Bivariate densities in Bayes spaces: orthogonal decomposition and spline representation

Karel Hron · Jitka Machalová · Alessandra Menafoglio

Received: 23 September 2021 / Revised: 7 April 2022 / Accepted: 7 September 2022 / Published online: 22 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
A new orthogonal decomposition for bivariate probability densities embedded in Bayes Hilbert spaces is derived. It allows representing a density into independent and interactive parts, the former being built as the product of revised definitions of marginal densities, and the latter capturing the dependence between the two random variables being studied. The developed framework opens new perspectives for dependence modelling (e.g., through copulas), and allows the analysis of datasets of bivariate densities, in a Functional Data Analysis perspective. A spline representation for bivariate densities is also proposed, providing a computational cornerstone for the developed theory.

Keywords Compositional data · Functional data · Tensor product splines · Anthropometric data

1 Introduction

The analysis of distributional data is gaining an increasing interest in the applied sciences. Distributional data, such as probability density functions (PDFs) or cumulative distribution functions, are routinely collected in social sciences (e.g., population pyramids (Delicado 2011, Hron et al. 2016) and geosciences (e.g., particle-size distributions (Menafoglio et al. 2014, 2016)). Analyses of distributional data based on
methods designed for functional data in $L^2$ often lead to inappropriate results, such as negative predictions (Menafoglio et al. 2018; Talská et al. 2018).

It is now widely recognized that an appropriate statistical analysis of PDF data should be precisely based on their characterizing properties (e.g., (Nerini and Ghattas 2007; Petersen and Müller 2016; Hron et al. 2016; Menafoglio et al. 2014)). In the literature, several approaches have been proposed to serve the purpose of analysing datasets of PDFs. Most works propose to analyse PDF data through a prior data transformation. For instance, Delicado (2011) considers a transformation approach to the principal component analysis of a dataset of PDFs. (Srivastava et al. 2007) use a square root transformations of densities to deal with a time-warping function in registration. (Petersen and Müller 2016) propose a set of transformations that map the PDF data to a Hilbert space, where further statistical analyses are possible; this setting allows for, e.g., principal component analysis, classification, regression. A relatively large body of recent literature proposes the use of the Wasserstein metric to define a notion of distance for density data. Such metric has appealing interpretations, being related to the problem of optimal transport. However, it defines a non-linear space (i.e., a Riemannian manifold), thus requiring the development of ad hoc methods for this setting (e.g., (Bigot et al. 2019; Panaretos and Zemel 2019; Petersen et al. 2019)), based on Frechét statistics. A different approach is that relying on the theory of Bayes linear spaces, that represent a generalization to the infinite-dimensional setting of the Compositional Data Analysis (CoDa, (Pawlowsky-Glahn et al. 2015)) approach. In this setting, PDFs are considered as infinite-dimensional objects that provide relative information van den Boogaart et al. (2010), van den Boogaart et al. (2014). Bayes Hilbert spaces were built as to represent the so-called principles of CoDa (i.e., scale invariance, relative scale, sub-compositional coherence, see (Pawlowsky-Glahn et al. 2015)), through a Hilbert geometry for PDFs. The Hilbert structure of the space allows one to develop most methods of functional data analysis, while accounting for the peculiar nature of PDFs. These include principal component analysis Hron et al. (2016), functional regression Talská et al. (2018), spatial prediction Mena foglio et al. (2014), profile monitoring Mena foglio et al. (2018), time-series analysis Kokoszka et al. (2019), Seo and Beare (2019). Even though the statistical literature is nowadays well-developed for distributional data, little attention has been paid so far to the setting of multivariate densities, whose study is of paramount importance in the applications. A first contribution in the direction of bivariate densities was recently provided in the preprint by (Guégan and Iacopini 2019), which uses the theory of Bayes Hilbert spaces over bivariate domains to study the temporal dynamic of coupled time series, modelled through copulas. As a key element of innovation with respect to previous literature, the present work proposes a novel statistical framework for bivariate PDFs that allows studying the dependence between the target random variables, grounding on the geometry of the Bayes space. We provide new meaningful notions of compositional marginals (so-called geometric marginals), which play the roles of the marginal distributions, consistent with the Bayes geometry. We further derive an orthogonal decomposition of bivariate PDFs in terms of independence and interaction parts, generalizing the well-known results developed in the discrete case (i.e., for compositional tables, (Egozcue et al. 2008, 2015)). To allow for explicit computations of the marginals and of the
latter representation, we develop a novel $B$-spline representation for bivariate PDFs, compatible with the compositional nature of the data.

The methodological results of our work shed light on the structure of multivariate Bayes spaces, suggesting a direction to create connections between the theory of Bayes spaces and the theory of copulas Nelsen (2006), which are widely used to build multivariate PDFs from marginals. Note that the theory of copulas is well-established and allows one to describe the joint distribution function of two random objects under very general assumptions. Our work is mostly focused on density functions (PDFs) instead, entailing a difference in the approaches in terms of (i) the assumptions made on the distribution at hand and (ii) the theoretical properties of the object being studied. In this sense, this work presents the initial steps in the direction of a new framework for dependence modeling, for which the extension to more general distributions (e.g., not absolutely continuous) is foreseen. On the other hand, this work is primarily aimed to build an analytical framework for datasets made of multivariate distributional objects, within the context of Functional Data Analysis (FDA, (Ramsay and Silverman 2005)). In this view, building the dependence modelling on PDFs might be preferable, and this would be consistent with the usual practice in FDA, where regularity assumptions (continuity, boundedness, squared-integrability) are typically made on data. In this context, the appealing properties of the Bayes space approach which are discussed in this work (resulting, e.g., from the orthogonal decomposition of the bivariate density into independent and interactive parts), are seen as key factors potentially fostering the development and interpretation of new FDA methods for multivariate distributional observations. In this sense, the methodology presented in the paper is going to offer an alternative viewpoint to the standard copula theory by providing an orthogonal decomposition of bivariate PDFs, while opening a novel frontier to analyse samples of bivariate densities using the methods of FDA.

The remaining part of this work is organized as follows. In Section 2, the Bayes space methodology is recalled from (van den Boogaart et al. 2014), with particular reference to bivariate densities. This enables us to develop an orthogonal decomposition of bivariate densities into independent and interactive parts, thoroughly discussed in Section 3 and demonstrated with simulated truncated Gaussian densities in Section 4. In Section 5, a spline representation for bivariate densities mapped in the $L^2$ space is introduced; such representation is relevant to allow processing raw data and to develop efficient computational methods. In Section 6 the theoretical framework is applied to a time series of bivariate densities coming from an anthropometric cross-sectional study. The final Section 7 concludes with some overview, comments, and further perspective.

2 PDFs as elements of a Bayes space

Bayes spaces are designed to provide a geometrical representation for density functions characterized by the property of scale invariance (van den Boogaart et al. 2014). The latter property assumes that, given a domain $\Omega$ and a positive real multiple $c$, two proportional positive functions $f(x)$ and $g(x)$ (i.e., such that $g(x) = cf(x)$, for $c > 0$) carry essentially the same, relative information van den Boogaart et al. (2014). This
follows also the common strategy used in Bayesian statistics where multiplying factors are typically dropped from computations, as these are not essential to the definition of the distributions at hand. Note that the scale invariance of a density $f$ is a direct consequence of the same property of the associated measure $\mu$, i.e., of the $\sigma$-finite measure $\mu$ such that $f = d\mu/dP$ for a reference measure $P$. In this context, we refer to the so-called $B$-equivalence of measures (and densities): two measures $\mu$ and $\nu$ are $B$-equivalent if they are proportional, i.e., there exists a positive real multiple $c$ such that $\nu(A) = c \cdot \mu(A)$ for any $A \in A$, $A$ being a sigma-algebra on $\Omega$.

Given a $\sigma$-finite measure $P$, the Bayes space $B^2(P)$ is a space of $B$-equivalence classes of $\sigma$-finite positive measures $\mu$ with square-integrable log-density w.r.t. $P$, i.e.,

$$B^2(P) = \left\{ \mu \in B^2(P) : \int \left| \ln \frac{d\mu}{dP} \right|^2 dP < +\infty \right\}.$$

From the practical point of view, an important role is played by the reference measure $P$, as thoroughly investigated in (Talská et al. 2020). The choice of the reference measure determines a weighting of the domain $\Omega$ of the PDF, which can be used to give more relevance to certain regions of $\Omega$ when conducting FDA, according to the purpose of the analysis van den Boogaart et al. (2014), Egozcue and Pawlowsky-Glahn (2016). Given that the weighting of the domain is not of primary interest here and one would intuitively resort simply to the Lebesgue reference measure, the discussion on $P$ might seem somehow lateral to the main focus of this work. Nevertheless, as we will see already in Theorem 2, the scale of $P$ indeed matters for a meaningful decomposition of a bivariate density into independent and interactive parts. For this reason, we here limit to mention two key points which shall be useful in the following. First, in general, an analysis based on a reference measure $P$ does not provide the same results as an analysis based on $cP$, for $c > 0$. Indeed, using $P$ or $cP$ typically leads to a difference in the scale of the result. Second, to change the reference measure from $\lambda$ to a measure $P$ with strictly positive $\lambda$-density $p = dP/d\lambda$, the well-known chain rule can be used. For a generic measure $\mu$ one has

$$\mu(A) = \int_A \frac{d\mu}{d\lambda} d\lambda = \int_A \frac{d\mu}{d\lambda} \cdot \frac{d\lambda}{dP} dP = \int_A \frac{d\mu}{d\lambda} \cdot \frac{1}{p} dP.$$

In the following, some of the presented results will only be valid for normalized reference measures, as we will specify below; nonetheless, we prefer to work in a complete generality, to allow for the use of non-normalized references which are preferred in some settings Egozcue and Pawlowsky-Glahn (2016).

The Bayes space, as described above, can also be defined for the case when the domain $\Omega$ is a Cartesian product of two domains $\Omega_X$ and $\Omega_Y$, i.e., $\Omega = \Omega_X \times \Omega_Y$. In this case, the reference measure $P$ can be decomposed as a product measure $P = P_X \times P_Y$ and the Hilbert space structure of the Bayes space $B^2(P)$ van den Boogaart et al. (2014), Egozcue and Pawlowsky-Glahn (2016) can be built accordingly. In this case,
the operations of **perturbation** and **powering** can be defined for any two bivariate densities \( f, g \) with respect to \( P \), i.e., \( f, g \in B^2(P) \), and a real constant \( \alpha \) as

\[
(f \oplus g)(x, y) = B^2(P) f(x, y) \cdot g(x, y) \quad \text{and} \quad (\alpha \odot f)(x, y) = B^2(P) f(x, y)^\alpha,
\]

respectively. The lower index in \( = B^2(P) \) means that the right hand side of the equations can be arbitrarily rescaled without altering the relative information that the resulting density in \( B^2(P) \) contains. The Hilbert space structure is completed by defining the inner product,

\[
\langle f, g \rangle_{B^2(P)} = \frac{1}{2P(\Omega)} \int_{\Omega_x} \int_{\Omega_y} \ln \frac{f(x, y)}{f(s, t)} \ln \frac{g(x, y)}{g(s, t)} \, dP(x, y) \, dP(s, t) = \frac{1}{2P(\Omega)} \int_{\Omega_x} \int_{\Omega_y} \ln \frac{f(x, y)}{f(s, t)} \ln \frac{g(x, y)}{g(s, t)} \, dP_X(x) \, dP_Y(y) \, dP_X(s) \, dP_Y(t),
\]

which implies in the usual way also the norm and the distance,

\[
\| f \|_{B^2(P)} = \sqrt{\langle f, f \rangle_{B^2(P)}}, \quad d_{B^2(P)}(f, g) = \| f \ominus g \|_{B^2(P)},
\]

where \( f \ominus g = f \oplus \left[ (-1) \odot g \right] \) is the perturbation-subtraction of densities. Here, the definition of the inner product (1) is presented according to (Egozcue and Pawlowsky-Glahn 2016). While the scale of the reference measure \( P \) does not have any impact for the operations of perturbation and powering, it does influence the inner product because the scale corresponds to shrinkage (or expansion) of the Bayes space (for details, see (Talská et al. 2020)).

