Compositional splines for representation of density functions

Jitka Machalová · Renáta Talská · Karel Hron · Aleš Gába

Received: 20 December 2019 / Accepted: 12 October 2020 / Published online: 19 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
In the context of functional data analysis, probability density functions as non-negative functions are characterized by specific properties of scale invariance and relative scale which enable to represent them with the unit integral constraint without loss of information. On the other hand, all these properties are a challenge when the densities need to be approximated with spline functions, including construction of the respective spline basis. The Bayes space methodology of density functions enables to express them as real functions in the standard $L^2$ space using the centered log-ratio transformation. The resulting functions satisfy the zero integral constraint. This is a key to propose a new spline basis, holding the same property, and consequently to build a new class of spline functions, called compositional splines, which can approximate probability density functions in a consistent way. The paper provides also construction of smoothing compositional splines and possible orthonormalization of the spline basis which might be useful in some applications. Finally, statistical processing of densities using the new approximation tool is demonstrated in case of simplicial functional principal component analysis with anthropometric data.

Keywords Spline representation · Constrained approximation · Smoothing spline · Simplicial functional principal component analysis

1 Introduction

Probability density functions are non-negative functions satisfying the unit integral constraint. This clearly inhibits their direct processing using standard methods of
functional data analysis (Ramsay and Silverman 2005) since unconstrained functions are assumed there. The same holds also for approximation of the raw input data using splines which is commonly considered to be a key step in functional data analysis. But more severely, in addition to the apparent unit integral constraint of densities which might seem to represent just a kind of numerical obstruction, density functions are rather characterized by deeper geometrical properties that need to be taken into account for any reliable analysis (Egozcue et al. 2006; Van den Boogaart et al. 2010, 2014). Specifically, in contrast to functions in the standard $L^2$ space, densities obey the scale invariance and relative scale properties (Hron et al. 2016). Scale invariance means that not just the representation of densities with the unit integral constraint, but any its positive multiple conveys the same information about relative contributions of Borel sets on the whole probability mass. Relative scale can be explained directly with an example: the relative increase of a probability over a Borel set from 0.05 to 0.1 (2 multiple) differs from the increase 0.5 to 0.55 (1.1 multiple), although the absolute differences are the same in both cases. If we restrict to a bounded support $I = [a, b] \subset \mathbb{R}$ that is mostly used in practical applications (Delicado 2011; Hron et al. 2016; Menafoglio et al. 2014, 2016), density functions can be represented with respect to the Lebesgue reference measure using the Bayes space $B^2(I)$ of functions with square-integrable logarithm (Egozcue et al. 2006; Van den Boogaart et al. 2014).

The Bayes space $B^2(I)$ has structure of a separable Hilbert space. This space was developed originally for measures which can be, however, equivalently expressed using densities. Here we consider them as scale invariant objects, which means that a density $f \in B^2(I)$ and its positive constant multiple carries essentially the same information (Van den Boogaart et al. 2014; Talská et al. 2020), the feature which is widely accepted, e.g., in Bayesian statistics. Accordingly, the respective sample space is formed by equivalence classes of proportional densities (in this broader sense, thus not necessarily densities with a unit integral representation) which forms a clear difference from the standard $L^2$ space.

An important property of the Bayes space is that it enables construction of an isometric isomorphism between $B^2(I)$ and $L^2(I)$, the $L^2$ space restricted to $I$. Accordingly, analogies of summing two functions and multiplication of a function by a real scalar in the $L^2$ space together with an inner product between two densities are required. Given two absolutely integrable density functions $f, g \in B^2(I)$ and a real number $\alpha \in \mathbb{R}$ we indicate with $f \oplus g$ and $\alpha \odot f$ the perturbation and powering operations, defined as

$$
(f \oplus g)(x) = \frac{f(x)g(x)}{\int_I f(y)g(y) \, dy}, \quad (\alpha \odot f)(x) = \frac{f(x)^\alpha}{\int_I f(y)^\alpha \, dy}, \quad x \in I,
$$

(1)

respectively. The resulting functions are readily seen to be probability density functions, though, note that the unit integral constraint representation was chosen just for the sake of convenient interpretation. In Egozcue et al. (2006), it is proven that $B^2(I)$ endowed with the operations ($\oplus, \odot$) is a vector space. Note that the neutral elements of perturbation and powering are the uniform density $e(x) = 1/\eta$, with $\eta = b - a$, and 1, respectively. The difference between two elements $f, g \in B^2(I)$, denoted by $f \ominus g$, is obtained as perturbation of $f$ with the reciprocal of $g$, i.e.,
\((f \oplus g)(x) = (f \oplus [(−1) \odot g])(x), x \in I\). Finally, to complete the Hilbert space structure, the inner product is defined as

\[
\langle f, g \rangle_B = \frac{1}{2\eta} \int_I \int_I \ln \frac{f(x)}{f(y)} \ln \frac{g(x)}{g(y)} \, dx \, dy, \quad f, g \in B^2(I);
\]

consequently, the norm and the distance are obtained as \(\|f\|_B = \sqrt{\langle f, f \rangle_B}\) and \(d_B(f, g) = \|f \ominus g\|_B\), respectively. Form of the inner product clearly indicates that the relevant information in densities is contained in (log-)ratios between elements from the support \(I\).

Density functions can be considered as functional counterparts to compositional data, positive vectors carrying relative information (Aitchison 1982; Pawlowsky-Glahn et al. 2015) that are driven by the Aitchison geometry (Pawlowsky-Glahn and Egozcue 2001). In order to enable their statistical processing using standard multivariate methods in real space (Eaton 1983), the preferred strategy is to express them either in centered log-ratio (clr) coefficients (Aitchison 1982) with respect to a generating system, or in logratio coordinates, preferably with respect to an orthonormal basis (Egozcue et al. 2003). The latter coordinates (called also isometric log-ratio coordinates), as well as the clr coefficients, provide isometry between the Aitchison geometry and the real Euclidean space. A similar strategy is used also for densities in the Bayes space (Van den Boogaart et al. 2014). An isometric isomorphism between \(B^2(I)\) and \(L^2(I)\) is represented by the centered log-ratio (clr) transformation (Van den Boogaart et al. 2014; Menafoglio et al. 2014), defined for \(f \in B^2(I)\) as

\[
\text{clr}(f)(x) \equiv f_c(x) = \ln f(x) - \frac{1}{\eta} \int_I \ln f(y) \, dy.
\]

We remark that such an isometry allows to compute operations and inner products among the elements in \(B^2(I)\) in terms of their counterpart in \(L^2(I)\) among the clr-transformations, i.e.

\[
\text{clr}(f \oplus g)(x) = f_c(x) + g_c(x), \quad \text{clr}(\alpha \odot f)(x) = \alpha \cdot f_c(x)
\]

and

\[
\langle f, g \rangle_B = \langle f_c, g_c \rangle_2 = \int_I f_c(x) g_c(x) \, dx,
\]

the latter being the inner product in the \(L^2\) space. However, the clr transformation induces an additional constraint,

\[
\int_I \text{clr}(f)(x) \, dx = \int_I \ln f(x) \, dx - \int_I \frac{1}{\eta} \int_I \ln f(y) \, dy \, dx = 0,
\]

that needs to be taken into account for the computation and analysis on clr-transformed density functions. As the clr space is clearly a subspace of \(L^2(I)\), hereafter it is denoted as \(L^2_0(I)\). The inverse clr transformation is obtained as
\[
\text{clr}^{-1}(f_c)(x) = \frac{\exp(f_c(x))}{\int_I \exp(f_c(y)) \, dy};
\]

again as before, the denominator is used just to achieve the unit integral constraint representation of the resulting density (without loss of relative information, carried by the density function).

According to Van den Boogaart et al. (2014), it is not necessary to restrict ourselves to the constrained clr space, because a basis in \( B^2(I) \) can be easily constructed. Specifically, let \( \psi_0(x), \psi_1(x), \psi_2(x), \ldots \) be a basis in \( L^2(I) \) and assume that \( \psi_0(x) \) is a constant function, then \( \varphi_1(x) := \exp(\psi_1(x)), \varphi_2(x) := \exp(\psi_2(x)), \ldots \) form a basis in \( B^2(I) \). Of course, also here an orthonormal basis is preferable, but it is not always possible in applications. Nevertheless, if this would be so, then a function \( f \in B^2(I) \) can be projected orthogonally to the space spanned, e.g., by the first \( r \) functions \( \varphi_1(x), \varphi_2(x), \ldots, \varphi_r(x) \). This is done through the respective coefficients \( c_1, \ldots, c_r \) in the basis expansion

\[
f(x) = c_1 \circ \varphi_1(x) \oplus c_2 \circ \varphi_2(x) \oplus \ldots \oplus c_r \circ \varphi_r(x) \oplus \ldots = \bigoplus_{i=1}^{\infty} c_i \circ \varphi_i(x), \, x \in I.
\]

Functional data analysis relies strongly on approximation of the input functions using splines (Ramsay and Silverman 2005). However, splines are mostly utilized purely as an approximation tool, without considering further methodological consequences. Because statistical processing of density functions requires a deeper geometrical background, provided by the Bayes spaces, this should be followed also by the respective spline representation, performed preferably in the clr space \( L^2_0(I) \). In Machalová et al. (2016), a first attempt of constructing a spline representation that would honor the zero integral constraint (4) was performed. The problem is that B-splines that form basis for the spline expansion in Machalová et al. (2016) come from \( L^2(I) \), but not from \( L^2_0(I) \). This paper presents an important step ahead – such splines are constructed that form basis functions in the clr space \( L^2_0(I) \). Consequently, the splines can be expressed also directly in \( B^2(I) \) and the spline representation formulated in terms of the Bayes space which can be used for interpretation purposes; hereafter we refer to compositional splines. Apart from methodological advantages, using compositional splines simplifies construction and interpretation of spline coefficients that can be considered as coefficients of a (possibly orthonormal) basis in \( B^2(I) \).

