DISTRIBUTIONS ON THE SIMPLEX

G. Mateu-Figueras, V. Pawlowsky-Glahn, and C. Barceló-Vidal
Universitat de Girona, Girona; *gloria.mateu@udg.es*

Abstract

The simplex, the sample space of compositional data, can be structured as a real Euclidean space. This fact allows to work with the coefficients with respect to an orthonormal basis. Over these coefficients we apply standard real analysis, in particular, we define two different laws of probability through the density function and we study their main properties.

1 Algebraic geometric structure on $\mathcal{S}^D$

Any vector $\mathbf{x} = (x_1, x_2, \ldots, x_D)'$ representing proportions of some whole can be expressed as subject to the unit-sum constraint $\sum_{i=1}^{D} x_i = 1$. Therefore, a suitable sample space for compositional data, consisting of such vectors of proportions (compositions), can be taken to be the unit simplex defined as

$$\mathcal{S}^D = \{(x_1, x_2, \ldots, x_D)' : x_1 > 0, x_2 > 0, \ldots, x_D > 0; \sum_{i=1}^{D} x_i = 1\}.$$

A real vector space structure is induced by the perturbation operation, defined for any two compositions $\mathbf{x}, \mathbf{x}^* \in \mathcal{S}^D$, as $\mathbf{x} \oplus \mathbf{x}^* = \mathcal{C}\,(x_1x_1^*, x_2x_2^*, \ldots, x_Dx_D^*)'$, and the power transformation, defined for any $\mathbf{x} \in \mathcal{S}^D$ and $\alpha \in \mathbb{R}$ as $\alpha \otimes \mathbf{x} = \mathcal{C}(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha)'$. $\mathcal{C}(\cdot)$ denotes the closure operation, a transformation from $\mathbb{R}_+^D$ to $\mathcal{S}^D$ that converts each vector into its composition.

The Euclidean space structure is induced by the inner product defined by Aitchison (2002) for any two compositions $\mathbf{x}, \mathbf{x}^* \in \mathcal{S}^D$ as

$$\langle \mathbf{x}, \mathbf{x}^* \rangle_a = \frac{1}{D} \sum_{i<j} \ln \frac{x_i}{x_j} \ln \frac{x_i^*}{x_j^*}. \tag{1}$$

The associated norm is $\|\mathbf{x}\|_a = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a}$ and the associated distance, which we know as Aichison distance, is

$$d_a(\mathbf{x}, \mathbf{x}^*) = \sqrt{\frac{1}{D} \sum_{i<j} \left(\ln \frac{x_i}{x_j} - \ln \frac{x_i^*}{x_j^*} \right)^2}. \tag{2}$$

The distance (2) follows the standard properties of a distance. Given $\mathbf{x}, \mathbf{x}^*, \mathbf{x}' \in \mathcal{S}^D$, and $\alpha \in \mathbb{R}$, it is easy to prove that $d_a(\mathbf{x}, \mathbf{x}^*) = d_a(\mathbf{x}' \oplus \mathbf{x}, \mathbf{x}' \oplus \mathbf{x}^*)$, and $d_a(\alpha \otimes \mathbf{x}, \alpha \otimes \mathbf{x}^*) = |\alpha| d_a(\mathbf{x}, \mathbf{x}^*)$. Thus we say that the distance d_a is coherent or compatible with the structure of $\mathcal{S}^D$. Also, the distance (2) does not depend on the particular order of the parts because $d_a(\mathbf{P}\mathbf{x}, \mathbf{P}\mathbf{x}^*) = d_a(\mathbf{x}, \mathbf{x}^*)$, for any permutation matrix $\mathbf{P}$. These properties had been studied by Aitchison (1992) and Martín-Fernández (2001); for a proof see Pawlowsky-Glahn and Egozcue (2002, p. 269).