The usual strategy when dealing with the Bayes spaces van den Boogaart et al. (2014), Menafoglio et al. (2014), Hron et al. (2016) is not to process densities directly in the original space but to map them into (a subset of) the standard \( L^2 \) space where most of the widely used methods of functional data analysis (FDA, (Ramsay and Silverman 2005)) can be employed. The **clr transformation** of a bivariate density \( f(x, y) \in B^2(P) \) is a real function \( \text{clr}(f) : \Omega \to \mathbb{R} \), \( \text{clr}(f) \in L^2_0(P) \), defined—using Fubini’s theorem—as

\[
\text{clr}(f)(x, y) = \ln f(x, y) - \frac{1}{P(\Omega)} \int_{\Omega} \ln f(x, y) \, dP = \ln f(x, y) - \frac{1}{P(\Omega)} \int_{\Omega_X} \int_{\Omega_Y} \ln f(x, y) \, dP_X \, dP_Y.
\]

Similarly as for perturbation and powering, the scale of \( P \) does not play any role in (3), too. On the other hand, one should note that the resulting function \( \text{clr}(f) \) is expressed with respect to reference \( P \). As a consequence, using any measure other than the Lebesgue \( \lambda \) leads to clr-transformations defined over a weighted \( L^2 \) space \( L^2(P) \) Egozcue and Pawlowsky-Glahn (2016). Given that, in the following, the reference measure will always be known from the context, we drop it from notation of
the clr transformation. Moreover, one should also take into account the zero-integral 
constraint of clr transformed densities, i.e.,
\[
\int_{\Omega_X} \int_{\Omega_Y} \text{clr}(f)(x, y) \, dP_X dP_Y = 0. \tag{4}
\]
In the following, we shall indicate by \( L^2_0(P) \) the subspace of the \( L^2(P) \) space of 
(equivalence classes of) functions having zero integral; in particular, one clearly has 
that \( \text{clr}(f)(x, y) \in L^2_0(P) \). Nevertheless, previous works focused on the univariate 
case demonstrate that this constraint usually does not represent any serious obstacle 
for the application of FDA methods, especially if a proper spline representation of the 
densities is used Hron et al. (2016), Machalová et al. (2016), Talská et al. (2018). Since 
a reliable and flexible spline representation forms a cornerstone in a large number of 
computational methods for FDA Ramsay and Silverman (2005), we shall pay special 
attention in developing a bivariate \( B \)-splines basis suited to represent clr transformation 
of bivariate densities in Sect. 5.

3 Decomposition of bivariate densities

One of the key goals in probability theory is to study the dependence structure between 
two random variables. A systematic approach to the analysis of dependence structure 
is represented by the theory of copulas Nelsen (2006), firstly introduced by Sklar Sklar 
(1959). The well-known Sklar’s theorem provides a decomposition of any PDF into 
its interactive and independent parts, the latter being built as the product of the respec-
tive marginal PDFs. Relying on the Bayes space methodology allows one to provide 
a similar decomposition which is now, however, orthogonal. This important property 
enables for an elegant geometrical representation of the decomposition, and for a pow-
erful probabilistic interpretation if a normalized reference measure is used, with direct 
consequences from the statistical viewpoint. For example, the proposed decomposi-
tion allows one to derive a measure of dependence called simplicial deviance, defined 
as the squared norm of the density expressing (solely) the relationships between both 
variables (factors).

The orthogonal decomposition of bivariate densities grounds into a novel definition 
of marginals, named geometric marginals, which are built upon marginalizing the 
bivariate clr transformation as follows. Given \( x \in \Omega_X \) and \( y \in \Omega_Y \), we define the clr 
marginals as
\[
\text{clr}(f_{X,Y})(x) := \int_{\Omega_Y} \text{clr}(f)(x, y) \, dP_Y \\
= \int_{\Omega_Y} \ln f(x, y) \, dP_Y - \frac{P_X(\Omega_X)}{P(\Omega)} \int_{\Omega_X} \int_{\Omega_Y} \ln f(x, y) \, dP_X dP_Y \tag{5}
\]
and

\[
\text{clr}(f_{Y,g})(y) = \int_{\Omega_X} \text{clr}(f)(x, y) \, dP_X \\
= \int_{\Omega_X} \ln f(x, y) \, dP_X - \frac{P_Y(\Omega_Y)}{P(\Omega)} \int_{\Omega_X} \int_{\Omega_Y} \ln f(x, y) \, dP_X \, dP_Y,
\]

(6)

respectively. It is easily seen that \(\text{clr}(f_{X,g}) \in L_0^2(P_X)\) and \(\text{clr}(f_{Y,g}) \in L_0^2(P_Y)\), where \(L_0^2(P_i)\) stands for the subspace of \(L^2(P_i)\) whose elements integrate to zero. We define the geometric marginals \(f_{X,g} \in B^2(P_X)\) and \(f_{Y,g} \in B^2(P_Y)\) as the elements of \(B^2(P_X)\) and \(B^2(P_Y)\) associated with the clr-marginals \(\text{clr}(f_{X,g}), \text{clr}(f_{Y,g})\), respectively, i.e.,

\[
f_{X,g}(x) =_{B(P_X)} \exp \left\{ \text{clr}(f_{X,g})(x) \right\} =_{B(P_X)} \exp \left\{ \int_{\Omega_Y} \ln f(x, y) \, dP_Y \right\},
\]

\[
f_{Y,g}(y) =_{B(P_Y)} \exp \left\{ \text{clr}(f_{Y,g})(y) \right\} =_{B(P_Y)} \exp \left\{ \int_{\Omega_X} \ln f(x, y) \, dP_X \right\}.
\]

(7)

In the following, the terms marginal, \(X\)-marginal and \(Y\)-marginal will always refer to the geometric notion of marginals given in (7).

In probability theory, the independence of random variables corresponds to the possibility of expressing a joint density as a product of its marginals. In a setting where the latter are defined as the geometric marginals (7), the independent and interactive parts of \(f(x, y) \in B^2(P)\) can be defined, respectively, as

\[
f_{\text{ind}}(x, y) = f_{X,g}(x)f_{Y,g}(y), \quad (x, y) \in \Omega
\]

(8)

and

\[
f_{\text{int}}(x, y) = \frac{f(x, y)}{f_{X,g}(x)f_{Y,g}(x)} = f(x, y) \ominus f_{\text{ind}}(x, y),
\]

(9)

where \(f_{X,g}(x)\) and \(f_{Y,g}(y)\) are the geometrical marginals defined above. The first and foremost important property of the proposed decomposition

\[
f(x, y) = f_{\text{ind}}(x, y) \oplus f_{\text{int}}(x, y)
\]

(10)

for a bivariate density \(f(x, y)\) is the orthogonality its parts. In the following, the geometrical marginals will be formally taken as bivariate functions, i.e. \(f_{X,g}(x) = f_{X,g}(x, y)\) and \(f_{Y,g}(y) = f_{Y,g}(x, y)\), and considered as elements of \(B^2(P)\); similarly for their clr counterparts. This enables, among others, to express the independence density \(f_{\text{ind}}\) as sum (perturbation) of the geometric marginals, i.e.,

\[
f_{\text{ind}}(x, y) = f_{X,g}(x) \oplus f_{Y,g}(y).
\]

(11)
In this context, note that the geometric marginals are also orthogonal projections on the respective marginal spaces $B^2(\mathcal{P}_X)$ and $B^2(\mathcal{P}_Y)$ spanning the product space $B^2(\mathcal{P}_X \otimes \mathcal{P}_Y) = B^2(\mathcal{P}_X) \otimes B^2(\mathcal{P}_Y)$. Likewise, $f_{\text{ind}}$ is the orthogonal projection on $B^2(\mathcal{P}_X \otimes \mathcal{P}_Y)$ and hence $f_{\text{int}}$ the one of $f$ on the orthogonal complement of it.

**Theorem 1** For the independent and interactive parts of a bivariate density $f(x, y)$, it holds that

(i) $\langle f_{\text{ind}}, f_{\text{int}} \rangle_{B^2(\mathcal{P})} = 0$, or equivalently that

(ii) $\langle \text{clr}(f_{\text{ind}}), \text{clr}(f_{\text{int}}) \rangle_{L^2_0(\mathcal{P})} = 0$.

The proof of Theorem 1—as well as those of the following theorems—is reported in Appendix B. Note that, from the orthogonality of the decomposition $f = f_{\text{ind}} \oplus f_{\text{int}}$, the Pythagorean theorem follows directly, i.e., $\| f \|^2_{B^2(\mathcal{P})} = \| f_{\text{ind}} \|^2_{B^2(\mathcal{P})} + \| f_{\text{int}} \|^2_{B^2(\mathcal{P})}$.

A further important property of independence densities $f_{\text{ind}}$ is the following. Call **arithmetic marginals** the usual marginal distributions (a similar notation is used in the discrete case of compositional tables Egozcue et al. (2015))

$$f_{X,a}(x) = \int_{\Omega_Y} f(x, y) d\mathcal{P}_Y, \quad f_{Y,a}(y) = \int_{\Omega_X} f(x, y) d\mathcal{P}_X.$$  

It is clear that, if the theory were built on arithmetic marginals, the above decompositions (10) and (11) together with the statement of Theorem 1 would not be achieved. We remark that the previous results hold true both for normalized and non-normalized reference measures. On the other hand, if the reference measure is normalized, an interesting link can be derived between the two types of marginals (geometric or arithmetic) when the independent part is concerned. Indeed, the following result states that, whenever the random variables $X$, $Y$ are independent and the reference measure is a probability measure, the bivariate PDF coincides with its independent part defined in (8). In this case, the arithmetic and the geometric marginals coincide.

**Theorem 2** Let $f$ be an independence density and let the reference measure $\mathcal{P} = \mathcal{P}_X \times \mathcal{P}_Y$ be the product measure of probability measures $\mathcal{P}_X$, $\mathcal{P}_Y$. Then the arithmetic and geometric marginals of $f$ coincide.

As such, when the measure $\mathcal{P}$ is normalized, the independent part built through the geometric marginals enables one to fully capture the joint distribution of two random variables when these are independent.

The next theorem states the mutual orthogonality between the geometric marginals $(f_{X,g}(x)$ and $f_{Y,g}(y))$ and the interaction density $f_{\text{int}}(x, y)$.

**Theorem 3** The $X$-marginal and $Y$-marginal are orthogonal with respect to the Bayes space $B^2(\mathcal{P})$, i.e., $\langle f_{X,g}, f_{Y,g} \rangle_{B^2(\mathcal{P})} = 0$. Moreover, the marginals are also orthogonal to the interaction density, i.e., $\langle f_{X,g}, f_{\text{int}} \rangle_{B^2(\mathcal{P})} = 0$ and $\langle f_{Y,g}, f_{\text{int}} \rangle_{B^2(\mathcal{P})} = 0$.

The relations $\langle \text{clr}(f), \text{clr}(f_{X,g}) \rangle_{L^2_0(\mathcal{P})} = \| \text{clr}(f_{X,g}) \|^2_{L^2_0(\mathcal{P})}$, $\langle \text{clr}(f), \text{clr}(f_{Y,g}) \rangle_{L^2_0(\mathcal{P})} = \| \text{clr}(f_{Y,g}) \|^2_{L^2_0(\mathcal{P})}$ and $\langle \text{clr}(f_{X,g}), \text{clr}(f_{Y,g}) \rangle_{L^2_0(\mathcal{P})} = 0$ nicely illustrate that $X$- and $Y$-marginals of the density $f(x, y)$ represent its orthogonal projections. In addition, the

 Springer
Pythagorean theorem between the independence density and its projections holds,
\[ \| f_{\text{ind}} \|_{B^2(P)}^2 = \| f_{X,g} \|_{B^2(P)}^2 + \| f_{Y,g} \|_{B^2(P)}^2. \]

As a consequence of Theorems 2-3, one can conclude that, in the case of independence and normalized \( P \), the arithmetic and geometric marginals coincide, and the interaction part is null (i.e., it is the neutral element of perturbations). More in general, the next result states that the geometric marginals are completely determined by the independent part of the bivariate density. Here, the clr marginals of \( f_{\text{int}} \) are defined as

\[
\text{clr}(f_{\text{int},X}) := \int_{\Omega} \ln f_{\text{int}}(x,y) dP_Y - \frac{P_Y(\Omega_Y)}{P(\Omega)} \int_\Omega f_{\text{int}}(x,y) dP,
\]

\[
\text{clr}(f_{\text{int},Y}) := \int_{\Omega} \ln f_{\text{int}}(x,y) dP_X - \frac{P_X(\Omega_X)}{P(\Omega)} \int_\Omega f_{\text{int}}(x,y) dP,
\]

and the geometric marginals \( f_{\text{int},X,g}, f_{\text{int},Y,g} \) are the associated densities in \( B^2(P) \).