The paper is organized as follows. In the next section the construction of splines basis in \( L^2_0(I) \) is presented together with a comparison to spline functions introduced in Machalová et al. (2016). Section 3 is devoted to smoothing splines in \( L^2_0(I) \) and Sect. 4 discusses orthogonalization of basis functions (that form, by construction, an oblique basis). Section 6 introduces a new class of splines that reflects the Bayes spaces methodology, compositional splines. Section 7 demonstrates usefulness of the new
approximation tool in context of simplicial functional principal component analysis with anthropometric data and the final Sect. 8 concludes.

2 Construction of spline in $L_0^2(I)$

Because the clr transformation enables to process density functions in the standard $L^2$ space, just restricted according to zero integral constraint (4), it is natural that also construction of compositional splines should start in $L_0^2(I)$. Nevertheless, before doing so, some basic facts about B-spline representation of splines are recalled, see De Boor (1978); Dierckx (1993); Schumaker (2007) for details. Let the sequence of knots

$$\lambda_0 = a < \lambda_1 < \cdots < \lambda_g < b = \lambda_{g+1}$$

be given. The (normalized) B-spline of degree 0 (order 1) is defined as

$$B_1^i(x) = \begin{cases} 1 & \text{if } x \in [\lambda_i, \lambda_{i+1}) \\ 0 & \text{otherwise} \end{cases}$$

for $i = 0, \ldots, g$, and the (normalized) B-spline of degree $k$, $k \in \mathbb{N}$, (order $k + 1$) is defined by

$$B_{k+1}^i(x) = \frac{x - \lambda_i}{\lambda_{i+k} - \lambda_i} B_k^i(x) + \frac{\lambda_{i+k+1} - x}{\lambda_{i+k+1} - \lambda_{i+1}} B_{k+1}^{i+1}(x).$$

Now let the functions $Z_{i+1}^k(x)$ for $k \geq 0$, $k \in \mathbb{N}$, which are the first derivatives of the B-splines, be defined

$$Z_{i+1}^k(x) := \frac{d}{dx} B_{i+1}^{k+2}(x),$$

i.e., more precisely for $k = 0$

$$Z_1^i(x) = \begin{cases} \frac{1}{\lambda_{i+1} - \lambda_i} & \text{if } x \in [\lambda_i, \lambda_{i+1}) \\ - \frac{1}{\lambda_{i+2} - \lambda_{i+1}} & \text{if } x \in (\lambda_{i+1}, \lambda_{i+2}] \end{cases}$$

and for $k \geq 1$

$$Z_{i+1}^k(x) = (k + 1) \left( \frac{B_{i+1}^{k+1}(x)}{\lambda_{i+k+1} - \lambda_i} - \frac{B_{i+1}^{k+1}(x)}{\lambda_{i+k+2} - \lambda_{i+1}} \right).$$

Noteworthy, functions $Z_{i+1}^k(x)$ have similar properties as B-splines $B_{i+1}^{k+1}(x)$. 
1. They are piecewise polynomials of degree $k$. Particularly, $Z^1_i(x)$ is a piecewise constant polynomial, $Z^2_i(x)$ is a piecewise linear polynomial, see Fig. 1a, $Z^3_i(x)$ is a piecewise quadratic polynomial, see Fig. 1b. For other examples see Fig. 1c–e.

2. It is evident that for $k \geq 1$ the function $Z^{k+1}_i(x)$ and its derivatives up to order $k-1$ are all continuous.

3. It is easy to check that for $k \geq 0$

$$\text{supp } Z^{k+1}_i(x) = \text{supp } B^{k+2}_i(x) = [\lambda_i, \lambda_{i+k+2}],$$

and of course

$$Z^{k+1}_i(x) = 0 \text{ if } x \notin [\lambda_i, \lambda_{i+k+2}].$$

4. From the perspective of $L_2^0(I)$, a crucial point is that the integral of $Z^{k+1}_i(x)$ equals to zero. If we consider Curry-Schoenberg B-splines $M^{k+1}_i(x)$ (De Boor 1978), which are defined as

$$M^{k+1}_i(x) := \frac{k+1}{\lambda_{i+k+1} - \lambda_i} B^{k+1}_i(x),$$

with property

$$\int_{\mathbb{R}} M^{k+1}_i(x) \, dx = 1,$$

than it is clear that

$$Z^{k+1}_i(x) = M^{k+1}_i(x) - M^{k+1}_{i+1}(x) \quad (9)$$

and

$$\int_{\mathbb{R}} Z^{k+1}_i(x) \, dx = 0. \quad (10)$$

It is known that for the vector space $S^{\Delta \lambda}_k[a, b]$ of polynomial splines of degree $k > 0$, $k \in \mathbb{N}$, defined on a finite interval $I = [a, b]$ with the sequence of knots $\Delta \lambda = \{\lambda_i\}_{i=0}^{g+1}$, $\lambda_0 = a < \lambda_1 < \ldots < \lambda_g < b = \lambda_{g+1}$, the dimension is

$$\dim(S^{\Delta \lambda}_k[a, b]) = g + k + 1.$$

For the construction of all basis functions $B^{k+1}_i(x)$, it is necessary to consider some additional knots. Without loss of generality we can add coincident knots

$$\lambda_{-k} = \cdots = \lambda_{-1} = \lambda_0 = a, \quad b = \lambda_{g+1} = \lambda_{g+2} = \cdots = \lambda_{g+k+1}. \quad (11)$$
Compositional splines for representation of density functions

(a) Linear B-spline $Z_i^1(x) = \frac{d}{dx} B_i^1(x)$ with equidistant knots 0, 1, 2, 3.

(b) Quadratic B-spline $Z_i^2(x) = \frac{d}{dx} B_i^2(x)$ with nonequidistant knots 0, 1, 2, 3, 4.

(c) Cubic B-spline $Z_i^3(x) = \frac{d}{dx} B_i^3(x)$ with equidistant knots 0, 1, 2, 3, 4, 5.

(d) Linear B-spline $Z_i^2(x) = \frac{d}{dx} B_i^1(x)$ with nonequidistant knots 0, 1, 10, 30.

(e) Quadratic B-spline $Z_i^3(x) = \frac{d}{dx} B_i^2(x)$ with nonequidistant knots 0, 1, 10, 30, 50.

Fig. 1 Example of piecewise polynomial functions $Z_i^{k+1}(x) := \frac{d}{dx} B_i^{k+2}(x)$ with $k \in \{1, 2, 3\}$. Vertical dashed gray lines indicate knot positions.
Then every spline $s_k(x) \in S_k^{\Delta \lambda}[a, b]$ in $L^2(I)$ has a unique representation

$$s_k(x) = \sum_{i=-k}^{g} b_i B_i^{k+1}(x).$$

In Machalová et al. (2016) and Talská et al. (2018), the splines with zero integral are studied. There is given the necessary and sufficient condition for B-splines coefficients of these splines. However, standard B-splines $B_i^{k+1}(x)$ ignoring the constraint (4) in $L^2(I)$ were considered here for the construction of the B-spline basis.

Now, regarding the definition (7), we are able to use spline functions $Z_i^{k+1}(x)$ which have zero integral on $I$ (denoted ZB-splines in the sequel). In the following, $Z_k^{\Delta \lambda}[a, b]$ denotes the vector space of polynomial splines of degree $k > 0$, defined on a finite interval $[a, b]$ with the sequence of knots $\Delta \lambda$ and having zero integral on $[a, b]$, it means

$$Z_k^{\Delta \lambda}[a, b] := \left\{ s_k(x) \in S_k^{\Delta \lambda}[a, b] : \int_I s_k(x) \, dx = 0 \right\}.$$  

**Theorem 1** The dimension of the vector space $Z_k^{\Delta \lambda}[a, b]$ defined by the formula (13) is $g + k$.

**Proof** For spline $s_k(x) \in S_k^{\Delta \lambda}[a, b]$, $s_k(x) = \sum_{i=-k}^{g} b_i B_i^{k+1}(x)$, with the coincident additional knots, it is known that (Dierckx 1993)

$$\int_I s_k(x) \, dx = \frac{1}{k+1} \sum_{i=-k}^{g} b_i (\lambda_{i+k+1} - \lambda_i).$$

It means that B-spline coefficients of $s_k(x) \in Z_k^{\Delta \lambda}[a, b] \subset S_k^{\Delta \lambda}[a, b]$ satisfy condition

$$0 = \sum_{i=-k}^{g} b_i (\lambda_{i+k+1} - \lambda_i) = A \mathbf{b}$$

with $A = (\lambda_1 - \lambda_{-k}, \ldots, \lambda_{g+k+1} - \lambda_g)$, $\mathbf{b} = (b_{-k}, \ldots, b_g)^T$. And it is obvious that $\text{codim}(Z_k^{\Delta \lambda}[a, b]) = 1$, thus

$$\dim(Z_k^{\Delta \lambda}[a, b]) = \dim(S_k^{\Delta \lambda}[a, b]) - \text{codim}(Z_k^{\Delta \lambda}[a, b]) = g + k.$$

\[ \square \]

**Theorem 2** For the coincident additional knots (11), the functions $Z_{-k}^{k+1}(x)$, $\ldots$, $Z_{g-1}^{k+1}(x)$ form a basis for the space $Z_k^{\Delta \lambda}[a, b]$.