A D-part composition is usually expressed in terms of the canonical basis of $\mathbb{R}^D$, using the sum and the scalar product operations. In fact, any $\mathbf{x} \in \mathcal{S}^D$ can be written as $\mathbf{x} = (x_1, x_2, \ldots, x_D)' = x_1(1, 0, \ldots, 0)' + x_2(0, 1, 0, \ldots, 0)' + \cdots + x_D(0, \ldots, 0, 1)'$. Note that the canonical basis of $\mathbb{R}^D$ is not a basis on $\mathcal{S}^D$ and this expression is not a linear combination on $\mathcal{S}^D$ with respect to its vector space structure. But, given that $\mathcal{S}^D$ is a vector space with dimension $D-1$, general algebra theory assures the existence of a basis. Before we introduce a basis of $\mathcal{S}^D$, note that the set $B^* = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_D\}$, where $\mathbf{w}_i = \mathcal{C}(1, 1, \ldots, e, \ldots, 1)'$ with the element e placed in the i-th row, is a generating set of $\mathcal{S}^D$. Certainly, any composition $\mathbf{x} \in \mathcal{S}^D$ can be expressed as, for example, $\mathbf{x} = (\ln x_1 \otimes \mathbf{w}_1) \oplus (\ln x_2 \otimes \mathbf{w}_2) \oplus \cdots \oplus (\ln x_D \otimes \mathbf{w}_D)$. Obviously there are more possible expressions as $\mathbf{x} = (\ln(x_1/g(\mathbf{x})) \otimes \mathbf{w}_1) \oplus (\ln(x_2/g(\mathbf{x})) \otimes \mathbf{w}_2) \oplus \cdots \oplus (\ln(x_D/g(\mathbf{x})) \otimes \mathbf{w}_D)$, where

$g(\mathbf{x})$ is the geometric mean of composition $\mathbf{x}$. Given $\mathbf{x} \in \mathcal{S}^D$, we say that one possible vector of coefficients with respect to the generating set B^* is

$$\mathrm{clr}(\mathbf{x}) = \left(\ln\left(\frac{x_1}{g(\mathbf{x})}\right), \ln\left(\frac{x_2}{g(\mathbf{x})}\right), \ldots, \ln\left(\frac{x_D}{g(\mathbf{x})}\right)\right)'. \tag{3}$$

We will use the notation $\mathrm{clr}(\mathbf{x})$ to emphasize the similarity with the vector obtained applying the centred logratio transformation to composition $\mathbf{x}$ (Aitchison, 1986, p. 94). Note that, like for the centred logratio vector, the sum of the components of the vector $\mathrm{clr}(\mathbf{x})$ is equal to 0.

As the dimension of $\mathcal{S}^D$ is $D-1$, we can eliminate any composition of B^* to obtain a basis of $\mathcal{S}^D$. If we eliminate the last composition $\mathbf{w}_D$ we obtain the basis $B = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{D-1}\}$ and now a composition $\mathbf{x}$ has a unique expression as a linear combination $\mathbf{x} = (\ln(x_1/x_D) \otimes \mathbf{w}_1) \oplus (\ln(x_2/x_D) \otimes \mathbf{w}_2) \oplus \cdots \oplus (\ln(x_{D-1}/x_D) \otimes \mathbf{w}_{D-1})$. Thus we say that the coefficients of $\mathbf{x}$ with respect to the B basis are

$$\mathrm{alr}(\mathbf{x}) = \left(\ln\left(\frac{x_1}{x_D}\right), \ln\left(\frac{x_2}{x_D}\right), \ldots, \ln\left(\frac{x_{D-1}}{x_D}\right)\right)'. \tag{4}$$

We will use the notation $\mathrm{alr}(\mathbf{x})$ to emphasize the similarity with the vector obtained applying the additive logratio transformation to a composition $\mathbf{x}$ (Aitchison, 1986, p. 113). We can suppress any other composition $\mathbf{w}_i$ $(i = 1, 2, \ldots, D-1)$ of the set B^* instead of $\mathbf{w}_D$, to obtain another basis. In this case, the components of $\mathbf{x}$ with respect to this new basis are $(\ln(x_1/x_i), \ldots, \ln(x_{i-1}/x_i), \ln(x_{i+1}/x_i), \ldots, \ln(x_D/x_i))'$.