**Theorem 4** Whenever the reference measure is the product measure of probability measures \( P_X, P_Y \), the geometric marginals \( f_{\text{int},X,g}, f_{\text{int},Y,g} \) of the interaction part \( f_{\text{int}} \) coincide with the neutral element of perturbation, i.e., for any \( f \) in \( B^2(P) \) one has

\[ f \oplus f_{\text{int},X,g} = f; \quad f \oplus f_{\text{int},Y,g} = f, \tag{12} \]

or equivalently

\[ \text{clr}(f_{\text{int},X,g}) = \text{clr}(f_{\text{int},Y,g}) = 0. \tag{13} \]

Theorem 4 motivates the name *interaction* density. Indeed, decomposition (10) applied to \( f_{\text{int}} \) reads

\[ f_{\text{int}} = 0_{\oplus} \oplus f_{\text{int}}, \]

where \( 0_{\oplus} \) is the neutral element of perturbation (with respect to probability reference measure \( P \)). Accordingly, the independent part of an interaction density is the null element \( 0_{\oplus} \). On the other hand, for an independent density, the interaction part is null. More in general, for any bivariate density \( f \), the nearest independence density is \( f_{\text{ind}} \), and its distance from it is precisely \( \| f_{\text{int}} \|_{B^2(P)} \). The squared norm \( \| f_{\text{int}} \|_{B^2(P)}^2 \) can be thus taken as a proper measure of dependence. For consistency with the discrete case Egozcue et al. (2015), we shall name it *simplicial deviance*, \( \Delta^2(f) = \| f_{\text{int}} \|_{B^2(P)}^2 \). Dividing the simplicial deviance by the squared norm of the bivariate density, one obtains a relative measure of dependence, hereafter named *relative simplicial deviance*,

\[ R^2_{B^2(P)}(f) = \frac{\| f_{\text{int}} \|_{B^2(P)}^2}{\| f \|_{B^2(P)}^2}, \quad 0 \leq R^2_{B^2(P)}(f) \leq 1. \tag{14} \]
Note that $R^2_{B^2(P)}(f)$ captures the amount of information contained in the interaction part with respect to the overall information within the density. If $R^2_{B^2(P)}(f)$ is small ($R^2_{B^2(P)}(f) \sim 0$), it means that most of the density is described by the independent part, and vice versa. Likewise, $R^2_{B^2(P)}(f) = 1$ implies that the geometric marginals coincide with densities of the reference measures $P_X$ and $P_Y$, which follows from the Pythagorean decomposition $\|f\|^2_{B^2(P)} = \|f_{X,g}\|^2_{B^2(P)} + \|f_{Y,g}\|^2_{B^2(P)} + \|f_{int}\|^2_{B^2(P)}$ – in other words, that the whole information in the bivariate density is contained in its interactive part. A further advantage of the use of $R^2_{B^2(P)}(f)$ is its relative character: $R^2_{B^2(P)}(f)$ does not rely on the norm of the bivariate density which might be in practice influenced by the sample size of data being aggregated in the density.

Furthermore, under normalized reference $P$, it can be proven that $f_{int}$ is marginal invariant, i.e., when the bivariate density $f$ is perturbed marginally (i.e., by marginal densities $g_X$ and $g_Y$), the interaction part $f_{int}$ is not changed. This important property Yule (1912) is formulated in the next theorem.

**Theorem 5** Let $P = P_X \times P_Y$ be a probability measure, $f \in B^2(P)$ a bivariate density with the orthogonal decomposition $f = f_{ind} \oplus f_{int}$ and $g_X, g_Y$ marginal densities, in the sense that these latter are bivariate densities in $B^2(P)$, constant in one argument, i.e.,

$$g_{X,g}(x, y) = \tilde{g}_{X,g}(x), \quad g_{Y,g}(x, y) = \tilde{g}_{Y,g}(y), \quad (x, y) \in \Omega.$$

Then, the marginally perturbed density, $h = g_{X,g} \oplus g_{Y,g} \oplus f$, has the orthogonal decomposition $h = h_{ind} \oplus h_{int}$, where $h_{int} = f_{int}$ and $h_{ind} = g_{X,g} \oplus g_{Y,g} \oplus f_{ind}$.

In conclusion, we remark that, in case of a normalized reference measure $P$, the complete theory holds true, including orthogonality in decomposition (10), and all the stated theorems. Whenever the reference measure is not normalized, one can still define geometric marginals, independent and interaction parts, as well as derive the orthogonal decomposition (10), but the properties available under independence are weaker and the interpretability of independence and interaction densities is more limited. Indeed, for a non-normalized reference, the equivalence under independence between arithmetic and geometric marginals is no longer true, given that Theorem 2 and its consequences (Theorems 4 and 5) do not hold in that case. In the following section, an illustration of the difference of using normalized or not normalized reference measures is given.

**4 An example with a truncated Gaussian Density**

For the sake of illustration, we present an example of application of the proposed framework to densities in the Gaussian family, where computations can be made explicitly. Given that, in general, one may not expect to be able to perform this type of computations explicitly, in Section 5 we develop a B-spline basis representation for bivariate distributions, from which the interactive and independent parts can be...
directly computed. We first consider a univariate Gaussian density, similarly as in Hron et al. (2016), Delicado (2011). For the sake of simplicity, we set the reference measure to the Lebesgue measure and consider a zero-mean Gaussian density, truncated over the interval \( I = [-T, T] \), \( T = 5 \). In this case, the density \( f \) reads

\[
f(x) = \mathcal{N}(\mu, \Sigma) \exp\left\{ -\frac{x^2}{2\sigma^2}\right\}, \quad x \in I.
\]

The (univariate) clr-transformation of \( f \) is defined as

\[
\text{clr}(f)(x) = -\frac{x^2}{2\sigma^2} + \frac{T^2}{6\sigma^2}, \quad x \in I.
\]

Increasing the dimensionality of the sample space, we consider a zero-mean bivariate Gaussian density \( \mathcal{N}_2(\mu, \Sigma) \) with respect to the (product) Lebesgue measure \( \lambda[I] = \lambda[I_1] \times \lambda[I_2] \), truncated on a rectangular domain \( I = I_1 \times I_2 \subset \mathbb{R}^2 \), with \( I_1 = I_2 = [-T, T], \ T = 5 \). In this case, the density is defined, for \( x = (x, y) \in I \), as

\[
f(x, y) = \mathcal{B}_2(\rho) \exp\left\{ x^\top \Sigma^{-1} x\right\} = \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_1^2} - 2\rho \frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right\},
\]

with \( \sigma_i^2 = \Sigma_{ii} \) and \( \rho \in [0, 1] \) being the correlation coefficient. In this setting, the clr transformation of \( f \) is

\[
\text{clr}(f)(x, y) = -\frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_1^2} - 2\rho \frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right] + \frac{T^2}{6(1-\rho^2)} \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right).
\]

Marginalizing the clr transformation with respect to \( x \) and \( y \) yields the clr-marginals

\[
\text{clr}(f_{X,g})(x) = -\frac{1}{2(1-\rho^2)} \cdot \frac{2Tx^2}{\sigma_1^2} + \frac{T^3}{3(1-\rho^2)} \cdot \frac{1}{\sigma_1^2}, \quad x \in I_1,
\]

\[
\text{clr}(f_{Y,g})(y) = -\frac{1}{2(1-\rho^2)} \cdot \frac{2Ty^2}{\sigma_2^2} + \frac{T^3}{3(1-\rho^2)} \cdot \frac{1}{\sigma_2^2}, \quad y \in I_2.
\]

On this basis, the geometric marginals are easily obtained – following (5) and (6)—as

\[
f_{X,g}(x) = \mathcal{B}_1(\rho) \exp\left\{ -\frac{1}{2(1-\rho^2)} \cdot \frac{2T x^2}{\sigma_1^2}\right\}, \quad x \in I_1,
\]

\[
f_{Y,g}(y) = \mathcal{B}_1(\rho) \exp\left\{ -\frac{1}{2(1-\rho^2)} \cdot \frac{2T y^2}{\sigma_2^2}\right\}, \quad y \in I_2.
\]

Note that both marginals still belong to a Gaussian family, with parameters \( \mu_X = \mu_Y = 0 \) and \( \sigma_X^2 = \sigma_1^2(1-\rho^2)/2T \), \( \sigma_Y^2 = \sigma_2^2(1-\rho^2)/2T \).
Given the marginals, the independence and interactive parts are built as in (8) and (9), leading to

\[
f_{\text{ind}}(x, y) = B(x, y) \exp \left\{ -\frac{2T}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} \right] \right\};
\]

\[
f_{\text{int}}(x, y) = B(x, y) \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ (1-2T)x^2 - 2\rho \frac{xy}{\sigma_1 \sigma_2} + (1-2T)y^2 \right] \right\}.
\]

The clr transformations of the latter parts are found as

\[
\text{clr}(f_{\text{ind}})(x, y) = -\frac{T}{1-\rho^2} \left( \frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} \right) + \frac{T^3}{3(1-\rho^2)} \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right);
\]

\[
\text{clr}(f_{\text{int}})(x, y) = -\frac{1}{2(1-\rho^2)} \left( (1-2T)x^2 - 2\rho \frac{xy}{\sigma_1 \sigma_2} + (1-2T)y^2 \right) +
\]

\[
+ \frac{T^2(1-2T)}{6(1-\rho^2)} \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right).
\]

Note that, in the case of independence (\(\rho = 0\)),

\[
\text{clr}(f_{\text{int}})(x, y) = \frac{3}{2} \left( \frac{3x^2 + T^2}{\sigma_1^2} + \frac{3y^2 + T^2}{\sigma_2^2} \right) = \frac{9}{2} \left( \frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} \right) + \frac{3}{2} \left( \frac{T^2}{\sigma_1^2} + \frac{T^2}{\sigma_2^2} \right),
\]

which is non-zero. This does not stand in contradiction with Theorem 2, since the previous computations are indeed referred to the Lebesgue measure, which is not a probability measure (it is not normalized). Analogous computations are made in the case of a uniform measure (i.e., the product measure \(U[I_1]\), \(U[I_2]\)) lead to a null clr\((f_{\text{int}})\) for \(\rho = 0\). Indeed, in this case one has that the clr geometric marginals are defined as

\[
\text{clr}(f_{X, g})(x) = -\frac{1}{2(1-\rho^2)} \cdot \frac{x^2}{\sigma_1^2} + \frac{T^2}{6(1-\rho^2)} \cdot \frac{1}{\sigma_1^2}, \quad x \in I_1,
\]

\[
\text{clr}(f_{Y, g})(y) = -\frac{1}{2(1-\rho^2)} \cdot \frac{y^2}{\sigma_2^2} + \frac{T^2}{6(1-\rho^2)} \cdot \frac{1}{\sigma_2^2}, \quad y \in I_2,
\]

leading to the following forms for the independent and interaction clr-densities

\[
\text{clr}(f_{\text{ind}})(x, y) = -\frac{1}{2(1-\rho^2)} \left( \frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} \right) + \frac{T^2}{6(1-\rho^2)} \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right);
\]

\[
\text{clr}(f_{\text{int}})(x, y) = \frac{1}{(1-\rho^2)} \left( \rho \frac{xy}{\sigma_1 \sigma_2} \right).
\]
It is then clear that the interaction part precisely captures the terms in $\text{clr}(f)$ depending on the mixed polynomial $xy$ (i.e., the interaction between $x$ and $y$), and its magnitude is controlled by the magnitude of $\rho$. In case of independence ($\rho = 0$), $\text{clr}(f_{\text{int}})$ is null, and $f_{\text{int}} = 0_{\mathcal{I}}$. Moreover, in this case, the geometric marginals and the arithmetic marginals coincide. Note that the former are found by normalizing the exponential of the first terms of $\text{clr}(f_{X,g})$ and $\text{clr}(f_{Y,g})$ in (15)-(16). In the degenerate case of a perfect linear dependence between the marginal variables $X$ and $Y$, $|\rho| = 1$, $\text{clr}(f_{\text{int}})$ is indeed degenerate as well. In fact, for $|\rho| = 1$ not only $\text{clr}(f_{\text{int}})$ is not defined, but $f$ does not belong to $B^2(U[1])$, nor to $B^2(\lambda[1])$ (the logarithms of the corresponding densities are not in $L^2(U[1])$ nor in $L^2(\lambda[1])$).

Figure 1 reports the contour plots associated with the bivariate Gaussian density with $\sigma_1 = 2$, $\sigma_2 = 3$ and $\rho = 0.75$, when the reference measure is the Lebesgue measure. Figure 2 reports the analogue contour plots when the quantities are computed w.r.t. a Uniform measure. For the sake of clarity, the quantities referred to the Uniform reference are reported with a subscript $w$ in Figure 2. The figures clearly show that the scale of the reference measure plays indeed a role, particularly for the shape of $f_{\text{int}}$ (Figures 1c-f and 2c-f). This is in agreement with the conclusions of (Talská et al. 2020), where the effect of the reference measure on the geometry of (univariate) Bayes spaces is discussed. Given the statistical consequences of Theorems 2 and 4, the representation based on a normalized reference shall be here preferred. In the latter case (Figure 2), the independent part represents the (unique) distribution which would be built upon the geometric marginals $f_{X,g}$ being $B^2$-equivalent to a truncated $N(0, 4(1-0.75^2))$, and $f_{Y,g}$ the $B^2$-equivalent to a truncated $N(0, 9(1-0.75^2))$. The simplicial deviance $\Delta^2(f) = \|f_{\text{int}}\|^2_{B^2(P)}$ is in this case $\Delta^2(f) = 5.66$. The value of the relative simplicial deviance $R^2_{B^2(P)}(f)$ represents the proportion of the norm of $f$ which can be attributed to the interaction part (i.e., to the deviation from independence). In this example, such proportion is 51%, indicating that the dependence between the two variables is indeed relevant in the definition of the bivariate distribution.

5 A spline representation for bivariate densities and their decompositions

Computational methods of FDA for the statistical analysis of datasets of bivariate densities are often based on basis representations of the data. In this section, we develop a spline representation for densities, which is based on a $B$-spline approximation for clr transformed data. This will allow for the smoothing of bivariate splines, and direct computations of geometric marginals, independence, and interaction parts, as well as of the relative simplicial deviance. On one hand, this avoids the necessity of developing splines directly in $B^2(P)$; on the other one, it implies that the zero integral constraint needs to be taken into account.