**Proof** Since $M_i^{k+1}(x)$ form a basis for the spline space $S_k^{\Delta \lambda}[a, b]$ and $Z_i^{k+1}(x) = M_i^{k+1}(x) - M_{i+1}^{k+1}(x)$, the functions $Z_i^{k+1}(x)$, $i = -k, \ldots, g - 1$, are linearly independent and lie in $Z_k^{\Delta \lambda}[a, b]$ with $\dim(Z_k^{\Delta \lambda}[a, b]) = g + k$. Therefore $Z_i^{k+1}(x)$, $i = -k, \ldots, g - 1$, form a basis for the $Z_k^{\Delta \lambda}[a, b]$. \[ \square \]
With regard to this theorem, every spline $s_k(x) \in \mathcal{Z}_{k}^{A\lambda}[a, b]$ has a unique representation

$$s_k(x) = \sum_{i=-k}^{g-1} z_i Z_{i+1}^{k+1}(x). \quad (14)$$

Now we can proceed to matrix notation of $s_k(x) \in \mathcal{Z}_{k}^{A\lambda}[a, b]$. With respect to (8) and (9), we are able to write the functions $Z_{i+1}^{k+1}(x)$ in matrix notation as

$$Z_{i+1}^{k+1}(x) = (k + 1) \left( B_{i+1}^{k+1}(x), B_{i+1}^{k+1}(x) \right) \begin{pmatrix} 1 & 0 & 0 \\ \frac{\lambda_i + k + 1 - \lambda_i}{\lambda_i + k + 2 - \lambda_{i+1}} & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$ (15)

Then it is clear that

$$(Z_{-k}^{k+1}(x), \ldots, Z_{g-1}^{k+1}(x)) = (B_{-k}^{k+1}(x), \ldots, B_{g}^{k+1}(x))DK = B_{k+1}(x)DK,$$

where

$$D = (k + 1) \text{diag} \left( \frac{1}{\lambda_1 - \lambda_{-k}}, \ldots, \frac{1}{\lambda_{g+k+1} - \lambda_g} \right) = (k + 1) \text{diag} \left( \frac{1}{l_1}, \ldots, \frac{1}{l_{g+k+1}} \right)$$ (16)

and

$$K = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix} \in \mathbb{R}^{g+k+1, g+k}.$$ (17)

Therefore the spline $s_k(x) \in \mathcal{Z}_{k}^{A\lambda}[a, b]$, $s_k(x) = \sum_{i=-k}^{g-1} b_i Z_{i+1}^{k+1}(x)$ can be written in matrix notation as

$$s_k(x) = Z_{k+1}(x)z = B_{k+1}(x)DKz,$$ (17)

where $Z_{k+1}(x) = (Z_{-k}^{k+1}(x), \ldots, Z_{g-1}^{k+1}(x))$ and $z = (z_{-k}, \ldots, z_{g-1})^\top$.

**Remark 1** The formula (17) is very useful, because we can use the standard B-spline basis for working with splines honoring the zero integral constraint, which is very convenient from a computational point of view.
Fig. 2 Cubic spline function $s_3(x)$ with the given coefficients $z = (0.5, -1, 2, 3, -8, 9, 1)^T$ in $Z_3^{Δλ}[0, 20]$ (left) and ZB-spline basis system of $Z_3^{Δλ}[0, 20]$ (right). Vertical dashed gray lines indicate knot positions (color figure online).

Example 1 We consider the knots $Δλ = \{λ_i\}_{i=0}^{g+1}$, $λ_0 = 0 = a < 2 < 5 < 9 < 14 < b = 20 = λ_5$. The task is to find a cubic spline with the given sequence of knots and which has zero integral on the interval $[0, 20]$. It is evident that $k = 3$, $g = 4$. We consider the additional knots

$$λ_{-3} = λ_{-2} = λ_{-1} = λ_0 = a = 0, \quad 20 = b = λ_5 = λ_6 = λ_7 = λ_8.$$ 

The basis functions of the space $Z_3^{Δλ}[0, 20]$ are plotted in Fig. 2 (right). Every spline $s_3(x) \in Z_3^{Δλ}[0, 20]$ can be written as

$$s_3(x) = \sum_{i=-3}^{3} z_i Z_i^4(x). \quad (18)$$

Thus, e.g., for $z = (z_{-3}, \ldots, z_3)^T = (0.5, -1, 2, 3, -8, 9, 1)^T$ the cubic spline $s_3(x)$ with zero integral is plotted in Fig. 2 (left).

Remark 2 In some applications it can be useful to consider splines on the unbounded interval $(-∞, ∞)$. For this purpose the same idea can be used. The only difference is that for given sequence of knots $λ_0 < λ_1 < \cdots < λ_g < λ_{g+1}$ we have to consider noncoincident additional knots (Schumaker 2007), i.e.

$$-∞ < λ_{-k} < \cdots < λ_{-1} < λ_0, \quad λ_{g+1} < λ_{g+2} < \cdots < λ_{g+k+1} < ∞.$$ 

Then every spline, i.e. every linear combination of the functions $Z_i^{k+1}(x)$, has zero integral over $\mathbb{R}$ because of (10).

3 Smoothing spline in $L_0^2(I)$

In Machalová et al. (2016), the construction of smoothing splines in the space $L_0^2(I)$ was studied, albeit using standard B-spline basis functions $B_i^{k+1}(x)$. Now we are able to construct smoothing splines in this space with new basis functions $Z_i^{k+1}(x)$. For this purpose, let data $(x_i, y_i), a \leq x_i \leq b$, weights $w_i > 0, i = 1, \ldots, n$, sequence
of knots $\Delta \lambda = \{\lambda_i\}_{i=0}^{g+1}, \lambda_0 = a < \lambda_1 < \ldots < \lambda_g < b = \lambda_{g+1}, n \geq g + 1$ and a parameter $\alpha \in (0, 1]$ be given. For arbitrary $l \in \{1, \ldots, k-1\}$ our task is to find a spline $s_k(x) \in Z^\Delta \lambda_k[a,b] \subset L_2^0(I)$, which minimizes the functional

$$J_l(s_k) = (1 - \alpha) \int_a^b \left[ s_k^{(l)}(x) \right]^2 dx + \alpha \sum_{i=1}^n w_i [y_i - s_k(x_i)]^2.$$  

Note that the choice of parameters $\alpha$ and $l$, where $l$ stands for $l$th derivation, affects smoothness of the resulting spline. The parameter $\alpha$ determines a kind of compromise between two complementary goals, to stay close to the given data or to obtain a smooth spline $s_k(x)$. As $\alpha$ approaches 1 the spline $s_k(x)$ converges to the least square spline and if $\alpha$ approaches 0 then $s_k(x)$ tends to the $L_2$ approximant (De Boor 1978). We consider cross-validation procedure in selecting the smoothing parameter $\alpha$, as detailed further in this section.

Let us denote $x = (x_1, \ldots, x_n)^\top, y = (y_1, \ldots, y_n)^\top, w = (w_1, \ldots, w_n)^\top$ and $W = diag(w)$. Regarding the representation (14) and matrix notation (17), the functional $J_l(s_k)$ can be written as a quadratic function

$$J_l(z) = (1 - \alpha) z^\top K^\top D S_l^\top M_{kl} S_l D^\top Kz + \alpha [y - B_{k+1}(x)D^\top Kz]^\top W [y - B_{k+1}(x)D^\top Kz],$$

see Machalová (2002a, b); Machalová et al. (2016) for details. In fact the matrix

$$M_{kl} = \left( m_{ij}^{kl} \right)_{i,j=-k+l}^g,$$

with

$$m_{ij}^{kl} = \int_a^b B_i^{k+1-l}(x) B_j^{k+1-l}(x) dx$$

is positive definite, because $B_i^{k+1-l}(x) \geq 0, i = -k + l, \ldots, g$ are basis functions. Upper triangular matrix $S_l = D_l L_l \ldots D_1 L_1 \in \mathbb{R}^{g+k+1-l,g+k+1}$ has full row rank. $D_j \in \mathbb{R}^{g+k+1-j,g+k+1-j}$ is a diagonal matrix such that

$$D_j = (k + 1 - j) diag(d_{-k+j}, \ldots, d_g)$$

with

$$d_i = \frac{1}{\lambda_{i+k+1-j} - \lambda_i}, \quad i = -k + j, \ldots, g,$$

and

$$L_j := \begin{pmatrix} -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{g+k+1-j,g+k+2-j}.$$
Finally, \( B_{k+1}(x) \in \mathbb{R}^{n \cdot g + k + 1} \) stands for the collocation matrix, i.e.

\[
B_{k+1}(x) = \left( B_{i+1}^{k+1}(x_j) \right)_{j=1, i=-k}^{n, g}.
\]

Using the notation \( U := DK \),

\[
G := U^\top \left[ (1 - \alpha) S_l^\top M_{kl} S_l + \alpha B_{k+1}^\top(x) W B_{k+1}^\top(x) \right] U
\]  
(21)

and

\[
g := \alpha K^\top D B_{k+1}^\top(x) W y,
\]

it is possible to rewrite the quadratic function \( J_l(z) \) as

\[
J_l(z) = z^\top G z - 2z^\top g + y^\top W y.
\]  
(22)

Our task is to find a spline \( s_k(x) \in Z_k^{\Delta \lambda} [a, b] \) which minimizes the functional \( J_l(s_k) \), in other words, we want to find a minimum of the function (22). It is obvious that this function has just one minimum if and only if the matrix \( G \) is positive definite (p.d.). From (21) it can be easily seen that

\[
G \text{ is p.d. } \iff B_{k+1}(x) \text{ is of full column rank.}
\]

From Schoenberg-Whitney theorem and its generalization, see De Boor (1978) and Machalová (2002a), it is known that matrix \( B_{k+1}(x) \) is of full column rank if and only if

\[
\text{there exists } \{u_{-k}, \ldots, u_g\} \subset \{x_1, \ldots, x_n\} \\
\text{with } u_i < u_{i+1}, \ i = -k, \ldots, g - 1, \\
\text{such that } \lambda_i < u_i < \lambda_{i+k+1}, \ i = -k, \ldots, g.
\]

(23)

In this case from the necessary and sufficient condition for a unique minimum of quadratic function, i.e.

\[
\frac{\partial J_l(z)}{\partial z^\top} = 0,
\]

we get a system of linear equations \( Gz = g \) and then the unique solution of this system is given by

\[
z^* = G^{-1} g.
\]  
(24)
Consequently, the resulting smoothing spline is obtained by the formula

\[ s_k^*(x) = \sum_{i=-k}^{g-1} z_i^* Z_i^{k+1}(x), \]

in matrix notation using standard B-splines \( B_{k+1}(x) \) as

\[ s_k^*(x) = B_{k+1}(x)DKz^*, \]

where the vector \( z^* = (z_{-k}^*, \ldots, z_{g-1}^*)^\top \) is given in (24).