The inner product (1) and its associated norm ensure the existence of an orthonormal basis. With simple algebraic operations we can check that the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{D-1}\}$ is not orthonormal. Using the Gram-Schmidt method we can obtain from this basis an orthonormal basis denoted as $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$. A composition $\mathbf{x}$ has a unique expression as a linear combination $\mathbf{x} = (\langle \mathbf{x}, \mathbf{e}_1\rangle_a \otimes \mathbf{e}_1) \oplus (\langle \mathbf{x}, \mathbf{e}_2\rangle_a \otimes \mathbf{e}_2) \oplus \ldots \oplus (\langle \mathbf{x}, \mathbf{e}_{D-1}\rangle_a \otimes \mathbf{e}_{D-1})$. Thus we say that the coefficients of any composition $\mathbf{x} \in \mathcal{S}^D$ with respect to the orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ are:

$$\mathrm{ilr}(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{e}_1\rangle_a, \langle \mathbf{x}, \mathbf{e}_2\rangle_a, \ldots, \langle \mathbf{x}, \mathbf{e}_{D-1}\rangle_a)'. \tag{5}$$

We will use the notation $\mathrm{ilr}(\mathbf{x})$ to emphasize the similarity with the vector obtained applying the isometric logratio transformation to composition $\mathbf{x}$ (Egozcue et al., 2003, p. 294). As in real space, there exist an infinite number of orthonormal basis on $\mathcal{S}^D$. Aitchison (1986, p. 92) provides the relationship between the alr and clr transformations. Later, Egozcue et al. (2003, p. 298) provide the relationship between the alr, clr and ilr transformations. Now they can be interpreted as a change of basis or of generating system that relates the coefficients $\mathrm{alr}(\mathbf{x}), \mathrm{clr}(\mathbf{x})$ and $\mathrm{ilr}(\mathbf{x})$. These relations are:

$$\begin{array}{ll} \mathrm{alr}(\mathbf{x}) = \mathbf{F}_{D-1,D}\mathrm{clr}(\mathbf{x}); & \mathrm{clr}(\mathbf{x}) = \mathbf{F}^*_{D,D-1}\mathrm{alr}(\mathbf{x}); \\ \mathrm{clr}(\mathbf{x}) = \mathbf{U}_{D,D-1}\mathrm{ilr}(\mathbf{x}); & \mathrm{ilr}(\mathbf{x}) = \mathbf{U}'_{D-1,D}\mathrm{clr}(\mathbf{x}); \\ \mathrm{alr}(\mathbf{x}) = \mathbf{F}_{D-1,D}\mathbf{U}_{D,D-1}\mathrm{ilr}(\mathbf{x}); & \mathrm{ilr}(\mathbf{x}) = \mathbf{U}'_{D-1,D}\mathbf{F}^*_{D,D-1}\mathrm{alr}(\mathbf{x}), \end{array} \tag{6}$$

where $\mathbf{F}_{D-1,D} = [\mathbf{I}_{D-1} : \mathbf{j}_{D-1}]$, with $\mathbf{I}_{D-1}$ representing the identity matrix of order $D-1$ and $\mathbf{j}_{D-1}$ the column vector of units; $\mathbf{U}_{D,D-1}$ is a $D \times (D-1)$ matrix with the coefficients $\mathrm{clr}(\mathbf{e}_i)$ $(i = 1, 2, \ldots, D-1)$ as columns; and $\mathbf{F}^*_{D,D-1}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{F}_{D-1,D}$ given by

$$\mathbf{F}^* = \frac{1}{D}\begin{pmatrix} D-1 & -1 & \cdots & -1 \\ -1 & D-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & D-1 \\ -1 & -1 & \cdots & -1 \end{pmatrix}.$$

Over the coefficients with respect to an orthonormal basis we can apply standard real analysis. Certainly, it is easy to see that the operations $\oplus$ and $\otimes$ are equivalent to the sum and the scalar product of the respective coefficients with respect to any basis (not necessarily orthonormal). In the particular case of coefficients with respect to an orthonormal basis, we can apply the standard inner product and the Euclidean distance in $\mathbb{R}^{D-1}$. We cannot assure this properties for the coefficients with respect to a generating set in general, but in the particular case of the coefficients (3), all properties are fulfilled.

