sum_{k=1}^{K} \langle X_i, \xi_k \rangle_2 \xi_k$

Functional principal component analysis (FPCA)

- The remaining FPCs, $\{\xi_j\}_{j\geq 2}$, capture the remaining modes of variability subject to be mutually orthogonal, and are thus obtained by solving problem the previous maximization problem with the additional orthogonality constraint $\langle \xi_k, \xi \rangle_2 = 0, k < j$

Functional principal component analysis (FPCA)

- The remaining FPCs, $\{\xi_j\}_{j\geq 2}$, capture the remaining modes of variability subject to be mutually orthogonal, and are thus obtained by solving problem the previous maximization problem with the additional orthogonality constraint $\langle \xi_k, \xi \rangle_2 = 0, k < j$

→ **Outputs**: eigenfunctions of the covariance operator/harmonics ξ_j (interpreted in terms of the original data) and scores (coefficients, representing data structure of the original observations)

FPCA: computational details

- Dealing with FPCA is analogous to the multivariate PCA
- The FPCs $\{\xi_j\}_{j\geq 1}$ coincide with the eigenfunctions of the sample covariance operator $V : L^2(I) \to L^2(I)$, acting on $x \in L^2(I)$ as

$$Vx = \frac{1}{N}\sum_{i=1}^{N}\langle X_i, x\rangle_2 X_i$$

→ The j-th FPC ξ_j and the associated scores $\Psi_{ij} = \langle X_i, \xi_j\rangle_2$, $i = 1, ..., N$, are obtained by solving the **eigenvalue equation**

$$V\xi_j = \rho_j \xi_j;$$

ρ_j denotes the j-th eigenvalue, with $\rho_1 \geq \rho_2 \geq ...$.

FPCA: computational details

- For each j, the term $\rho_j / \sum_j \rho_j$ is associated with the proportion of total variability explained by the FPC ξ_j.
- The eigenvalue equation is solved using basis expansion of each datum X_i, $i = 1, ..., N$ using K known basis functions $\phi_1, ..., \phi_K$:

$$X_i(\cdot) = \sum_{k=1}^{K} c_{ik} \phi_k(\cdot),$$

where $c_{ik} = \langle X_i, \phi_k \rangle_2$, $k = 1, ..., K$

→ Commonly, smoothing splines are used for this purpose

Simplicial functional principal component analysis

→ **SFPCA**: Reformulate FPCA in terms of Bayes spaces for $X_1, \ldots, X_N$ being a (centred) sample in $\mathcal{B}^2(I)$, i.e., we performed perturbation-subtraction by $\overline{X} = \frac{1}{N} \odot \bigoplus_{i=1}^{N} X_i$

- Maximizing over $\zeta \in \mathcal{B}^2(I)$

$$\frac{1}{N}\sum_{i=1}^{N} \langle X_i, \zeta \rangle_B^2 \text{ subject to } \|\zeta\|_B = 1;\ \langle \zeta_j, \zeta_k \rangle_B = 0,\ k < j$$

→ We can formulate the problem and find the unique solution because $\mathcal{B}^2(I)$ is a separable Hilbert space

→ **Problem**: how to efficiently implement all of this?

Clr transformation and SFPCA

- **Goal**: To perform SFPCA exploiting the efficient routines available in L^2 space (i.e., avoid computations in Bayes spaces)

Clr transformation and SFPCA

- **Goal**: To perform SFPCA exploiting the efficient routines available in L^2 space (i.e., avoid computations in Bayes spaces)

→ Through centred logratio (clr) transformation:

$$\mathrm{clr}(f)(t) = f^c(t) = \ln f(t) - \frac{1}{\eta}\int_I \ln f(s)\,\mathrm{d}s, \ \int_I f^c(t)\,\mathrm{d}t = 0;$$

Clr transformation and SFPCA

- **Goal**: To perform SFPCA exploiting the efficient routines available in L^2 space (i.e., avoid computations in Bayes spaces)

→ Through centred logratio (clr) transformation:

$$\mathrm{clr}(f)(t) = f^c(t) = \ln f(t) - \frac{1}{\eta}\int_I \ln f(s)\,\mathrm{d}s, \quad \int_I f^c(t)\,\mathrm{d}t = 0;$$