It is obvious that the quality of the resulting smoothing spline is substantially determined also by the number of knots and their locations. Importantly, the choice of knots should reflect the anticipated course of the function. In the context of functional data analysis of density functions, so far a heuristic approach was utilized (Machalová et al. 2016; Hron et al. 2016; Talšká et al. 2018) and it is followed also here. However, there are also some recent papers (Goepp et al. 2018; Liu et al. 2019) considering an automatic choice of knots which can be applied in the presented theory and will be further investigated in the near future.

To use smoothing splines in practice, it is necessary to choose the smoothing parameter \( \alpha \). Although this choice can be done subjectively, it is desirable to have an automatic method for this purpose, especially when large amounts of data sets need to be processed. One of the most popular methods is cross-validation (CV) (Ramsay and Silverman 2005). The basic idea of (leave-one-out) CV is to leave \( i \)th observation out and use remaining \( n-1 \) observations to fit a smoothing spline with respect to \( \alpha \). Denote by \( s_k^{*(i-1)} \) the estimated curve using all data but the \( i \)th (in contrast to \( s_k^* \), the smoothing spline calculated from all data). This method gives an automatic choice of \( \alpha \) by minimizing the cross-validation criterion \( CV(\alpha) \) on a grid of values of \( \alpha \) from \((0, 1]\), given by

\[ CV(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - s_k^{*(i-1)}(x_i) \right)^2. \]

That is, the selection of \( \alpha \) is based on prediction error, i.e., how well \( s_k^{*(i-1)}(x_i) \) predicts \( y_i \).

In order to calculate \( CV(\alpha) \), it is necessary to fit the smoothing spline \( n \) times, once for each omitted data point \((x_i, y_i)\). Nevertheless, the criterion (25) can be equivalently written as (Ramsay and Silverman 2005),

\[ CV(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i - s_k^*(x_i)}{1 - h_{ii}} \right)^2, \]

\( \odot \) Springer
where $h_{ii}$ is the $i$th diagonal element of the projection matrix $H(\alpha)$ (called hat matrix), given by
\[
H(\alpha) = \alpha B_{k+1}(x) U G^{-1} U^\top B_{k+1}(x)^\top W. \tag{27}
\]
Thus one needs to calculate the smoothing spline only once together with the diagonal elements of the hat matrix (27). Finally, the criterion (26) can be modified by replacing the values $h_{ii}$ by their average. This leads to the generalized cross-validation (GCV), originally introduced by Craven and Wahba (1976), and defined as
\[
GCV(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - s_k^*(x_i))^2}{(1 - \text{trace}(H(\alpha))/n)^2}. \tag{28}
\]
As shown in Wahba (1990), CV may have a tendency to under-smooth, therefore GCV is sometimes preferred in practice.

4 Orthogonalization of basis functions

A further step is to orthogonalize the basis
\[
Z_{k+1}(x) = (Z_{-k}^{k+1}(x), \ldots, Z_{g-1}^{k+1}(x))^\top
\]
of the space $Z_{\Delta\lambda}^k[a, b]$ that is by construction oblique with respect to the $L^2$ space metric. For this purpose the idea presented in Redd (2012) is used. We search for a linear transformation $\Phi$ such that
\[
O_{k+1}(x) = \Phi Z_{k+1}(x)
\]
forms an orthogonal set of basis functions of the space $Z_{\Delta\lambda}^k[a, b]$, i.e.
\[
\int_a^b O_{k+1}(x) O_{k+1}^\top(x) \, dx = I.
\]
Regarding the lemma presented in Redd (2012) and the notation used here, we can formulate the following statement.

**Lemma 1** An invertible transformation $\Phi$ orthogonizes the basis functions $Z_{k+1}(x)$ if and only if it satisfies the condition that
\[
\Phi^\top \Phi = \Sigma^{-1},
\]
where $\Sigma$ represents the positive definite matrix

$$
\Sigma = \int_a^b Z_{k+1}(x)Z_{k+1}^\top(x) \, dx = \left( \int_a^b Z_{i+1}^{k+1}(x)Z_{j+1}^{k+1}(x) \, dx \right)_{i,j=-k}^{g-1}.
$$

With respect to the definition of basis functions $Z_{k+1}(x) = \left( Z_{-k+1}^{k+1}(x), \ldots, Z_{g-1}^{k+1}(x) \right)^\top$ the matrix $\Sigma$ can be expressed as

$$
\Sigma = K^\top D M K,
$$

(29)

where $M := M_{k0}$. The linear transformation $\Phi$ is not unique and can be computed for example by the Cholesky decomposition. The basis functions

$$
O_{k+1}(x) = \Phi Z_{k+1}(x), \quad O_{k+1}(x) = (O_{-k}^{k+1}(x), \ldots, O_{g-1}^{k+1}(x))^\top
$$

are orthogonal and have a zero integral. The linear and quadratic ZB-splines with zero integral and their orthogonalization are plotted in Fig. 3.

To sum up, the spline $s_k(x)$ with zero integral can be constructed as a linear combination of orthogonal splines $O_{k+1}^{k+1}(x)$ having zero integral in a form

$$
s_k(x) = \sum_{i=-k}^{g-1} z_i O_{i+1}^{k+1}(x) = O_{k+1}(x)z.
$$

On the other hand, the standard and well-known B-splines $B_{k+1}^{k+1}(x)$ can be used to represent $s_k(x) \in \bar{Z}_{k}^{\Delta^k}[a, b]$ in matrix form

$$
s_k(x) = \Phi B_{k+1}(x)D K z.
$$

This formulation seems to be very useful because it allows us to use existing B-spline codes in software R or Matlab, for example, the collocation matrix or computation of integrals in (20) (Ramsay et al. 2019).

The orthogonalization of the splines leads to an orthonormal basis - as well as the simplicial functional principal component analysis (SFPCA) (Hron et al. 2016), used in Appendix for anthropometric data, does. Furthermore, the projection over these bases are in fact isometric log-ratio coordinates (Van den Boogaart et al. 2014).

5 Compositional splines in the Bayes space

Construction of splines directly in $L_0^2(I)$ has important practical consequences, however, it is crucial also from the theoretical perspective. Expressing splines as functions
(a) Linear ZB-splines $Z_2^i(x)$ with given knots 0, 0, 1, 2, 3, 3 (left), linear orthogonal ZB-splines $O_2^i(x)$ (right).

(b) Quadratic ZB-splines $Z_3^i(x)$ with given knots 0, 0, 0, 1, 2, 3, 3, 3 (left), quadratic orthogonal ZB-splines $O_3^i(x)$ (right).

Fig. 3 Example of ZB-spline functions of degree 1 and 2 and corresponding orthogonal ZB-splines (a and b respectively). Vertical dashed lines indicate knot positions.

in $L_0^2(I)$ enables to back-transform them to the original Bayes space $B^2(I)$ using (5). It results in compositional B-splines (CB-splines), obtained from (8) as

$$
\zeta^{k+1}_i(x) = \frac{\exp[Z^{k+1}_i(x)]}{\int_j \exp[Z^{k+1}_i(y)] \, dy}, \quad i = -k, \ldots, g - 1, \quad k \geq 0. \quad (30)
$$

Accordingly, for instance ZB-splines from Fig. 1a, b and e can be now expressed directly in the Bayes space $B^2(I)$ as CB-splines, see Fig. 4. Note that CB-splines $\zeta^{k+1}_i(x)$ fulfill the unit integral constraint which is, however, not necessary for further considerations. As a consequence, it is immediate to define a vector space $C^\Delta_\lambda[a, b]$ of compositional polynomial splines of degree $k \geq 0$, defined on a finite interval $[a, b]$ with the sequence of knots $\Delta\lambda$. From isomorphism between $\mathcal{C}^\Delta_\lambda[a, b]$ and $Z^\Delta_\lambda[a, b]$ it holds that

$$
\dim \left( \mathcal{C}^\Delta_\lambda[a, b] \right) = g + k.
$$
Moreover, from isometric properties of clr transformation it follows that every compositional spline \( \xi_k(x) \in C^\Delta_k[a, b] \) in \( B^2(I) \) can be uniquely represented as

\[
\xi_k(x) = \bigoplus_{i=-k}^{g-1} z_i \odot \xi_i^{k+1}(x).
\]  

CB-splines \( \xi_i^{k+1}(x) \) forming the basis are by the default setting (8) not orthogonal. Their orthogonalization can be done as described in Sect. 4, i.e. by employing \( L_0^2(I) \) with the back-transformation to \( B^2(I) \).
The resulting compositional splines (with either orthogonal, or non-orthogonal CB-spline basis) can be used for representation of densities directly in $B^2(I)$. This is an important step in construction of methods of functional data analysis involving density functions, like for ANOVA modeling (Van den Boogaart et al. 2014) or for the SFPCA method introduced in Hron et al. (2016) and demonstrated in the next section. In the latter case CB-splines were indeed already used for construction of the procedure, although at the respective level of development the authors were not aware of that. With CB-splines one has a guarantee that methods are developed consistently in the Bayes space. Moreover, the possibility of having an orthogonal basis enables to gain additional features resulting from orthogonality of finite dimensional projection in combination with approximate properties of spline functions.