Following Aitchison (1986), given a composition $\mathbf{x} \in \mathcal{S}^D$ we may wish to focus attention on the relative magnitude of a subset of components, so we need the definition of a subcomposition. As stated in Aitchison (1986), the formation of a C-part subcomposition, $\mathbf{s}$, from a D-part composition, $\mathbf{x}$, can be achieved as $\mathbf{s} = \mathcal{C}(\mathbf{S}\mathbf{x})$, where $\mathbf{S}$ is a $C \times D$ selection matrix with C elements equal to 1 (one in each row and at most one in each column) and the remaining elements equal to 0. A subcomposition can be regarded as a composition in a simplex of lower dimension.

When we work with large dimensional compositions, it may be of interest to amalgamate components to form a new composition. An amalgamation is any composition in a simplex of lower dimension obtained from the sum of some groups of components of the original composition.

A random composition $\mathbf{x}$ is a random vector with $\mathcal{S}^D$ as domain. Aitchison (1997) uses the geometric interpretation of the expected value of a random vector to define the center of a random composition as the composition $\mathrm{cen}(\mathbf{x})$ which minimizes the expression $\mathrm{E}[d_a^2(\mathbf{x}, \mathrm{cen}[\mathbf{x}])]$. We obtain $\mathrm{cen}[\mathbf{x}] = \mathcal{C}(\exp(\mathrm{E}[\ln \mathbf{x}]))$ or equivalently $\mathrm{cen}[\mathbf{x}] = \mathcal{C}(\exp(\mathrm{E}[\ln(\mathbf{x}/g(\mathbf{x}))]))$ (Aitchison, 1997, p. 10, Pawlowsky-Glahn and Egozcue, 2002, p. 270), who also include the equality

$$\mathrm{alr}(\mathrm{cen}[\mathbf{x}]) = \mathrm{E}[\mathrm{alr}(\mathbf{x})]. \tag{7}$$

Using (6) and (7), we can prove the equalities

$$\begin{aligned} \mathrm{clr}(\mathrm{cen}[\mathbf{x}]) &= \mathrm{E}[\mathrm{clr}(\mathbf{x})], \\ \mathrm{ilr}(\mathrm{cen}[\mathbf{x}]) &= \mathrm{E}[\mathrm{ilr}(\mathbf{x})] \end{aligned} \tag{8}$$

This means that the coefficients of composition $\mathrm{cen}(\mathbf{x})$ with respect to the B basis, to the generating set B^* and to an orthonomal basis of $\mathcal{S}^D$ are equal to the expected value of the coefficients of $\mathbf{x}$ with respect to the B basis, the generating set B^* and an orthonormal basis of $\mathcal{S}^D$, respectively. Observe that we can compute $\mathrm{E}[\mathrm{alr}(\mathbf{x})], \mathrm{E}[\mathrm{clr}(\mathbf{x})]$ and $\mathrm{E}[\mathrm{ilr}(\mathbf{x})]$ applying the standard definition. Afterwards we can obtain the related composition $\mathrm{cen}(\mathbf{x})$ through the corresponding linear combination. This result confirms that working with the coefficients with respect to a basis and applying the standard methodology is equivalent to work directly with compositions. In particular, for the center of a random composition, we can use the coefficients alr, clr or ilr without distinction. More generally, Pawlowsky-Glahn and Egozcue (2001, p.388) prove the equality $h(\mathrm{cen}[\mathbf{x}]) = \mathrm{E}[h(\mathbf{x})]$ for any isomorphism h.