This goal is here achieved by using the tensor product splines (de Boor 1978; Dierckx 1993; Schumaker 2007) which are an established tool in the field and, in principle, enable a generalization to $k$ dimensions. However, for the purpose of this paper and for ease of notation, we shall focus on bivariate splines only. We also avoid considering...
Fig. 1 A simulated example with a truncated Gaussian density, with respect to the Lebesgue product measure \( \lambda[I] = \lambda[I_1] \times \lambda[I_2] \), with \( I_1 = I_2 = [-5, 5] \), \( \sigma_1 = 2, \sigma_2 = 3, \rho = 0.75 \)

Fig. 2 A simulated example with a Gaussian density, with respect to the Uniform product measure \( P = U[I_1] \times U[I_2] \), with \( I_1 = I_2 = [-5, 5] \), \( \sigma_1 = 2, \sigma_2 = 3, \rho = 0.75 \)
the general reference measure $P$ and focus on the Lebesgue measure (or its normalized counterpart, the uniform measure), although the case of a generic $P$-reference can be reformulated as well Talská et al. (2020). We shall base our developments on (Machalová 2002a, b)—the same representation being used for approximation of PDFs, e.g., in (Machalová et al. 2016; Hrone et al. 2016; Talská et al. 2018; Menafoglio et al. 2018); this setting is recalled in the Appendix A. Hereafter, in this section, we limit to present the key points and results of our construction, leaving the details and proofs to the Appendix B.

We consider two strictly increasing sequences of knots

$$\Delta \lambda := \{\lambda_i\}_{i=0}^{g+1}, \quad \lambda_0 = a < \lambda_1 < \ldots < \lambda_g < b = \lambda_{g+1},$$

$$\Delta \mu := \{\mu_j\}_{j=0}^{h+1}, \quad \mu_0 = c < \mu_1 < \ldots < \mu_h < d = \mu_{h+1}$$

and denote by $S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega)$ the vector space of tensor product splines on $\Omega = [a, b] \times [c, d]$ of degree $k > 0$ in $x$ and $l > 0$ in $y$, with knots $\Delta \lambda$ in the $x$-direction and $\Delta \mu$ in the $y$-direction. As usual in spline theory, to get a unique representation, additional knots are considered, namely

$$\lambda_{-k} = \ldots = \lambda_{-1} = \lambda_0 = a, \quad b = \lambda_{g+1} = \lambda_{g+2} = \ldots = \lambda_{g+k+1},$$

$$\mu_{-l} = \ldots = \mu_{-1} = \mu_0 = c, \quad d = \mu_{h+1} = \mu_{h+2} = \ldots = \mu_{h+l+1}.$$  

The general goal here explored is that of smoothing the values $f_{ij}$ at points $(x_i, y_j) \in \Omega = [a, b] \times [c, d], \ i = 1, \ldots, n, \ j = 1, \ldots, m$ using a tensor-product spline. The values $f_{ij}$ will be the clr-transformation of a discrete representation of the bivariate densities (i.e., histogram data), as showcased in Section 6. For the strictly increasing sequences of knots (17) and (18), a parameter $\alpha \in (0, 1)$ and arbitrary $u \in \{0, 1, \ldots, k - 1\}$ and $v \in \{0, 1, \ldots, l - 1\}$, we aim to find a spline $s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega)$ which minimizes the functional

$$J_{uv}(s_{kl}) = \alpha \sum_{i=1}^{n} \sum_{j=1}^{m} [f_{ij} - s_{kl}(x_i, y_j)]^2 + (1 - \alpha) \iint_{\Omega} \left[ s_{kl}^{(u,v)}(x, y) \right]^2 \ dx \ dy,$$

where the upper index $(u, v)$ stands for the derivative, specifically

$$s_{kl}^{(u,v)}(x, y) = \frac{\partial^{u}}{\partial x^{u}} \frac{\partial^{v}}{\partial y^{v}} s_{kl}(x, y).$$

Clearly, the choice of the parameter $\alpha$ and of the derivative orders $(u, v)$ affects the smoothness of the resulting spline. For the optimal choice of $\alpha$, the generalized cross-validation (GCV) criterion is used here, similarly as in (Machalová et al. 2020).

Given that we aim to reconstruct clr-transformed PDFs, the zero-integral constraint needs to be incorporated into the tensor product splines. Accordingly, we here aim to find a spline $s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega), \ \Omega = [a, b] \times [c, d]$, which minimizes the functional (21) and satisfies the additional condition
\[
\int_{\Omega} \int s_{kl}(x, y) \, dx \, dy = 0. \tag{22}
\]

We thus generalize to the tensor product splines the idea presented in (Machalová et al. 2016) for one-dimensional splines. To state the solution of the problem and the conditions for its well-posedness, we need to introduce an additional notation that follows.

We express the tensor spline \( s_{kl}(x, y) \) appearing in (21) as

\[
s_{kl}(x, y) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B^{k+1}_{i}(x) B^{l+1}_{j}(y), \tag{23}
\]

where \( B^{k+1}_{i}(x) \), \( B^{l+1}_{j}(y) \) are (univariate) \( B \)-splines defined on the sequence of knots \( \{\lambda_{i}\} \) or \( \{\mu_{j}\} \) and \( b_{ij} \) are the coefficients of this spline. The tensor spline in (23) can be expressed in matrix notation as

\[
s_{kl}(x, y) = \mathbf{B} B^{k+1}_{x}(x) \mathbf{B}^{T} B^{l+1}_{y}(y),
\]

where \( \mathbf{B} \) is a matrix of \( B \)-spline coefficients \( \mathbf{B} = (b_{ij})_{i=-k, j=-l}^{g, h} \). \( B^{k+1}_{x}(x) \) is the collocation matrix of the \( B \)-splines \( B^{k+1}_{i}(x) \), and \( B^{l+1}_{y}(y) \) is the collocation matrix of the \( B \)-splines \( B^{l+1}_{j}(y) \). This admits also a tensor product representation, as

\[
s_{kl}(x, y) = \mathbf{B}^{k+1}_{x}(x) \mathbf{B}^{T} B^{l+1}_{y}(y) = (\mathbf{B}^{k+1}_{x}(y) \otimes \mathbf{B}^{T} B^{l+1}_{y}(x)) \mathbf{c s}(\mathbf{B}) = \mathbb{B}(x, y) \mathbf{c s}(\mathbf{B}), \tag{24}
\]

where \( \mathbb{B}(x, y) := \mathbf{B}^{k+1}_{x}(y) \otimes \mathbf{B}^{T} B^{l+1}_{y}(x) \) and \( \mathbf{c s}(\mathbf{B}) \) is the vectorized form of the matrix \( \mathbf{B} \) (columnwise).

Let

\[
\mathbf{S}_{u} = \mathbf{D}_{u}^{x} L_{u}^{x} \cdots \mathbf{D}_{1}^{x} L_{1}^{x} \in \mathbb{R}^{g+k+1-u, g+k+1}, \quad \text{with} \quad \mathbf{D}_{j}^{x} = (k + 1 - j) \ \text{diag}(d_{-k+j}^{x}, \ldots, d_{h}^{x}),
\]

\[
d_{i}^{x} = \frac{1}{\lambda_{i-k+1} - \lambda_{i}}, \quad i = -k + j, \ldots, g,
\]

\[
\mathbf{L}_{j}^{x} = \begin{pmatrix}
-1 & 1 & \cdots & \\
& & \ddots & \\
& & & -1 & 1
\end{pmatrix} \in \mathbb{R}^{g+k+1-j, g+k+2-j}.
\]

Similarly, let

\[
\mathbf{S}_{v} = \mathbf{D}_{v}^{y} L_{v}^{y} \cdots \mathbf{D}_{1}^{y} L_{1}^{y} \in \mathbb{R}^{h+l+1-v, h+l+1}, \quad \text{with} \quad \mathbf{D}_{j}^{y} = (l + 1 - j) \ \text{diag}(d_{-l+j}^{y}, \ldots, d_{h}^{y}),
\]

\[
d_{i}^{y} = \frac{1}{\mu_{i-l+1} - \mu_{i}}, \quad i = -l + j, \ldots, h,
\]

\[
\mathbf{L}_{j}^{y} = \begin{pmatrix}
-1 & 1 & \cdots & \\
& & \ddots & \\
& & & -1 & 1
\end{pmatrix} \in \mathbb{R}^{h+l+1-j, h+l+2-j}.
\]
Denote by $\mathbb{S}$ the tensor product between $S_v$ and $S_u$, $\mathbb{S} := S_v \otimes S_u$ and define the matrices of inner products between $u$-th and $v$-th derivatives of the spline basis elements, $M_{k,u}^x = \left( m_{ij}^x \right)_{i,j=-k+u}$, $M_{l,v}^y = \left( m_{ij}^y \right)_{i,j=-l+v}$, with

$$m_{ij}^x := \int_a^b B_i^{k+1-u}(x) B_j^{k+1-u}(x) \, dx, \quad m_{ij}^y := \int_a^b B_i^{l+1-v}(y) B_j^{l+1-v}(y) \, dy.$$ 

Finally, let $\mathbb{M} := M_{l,v}^y \otimes M_{k,u}^x$. Regarding the condition (22), by using the well-known properties of the splines (see, e.g., (de Boor 1978; Schumaker 2007)), it is possible to write

$$\int_a^b \int_c^d s_{kl}(x,y) \, dy \, dx = \int_a^b \left[ s_{k,l+1}(x,y) \right]_c^d \, dx - \int_a^b s_{k,l+1}(x,c) \, dx = \left[ s_{k+1,l+1}(x,d) \right]_a^b - \left[ s_{k+1,l+1}(x,c) \right]_a^b = s_{k+1,l+1}(b,d) - s_{k+1,l+1}(a,d) - s_{k+1,l+1}(a,c) + s_{k+1,l+1}(a,c) = c_{g,h} - c_{-k-1,h} - c_{g,-l-1} + c_{-k-1,-l-1}$$

using the notation $s_{k+1,l+1}(x,y) = \sum_{i=-k}^g \sum_{j=-l}^h c_{ij} B_i^{k+2}(x) B_j^{l+2}(y)$ and the coincident additional knots (19), (20). Accordingly, the condition (22) is fulfilled if and only if

$$c_{-k-1,-l-1} = c_{-k-1,h} + c_{g,-l-1} - c_{g,h}. \quad (25)$$

There is a useful relation between the $B$-spline coefficients of $s_{kl}(x,y)$ and $s_{k+1,l+1}(x,y)$, which can be expressed in matrix notation as $B = D_x K_x C K_y^\top D_y^\top$, where $C = (c_{ij}) \in \mathbb{R}^{g+k+2,h+l+2}$,

$$D_x = (k+1) \text{ diag} \left\{ \frac{1}{\lambda_1 - \lambda_{-k}}, \ldots, \frac{1}{\lambda_g + k + 1 - \lambda_g} \right\},$$

$$D_y = (l+1) \text{ diag} \left\{ \frac{1}{\mu_1 - \mu_{-l}}, \ldots, \frac{1}{\mu_h + l + 1 - \mu_h} \right\},$$

$$K_x = \begin{pmatrix} -1 & 1 \\ & \ddots & \ddots \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{g+k+1,g+k+2},$$

$$K_y = \begin{pmatrix} -1 & 1 \\ & \ddots & \ddots \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{h+l+1,h+l+2}.$$
By using notation $D = D_y \otimes D_x$, $K = K_y \otimes K_x$, and relation (25) to elide the coefficient $c_{-k-1, -l-1}$ we have

$$cs(B) = D \tilde{K} cs(\tilde{C}),$$

where $cs(\tilde{C}) = (c_{-k, -l-1}, \ldots, c_{g, -l-1}, \ldots, c_{-k-1, h}, \ldots, c_{g, h})^T$. Having set this notation, we can now state explicitly the minimizer of (21), under the zero-integral constraint.

**Theorem 6** The tensor smoothing spline $s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega)$, which minimizes the functional (21) under the condition (22) is obtained as

$$s_{kl}(x, y) = B(x, y) D \tilde{K} cs(\tilde{C}^*),$$

where

$$cs(\tilde{C}^*) = \left[\tilde{K}^T D^T \left[(1 - \alpha) S^T M S + \alpha B^T B\right] D \tilde{K}\right] \alpha \tilde{K}^T D^T B^T cs(F),$$

$$F = (f_{ij})$$

and $cs(F)$ denotes its vectorized form.

As a by-product of the derivations leading to Theorem 6, one indeed obtains the following result, which states the necessary and sufficient condition for bivariate splines to have zero integral (the proof is provided in the Appendix B).

**Theorem 7** For every spline $s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega)$, with the representation

$$s_{kl}(x, y) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B_{i}^{k+1}(x) B_{j}^{l+1}(y),$$

the condition $\iint_{\Omega} s_{kl}(x, y) \, dx \, dy = 0$ is fulfilled if and only if

$$\sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} (\lambda_{i+k+1} - \lambda_{i}) (\mu_{j+l+1} - \mu_{j}) = 0.$$

The next important result is that, if using the proposed spline representation for the bivariate densities, the spline representations of the corresponding geometrical marginals in the clr space can be explicitly computed, and carry automatically the zero integral constraint, as stated in the next theorem.

**Theorem 8** Let $s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega)$ such that $\iint_{\Omega} s_{kl}(x, y) \, dx \, dy = 0$ be given.