As usual, compositional splines can be tuned according to concrete problem, with the advantage of their direct formulation in the Bayes space sense.

**Example 2** To illustrate smoothing of concrete data with a compositional spline, 1000 values from standard normal distribution were simulated and the support was determined by minimum and maximum simulated values, $I = [x_{\text{min}}, x_{\text{max}}]$. Data were collected in a form of histogram, where the breakpoints are equidistantly spaced. The representative points of the histogram cells are denoted as asterisks in Fig. 5 (right) and these points form the input data $(x_i, y_i), i = 1, \ldots, n$ for smoothing purpose. The $y$-values stand for discretized relative contributions to the overall probability mass, therefore discrete clr transformation (Aitchison 1982) is needed to obtain a real vector with the zero sum constraint (Fig. 5, left). These data points were smoothed using the procedure described in Sect. 3 and back-transformed to the original space. In the concrete setting $k = 2, l = 1, \alpha = 0.9, \Delta \lambda = [x_{\text{min}}, -2, -1, 0, 1, 2, x_{\text{max}}], w_i = 1, \forall i = 1, \ldots, n$ were considered. The resulting spline $s_2(x)$ with the zero integral on interval $[x_{\text{min}}, x_{\text{max}}]$ is plotted in Fig. 5 (left). In the right plot the compositional spline $\xi_2(x)$ with the unit integral by using (5) is depicted. We might also use GCV to determine the optimal value of $\alpha$. Due to smoothness and a characteristic symmetric “bell curve” shape of the underlying Gaussian density, from which the data points were sampled, the criterion values get monotonically smaller for $\alpha \to 1$ (see Fig. 6 with the GCV error values for 30 equally spaced values of $\alpha$ from $[0.01, 1])$, thus the least
square spline approximation would be preferred. This may not necessarily be the case in practice, as demonstrated in application to anthropometric data in Sect. 7.

6 Simulation study

A simulation study was designed to examine the sensitivity of compositional smoothing splines to the setting of selected input parameters. We focus here on the number of classes, knots and sampled data. For this purpose a (truncated) gamma density with the scale parameter $\theta = 3$ and the shape parameter $\kappa = 2$ as reference density on the interval $I = [0.5, 15]$ was considered, i.e.,

$$f(x) = B x^\kappa \exp \left( -\frac{x}{2} \right), \quad x \in I; \quad (32)$$

Note that due to the scale invariance of densities, the normalization constant in (32) was omitted and simply a representant within the equivalence class of proportional densities was taken (this is reflected by the notation “$= B$”). We started with 600 realizations from this distribution and focused at first on the sensitivity of results to the number of classes $q$ upon which histogram data are built. Specially, the number $q$ was determined by the Sturges’ rule, suggesting an optimal value of $q = 10$ classes, and its lower ($q = 8$) and higher ($q = 13$) values were taken additionally. According to the shape of the gamma distribution, we consider cubic smoothing splines ($k = 3, l = 2$) on the restricted interval $I = [1, 14.5]$; this step prevents the undesirable impact of the resulting approximation at the boundaries of the respective domain. Accordingly, the total of five equally spaced knots were set as $\{\lambda_0 = 1, \ldots, \lambda_4 = 14.5\}$. The smoothing parameter $\alpha$ was determined by a generalized cross-validation; the cross-validation error (28) was minimized on a grid of 30 equally spaced values from $[0.01, 1]$. Finally, the simulation was replicated $R = 100$ times for each set of parameters and the quality of the resulting approximations of gamma density function was assessed in terms of the integrated square error (ISE) between the true (reference) and estimated density functions, i.e.,

$$\text{ISE} = \left\| f(x) \otimes \hat{f}_r(x) \right\|_B^2, \quad r = 1, \ldots, R.$$
The panel A of Fig. 7 displays boxplots of integrated square errors for \( q = 8, 10, 13 \). It seems that the resulting approximation is not much sensitive to that parameter setting so that the Sturges’ rule can be viewed in general as a reasonable choice.

Having fixed the number of classes according to the Sturges’ rule, the sensitivity of the result to the number of knots was assessed considering both equally and non-equally spaced knots in the same simulation setting (Fig. 7, panel B). For their non-equal spacing, we took into account the fact that the reference density is right-skewed and customized the placing of knots accordingly in the left part of the domain. Considering the number of knots in \( \{3, 5, 7\} \), the simulation suggests to use its moderate number in both cases. Moreover it shows that placing knots with respect to the behaviour of data can lead to better results.

Finally, effect of the sample size of the input (discretized) distributional data on the estimation of the gamma density function was tested, taking 300 and 1000 of them. The results are consistent with the previous ones (hence omitted): they confirm the overall good performances of Sturges’ rule and suggest to use a moderate number of knots in all the cases. Moreover, from the panel C of Fig. 7 it can be observed that with an increasing number of sampled data we get lower ISE, thus more accurate estimates.

To asses the sensitivity of the parameters for symmetric densities, we further employed a (truncated) Gaussian density with the mean \( \mu = 0 \) and the standard deviation \( \sigma = 1.5 \) on the interval \( I = [-5, 5] \), i.e.,

\[
f(x) = 8 \exp \left\{-\frac{x^2}{2\sigma^2}\right\}, \quad x \in I.
\] (33)

The simulation design is the same as before. Nevertheless, in this case, the simulations suggest to set the number of classes by using Sturges’ rule only for the lower number of sampled data (i.e., 300), and to use higher number of classes for a higher number of sampled data (i.e., 600 and 1000), see the first column of Fig. 8. The second column of Fig. 8 suggests to use a lower number of knots, as a higher number of knots may lead to overfitting of the data. Finally, Fig. 8d clearly indicates that increasing sample size of the input discretized distributional data leads to better results in terms of ISE.
(a) Sensitivity to the number of classes (left): 7, 9 (Sturges’ rule), 12. Sensitivity to the number of knots in \{3, 5, 7\} using 9 classes (right). Sample size of the sampled data: 300.

(b) Sensitivity to the number of classes (left): 8, 10 (Sturges’ rule), 13. Sensitivity to the number of knots in \{3, 5, 7\} using 13 classes (right). Sample size of the sampled data: 600.

(c) Sensitivity to the number of classes (left): 8, 11 (Sturges’ rule), 14. Sensitivity to the number of knots in \{3, 5, 7\} using 14 classes (right). Sample size of the sampled data: 1000.

(d) Sensitivity to the number of starting data with optimal parameter setting: 300 (9 classes, 3 knots), 600 (13 classes, 3 knots) and 1000 (14 classes, 3 knots).

**Fig. 8** Boxplots of ISE between the true normal density and its estimates. Panel (a)–(c): sensitivity to the number of classes (left) and to the number of equidistant knots (right). Panel (d): sensitivity to the sample size of the input data.
Finally, for the optimal setting for both the gamma density (600 sampled data, 10 classes, 3 knots) and the normal density (600 sampled data, 13 classes, 3 knots) also empirical error bounds were constructed by depicting upper and lower borders of estimated compositional splines together with the respective true density functions, see Fig. 9. For their visual inspection the clr space case is more appropriate which shows heavier tails for both densities. Nevertheless, due to relative scale of the Bayes space geometry, only for the left tail of the gamma density the error bound indicates some potentially problematic behavior in the original space. The heavier tails might occur due to keeping the zero integral constraint and could be suppressed by a proper choice of knots, see the next section.

7 Application to anthropometric data

For the purpose of illustrating the smoothing procedure outlined in Sect. 3, a real-world data set dealing with the most commonly used anthropometric measure relating to body weight is presented. The data set we consider collects the body weight of apparently healthy Czech adolescents and young adults aged 15-31 years (the total of 4436 records) which were recruited non-randomly by offering free body composition assessment. Body weight was measured by the InBody 720 device (Biospace Co., Ltd, Seoul, Korea), recorded as the total body mass rounded to the nearest 0.1 kg.

The raw data for each of \( N = 16 \) age groups, i.e. \([15, 16), [16, 17), \ldots, [30, 31)\], were turned into a form of histogram data as follows. The sampled values of the body weight in each age group were divided into equally-spaced classes of the united range 40-110 kg and the optimal number of classes, denoted by \( q_i, i = 1, \ldots, 16 \), was set according to the Sturje’s rule separately across the age groups. Because of
the insufficient number of sampled data for males and females in each age group, gender information was ignored. Although there might be some differences between male and female samples due to sex dimorphism, they are not that dramatic, e.g., contrary to body height, because the weight is influenced also by external factors (nutrition, physical activity), and still allow for a reasonable aggregation of data. Subsequently, the proportions in classes within each age group were computed and present zero-values (zero counts in the respective classes) were imputed by values \((2/3) \cdot (1/n_i), i = 1, \ldots, N\) according to Martín-Fernández et al. (2015), where \(n_i\) stands for the number of observations in \(i\)th age group. Finally, the raw discretized density data \(f_{i,j}, i = 1, \ldots, N, j = 1, \ldots, q_i\) which correspond to the midpoints \(t_{i,j}\) of classes, \(i = 1, \ldots, N, j = 1, \ldots, q_i\) (i.e., \(f_{i,j} = f(t_{i,j})\)), were obtained by dividing (not necessary normalized) proportions \(p_i = (p_{i,1}, \ldots, p_{i,q_i})^\top\) of counts in classes by the length of the respective intervals resulting from partition of the weight range in each of age groups.