The variance of a real random vector can be interpreted as the expected value of the squared Euclidean distance around its expected value. This interpretation is used by Pawlowsky-Glahn and Egozcue (2002, p. 264) to define the metric variance of a random composition $\mathbf{x}$ as $\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_a^2(\mathbf{x}, \mathrm{cen}[\mathbf{x}])]$. The coefficients with respect to an orthonormal basis and with respect to the generating set B^* provide two equivalent expressions for the metric variance as

$$\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_{eu}^2(\mathrm{ilr}(\mathbf{x}), \mathrm{ilr}(\mathrm{cen}[\mathbf{x}]))] = \mathrm{E}[d_{eu}^2(\mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathrm{cen}[\mathbf{x}]))].$$

In this case it is not possible to use the Euclidean distance between the alr coefficients because the basis B is not orthonormal. Consequently, the Aitchison distance between $\mathbf{x}$ and $\mathrm{cen}(\mathbf{x})$ is not equal to the Euclidean distance between the respective alr coefficients. Certainly, Pawlowsky-Glahn and Egozcue (2001) prove the equality $\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_{eu}^2(h(\mathbf{x}), \mathrm{E}[h(\mathbf{x})])]$ for any h isometry but we know that the alr transformation is not an isometry.

Aitchison(1997, p. 13) defines the total variability as $\mathrm{totvar}(\mathbf{x}) = \mathrm{trace}(\mathbf{\Gamma})$ where $\mathbf{\Gamma}$ is the standard covariance matrix of coefficients (3). Later, Pawlowsky-Glahn and Egozcue (2002, p. 265) obtain the equality Mvar[$\mathbf{x}$] = totvar[$\mathbf{x}$]. Using equalities (6) we can prove that totvar($\mathbf{x}$) = trace($\boldsymbol{\Upsilon}$) where $\boldsymbol{\Upsilon}$ is the standard covariance matrix of coefficients (5). As $\boldsymbol{\Upsilon} = \mathrm{var}(\mathrm{ilr}(\mathbf{x})) = \mathrm{var}(\mathbf{U}'\mathrm{clr}(\mathbf{x})\mathbf{U}) = \mathbf{U}'\mathrm{var}(\mathrm{clr}(\mathbf{x}))\mathbf{U} = \mathbf{U}'\boldsymbol{\Gamma}\mathbf{U}$, we obtain $\mathrm{trace}(\boldsymbol{\Upsilon}) = \mathrm{trace}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U}) = \mathrm{trace}(\boldsymbol{\Gamma}\mathbf{U}\mathbf{U}') = \mathrm{trace}(\boldsymbol{\Gamma})$ because matrix $\mathbf{U}\mathbf{U}'$ has the rows of $\boldsymbol{\Gamma}$ as eigenvectors with eigenvalue 1. In this case it is not possible to use the covariance matrix of coefficients (4) because when we compute a covariance we use implicitly a distance and we know that the Aitchison distance between two compositions is not equal to the Euclidean distance between the respective alr coefficients.

We know that $\mathcal{S}^D \subseteq \mathbb{R}^D$ and consequently we can apply the traditional approach to define laws of probability. The additive logistic normal distribution defined by Aitchison and Shen (1980) and studied by Aitchison (1986) and the additive logistic skew-normal distribution introduced by Mateu-Figueras et al. (1998) and studied in Mateu Figueras (2003) are two laws of probability defined through transformations from $\mathcal{S}^D$ to the real space. In the next section, we consider the Euclidean space structure of $\mathcal{S}^D$ and we define a law of probability over the coefficients of the random composition with respect to an orthonormal basis. In particular, we define the normal and the skew-normal distributions on $\mathcal{S}^D$.

To state it clearly: the essential difference between the traditional methodology and the approach we are going to present is the measure assumed on $\mathcal{S}^D$. In the traditional approach the measure used is the usual Lebesgue measure on $\mathcal{S}^D$, whereas in our approach, the measure is the Lebesgue measure on the coefficients with respect to an orthonormal basis on $\mathcal{S}^D$.