Let $s_k(x) \in S_k^{\Delta \lambda}[a, b]$, $s_l(y) \in S_l^{\Delta \mu}[c, d]$ be defined as $s_k(x) = \int_{c}^{d} s_{kl}(x, y) dy$, and $s_l(y) = \int_{a}^{b} s_{kl}(x, y) dx$. Then

$$s_k(x) = \sum_{i=-k}^{g} v_i B_i^{k+1}(x), \quad s_l(y) = \sum_{j=-l}^{h} u_j B_j^{l+1}(y)$$
with
\[ v_i = \frac{b_{ih}}{t_l} + \cdots + \frac{b_{i,-l}}{t_{-l}}; \quad t_{j'} = \frac{l + 1}{\mu_{j'+l} - \mu_j}, \quad j' = -l, \ldots, h, \]
\[ u_j = \frac{b_{jg}}{d_k} + \cdots + \frac{b_{j,-k}}{d_{-k}}; \quad d_{j'} = \frac{k + 1}{\lambda_{j'+k} - \lambda_j}, \quad j' = -k, \ldots, g. \]

Moreover, the splines \( s_k(x), s_l(y) \) fulfil the zero-integral constraint, i.e.,
\[ \int_a^b s_k(x)dx = 0 \quad \text{and} \quad \int_c^d s_l(y)dy = 0. \]

The proof of Theorem 8 is reported in the Appendix B. We finally introduce a spline representation for the independent and interactive parts of the bivariate densities.

Theorem 9 Let \( s_{kl}(x, y) \in S^{\Lambda_k, \Lambda_l}(\Omega), s_{kl}(x, y) = \sum_{i=-k}^g \sum_{j=-l}^h b_{ij} B_i^{k+1}(x)B_j^{l+1}(y), \) such that \( \int_\Omega s_{kl}(x, y) dx \, dy = 0 \) be given. Let \( s_k(x) \in S^{\Lambda_k}[a, b], s_l(y) \in S^{\Lambda_l}[c, d] \) be defined as
\[ s_k(x) := \int_c^d s_{kl}(x, y)dy, \quad s_l(y) := \int_a^b s_{kl}(x, y)dx \]
with representation in the form
\[ s_k(x) = \sum_{i=-k}^g v_i B_i^{k+1}(x), \quad s_l(y) = \sum_{j=-l}^h u_j B_j^{l+1}(y). \]

Then the independent part of the bivariate density \( s_{kl}(x, y) \) admits the spline representation
\[ s_{kl}^{ind}(x, y) = s_k(x) + s_l(y) = \sum_{i=-k}^g \sum_{j=-l}^h (v_{ij} + u_{ij}) B_i^{k+1}(x)B_j^{l+1}(y), \]
and the interactive part is expressed as the spline
\[ s_{kl}^{int}(x, y) = s_{kl}(x, y) - s_{kl}^{ind}(x, y) = \sum_{i=-k}^g \sum_{j=-l}^h (b_{ij} - v_{ij} - u_{ij}) B_i^{k+1}(x)B_j^{l+1}(y), \]
where \( v_{ij} := \frac{1}{d-c} v_i, \quad u_{ij} := \frac{1}{b-a} u_j \forall i, j. \)
Table 1  Sample sizes for age groups in anthropometric data

| Age interval   | Sample size | Age interval   | Sample size | Age interval   | Sample size | Age interval   | Sample size | Age interval   | Sample size |
|----------------|-------------|----------------|-------------|----------------|-------------|----------------|-------------|----------------|-------------|
| [15, 16)        | 95          | [16, 17)       | 126         | [17, 18)       | 234         | [18, 19)       | 492         | [19, 20)       | 686         |
| [21, 22)        | 443         | [22, 23)       | 450         | [23, 24)       | 385         | [24, 25)       | 318         | [25, 26)       | 220         |
| [27, 28)        | 108         | [28, 29)       | 99          | [29, 30)       | 79          | [30, 31)       | 90          |                |             |

The proof of Theorem 9 is again reported in the Appendix B.

We remark that the results presented in this section form a computational cornerstone for the theoretical framework proposed in this work. Indeed, they not only allow for a complete characterization of bivariate densities although splines, but also for an explicit spline representation of the geometric marginals, as well as of the independence and interaction densities. Lastly, the spline representation enables one to compute the deviance and relative deviance from the interaction density $s_{kl}^{int}(x, y)$ simply as

$$\Delta^2(s_{kl}^{int}(x, y)) = \int_a^b \int_c^d [s_{kl}^{int}(x, y)]^2 \, dx \, dy.$$ 

This result follows from an analogous development as that leading to the proof of Theorem 6 (namely, the derivation of $J_1$ by setting $u = 0, v = 0$, see the Appendix B).

The next section showcases the application of the proposed methodology to a real dataset dealing with anthropometric measurements.

6 An application to anthropometric densities

Periodic collection and reporting of anthropometric data such as body height and weight is essential to measure time trends in the prevalence of overweight and obesity at the population level. To this aim, a representative dataset of 4,436 Czech adolescents and young adults aged 15–31 years was collected as part of a large cross-sectional study (the reader may refer to (Gába and Přidalová 2014, 2016) for further details on the study). Participants to the study were selected on a volunteer basis among university students, staff, and attendants to university open-house days and education exhibitions. The sample sizes were, however, not distributed uniformly throughout the age intervals, mainly due to broader participation by university students, see Table 1.

Body height was measured with a precision of 0.1 cm by an anthropometer P-375 (Trystom, Olomouc, Czech Republic) and body weight was measured using the InBody 720 device (Biospace Co., Ltd.; Seoul, Korea). Histogram data were then obtained from the raw data separately in each of $N = 16$ age groups, i.e. [15, 16), [16, 17), ..., [30, 31). Note that the same age range was used also in (Machalová et al. 2020).
The support of the marginal distribution of weights \(X\) and heights \(Y\) was set to the respective interval of observation, namely to \(I_1 = [40, 100]\) for \(X\) and \(I_2 = [155, 195]\) for \(Y\); the Sturge’s rule was used to select the number of classes \(K, L\) along the variable \(X,Y\), respectively, in each of the histograms. Possible (count) zeros in the histograms were imputed as advocated in (Martín-Fernández et al. 2015); more precisely, a zero value in the class \(k, l\) of a histogram was set to \(2/(3n_{kl})\), where \(n_{kl}\) stands for the number of observations within the class. For each age group, this led to the discrete representation \(\{f_{kl}, k = 1, \ldots, K, l = 1, \ldots, L\}\) of the bivariate distributions, which were referred to the midpoints of the classes \(\{t_{kl} = (x_k, y_l), k = 1, \ldots, K, l = 1, \ldots, L\}\). The associated discrete bivariate clr transformations were computed as

\[
\text{clr}(f)_{kl} = \ln f_{kl} - \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \ln f_{kl}.
\]

These clr transformations were smoothed by using the tensor product smoothing splines with zero integral introduced in Section 5, considering a rectangular domain \(\Omega = [40; 100] \times [155; 195]\). For each age class, the following strategy was considered to set the parameters for the smoothing procedure. Quadratic smoothing splines were employed in each direction, setting the knots to equispaced sequences in both directions, with a spacing of 15 kg along \(X\) and 10 cm along \(Y\). The order of derivative in the penalty term was set to \(u = 1, v = 1\). The smoothing parameter \(\alpha\) was determined by means of GCV errors over all sampled bivariate densities, resulting in \(\alpha = 0.0496\), see Figure 3.

The matrix \(B^{*} = (b_{ij}^{t}), t = 1, \ldots, 16\), of coefficients for the smoothing spline \(s_{kl}^{t}(x, y)\) with zero integral were finally obtained by (27). A subset of the resulting clr densities, expressed with respect to the uniform measure on \(\Omega\), are displayed in Figure 4a. The corresponding densities (obtained by exponentiating the clr-transformations) are reported in Figure 4b. The complete set of smoothed data and clr-transform is available in the Appendix D. One can clearly see the bimodal character of the densities, which is probably due to the presence of both males and females in the sample. In addition, some dependence between heights and weights is apparent, without a substantial difference between the two modes for most densities.
From the smoothed data, the decomposition of the bivariate densities into their independent and interactive parts was computed using the results detailed in Section 5, based on the corresponding B-spline coefficients (see Theorem 9). The independent and interactive parts of the densities in Figure 4b are reported in Figures 4c and 4d. It is interesting to observe that the bimodal character of the densities almost disappears in the independent densities, as this feature is mostly captured by the interaction densities. Apart from that, it is obvious that the upper/lower/combined extreme values of the variables heights and weights have a relevant contribution to the dependence between the variables in the data set. This can be seen from the uplifted values appearing in corners of the majority of the interaction densities (Figure 4d).

We now aim to investigate whether the dependence between height and weight changes with ageing of the population. For this purpose, the simplicial deviances $\Delta^2(f_i)$ and the relative simplicial deviances $R^2_{E^2(P_i)}(f_i), i = 1, \ldots, 16$, were computed as described in Section 5, see Table 2. Inspection of Table 2 suggests that simplicial deviances are clearly influenced by the sample sizes in the age intervals, yielding higher values of $\Delta^2(f_i)$ between ages 18 and 24, due more local effects resulting from the smoothing of histograms with more classes. These effects are filtered out in the relative simplicial deviances, whose time series is reported in Figure 5 (upper figure).

A stationarity check performed with the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test Kwiatkowski et al. (1992) yields a narrow rejection of the stationarity assumption ($p$-value $p = 0.0482$). As a consequence, (slightly) non-stationary effects emerge in the time series of relative simplicial deviances. A major effect can be observed at the beginning of the time series; here the relative simplicial deviance
Table 2 Values of norms of the bivariate densities and their decomposition together with the derived coefficients for all age groups in anthropometric data

| $i$ | Age group | $\|f_i\|_{B^2(P)}$ | $\|f^i_{\text{ind}}\|_{B^2(P)}$ | $\|f^i_{\text{int}}\|_{B^2(P)}$ | $\Delta^2(f_i)$ | $R^2_{B^2(P)}(f_i)$ |
|-----|-----------|----------------|----------------|----------------|----------------|----------------|
| 1   | [15,16)  | 44.786         | 34.853         | 28.125         | 791.007        | 0.394         |
| 2   | [16,17)  | 40.437         | 31.518         | 25.333         | 641.772        | 0.392         |
| 3   | [17,18)  | 49.403         | 32.794         | 36.948         | 1365.180       | 0.559         |
| 4   | [18,19)  | 55.132         | 34.132         | 43.296         | 1874.570       | 0.617         |
| 5   | [19,20)  | 63.097         | 40.634         | 48.271         | 2330.051       | 0.585         |
| 6   | [20,21)  | 58.650         | 36.198         | 46.147         | 2129.520       | 0.619         |
| 7   | [21,22)  | 54.236         | 32.744         | 43.236         | 1869.361       | 0.636         |
| 8   | [22,23)  | 54.305         | 34.407         | 42.014         | 1765.162       | 0.599         |
| 9   | [23,24)  | 52.934         | 31.971         | 42.189         | 1779.876       | 0.635         |
| 10  | [24,25)  | 52.647         | 29.572         | 43.557         | 1897.188       | 0.684         |
| 11  | [25,26)  | 45.313         | 26.604         | 36.680         | 1345.455       | 0.655         |
| 12  | [26,27)  | 40.893         | 24.782         | 32.528         | 1058.088       | 0.633         |
| 13  | [27,28)  | 34.739         | 21.963         | 26.915         | 724.405        | 0.600         |
| 14  | [28,29)  | 33.712         | 19.344         | 27.610         | 762.289        | 0.671         |
| 15  | [29,30)  | 31.862         | 20.223         | 24.621         | 606.206        | 0.597         |
| 16  | [30,31)  | 30.849         | 17.031         | 25.722         | 661.618        | 0.695         |

Fig. 5 Time series of the relative simplicial deviances (solid line), the Spearman (dashed line) and the Kendall (dotted line) correlation coefficients for anthropometric data

slightly increases and gets stabilized around the nineteenth year. This development can be easily explained by pubertal and postpubertal changes in height and weight, which still occur in the mentioned time period.

It is interesting to compare the relative simplicial deviance with the well-known Spearman and Kendall correlation coefficients, which both are closely connected to the copula theory (Nelsen 2006). Their time series are displayed in Figure 5 (middle and lower figure) and they look pretty similar. The effect of (post)pubertal changes is no more visible here; we can rather observe a slightly decreasing trend from around 24 years, which, interestingly, corresponds to border age of the “Youth” age group.
according to World Health Organization WHO (2020). This would indicate that since then the strength of the monotonic dependence between height and weight slightly weakens. This result reflects the fact that unlike both the mentioned correlation coefficients, the relative simplicial deviance captures the whole “mass” of the interactions between height and weight distributions, thus also including effects like possible tail dependence. Note that a similar idea of using a norm for development of a dependence measure is followed in (Tran et al. 2015) with the Sobolev metric, thought there with the aim to capture rather solely monotone dependence.