- Consequence for FPCA in clr space: $\xi_0 \equiv 1/\sqrt{b-a}$

Clr transformation and SFPCA

- **Goal**: To perform SFPCA exploiting the efficient routines available in L^2 space (i.e., avoid computations in Bayes spaces)

→ Through centred logratio (clr) transformation:

$$\mathrm{clr}(f)(t) = f^c(t) = \ln f(t) - \frac{1}{\eta}\int_I \ln f(s)\,\mathrm{d}s, \quad \int_I f^c(t)\,\mathrm{d}t = 0;$$

- Consequence for FPCA in clr space: $\xi_0 \equiv 1/\sqrt{b-a}$
- The zero integral contraint needs to be incorporated into the basis expansion → **compositional splines**

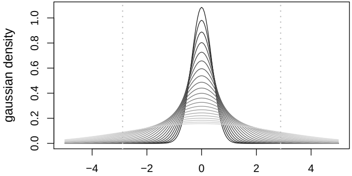
Example: Truncated normal PDFs

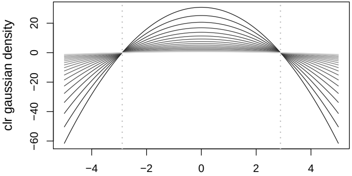
- Normal densities, $\mu = 0, \sigma_i = \exp(-1 + (i-1)/10)$, $i = 1, ..., 21$, $I = [-5, 5]$

$$f(t;\sigma_i) =_{\mathcal{B}^2} \exp\left\{-\frac{t^2}{2\sigma_i^2}\right\}, \quad t \in I, \tag{1}$$

$=_{\mathcal{B}^2(I)}$ denotes the equivalence in the space $\mathcal{B}^2(I)$

$$f^c(t;\sigma_i) = -\frac{t^2}{2\sigma_i^2} + \frac{25}{6\sigma_i^2}, \quad t \in I.$$





Dimensionality of PDFs from the exponential family

An important feature of (log-)normal densities in context of Bayes spaces is that they belong to the extended exponential family:

- Recall that a *k-parametric extended exponential family* on Ω, $Exp_{\mathcal{B}^2(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ is a collection of densities

$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}^2(I)} g(t) \cdot \exp\left\{ \sum_{j=1}^{k} \vartheta_j(\boldsymbol{\alpha}) T_j(t) \right\}, \quad t \in \Omega,$$

where $\boldsymbol{\alpha}$ denotes the k-dimensional vector of parameters in a k-dimensional parameter space A, while functions $g : \Omega \to \mathbb{R}$, $\vartheta_j : A \to \mathbb{R}$ and $T_j : \Omega \to \mathbb{R}$, $j = 1, \ldots, k$, are Borel-measurable

Dimensionality of PDFs from the exponential family

An important feature of (log-)normal densities in context of Bayes spaces is that they belong to the extended exponential family:

- Recall that a *k-parametric extended exponential family* on Ω, $Exp_{\mathcal{B}^2(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ is a collection of densities

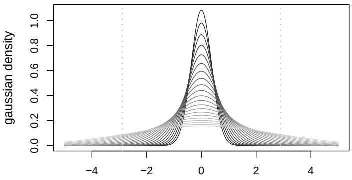
$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}^2(I)} g(t) \cdot \exp\left\{\sum_{j=1}^{k} \vartheta_j(\boldsymbol{\alpha}) T_j(t)\right\}, \quad t \in \Omega,$$

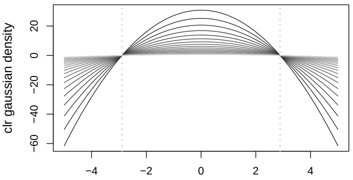
  where $\boldsymbol{\alpha}$ denotes the k-dimensional vector of parameters in a k-dimensional parameter space A, while functions $g : \Omega \to \mathbb{R}$, $\vartheta_j : A \to \mathbb{R}$ and $T_j : \Omega \to \mathbb{R}$, $j = 1, \ldots, k$, are Borel-measurable