Fig. 11 shows four examples of histograms with different number of classes together with raw data (Table 1) to be smoothed. To do so, their transformation into real vectors is needed. We note that if the histogram data are constructed on subintervals of the same length, i.e. with equally-spaced breakpoints, it enables to use the discrete version of the clr transformation (Aitchison 1982) directly on the vector of proportions \(p_i, i = 1, \ldots, N\) by considering the scale invariance property. Otherwise, the input of the clr transformation must be vectors with raw density data \(f_i = (f_{i,1}, \ldots, f_{i,q_i})^\top, i = 1, \ldots, N\) to avoid misleading results which would not reflect the actual behaviour of data. Vectors of clr transforms are hereafter denoted as clr\((f_i) = (\text{clr}(f_{i,1}), \ldots, \text{clr}(f_{i,q_i}))^\top\) for \(i = 1, \ldots, N\) and clr values are listed in Table 2.

Having the collected data \((t_{i,j}, \text{clr}(f_{i,j}))\), we proceed to smooth them with the compositional smoothing splines using a system of ZB-spline basis functions from the \(L_2^2(I)\). They are considered on domain \(I = [40, 107]\) which has been modified in order to avoid undesired artefacts in densities at their right-hand side. For all \(N\) observations, the same strategy was followed to set the values of the input parameters for the smoothing procedure. We employed cubic smoothing splines \((k = 3, l = 2)\) with the given sequence of knots \(\Delta \lambda = \{40, 62, 84, 107\}\) and set the vectors of weights \(w_i\) for all input data to vectors of ones. The value of the smoothing parameter \(\alpha\) was set by performing generalized CV: Fig. 10 plots the GCV error \((\text{28})\) against equally spaced 30 values of \(\alpha\) from \([0.01, 1]\); the minimum was reached for \(\alpha = 0.147\). The resulting compositional smoothing splines are obtained via their clr representation.
Fig. 11 Histograms for four age groups: [15,16), [22,23), [23,24) and [30,31) together with estimated probability density functions via compositional smoothing splines. Asterisks indicate discrete data \((t_i,j, f_i,j), i = 1, 8, 9, 16, j = 1, 2, \ldots, q_i\), and \(p_{ij}, i = 1, 8, 9, 16, j = 1, 2, \ldots, q_i\) indicate proportions of equidistant classes resulted for given partition of the range weight body values. The smoothing parameter \(\alpha\) was chosen by generalized cross-validation

\[
s_3^i(t) = \sum_{\nu = -3}^{1} z_{i,\nu} Z_{\nu}^4(t), \quad i = 1, \ldots, N, \quad t \in I; \tag{34}
\]

the corresponding ZB-spline coefficients are reported in Table 3.

Figure 12a displays an example of three raw density data from Fig. 11 together with smoothed curves in the \(L_2^4\) space (right) and after the inverse transformation (5) in the \(B^2\) space (left). The whole sample of smoothed density functions (Fig. 12b) is plotted on blue scale distinguishing the age groups. Two trends are apparent—in the younger age groups, the estimated density functions are right skewed and exhibit lower variability while with increasing age they become more symmetric followed by higher variability. Nevertheless, density function in age group [23, 24) does not fully respect this behaviour: the variability trend holds, but the distribution of weights is more similar to those in younger age groups as it is skewed more to the left. In general, it seems that adolescents appear to be predominantly of a lower body weight than the older persons whose weight is more spread over the weight classes and more pronounced in the middle part of the distribution. Accordingly, there is also a higher incidence of higher weights in comparison with the younger adolescents.

Importantly, the quality of smoothing is the same irrespective of the shape of the distribution. It might just happen that the smoothed densities exhibit heavier tails although they are not indicated by the data \(f_{i,j}\), even with some artefacts typical for overfitting. This is obviously due to keeping the zero integral constraint, which is more sensitive to deviations from monotonic character of densities at their tails.
Table 1  Histogram data for four age groups: [15, 16), [22, 23), [23, 24) and [30, 31). $f_i$ are raw density values at midpoints $t_i = (t_{i,1}, \ldots, t_{i,q_i})^T$ of weight classes with proportions $p_i$ for $i = 1, 8, 9, 16$; $q_i$ indicates the number of the weight classes

|          | $p_i$    | 0.0656  | 0.2625  | 0.3375  | 0.2156  | 0.0750  | 0.0281  | 0.0094  | 0.0062  |
|----------|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| $f_1$    | 0.0075   | 0.0300  | 0.0386  | 0.0246  | 0.0086  | 0.0032  | 0.0011  | 0.0007  |
| $t_1$    | 44.375   | 53.125  | 61.875  | 70.625  | 79.375  | 88.125  | 96.875  | 105.625 |
| $q_1 = 8$|          |         |         |         |         |         |         |         |
|          | $p_8$    | 0.0156  | 0.0869  | 0.1804  | 0.2138  | 0.2272  | 0.1514  | 0.0935  | 0.0200  | 0.0067  | 0.0045  |
| $f_8$    | 0.0022   | 0.0124  | 0.0258  | 0.0305  | 0.0325  | 0.0216  | 0.0134  | 0.0029  | 0.0010  | 0.0006  |
| $t_8$    | 43.5     | 50.5    | 57.5    | 64.5    | 71.5    | 78.5    | 85.5    | 92.5    | 99.5    | 106.5   |
| $q_8 = 10$|          |         |         |         |         |         |         |         |
|          | $p_9$    | 0.0078  | 0.0908  | 0.2100  | 0.1971  | 0.1659  | 0.1659  | 0.1011  | 0.0259  | 0.0537  | 0.0017  |
| $f_9$    | 0.0011   | 0.0130  | 0.0300  | 0.0282  | 0.0237  | 0.0237  | 0.0144  | 0.0037  | 0.0048  | 0.0002  |
| $t_9$    | 43.5     | 50.5    | 57.5    | 64.5    | 71.5    | 78.5    | 85.5    | 92.5    | 99.5    | 106.5   |
| $q_9 = 10$|          |         |         |         |         |         |         |         |
|          | $p_{16}$ | 0.0568  | 0.1023  | 0.2045  | 0.2386  | 0.2159  | 0.1364  | 0.0455  |
| $f_{16}$ | 0.0057   | 0.0102  | 0.0205  | 0.0239  | 0.0216  | 0.0136  | 0.0045  |
| $t_{16}$ | 45.0     | 55.0    | 65.0    | 75.0    | 85.0    | 95.0    | 105.0   |
| $q_{16} = 7$|        |         |         |         |         |         |         |         |
Table 2: Input data for smoothing procedure: clr\( (f_i) \) are raw clr density values at midpoints \( t_i = (t_{i,1}, \ldots, t_{i,q_i})^\top \) of the weight classes for \( i = 1, 2, \ldots, N \); \( q_i \) indicates the number of the weight classes.