We now further investigate the non-stationarity effects being observed for the time series of relative simplicial deviances. For this purpose, we formulate a compositional regression model with functional response (Talská et al. 2018) and scalar regressors (i.e., the time \( t \) of observations), which results in the linear model

\[
  f_{\text{int},i} = \beta_0 \oplus \beta_1 \odot t_i \oplus \varepsilon_i,
\]

for \( i = 1, \ldots, 16 \), with \( \beta_0, \beta_1 \) unknown coefficients in \( B^2(\mathcal{P}) \) and \( \varepsilon_i \) a zero-mean random error. Note that, by linearity, the properties of \( f_{\text{int},i} \) (as stated in Theorems 1, 3 and 4) are inherited by \( \beta_0, \beta_1 \), as \( \mathbb{E}[f_{\text{int}}] = \beta_0 \oplus \beta_1 \odot t \). Applying the clr transformation (3) to both sides of the model (29) yields

\[
  \text{clr}(f_{\text{int},i})(x, y) = \text{clr}(\beta_0)(x, y) + \text{clr}(\beta_1)(x, y)t_i + \text{clr}(\varepsilon_i)(x, y),
\]

\( i = 1, \ldots, 16, (x, y) \in I_1 \times I_2 \). For the estimation of the functional regression parameters using the least squares criterion, the smoothing tensor spline coefficients (Theorem 9) can be utilized. Similarly as in (Talská et al. 2018), the spline coefficients of the regression estimates (clr transformed densities) \( \text{clr}(\hat{\beta}_0)(x, y) \) and \( \text{clr}(\hat{\beta}_1)(x, y) \) fulfill the condition from Theorem 7. The estimated parameters \( \hat{\beta}_0, \hat{\beta}_1 \) are reported in Figure 6. A permutation test on the global significance of the parameter \( \beta_1 \)—run using a Freedman and Lane scheme (Freedman and Lane 1983; Pini et al. 2018) with test statistics \( T^2 = \| \hat{\beta}_1 \|^2_{B^2(\mathcal{P})} \) — confirms the statistical significance of this parameter (p-value 0.032), suggesting that the time variation in the interaction between the random variables is indeed relevant. In the light of the shape of \( \beta_1 \) (Figure 6) one can conclude that, along time, the interaction between weights and heights tends to get more concentrated at medium-high values or low values of height, whereas it deflates for medium-low values of height. A less pronounced time variation is instead observed for different values of the weight, suggesting that the highest variability in the dependence between these variables is indeed observed across the values of the other variables.

7 Conclusions

The Bayes spaces methodology provides a robust and flexible framework for modelling data with relative character, including measures, probability density functions as well as compositional data. It can serve for many different purposes, from the geometrical representation of the Bayes theorem to functional data analysis of a sample of densities. In this paper, its potential was further extended to bivariate density functions. Their
decomposition into independent and interactive parts has a solid geometrical basis and allows for an appealing probabilistic interpretation if a normalized reference measure is used. This opens new perspectives for both further generalization to multivariate densities as well as to dependence modelling, with the aim to provide an alternative viewpoint than that offered by the widely-used copula theory. Note that the Bayes space theory is built for general types of positive measures (not necessarily absolutely continuous); we here foresee clear perspectives of development for a general Bayes space approach for distributions, in contexts and assumptions even closer to those of the well-established theory of copulas. Other important envisioned impacts of this work are worth to be mentioned. For instance, the spline smoothing here developed may be used for a non-parametric estimate of PDFs, directly allowing for further data processing in view of FDA, although at the expense of a possible lower convergence rate than empirical distribution functions or empirical copulas constructed using ranks. From the application viewpoint, in the light of the promising theoretical properties presented in Section 3, the Bayes space approach could be used to develop novel FDA methods for bivariate densities, and provide a broader statistical framework with pioneering applications as those developed in (Guégan and Iacopini 2019). An instance of this has been shown in Section 6, where a linear regression model for interaction densities has been formulated to further investigate the variability of the interaction between two random variables along time. Such linear modelling would not be easy in other settings, based on non-linear and non-orthogonal relations between independent and interaction parts. More in general, in this first work, we outlined a number of novel views allowed by the proposed spline representation, with particular reference to the statistical analysis of bivariate densities in Bayes spaces. In fact, we here envision a great potential of this framework, which can be used to provide a mathematical setting for the statistical processing of samples of bivariate densities in varied contexts. Note that distributional datasets are becoming increasingly available in applications, as these could result from the aggregation of massive data coming from large-scale studies or automated collection of data. Despite this, the statistical methods available for their analysis (particularly in the multivariate case) are still limited. Depending on whether such densities form a random sample, regionalized observations or time series, appropriate methods of FDA far beyond those explicitly mentioned in this work can be built, precisely grounding upon the presented theory and the associated spline representations.
Acknowledgements  The authors were supported by Czech Science Foundation (GAČR), GA22-15684L.

Appendix A: Spline representation of univariate clr transformed densities

In this appendix section, the terminology and basics for the spline representation of clr transformed univariate densities as $L^2$ functions with zero integral are recalled. Let the sequence of knots $\Delta \lambda := \{\lambda_i\}_{i=0}^{g+1}$, $\lambda_0 = a < \lambda_1 < \ldots < \lambda_g < b = \lambda_{g+1}$ be given. The symbol $S^\Delta_k[a, b]$ denotes the vector space of polynomial splines of degree $k > 0$, defined on a finite interval $[a, b]$ with the sequence of knots $\Delta \lambda$. It is known that $\dim \left( S^\Delta_k[a, b] \right) = g + k + 1$. Then every spline $s_k(x) \in S^\Delta_k[a, b]$ has an unique representation

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

For generalization of splines to the bivariate density case, the following theorem, which was published in Talská et al. (2018), is of paramount importance.

**Theorem 10** For a spline $s_k(x) \in S^\Delta_k[a, b]$, $s_k(x) = \sum_{i=-k}^{g} b_i B_{k+1}^{i}(x)$, the condition

$$\int_a^b s_k(x) \, dx = 0$$

is fulfilled if and only if

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

**Proof** From the spline theory, it is known that $\int_a^b s_k(x) \, dx = s_{k+1}(x)$. If the notation $s_k(x) = \sum_{i=-k}^{g} b_i B_{k+1}^{i}(x)$ is used, then $s_{k+1}(x) = \sum_{i=-k-1}^{g} c_i B_{k+2}^{i}(x)$, there is known the relationship between their B-spline coefficients in the form

$$b_i = (k + 1) \frac{c_i - c_{i-1}}{\lambda_{i+k+1} - \lambda_i}, \quad \forall i = -k, \ldots, g.$$

Thus the coefficients $c_i$ can be expressed as

$$c_i = c_{i-1} + \frac{b_i}{d_i}, \quad \forall i = -k, \ldots, g$$

with $d_i = \frac{k + 1}{\lambda_{i+k+1} - \lambda_i}$ and it means that

$$c_g = \frac{b_g}{d_g} + \cdots + \frac{b_{-k}}{d_{-k}} + c_{-k-1}.$$
According to the coincident additional knots, see Machalová et al. (2016) for details, it holds
\[
\int_a^b s_k(x) \, dx = \left[ s_{k+1}(x) \right]_a^b = s_{k+1}(b) - s_{k+1}(a) = c_g - c_{-k-1}, \quad (A1)
\]
and it is obvious that
\[
0 = \int_a^b s_k(x) \, dx \iff c_g = c_{-k-1} \iff \frac{b_g}{d_g} + \cdots + \frac{b_{-k}}{d_{-k}} = 0.
\]
Finally, the definition of \( d_i \) implies that the following sequence of equivalences can be formulated,
\[
0 = \int_a^b s_k(x) \, dx \iff \sum_{i=-k}^{g} \frac{b_i}{d_i} = 0 \iff \sum_{i=-k}^{g} b_i (\lambda_{i+k+1} - \lambda_i) = 0.
\]

**Algorithm**
The algorithm to find a spline \( s_k(x) \in S^\Delta_k[a, b] \) with zero integral, i.e., the respective vector \( b = (b_{-k}, \ldots, b_g)^\top \), can be summarized as follows:
1. choose \( g + k \) arbitrary \( B \)-spline coefficients \( b_i \in \mathbb{R}, i = -k, \ldots, j-1, j+1, \ldots, g \),
2. compute
\[
b_j = \frac{-1}{\lambda_{j+k+1} - \lambda_j} \sum_{i=-k}^{g} b_i (\lambda_{i+k+1} - \lambda_i).
\]

**Appendix B: Proofs**

**Proof of Theorem 1** The clr transformation of the independence density \( f_{\text{ind}}(x, y) \) can be written as
\[
\text{clr}(f_{\text{ind}})(x, y) = \ln[f_X(x) f_Y(y)] - \frac{1}{\mathcal{P}(\Omega)} \int_{\Omega_X} \int_{\Omega_Y} \ln[f_X(x) f_Y(y)] \, dP_X dP_Y. \quad (B2)
\]
This is invariant under rescaling of the product \( f_{X,g}(x) f_{Y,g}(y) \). By choosing the following representations of \( f_{X,g}(x) \) and \( f_{Y,g}(y) \),
\[
f_{X,g}(x) = \exp[\text{clr}(f_{X,g})(x)], \quad f_{Y,g}(y) = \exp[\text{clr}(f_{Y,g})(y)],
\]
the second term in (B2) equals zero. Thus (B2) can be rewritten as
\[ \text{clr}(f_{\text{ind}})(x, y) = \ln\{\exp[\text{clr}(f_{X,g})(x) + \text{clr}(f_{Y,g})(y)]\} = \text{clr}(f_{X,g})(x) + \text{clr}(f_{Y,g})(y). \]

For the sake of simplicity in notation, arguments are hereafter omitted. Consider
\[ \text{clr}(f_{\text{in}}) = \text{clr}(f) - \text{clr}(f_{\text{ind}}) = \text{clr}(f) - \text{clr}(f_{X,g}) - \text{clr}(f_{Y,g}), \]
then
\[ \langle \text{clr}(f_{\text{in}}), \text{clr}(f_{\text{ind}}) \rangle_{L^2(P)} = \langle \text{clr}(f) - \text{clr}(f_{X,g}) - \text{clr}(f_{Y,g}), \text{clr}(f_{X,g}) + \text{clr}(f_{Y,g}) \rangle_{L^2(P)} = \]
\[ = \langle \text{clr}(f), \text{clr}(f_{X,g}) \rangle + \langle \text{clr}(f), \text{clr}(f_{Y,g}) \rangle_{L^2(P)} - \|\text{clr}(f_{X,g})\|_{L^2(P)}^2 - 2\langle \text{clr}(f_{X,g}), \text{clr}(f_{Y,g}) \rangle_{L^2(P)}. \]

For the first scalar product one has
\[ \langle \text{clr}(f), \text{clr}(f_{X,g}) \rangle_{L^2(P)} = \int \int_{\Omega_X \Omega_Y} \text{clr}(f)(x, y)\text{clr}(f_{X,g})(x) d\mathcal{P}_X d\mathcal{P}_Y = \]
\[ = \int_{\Omega_X} \text{clr}(f_{X,g})(x) \int_{\Omega_Y} \text{clr}(f)(x, y) d\mathcal{P}_X d\mathcal{P}_Y = \]
\[ = \int_{\Omega_X} [\text{clr}(f_{X,g})(x)]^2 d\mathcal{P}_X = \|\text{clr}(f_{X,g})\|_{L^2(P)}^2, \]
similarly also \( \langle \text{clr}(f), \text{clr}(f_{Y,g}) \rangle_{L^2(P)} = \|\text{clr}(f_{Y,g})\|_{L^2(P)}^2 \). Finally,
\[ \langle \text{clr}(f_{X,g}), \text{clr}(f_{Y,g}) \rangle_{L^2(P)} = \int \int_{\Omega_X \Omega_Y} \text{clr}(f_{X,g})(x)\text{clr}(f_{Y,g})(y) d\mathcal{P}_X d\mathcal{P}_Y = \]
\[ = \int_{\Omega_X} \text{clr}(f_{X,g})(x) d\mathcal{P}_X \int_{\Omega_Y} \text{clr}(f_{Y,g})(y) d\mathcal{P}_Y = 0, \]
which completes the proof.

\[ \square \]

**Proof of Theorem 2** In case of independence, one may decompose a bivariate density as the product of its arithmetic marginals as \( f(x, y) = f_{X,a}(x) f_{Y,a}(y) \). In Bayes spaces, this is reformulated as in (11). Call \( \text{clr}(f_{X,g})(x), \text{clr}(f_{Y,g})(y) \) the \( \text{clr} \)-representation of the marginals, i.e., \( f_{X,a}(x) = \exp[\text{clr}(f_{X,a})(x)] \) and similarly \( f_{Y,a}(y) = \exp[\text{clr}(f_{Y,a})(y)] \). Using (11), one may build the independent component as \( \text{clr}(f_{\text{ind}})(x, y) = \text{clr}(f_{X,a})(x) + \text{clr}(f_{Y,a})(y) \), which clearly coincides with \( f \) itself. The \( \text{clr} \) representation of the geometric \( X \)-marginal is derived—by definition (5)—as
\[ \int_{\Omega_Y} \text{clr}(f_{\text{ind}})(x, y) d\mathcal{P}_Y = \int_{\Omega_Y} \text{clr}(f_{X,a})(x) d\mathcal{P}_Y = P_Y(\Omega_Y)\text{clr}(f_{X,a})(x). \]
By considering that $P_Y(\Omega_Y) = 1$, the geometric $X$-marginal is obtained by applying the exponential as $f_{X,g}(x) = \exp[\text{clr}(f_{X,a})(x)]$, i.e., it coincides with the arithmetic marginal $f_{X,a}(x)$. The case of $Y$-marginals would be proven analogously.