2 Distributions on $\mathcal{S}^D$, some general aspects

Given any measurable space, the Radon-Nikodym derivative of a probability P with respect to a measure ν is a measurable and non-negative function $f(\cdot)$ such that the probability of any event A of the corresponding σ-algebra is

$$P(A) = \int_A f(\mathbf{x}) d\nu(\mathbf{x}).$$

We also name $f(\cdot)$ as the density function of probability P with respect to the measure ν. When we work with random variables or random vectors in real space, we use the density function with respect to the Lebesgue measure and we call it "the density function of the random variable". We do not mention the measure because it is understood that it is the Lebesgue measure. The Lebesgue measure has an important role in real analysis because it is invariant under translations.

On $\mathcal{S}^D$ we can define quite straightforwardly a law of probability through the density function with respect to a measure ν on $\mathcal{S}^D$. But we are interested in finding a measure similar to the Lebesgue measure in real space. In particular we look for a measure ν invariant under the operation $\oplus$, the internal operation in $\mathcal{S}^D$. Taking into account the distance d_a, we conclude that the Lebesgue measure in real space is not adequate in $\mathcal{S}^D$. Nevertheless, note that if ν is not the usual Lebesgue measure in real space, we cannot make effective the calculation of any integral.

As we have seen in the previous section, $\mathcal{S}^D$ has a $D-1$ dimensional Euclidean space structure and we could apply the isomorphism between $\mathcal{S}^D$ and $\mathbb{R}^{D-1}$. We have only to identify each element of $\mathcal{S}^D$ with its vector of coefficients with respect to an orthonormal basis. In this case we can introduce the density function of the coefficients with respect to the Lebesgue measure and apply all the standard probability theory. This function allows us to compute the probability of any event by means of an ordinate integral, i.e., if $f^*(\cdot)$ is the density function of the coefficients with respect to an orthonormal basis, we can compute the probability of any event $A \subseteq \mathcal{S}^D$ as

$$P(A) = \int_{A^*} f^*(v_1, v_2, \ldots, v_{D-1}) dv_1 dv_2 \ldots dv_{D-1},$$

where A^* and $(v_1, v_2, \ldots, v_{D-1})$ represent the coefficients with respect to the orthonormal basis that characterize the set A and the composition $\mathbf{x}$.

We have to bear in mind that if we use this methodology to compute any element of the support space, we will obtain the coefficients of this element with respect to the orthonormal basis used. Next, we could obtain the corresponding composition by means of a linear combination. Mateu-Figueras and Pawlowsky-Glahn (2003) and Pawlowsky-Glahn et al. (2003) compute, respectively, the expected value of a normal model on $\mathbb{R}^+$ and the expected value of a bivariate normal model on $\mathbb{R}^2_+$ using the principles of working on coefficients with respect to an orthonormal basis. In our case it will be easy to compute the expected value of the coefficients of any random composition $\mathbf{x}$, i.e. $\mathrm{E}[\mathrm{ilr}(\mathbf{x})]$. Interpreting them as coefficients of our basis of reference, we will obtain the composition that we call expected value and we denote as $\mathrm{E}[\mathbf{x}]$ through a linear combination.

Now we introduce two laws of probability on $\mathcal{S}^D$ indicating the expression of function f^*. In both cases, the families of distributions are closed by $\oplus$ and $\otimes$ operations and, moreover, the densities satisfy the equality

$$f^*_{\mathbf{x}}(\mathbf{x}) = f^*_{\mathbf{a}\oplus\mathbf{x}}(\mathbf{a}\oplus\mathbf{x}), \tag{9}$$

where $f^*_{\mathbf{x}}$ and $f^*_{\mathbf{a}\oplus\mathbf{x}}$ represent the densities of random compositions $\mathbf{x}$ and $\mathbf{a}\oplus\mathbf{x}$ respectively, with $\mathbf{a}$ a constant composition. This property has important consequences in $\mathcal{S}^D$ because we often apply the centering operation (Martín-Fernández, 2001).