| [15, 16) | clr\( [f_1] \) | 0.100 | 1.486 | 1.737 | 1.289 | 0.233 | −0.748 | −1.846 | −2.252 |
|----------|----------------|-------|-------|-------|-------|-------|--------|--------|--------|
|          | \( t_1 \)     | 44.375| 53.125| 61.875| 70.625| 79.375| 88.125 | 96.875 | 105.625|
| [16, 17) | clr\( [f_2] \) | −0.210| 1.217 | 1.636 | 0.396 | −0.392| −2.001 | −2.407 | −1.536 |
|          | \( t_2 \)     | 44.375| 53.125| 61.875| 70.625| 79.375| 88.125 | 96.875 | 105.625|
| [17, 18) | clr\( [f_3] \) | −1.375| 0.570 | 1.316 | 1.669 | 1.381 | −0.364| −2.069 | −1.663 |
|          | \( t_3 \)     | 43.889| 51.667| 59.444| 67.222| 75.000| 82.778 | 90.556 | 98.333 |
| [18, 19) | clr\( [f_4] \) | −1.354| 0.592 | 1.419 | 1.443 | 1.406 | 1.131 | 0.563  | −0.661 |
|          | \( t_4 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [19, 20) | clr\( [f_5] \) | −1.536| 0.628 | 1.408 | 1.555 | 1.535 | 1.209 | 0.302  | −0.774 |
|          | \( t_5 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [20, 21) | clr\( [f_6] \) | −1.341| 0.674 | 1.333 | 1.558 | 1.638 | 1.452 | 0.422  | −0.568 |
|          | \( t_6 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [21, 22) | clr\( [f_7] \) | −1.746| 0.451 | 1.185 | 1.463 | 1.411 | 1.131 | 0.531  | −0.242 |
|          | \( t_7 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [22, 23) | clr\( [f_8] \) | −1.168| 0.550 | 1.281 | 1.450 | 1.511 | 1.106 | 0.624  | −0.917 |
|          | \( t_8 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [23, 24) | clr\( [f_9] \) | −1.884| 0.573 | 1.412 | 1.348 | 1.177 | 1.177 | 0.681  | −0.417 |
|          | \( t_9 \)     | 43.5  | 50.5  | 57.5  | 64.5  | 71.5  | 78.5  | 85.5  | 92.5  |
| [24, 25) | clr\( [f_{10}] \) | −1.602| 0.595 | 1.186 | 1.274 | 1.106 | 0.796 | 0.056  | −0.423 |
|          | \( t_{10} \)  | 43.889| 51.667| 59.444| 67.222| 75.000| 82.778| 90.556 | 98.333 |
| [25, 26) | clr\( [f_{11}] \) | −1.401| 0.471 | 0.768 | 0.824 | 1.145 | 0.850 | 0.209  | −1.178 |
|          | \( t_{11} \)  | 43.889| 51.667| 59.444| 67.222| 75.000| 82.778| 90.556 | 98.333 |
| [26, 27) | clr\( [f_{12}] \) | −1.045| 0.513 | 0.901 | 1.180 | 1.258 | 0.513 | −0.485 | −2.836 |
| \( t_{12} \) | 44.375 | 53.125 | 61.875 | 70.625 | 79.375 | 88.125 | 96.875 | 105.625 |
|---|---|---|---|---|---|---|---|---|
| clr \([f_{13}]\) | -0.816 | 0.570 | 0.742 | 0.742 | 1.056 | 0.570 | -0.256 | -2.608 |
| \( t_{13} \) | 44.375 | 53.125 | 61.875 | 70.625 | 79.375 | 88.125 | 96.875 | 105.625 |
| \( q_{13} = 8 \) |
| \( t_{14} \) | 45.0 | 55.0 | 65.0 | 75.0 | 85.0 | 95.0 | 105.0 |
| clr \([f_{14}]\) | -1.155 | 0.579 | 0.790 | 0.965 | 0.690 | -0.308 | -1.561 |
| \( q_{14} = 7 \) |
| \( t_{15} \) | 45.0 | 55.0 | 65.0 | 75.0 | 85.0 | 95.0 | 105.0 |
| clr \([f_{15}]\) | -1.060 | 0.480 | 0.837 | 0.674 | 0.614 | -0.773 | -0.773 |
| \( q_{15} = 7 \) |
| \( t_{16} \) | 45.0 | 55.0 | 65.0 | 75.0 | 85.0 | 95.0 | 105.0 |
| clr \([f_{16}]\) | -0.756 | -0.168 | 0.525 | 0.679 | 0.579 | 0.120 | -0.979 |
| \( q_{16} = 7 \) |
possible way out, applied here, was to reduce slightly the range from \( I = [40, 110] \) to \( I = [40, 107] \) in order to keep predominantly the monotonic behaviour of (normalized) counts \( f_{i,j} \) at the right tails.

Of course, the smoothed data can be further analysed using methods of functional data analysis (Ramsay and Silverman 2005), adapted in order to respect specific properties of densities. It is demonstrated here for case of the compositional functional principal component analysis (SFPCA) (Hron et al. 2016) with the aim to reduce dimensionality and provide insight into the data structure. A detailed analysis is contained in Appendix.

### 8 Conclusions

The compositional splines, which enable to construct a spline basis in the clr space of density functions (ZB-spline basis) and consequently also in the original space of densities (CB-spline basis), might become an important contribution within the Bayes space methodology for processing of functional data carrying relative information. They provide a solid theoretical base for further developments of the approximation theory in context of the Bayes spaces, but even more importantly, compositional splines can be used also for adaptation of popular methods of functional data analysis for density functions. Here the case of compositional functional principal component analysis was presented, but similarly, e.g., regression analysis or classification methods could be developed. Also further tuning of the compositional splines is possible, here represented by the smoothing compositional splines or by orthonormalization of the

---

**Table 3** ZB-spline coefficients for clr transformed density functions of \( N = 16 \) age groups

| Age group | Spline coefficients, \( z_i = (z_{i,-3}, \ldots, z_{i,1})^\top, \ i = 1, \ldots, N \) |
|-----------|---------------------------------------------------------------|
| [15, 16]  | \(-5.963, 5.630, 47.697, 40.241, 13.340\)                   |
| [16, 17]  | \(-7.130, -1.337, 42.567, 44.401, 14.289\)                  |
| [17, 18]  | \(-15.521, -12.815, 20.197, 42.455, 9.424\)                 |
| [18, 19]  | \(-16.140, -9.981, 22.225, 33.419, 20.142\)                |
| [19, 20]  | \(-17.433, -11.153, 23.670, 37.668, 16.515\)               |
| [20, 21]  | \(-16.130, -8.419, 19.725, 45.969, 18.796\)                |
| [21, 22]  | \(-19.186, -12.452, 13.311, 36.055, 14.907\)               |
| [22, 23]  | \(-14.185, -6.793, 18.397, 45.321, 14.264\)                |
| [23, 24]  | \(-21.234, -13.636, 18.031, 23.526, 20.037\)               |
| [24, 25]  | \(-18.698, -15.234, 14.313, 20.413, 19.280\)               |
| [25, 26]  | \(-17.180, -8.200, -2.082, 29.486, 10.203\)               |
| [26, 27]  | \(-13.739, -11.717, 3.764, 28.106, 19.076\)                |
| [27, 28]  | \(-12.845, -5.971, 0.737, 21.801, 17.969\)                |
| [28, 29]  | \(-15.955, -13.331, -0.034, 20.649, 10.629\)              |
| [29, 30]  | \(-13.977, -11.869, 1.421, 21.514, 3.470\)                |
| [30, 31]  | \(-5.590, -14.968, -4.919, 8.745, 7.450\)                 |
(a) Example of smoothed raw density data.

(b) Smoothed raw density data.

**Fig. 12** Smoothed weight density functions via compositional smoothing splines in $B^2$ space (left) and its clr transformation in $L^2$ space (right). Data are displayed on blue scale distinguishing age groups: with increasing age, the intensity of blue color increases. Vertical dashed gray lines indicate knots position. The smoothing parameter $\alpha$ was chosen by generalized cross-validation.

ZB-basis. The latter case be used for an orthogonal projection of a density function on a subset of CB-splines, to further applications within the approximation theory or also for development of the theoretical framework of functional data analysis.

The pending challenge is to generalize the methodology introduced above also to $p$ dimensional density functions, $p > 1$, which can be formally extended from any univariate density $f(x)$, $x \in I = [a, b]$ to $f(x)$, where $x = (x_1, \ldots, x_p)\top \in I = I_1 \times \ldots \times I_p = [a_1, b_1] \times \ldots \times [a_p, b_p]$, in Equations (1) to (6); $\eta = b - a$ would be replaced by $H = \prod_{i=1}^p (b_i - a_i)$. Currently an approach which focuses on keeping the zero integral constraint of the clr transformed densities was developed in Guégan and Iacopini (2018) and Hron et al. (2020) as a generalization of Machalová et al. (2016), which, however, does not lead to a compositional counterpart of the B-spline basis. A consistent approach in this direction is currently under development.

**Acknowledgements** The authors gratefully acknowledge both the support by Czech Science Foundation GA19-01768S and the grant IGA_PrF_2018_024 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Appendix: SFPCA with anthropometric data**

The SFPCA has been recently designed based on the Bayes space methodology (Hron et al. 2016), so it enables to capture the main modes of relative variability in a data
set consisting of sampled density functions. Given a data set of \( N \) zero-mean functional observations \( X_1, \ldots, X_N \) in \( B^2(I) \), SFPCA aims to find the (normalized and orthogonal) directions of maximum variability in dataset, i.e., a collection of density functions \( \{ \theta_\kappa \}_{\kappa \geq 1} \) called simplicial functional principal components (SFPCs) which maximize the following objective function over \( \theta \in B^2(I) \),

\[
\sum_{i=1}^{N} \langle X_i, \theta \rangle_B^2 \quad \text{subject to} \quad \| \theta \|_B = 1; \quad \text{with} \quad \langle \theta, \theta_\kappa \rangle_B = 0, \ \kappa < K, \tag{35}
\]

where the orthogonality condition is assumed to be fulfilled for \( K \geq 2 \). Since \( \langle X_i, \theta \rangle_B \) represents a projection of \( X_i \) along the direction \( \theta \in B^2(I) \), in fact we look for orthogonal basis functions in \( B^2(I) \) maximizing the relative variability of these projections. The maximization task (35) is efficiently implemented by applying the clr transformation (3) and the output can be back-transformed from the \( L^2(I) \) to the \( B^2(I) \) space, as detailed in Hron et al. (2016).

The interpretation of SFPCs can be performed by displaying:

- individual SFPCs (as clr transformed density functions, SFPCs always represent contrasts between the parts of domain \( I \));
- overall mean density function \( \bar{X} \) along with its perturbation by SFPCs powered by a suitable constant,
  \[
  \bar{X} \oplus \sqrt{\rho_\kappa} \otimes \theta_\kappa, \quad \bar{X} \ominus \sqrt{\rho_\kappa} \otimes \theta_\kappa, \tag{36}
  \]

  where \( \rho_\kappa \) is an amount of the variability of the dataset along the direction \( x_\kappa \) and it holds \( \rho_1 \geq \rho_2 \geq \ldots \). This is a natural choice because SFPCs represent variation around the overall mean density function;
- the projection of dataset along the directions \( \theta_\kappa \),
  \[
  \langle X_i, \theta_\kappa \rangle_B \otimes \theta_\kappa = x_{i\kappa} \otimes \theta_\kappa, \quad i = 1, \ldots, N,
  \]

  where \( x_{i\kappa} = \langle X_i, \theta_\kappa \rangle_B, \ i = 1, \ldots, N \) are so called principal component scores associated with the \( \kappa \)th SFPCs \( \theta_\kappa \). The scores can be plotted for pairs of the first SFPCs to assess the relationship among sampled density functions or to reveal presence of outlying observations;
- to complete the above interpretation it is important to note that the original functions \( X_1, \ldots, X_N \) and the eigenfunctions \( \theta_\kappa, \ k \geq 1 \) are based on a CB-spline basis expansion

\[
X_i(\cdot) = \bigoplus_{v=-3}^{1} c_{i,v} \otimes \zeta_v(\cdot), \quad \theta_\kappa(\cdot) = \bigoplus_{v=-3}^{1} b_{\kappa,v} \otimes \xi_v(\cdot), \tag{37}
\]

respectively, by considering (34). Note that the CB-splines \( \xi_v(\cdot) \) are the same in both expansions.
SFPCA is also a statistical method for reducing dimensionality of dataset. The number of SFPCs can be determined from the scree plot which displays percentage of variability explained by each SFPC. That is, the dimensionality is identified by a point in scree plot at which explained variability drops off.