- An extended exponential family on Ω is a **finite dimensional affine subspace** of the Bayes space $\mathcal{B}^2(I)$

Dimensionality of PDFs from the exponential family

- Most routinely used distributions belong to the exponential family
- **Example**: a Gaussian density $N(0, \sigma^2)$ *restricted on* Ω belongs to a 1-parametric extended exponential family, with $\alpha = \sigma$, $\vartheta_1(\alpha) = 1/\sigma^2$, and $T_1(t) = -t^2$





Dimensionality of PDFs from the exponential family

- A PDF in $Exp_{\mathcal{B}(I)}(g, \boldsymbol{T}, \vartheta)$ can be expressed as a linear combination in $\mathcal{B}^2(I)$:

$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}^2(I)} g(t) \oplus \bigoplus_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \odot \exp\{T_j(t)\}\right], \quad t \in \Omega,$$

- Clr-transformed:

$$f^c(t, \boldsymbol{\alpha}) = \mathrm{clr}(g(t)) + \sum_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \cdot \mathrm{clr}(\exp\{T_j(t)\})\right], \quad t \in \Omega.$$

Dimensionality of PDFs from the exponential family

- A PDF in $Exp_{\mathcal{B}(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ can be expressed as a linear combination in $\mathcal{B}^2(I)$:

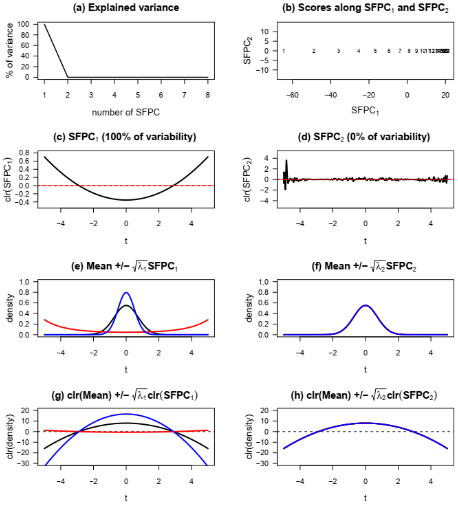
$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}^2(I)} g(t) \oplus \bigoplus_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \odot \exp\{T_j(t)\}\right], \quad t \in \Omega,$$

- Clr-transformed:

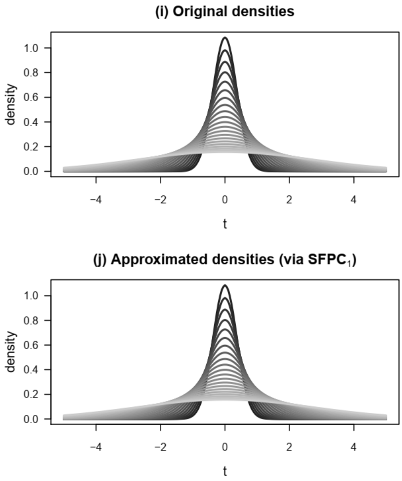
$$f^c(t, \boldsymbol{\alpha}) = \mathrm{clr}(g(t)) + \sum_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \cdot \mathrm{clr}(\exp\{T_j(t)\})\right], \quad t \in \Omega.$$

$\Rightarrow$ For $k_0 \leq k$ uncertain parameters, the SFPCA estimates an orthonormal basis of the corresponding k-dimensional affine space in $\mathcal{B}^2(I)$, which is associated to $k_0 \leq k$ non-zero eigenvalues