**Proof of Theorem 3** The orthogonality of the marginals is easy to be proven in the clr space. Specifically,

$$\langle \text{clr}(f_{X,g}), \text{clr}(f_{Y,g}) \rangle_{L_0^2(P)} = \left( \int_{\Omega_X} \text{clr}(f)(x, y) \, dP_X, \int_{\Omega_Y} \text{clr}(f)(x, y) \, dP_Y \right)_{L_0^2(P)} =$$

$$= \int_{\Omega_X} \int_{\Omega_Y} \int_{\Omega_X} \text{clr}(f)(x, y) \, dP_X \left[ \int_{\Omega_Y} \text{clr}(f)(x, y) \, dP_Y \right] \, dP_X \, dP_Y =$$

$$= \int_{\Omega_Y} \int_{\Omega_X} \text{clr}(f)(x, y) \, dP_X \cdot \int_{\Omega_X} \int_{\Omega_Y} \text{clr}(f)(x, y) \, dP_Y \, dP_X = 0$$

from the fact that $\text{clr}(f_{X,g}) \in L_0^2(\Omega_X)$ and $\text{clr}(f_{Y,g}) \in L_0^2(\Omega_Y)$. In the next step, the orthogonality between $f_{\text{int}} \equiv f_{\text{int}}(x, y)$ and the $X$-marginal is proven. Using the first part of this theorem and the relation $\langle \text{clr}(f), \text{clr}(f_{X,g}) \rangle_{L_0^2(P)} = \|\text{clr}(f_{X,g})\|_{L_0^2(P)}^2$ from the proof of Theorem 1 it holds

$$\langle \text{clr}(f_{\text{int}}), \text{clr}(f_{X,g}) \rangle_{L_0^2(P)} = \langle \text{clr}(f) - \text{clr}(f_{\text{ind}}), \text{clr}(f_{X,g}) \rangle_{L_0^2(P)} =$$

$$= \langle \text{clr}(f) - \text{clr}(f_{X,g}) - \text{clr}(f_{Y,g}), \text{clr}(f_{X,g}) \rangle_{L_0^2(P)} =$$

$$= \|\text{clr}(f_{X,g})\|_{L_0^2(P)}^2 - \|\text{clr}(f_{X,g})\|_{L_0^2(P)}^2 - \langle \text{clr}(f_{X,g}), \text{clr}(f_{Y,g}) \rangle_{L_0^2(P)} =$$

$$= \|\text{clr}(f_{X,g})\|_{L_0^2(P)}^2 - \|\text{clr}(f_{X,g})\|_{L_0^2(P)}^2 = 0.$$  

**Proof of Theorem 4** Equation (12) can be equivalently stated in terms of the clr marginals as

$$\text{clr}(f) + \text{clr}(f_{\text{int}}, X, g) = \text{clr}(f); \text{clr}(f) + \text{clr}(f_{\text{int}}, Y, g) = \text{clr}(f).$$  

(3)

In this case, one has

$$\text{clr}(f) + \text{clr}(f_{\text{int}}, X, g) = \text{clr}(f) + \int_{\Omega_X} \text{clr}(f_{\text{int}}) \, dP_X =$$

$$= \text{clr}(f) + \int_{\Omega_X} \text{clr}(f) \, dP_X - \int_{\Omega_X} \text{clr}(f_{X,g}) \, dP_X - \int_{\Omega_X} \text{clr}(f_{Y,g}) \, dP_X =$$

$$= \text{clr}(f) + \text{clr}(f_{Y,g}) - \text{clr}(f_{Y,g}) \cdot P_X(\Omega_X) = \text{clr}(f),$$

where the last equality holds true if the measure $P_X(\Omega_X)$ is normalized. With analogous argument, the same equality is proven for $f_{\text{int}, Y, g}$.  

$\square$
Proof of Theorem 5 From (11) and the expression $g_{\text{ind}} = (g_x \oplus f_x, g) \oplus (g_y \oplus f_y, g)$ it follows that $g_{\text{ind}}$ is an independence density of $g$. Therefore

$$g_{\text{int}} = g \ominus g_{\text{ind}} = f \ominus f_{\text{ind}} = f_{\text{int}}.$$ 

\[ \square \]

Proof of Theorem 6 Let the first term in (21) be denoted as

$$J_1 = \alpha \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ f_{ij} - s_{kl}(x_i, y_j) \right]^2$$

and the second one as

$$J_2 = (1 - \alpha) \iint_{\Omega} \left[ s_{kl}(u, v) \right]^2 \, dx \, dy$$

We can express the functional $J_1$ from (4) in matrix notation as

$$J_1 = \alpha \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ f_{ij} - s_{kl}(x_i, y_j) \right]^2 = \alpha \left[ cs(F) - B cs(B) \right]^\top \left[ cs(F) - B cs(B) \right] =$$

$$= \alpha \left( cs(F) \right)^\top cs(F) - 2\alpha \left( cs(B) \right)^\top B^\top cs(F) + \alpha \left( cs(B) \right)^\top B^\top B cs(F),$$

where $F = (f_{ij}), B := B_{l+1}(y) \otimes B_{k+1}(x), y = (y_1, \ldots, y_m), x = (x_1, \ldots, x_n).$ Now we consider the derivative of the spline. Similarly as in case of one-dimensional splines, Machalová et al. (2016); Machalová (2002a), the derivative can be expressed by using (23), (24) as

$$s_{kl}^{(u,v)}(x, y) = \frac{\partial^u}{\partial x^u} \frac{\partial^v}{\partial y^v} \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B_{i+1}^k(x) B_{j+1}^l(y) =$$

$$= \frac{\partial^u}{\partial x^u} \frac{\partial^v}{\partial y^v} (B_{l+1}^{(y)} \otimes B_{k+1}^{(x)}) cs(B) =$$

$$= \left[ B_{l+1-u}(y) S_v \otimes B_{k+1-u}(x) S_u \right] cs(B).$$

With respect to the properties of the tensor product, and using the notation $\mathbb{B}^{u,v}(x, y) := B_{l+1-u}(y) \otimes B_{k+1-u}(x)$, the derivative given in (6) can be reformulated as $s_{kl}^{(u,v)}(x, y) = \mathbb{B}^{u,v}(x, y) \otimes cs(B)$. Note that the flexibility in the choice of the orders $u, v$ in the derivatives $s_{kl}^{(u,v)}(x, y)$ can be considered as an element of innovation with respect to the classical tensor smoothing spline approach Dierckx (1993). Then the functional $J_2$ from (5) can be rewritten as

\[ \square \]
\[ J_2 = (1 - \alpha) \int_\Omega \left[ s_{kl}(u,v)(x, y) \right]^2 \mathrm{d}x \mathrm{d}y = \]

\[ = (1 - \alpha) \int_a^b \int_c^d \left[ B_{u,v}(x, y) \otimes s_c(B) \right]^\top B_{u,v}(x, y) \otimes s_c(B) \mathrm{d}y \mathrm{d}x = \]

\[ = (1 - \alpha) (cs(B))^\top S^\top S cs(B). \]

Furthermore,

\[ \int_a^b \int_c^d (B_{u,v}(x, y))^\top B_{u,v}(x, y) \mathrm{d}y \mathrm{d}x = \]

\[ = \int_a^b \int_c^d \left[ B_{l+1-v}(y) \otimes B_{k+1-u}(x) \right]^\top \left[ B_{l+1-v}(y) \otimes B_{k+1-u}(x) \right] \mathrm{d}y \mathrm{d}x = \]

\[ = \int_a^b \int_c^d \left[ B_{l+1-v}(y) B_{l+1-v}(y) \right] \otimes \left[ B_{k+1-u}(x) B_{k+1-u}(x) \right] \mathrm{d}y \mathrm{d}x = \]

\[ = M_{l,v}^{\top} \otimes M_{k,u}^{\top}. \]

This yields, \[ J_2 = (1 - \alpha) (cs(B))^\top S^\top M S cs(B). \] By putting together the matrix forms of \( J_1 \) and \( J_2 \), the functional \( J_{uv}(s_{kl}(x, y)) \) from (21) can be expressed as a function of unknown \( B \)-spline parameters \( b_{ij} \), specifically

\[ J_{uv} (cs(B)) = \alpha (cs(F))^\top cs(F) - 2\alpha (cs(B))^\top B^\top cs(F) + \alpha (cs(B))^\top B^\top B cs(B) + +(1 - \alpha) (cs(B))^\top S^\top M S cs(B). \] (7)

The fulfilment of the zero integral condition (22) is based on relation (26). By using this, the function \( J_{uv}(cs(B)) \) can be reformulated as

\[ J_{uv} (cs(\tilde{C})) = \alpha (cs(F))^\top cs(F) - 2\alpha (D \hat{K} cs(\tilde{C}))^\top B^\top cs(F) + \]

\[ + \alpha (D \hat{K} cs(\tilde{C}))^\top B^\top B D \hat{K} cs(\tilde{C}) + \]

\[ + (1 - \alpha) (D \hat{K} cs(\tilde{C}))^\top S^\top M S D \hat{K} cs(\tilde{C}). \] (8)

Thus, the necessary and sufficient condition for the minimum of function \( J_{uv}(cs(B)) \) is

\[ \frac{\partial J_{uv}(cs(B))}{\partial cs(B)} = 0. \] By applying this condition to (8) the following equation is obtained,

\[ \mathcal{K}^\top D^\top \left[ (1 - \alpha) S^\top M S + \alpha B^\top B \right] \mathcal{K} cs(\tilde{C}) = \alpha \mathcal{K}^\top D^\top B^\top cs(F). \]
Then the solution to this system is given by
\[
\mathbf{c}\mathbf{s}(\mathbf{C}^*) = \left[ \mathbf{K}^T \mathbf{D}^T \left( (1 - \alpha) \mathbf{S}^T \mathbf{M} \mathbf{S} + \alpha \mathbf{B}^T \mathbf{B} \right) \mathbf{D} \mathbf{K} \right]^+ \alpha \mathbf{K}^T \mathbf{D}^T \mathbf{B}^T \mathbf{c}\mathbf{s}(\mathbf{F}) \quad (9)
\]
And finally, the matrix \( \mathbf{B}^* \) of coefficients for the resulting smoothing spline with zero integral is obtained by
\[
\mathbf{c}\mathbf{s}(\mathbf{B}^*) = \mathbf{D} \mathbf{K} \mathbf{c}\mathbf{s}(\mathbf{C}^*). \quad (10)
\]

**Proof of Theorem 7**
The spline \( s_{kl}(x, y) \in S_{kl}^{\Delta_\lambda, \Delta_\mu}(\Omega) \) can be expressed as
\[
s_{kl}(x, y) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B_i^{k+1}(x) B_j^{l+1}(y) = \sum_{i=-k}^{g} s_i^j(y) B_i^{k+1}(x),
\]
where \( s_i^j(y) := \sum_{j=-l}^{h} b_{ij} B_j^{l+1}(y), i = -k, \cdots, g, \) are in fact one-dimensional splines of order \( l + 1 \) for the \( y \)-variable with coefficients \( b_{ij}, j = -l, \cdots, h. \) Then
\[
\int s_{kl}(x, y) \, dy = \int \sum_{i=-k}^{g} s_i^j(y) B_i^{k+1}(x) \, dy = \sum_{i=-k}^{g} B_i^{k+1}(x) \int s_i^j(y) \, dy
\]
and
\[
\int s_i^j(y) \, dy = s_i^{j+1}(y), \quad \text{with} \quad s_i^{j+1}(y) = \sum_{j=-l}^{h} u_{ij} B_j^{l+2}(y).
\]
By considering the case of one-dimensional splines, specifically the proof of Theorem 10, it holds
\[
u_{ij} = u_{i,j-1} + \frac{b_{ij}}{t_j}, \quad \text{where} \quad t_j = \frac{l + 1}{\mu_{j,l+1} - \mu_j}, \quad j = -l, \cdots, h,
\]
i.e.
\[
u_{ih} = \frac{b_{ih}}{t_h} + \cdots + \frac{b_{i,-l}}{t_{-l}} + u_{i,-l-1}, \quad \forall i.
\]
Altogether

\[
\int s_{kl}(x, y) \, dy = \sum_{i=-k}^{g} B_{i}^{k+1}(x) \sum_{j=-l-1}^{h} u_{ij} B_{j}^{l+2}(y) =
\]

\[
= \sum_{i=-k}^{g} \sum_{j=-l-1}^{h} u_{ij} B_{i}^{k+1}(x) B_{j}^{l+2}(y) =: s_{k,l+1}(x, y).
\]

Subsequently, using the last expression, the integral can be expressed as

\[
\int_{c}^{d} s_{kl}(x, y) \, dy = \left[ s_{k,l+1}(x, y) \right]_{c}^{d} = s_{k,l+1}(x, d) - s_{k,l+1}(x, c) =
\]

\[
= \sum_{i=-k}^{g} \sum_{j=-l-1}^{h} u_{ij} B_{i}^{k+1}(x) \left( B_{j}^{l+2}(d) - B_{j}^{l+2}(c) \right) = (13)
\]

\[
= \sum_{i=-k}^{g} B_{i}^{k+1}(x) \left( u_{ih} - u_{i,-l-1} \right) = \sum_{i=-k}^{g} v_{i} B_{i}^{k+1}(x) =: s_{k}(x),
\]

for

\[
v_{i} := u_{ih} - u_{i,-l-1} \quad \forall i = -k, \ldots, g,
\]