For the actual computation, CB-splines (37) of the input functional observations, represented by the corresponding ZB-splines (34), were expressed by B-splines with B-spline coefficients (listed in Table 4) using formula (17), so that the computations were done as in Hron et al. (2016). The CB-splines and ZB-splines are important for theoretical developments and interpretation. From this perspective they provide a clear update of the spline representation used in Hron et al. (2016) which has led to singularity constraint of the resulting B-spline coefficients (Talská et al. 2018).

The output of SFPCA for body weight density functions is reported in Fig. 13. According to scree plot (Fig. 13a), two or three SFPCs should be taken, but we resort to use only first two of them which capture together more than 86% of the total variability of the data set. The first SFPC (Fig. 13c) represents the contrast between the weight below and above 78 kg, which could be considered as a reference (average) weight. Hence, higher scores along the SFPC 1 are expected for age groups with higher incidence of individuals with higher body weight than the average, and, conversely, lower scores are associated with age groups with higher incidence of individuals with lower body weight then the average. The interpretation of SFPC 1 can be obviously linked with age, see Fig. 13b. The scree plot more or less separates rather right skewed weight density functions of younger age groups (located on the left in the scree plot) from those more symmetric ones associated with older age groups (located on the right in the scree plot).

The second SFPC (Fig. 13d) characterizes the variability within the tails of density functions, i.e. the main contribution to the variability along SFPC 2 is provided by the lowest and highest weight values ($\leq 51$ kg and $\geq 98$ kg respectively). It contrasts low and high weights (associated with high scores along the SFPC 2) against middle weight values (associated with low scores along the SFPC 2), see Fig. 13b. The consistent interpretation can be also observed from Fig. 13e which displays the variation along the first two directions—SFPC 1 and SFPC 2—with respect to sample mean $\bar{f}(t)$, $t \in I$ (i.e. $f \oplus / \ominus 2\sqrt{\kappa} \otimes \text{SFPC}_\kappa$, $\kappa = 1, 2$).

Figure 13f, g, respectively, represent two main modes of variability in the data set $((f_i, \theta_\kappa)_{B} \otimes \theta_\kappa, \kappa = 1, 2, i = 1, \ldots, N)$. For instance, the variation along SFPC 2 is confirmed to be exhibited in tails of density functions and the observations with lowest (gold curve) and highest score (red curve) further support the conclusions made so far. The high scores along the second direction thus reflect heavier tails and, conversely, the low scores along the second direction reflects low incidence of individuals with extreme (both small and high) weights. Nevertheless, the relationship of scores (Fig. 13b) is apparent: at the beginning, they continue to fall, reach a bottom and then continue to grow. The relationship might be partially explained by unequal representation of men and women in age groups and unequal number of observations in these age groups. Another reason might be that data corresponding to age groups with low SFPC 2 scores
were collected mostly from students of the Faculty of Physical Culture at Palacký University in Olomouc, Czech Republic, which form more homogeneous population than an average one. In any case, the second SFPC reveals an interesting feature which is worth to be further investigated.
Table 4  B-spline coefficients for clr transformed density functions of $N = 16$ age groups

| Age group       | Spline coefficients, $b_i = (b_{i,-3}, \ldots, b_{i,2})^T$, $i = 1, \ldots, N$ |
|-----------------|----------------------------------------------------------------------------------|
| [15, 16]        | $-1.084$ $1.054$ $2.511$ $-0.445$ $-2.391$ $-2.320$                            |
| [16, 17]        | $-1.296$ $0.527$ $2.621$ $0.109$ $-2.677$ $-2.485$                            |
| [17, 18]        | $-2.822$ $0.246$ $1.971$ $1.329$ $-2.936$ $-1.639$                            |
| [18, 19]        | $-2.935$ $0.560$ $1.923$ $0.668$ $-1.180$ $-3.503$                            |
| [19, 20]        | $-3.170$ $0.571$ $2.079$ $0.836$ $-1.880$ $-2.872$                            |
| [20, 21]        | $-2.933$ $0.701$ $1.680$ $1.567$ $-2.415$ $-3.269$                            |
| [21, 22]        | $-3.488$ $0.612$ $1.538$ $1.358$ $-1.880$ $-2.593$                            |
| [22, 23]        | $-2.579$ $0.672$ $1.504$ $1.607$ $-2.761$ $-2.481$                            |
| [23, 24]        | $-3.861$ $0.691$ $1.891$ $0.328$ $-0.310$ $-3.485$                            |
| [24, 25]        | $-3.400$ $0.315$ $1.764$ $0.364$ $-0.101$ $-3.353$                            |
| [25, 26]        | $-3.124$ $0.816$ $0.365$ $1.885$ $-1.714$ $-1.774$                            |
| [26, 27]        | $-2.498$ $0.184$ $0.924$ $1.453$ $-0.803$ $-3.317$                            |
| [27, 28]        | $-2.335$ $0.625$ $0.400$ $1.258$ $-0.341$ $-3.125$                            |
| [28, 29]        | $-2.901$ $0.239$ $0.794$ $1.235$ $-0.891$ $-1.849$                            |
| [29, 30]        | $-2.541$ $0.192$ $0.793$ $1.200$ $-1.604$ $-0.603$                            |
| [30, 31]        | $-1.016$ $-0.853$ $0.600$ $0.816$ $-0.115$ $-1.296$                            |

References

Aitchison J (1982) The statistical analysis of compositional data. J R Stat Soc Ser B 5:139–177
Craven P, Wahba G (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math 31:377–403
De Boor C (1978) A practical guide to splines, vol 27. Springer, New York
Delicado P (2011) Dimensionality reduction when data are density functions. Comput Stat Data Anal 55(1):401–420
Dierckx P (1993) Curve and surface fitting with splines. Oxford University Press, Oxford
Eaton ML (1983) Multivariate statistics: a vector space approach. Wiley, New York
Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barcelo-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Math Geol 35(3):279–300
Egozcue JJ, Díaz-Barrero JL, Pawlowsky-Glahn V (2006) Hilbert space of probability density functions based on Aitchison geometry. Acta Mathematica Sinica 22(4):1175–1182
Goepp V, Bouaziz O, Nuel G (2018) Spline Regression with Automatic Knot Selection. arXiv preprint arXiv:1808.01770v1
Guégan D, Iacopini M (2018) Nonparametric forecasting of multivariate probability density functions. arXiv preprint arXiv:1803.06823
Hron K, Menafoglio A, Templ M, Hruzová K, Filzmoser P (2016) Simplicial principal component analysis for density functions in Bayes spaces. Comput Stat Data Anal 94:330–350
Hron K, Machalová J, Menafolio A (2020) Bivariate densities in Bayes spaces: orthogonal decomposition and spline representation. Submitted
Liu X, Nassar H, PodgórsKI K (2019) Splinets-efficient orthogonalization of the B-splines. arXiv preprint arXiv:1910.07341v1
Machalová J (2002a) Optimal interpolating and optimal smoothing spline. J Electr Eng 53:79–82
Machalová J (2002b) Optimal interpolatory splines using B-spline representation. Acta Universitatis Palackianae Olomucensis Facultas Rerum Naturalium Mathematica 41(1):105–118
Machalová J, Hron K, Monti GS (2016) Preprocessing of centred logratio transformed density functions using smoothing splines. J Appl Stat 43(8):1419–1435
Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J (2015) Bayesian-multiplicative treatment of count zeros in compositional data sets. Stat Modell 15(2):134–158
Menafoglio A, Guadagnini A, Secchi P (2014) A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. Stoch Env Res Risk Assess 28(7):1835–1851
Menafoglio A, Guadagnini A, Secchi P (2016) Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach. Water Resour Res 52(8):5708–5726
Pawlowsky-Glahn V, Egozcue JJ (2001) Geometric approach to statistical analysis on the simplex. Stoch Env Res Risk Assess 15(5):384–398
Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data. Wiley, Chichester
Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer Series in Statistics. Springer, New York
Ramsay JO, Hooker G, Graves S (2019) Functional data analysis with R and MATLAB. Springer, New York
Redd A (2012) A comment on the orthogonalization of B-spline basis functions and their derivatives. Stat Comput 22(1):251–257
Schumaker L (2007) Spline functions: basic theory. Cambridge University Press, Cambridge
Talská R, Menafoglio A, Machalová J, Hron K, Fišerová E (2018) Compositional regression with functional response. Comput Stat Data Anal 123:66–85
Talská R, Menafoglio A, Hron K, Egozcue JJ, Palarea-Albaladejo J (2020) Weighting the domain of probability densities in functional data analysis. Stat, accepted for publication
Van den Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2010) Bayes linear spaces. Stat Oper Res Trans 34(4):201–222
Van den Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2014) Bayes Hilbert spaces. Aust New Zealand J Stat 56(2):171–194
Wahba G (1990) Spline models for observational data. SIAM, Philadelphia

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.