Dimensionality of PDFs: normal distribution



Dimensionality of PDFs: normal distribution

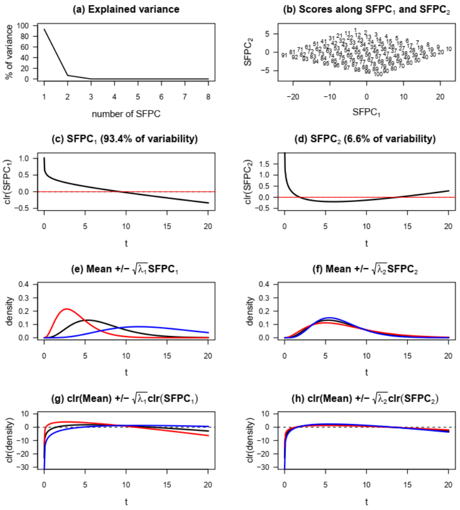


Dimensionality of PDFs: gamma distribution

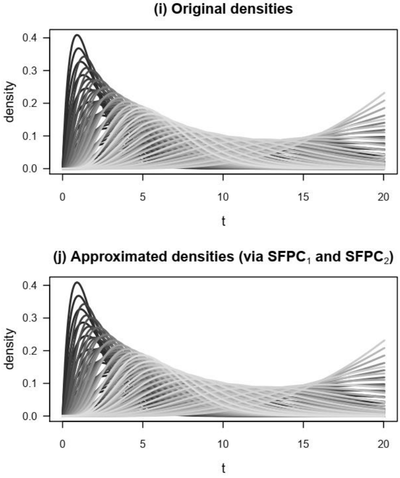
Data: $n = 100$ densities with kernel Gamma $\Gamma(\theta_i, \kappa_j)$, with $\theta_i = 1/9 + (i-1)/9$ and $\kappa_j = 2 + (j-1)/4$ for $i, j = 1, \ldots, 10$, and domain $I = [e^{-7}, e^3]$

- A Gamma distribution $\Gamma(\theta, \kappa)$ on I belongs to a 2-parametric extended exponential family with $\boldsymbol{\alpha} = (\theta, \kappa)$, $\vartheta_1(\boldsymbol{\alpha}) = \theta$, $\vartheta_2(\boldsymbol{\alpha}) = \kappa$, $T_1(t) = -t$, and $T_2(t) = \ln(t)$, for $t \in I$
- We expect now that a sensible dimensionality reduction method will single out the dimension $k = 2$ of these densities
- A comparison with FPCA for the original densities is performed as well

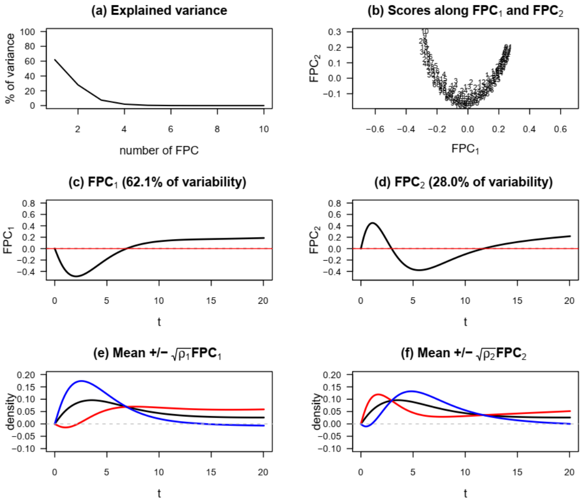
Dimensionality of PDFs: gamma distribution



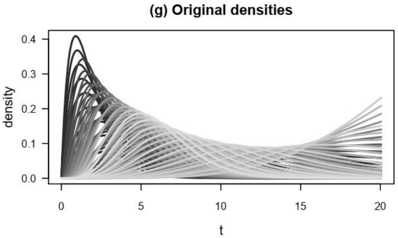
Dimensionality of PDFs: gamma distribution

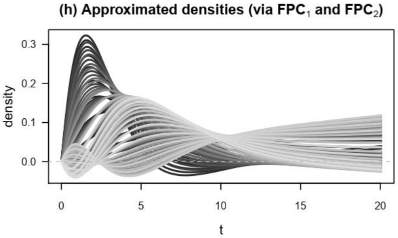


Dimensionality of PDFs: gamma distribution (L^2)

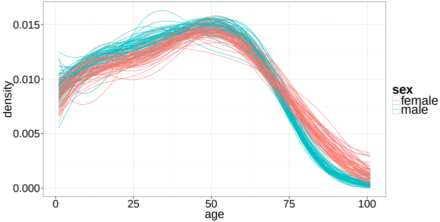


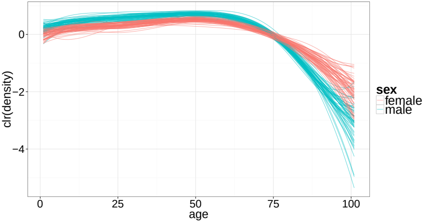
Dimensionality of PDFs: gamma distribution (L^2)