(14)

because of coincident additional knots (19), (20) it holds

\[
B_{j}^{l+2}(d) = \begin{cases} 1 & \text{if } j = h \\ 0 & \text{otherwise} \end{cases} \quad B_{j}^{l+2}(c) = \begin{cases} 1 & \text{if } j = -l - 1 \\ 0 & \text{otherwise} \end{cases}
\]

Finally, according to (13) and (A1), there is

\[
\int_{a}^{b} \int_{c}^{d} s_{kl}(x, y) \, dy \, dx = \int_{a}^{b} s_{k}(x) \, dx = \left[ s_{k+1}(x) \right]_{a}^{b} =
\]

\[
= s_{k+1}(b) - s_{k+1}(a) = w_{g} - w_{-k-1},
\]

(15)

where \( s_{k+1}(x) = \sum_{i=-k-1}^{g} w_{i} B_{i}^{k+2}(x) \) and

\[
w_{i} = w_{i-1} + \frac{v_{i}}{d_{i}}, \quad \text{with} \quad d_{i} = \frac{k + 1}{\lambda_{i+k+1} - \lambda_{i}} \quad \forall i = -k, \ldots, g,
\]

(16)

i.e.

\[
w_{g} = \frac{v_{g}}{d_{g}} + \cdots + \frac{v_{-k}}{d_{-k}} + w_{-k-1}.
\]

(17)
As a direct consequence, the following equivalence can be formulated

\[
\int_a^b \int_c^d s_{kl}(x, y) \, dy \, dx = 0 \quad \Leftrightarrow \quad w_g = w_{-k-1} \quad \Leftrightarrow \quad \frac{v_g}{d_g} + \cdots + \frac{v_{-k}}{d_{-k}} = 0.
\]

By using (14) and (12),

\[
\frac{v_g}{d_g} + \cdots + \frac{v_{-k}}{d_{-k}} = \sum_{i=-k}^{g} \frac{u_{ih} - u_{i,-l-1}}{d_i} = \sum_{i=-k}^{g} \frac{1}{d_i} \left(\frac{b_{ih}}{t_h} + \cdots + \frac{b_{i,-l}}{t_{-l}}\right) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} \frac{b_{ij}}{d_i t_j},
\]

and altogether

\[
\int_a^b \int_c^d s_{kl}(x, y) \, dy \, dx = 0 \quad \Leftrightarrow \quad \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} \left(\lambda_{i+k+1} - \lambda_i\right) \left(\mu_{j+k+1} - \mu_j\right) = 0.
\]

\(\square\)

**Proof of Theorem 8** Let \( s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega) \), with the given representation

\[
s_{kl}(x, y) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B_{i+1}^{k}(x) B_{j+1}^{l}(y),
\]

and let \( \int_{\Omega} s_{kl}(x, y) \, dx \, dy = 0 \). Then from Theorem 7 it is

\[
\sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} \left(\lambda_{i+k+1} - \lambda_i\right) \left(\mu_{j+k+1} - \mu_j\right) = 0. \tag{18}
\]

By using (13), (14) from the proof of Theorem 7 it is obtained that \( s_k(x) = \sum_{i=-k}^{g} v_i B_{i+1}^{k}(x) \), where \( v_i = u_{ih} - u_{i,-l-1} \). According to (12) it holds

\[
v_i = \frac{b_{ih}}{t_h} + \cdots + \frac{b_{i,-l}}{t_{-l}}. \tag{19}
\]

Next, by considering (15),

\[
\int_a^b s_k(x) \, dx = [s_{k+1}(x)]_a^b = w_g - w_{-k-1},
\]
where \( s_{k+1}(x) = \sum_{i=-k-1}^{g} w_i B_i^{k+2}(x) \). However, with respect to (16), (17), (19) and (18) this difference equals to

\[
w_g - w_{-k-1} = \frac{v_g}{d_g} + \ldots + \frac{v_{-k}}{d_{-k}} = \sum_{i=-k}^{g} \frac{v_i}{d_i} = \sum_{i=-k}^{g} \frac{1}{d_i} \sum_{j=-l}^{h} b_{ij} = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} \left( \lambda_{i+k+1} - \lambda_i \right) \left( \mu_{j+l+1} - \mu_j \right) = 0,
\]

and consequently also \( \int_{a}^{b} s_k(x)dx = 0 \). The second statement can be proven analogously.

\( \square \)

**Proof of Theorem 9** Every bivariate spline \( s_{kl}(x, y) \in S_{kl}^{\Delta \lambda, \Delta \mu}(\Omega) \) can be expressed as

\[
s_{kl}(x, y) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} B_i^{k+1}(x) B_j^{l+1}(y) = \sum_{i=-k}^{g} c_i B_i^{k+1}(x),
\]

where \( c_i = \sum_{j=-l}^{h} b_{ij} B_j^{l+1}(y) \). For a given univariate spline \( s_k(x) = \sum_{i=-k}^{g} v_i B_i^{k+1}(x) \) we can define coefficients

\[
v_{ij} := v_i, \quad \forall j = -l, \ldots, h.
\]

Then \( s_k(x) \) can be expressed as a bivariate spline which is constant in variable \( y \) and which uses \( B \)-spline bases functions \( B_j^{l+1}(y) \) in the form

\[
s_k(x) = \sum_{i=-k}^{g} \sum_{j=-l}^{h} v_{ij} B_i^{k+1}(x) B_j^{l+1}(y),
\]

since with respect to the properties of \( B \)-splines, de Boor (1978), Dierckx (1993), Schumaker (2007), we have

\[
\sum_{j=-l}^{h} v_{ij} B_j^{l+1}(y) = v_i \sum_{j=-l}^{h} B_j^{l+1}(y) = v_i \cdot 1 = v_i.
\]

The rest of the proof is obvious with respect to the addition or subtraction of two splines. \( \square \)
Appendix C: Algorithm

Theorem 7 enables to formulate an algorithm for finding a bivariate tensor spline
\( s_{kl}(x, y) \in S^{\Delta_{\lambda}, \Delta_{\mu}}(\Omega) \) with zero integral over \( \Omega \). This task is equivalent to finding
the matrix \( B = (b_{ij}), i = -k, \ldots, g, j = -l, \ldots, h \) of the \( B \)-spline coefficients:

1. choose \((g + k + 1)(h + l + 1) - 1\) arbitrary \( B \)-spline coefficients \( b_{ij} \in \mathbb{R} \), for
i = \(-k\), \ldots, \(\beta - 1\), \(\beta + 1\), \ldots, \(g\) and j = \(-l\), \ldots, \(\gamma - 1\), \(\gamma + 1\), \ldots, \(h\),
2. compute
\[
b_{\beta \gamma} = \frac{-1}{(\lambda_{\beta + k + 1} - \lambda_{\beta})(\mu_{\gamma + l + 1} - \mu_{\gamma})} \sum_{i=-k}^{g} \sum_{j=-l}^{h} b_{ij} (\lambda_{i+k+1} - \lambda_{i}) (\mu_{j+l+1} - \mu_{j}).
\]

Appendix D: Complete set of anthropometric data

See Figs. 7, 8, 9, and 10.

Fig. 7 Anthropometric data: smoothed clr transformed densities for all age intervals together with data points resulting from the discrete clr transformation at mid-points of histogram classes. The choice of the scale of the reference measure (uniform measure) does not play any role here.
Fig. 8  Anthropometric data: smoothed original bivariate densities for all age intervals

Fig. 9  Anthropometric data: smoothed independent densities for all age intervals
Fig. 10 Anthropometric data: smoothed interation densities for all age intervals

References

Bigot J, Gouet R, Klein T, López A (2019) Geodesic pca in the wasserstein space by convex pca. Ann. Inst. Henri Poincaré Probab. Stat. 53(1):1–26

de Boor C (1978) A practical guide to splines. Springer, New York

Delicado P (2011) Dimensionality reduction when data are density functions. Comput Stat Data Anal 55:401–420

Dierckx P (1993) Curve and surface fitting with splines. Oxford University Press, New York

Egozcue JJ, Diaz-Barrero JL, Pawlowsky-Glahn V (2008) Compositional analysis of bivariate discrete probabilities. In: Proceedings of CODAWORK 08

Egozcue JJ, Pawlowsky-Glahn V (2016) Changing the reference measure in the simplex and its weighting effects. Aust J Stat 45(4):25–44

Egozcue JJ, Pawlowsky-Glahn V, Templ M, Hron K (2015) Independence in contingency tables using simplicial geometry. Commn Stat Theory Methods 44:3978–3996

Freedman D, Lane D (1983) A nonstochastic interpretation of reported significance levels. J Bus Econ Stat 1(4):292–298

Gába A, Přidalová M (2014) Age-related changes in body composition in a sample of czech women aged 18–89 years: a cross-sectional study. Eur J Nutr 53(1):167–176

Gába A, Přidalová M (2016) Diagnostic performance of body mass index to identify adiposity in women. Eur J Clin Nutr 70:898–903

Guégan D, Iacopini M (2019) Nonparametric forecasting of multivariate probability density functions. ArXiv report arXiv:1803.06823v1

Hron K, Menafoglio A, Templ M, Hrůzová K, Filzmoser P (2016) Simplicial principal component analysis for density functions in bayes spaces. Comput Stat Data Anal 94:330–350

Kokoszka P, Miao H, Petersen A, Shang HL (2019) Forecasting of density functions with an application to cross-sectional and intraday returns. Int J Forecasting 35(4):1304–1317

Kwitowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. J Econ 54:159–178

Machalová J (2002) Optimal interpolatory splines using b-spline representation. Acta Univ Palacki Olomuc Fac rer nat Mathematica 41:105–118

Machalová J (2002) Optimal interpolatory and optimal smoothing spline. J Electr Eng 53(12/s):79–82
Machalová J, Hron K, Monti GS (2016) Preprocessing of centred logratio transformed density functions using smoothing splines. J Appl Stat 43(8):1419–1435
Machalová J, Talšká R, Hron K, Gába A (2020) Compositional splines for representation of density functions. Comput Stat. https://doi.org/10.1007/s00180-020-01042-7
Martín-Fernández JA, Hron K, Tempel M, Filzmoser P, Palarea-Albaladejo J (2015) Bayesian-multiplicative treatment of count zeros in compositional data sets. Stat Model 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 Environ Res Risk Assess 28(7):1835–1851
Menafoglio A, Grasso M, Secchi P, Colosimo BM (2016) A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. Math Geosci 48(4):463–485
Menafoglio A, Grasso M, Secchi P, Colosimo BM (2018) Monitoring of probability density functions via simplicial functional pca with application to image data. Technometrics 60(4):497–510
Menafoglio A, Gaetani G, Secchi P (2018) Random domain decompositions for object-oriented kriging over complex domains. Stochastic Environmental Research and Risk Assessment
Nelsen RB (2006) An introduction to copulas. Springer, New York
Nerini D, Ghattas B (2007) Classifying densities using functional regression trees: applications in oceanology. Comput Stat Data Anal 51(10):4984–4993
Panaretos VM, Zemel Y (2019) Statistical aspects of wasserstein distances. Annu. Rev. Stat. Appl. 6(1):405–431
Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data. Wiley, Chichester
Petersen A, Müller HG (2016) Unctional data analysis for density functions by transformation to a Hilbert space. Ann Stat 44(1):183–218
Petersen A, Xi L, Divani AA (2019) Wasserstein f-tests and confidence bands for the fréchet regression of density response curves. ArXiv report arXiv:1910.1341
Pini A, Stamm A, Vantini S (2018) Hotelling’s t2 in functional hilbert spaces. J Multiv Anal 167:284–305
Ramsay J, Silverman BW (2005) Functional data analysis. Springer, New York
Schumaker L (2007) Spline functions: basic theory. Cambridge University Press, Cambridge
Seo WK, Beare BK (2019) Cointegrated linear processes in Bayes Hilbert space. Stat Probab Lett 147:90–95
Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publ Inst Stat Univ Paris 8:229–231
Srivastava A, Jermyn I, Joshi S (2007) Riemannian analysis of probability density functions with applications in vision. IEEE Xplore. https://doi.org/10.1109/CVPR.2007.383188
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. https://doi.org/10.1002/sta4.283
Tran HD, Pham UH, Ly S, Vo-Duy T (2015) A new measure of monotone dependence by using sobolev norms for copula. In: Huynh V-N, Inuiuch M, Demouex T (eds) Integrated uncertainty in knowledge modelling and decision making. Springer, Cham, pp 126–137
van den Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2010) Bayes linear spaces. Stat Oper Res Trans 34(2):201–222
van den Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2014) Hilbert bayes spaces. Aust NZ J Stat 54(2):171–194
WHO (2020) Adolescent health. https://www.who.int/southeastasia/health-topics/adolescent-health. Accessed 27 Nov 2020
Yule GU (1912) On the methods of measuring association between two attributes. J R Stat Soc 75(6):579–642

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.