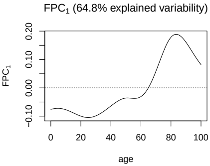


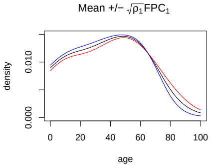


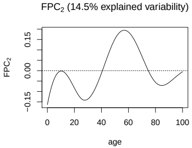
SFPCA: Population age distributions

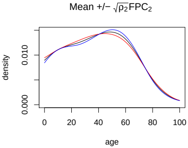




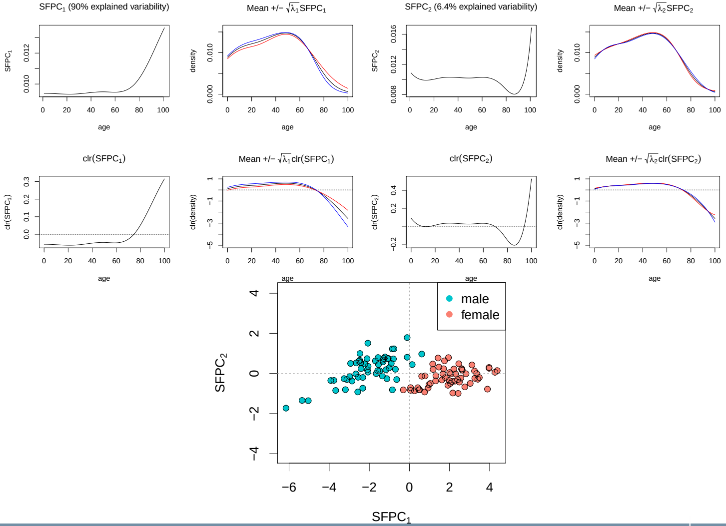




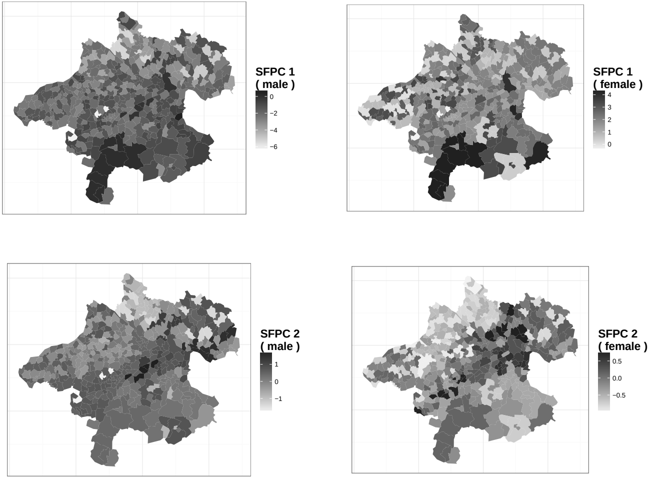




SFPCA: Population age distributions



SFPCA: Population age distributions



SFPCA: R code

https://github.com/AMenafoglio/BayesSpaces-codes

(with special thanks to Ivana Pavlů, *Palacký University*)

References

Aitchison, J.: *The statistical analysis of compositional data*. Chapman and Hall, 1986.

Delicado, P.: *Dimensionality reduction when data are density functions*. Computational Statistics and Data Analysis 55, 401-420, 2011.

Hron, K., Menafoglio, A., Templ, M., Hrůzová, K., Filzmoser, P.: Simplicial principal component analysis for density functions in Bayes spaces. Computational Statistics and Data Analysis 94, 330–350, 2016.

Ramsay, J., Silverman, B.W.: *Functional data analysis*, 2nd ed. Springer, New York, 2005.

Talská, R., Menafoglio, A., Hron, K., Egozcue, J.J., Palarea-Albaladejo, J.: *Weighting the domain of probability densities in functional data analysis*. Stat 9 (1), e283, 2020.

van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V.: *Bayes Hilbert spaces*. Australian & New Zealand Journal of Statistics 56, 171–194, 2014.