We also compute the expected value and the covariance matrix using the coefficients with respect to the orthonormal basis and the standard procedures in real space. The obtained results are coherent with the center and the metric variance of a random composition.

3 The normal distribution on $\mathcal{S}^D$

3.1 Definition and properties

Definition 1 Let be $(\Omega, \mathcal{F}, p)$ a probability space. A random composition $\mathbf{x} : \Omega \longrightarrow \mathcal{S}^D$ is said to have a normal on $\mathcal{S}^D$ distribution with parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Upsilon}$, if the density function of the coefficients with respect to an orthonormal basis of $\mathcal{S}^D$ is

$$f^*_{\mathbf{x}}(\mathbf{x}) = (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}\,(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})'\,\boldsymbol{\Upsilon}^{-1}\,(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right]. \tag{10}$$

We use the notation $\mathbf{x} \sim \mathcal{N}^D_{\mathcal{S}}(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$. The subscript $\mathcal{S}$ indicates that it is a model on the simplex and the superscript D indicates the number of parts of the composition. We also use the notation $\boldsymbol{\xi}$ and $\boldsymbol{\Upsilon}$ to indicate the expected value and the covariance matrix of the vector of coefficients $\mathrm{ilr}(\mathbf{x})$.

The density (10) corresponds to a density function of a normal random vector in $\mathbb{R}^{D-1}$. This is the reason why we call it the normal on $\mathcal{S}^D$ law. We want to insist that (10) is the density of the coefficients of $\mathbf{x}$ with respect to an orthonormal basis on $\mathcal{S}^D$, and therefore it is a Radon-Nikodym derivative with respect to the Lebesgue measure in $\mathbb{R}^{D-1}$, the space of coefficients. This allows us to compute the probability of any event $A \subseteq \mathcal{S}^D$ with an ordinary integral as

$$P(A) = \int_{A^*} (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}\,(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})'\,\boldsymbol{\Upsilon}^{\ -1}\,(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] d\lambda(\mathrm{ilr}(\mathbf{x})),$$

where A^* represents the coefficients of A with respect to the considered orthonomal basis, and λ is the Lebesgue measure in $\mathbb{R}^{D-1}$.

One important aspect is that the normal on $\mathcal{S}^D$ law is equivalent, on $\mathcal{S}^D$, in terms of probabilities, to the additive logistic normal law studied by Aitchison (1986) and defined using transformations from $\mathcal{S}^D$ to real space. Recall that to define the additive logistic normal model we consider $x_D = 1 - x_1 - x_2 - \cdots - x_{D-1}$ and $\mathcal{S}^D \subset \mathbb{R}^D$. Its density function is the Radon-Nikodym derivative with respect to the Lebesgue measure in $\mathbb{R}^{D-1}$ and it is obtained using the additive logratio transformation. Given relations (6) it is also possible to obtain the density function using the isometric logratio transformation. Then, we obtain the probability of any event $A \subseteq \mathcal{S}^D$ computing the standard integral

$$P(A) = \int_A \frac{D^{-1/2}\left(\prod_{i=1}^D x_i\right)^{-1}}{(2\pi)^{(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})' \boldsymbol{\Upsilon}^{-1} (\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] dx_1 dx_2 \cdots dx_{D-1}, \quad (11)$$

where now the vector $\mathrm{ilr}(\mathbf{x})$ denotes the isometric logratio transformation of composition $\mathbf{x}$ (Egozcue et al., 2003). The parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Upsilon}$ are now the standard expected value and the covariance matrix of the ilr transformed vector. Note that we use the same notation as in the normal on $\mathcal{S}^D$ case, because even though the interpretation is different, the expression of the vectors $\mathrm{ilr}(\mathbf{x})$, $\boldsymbol{\xi}$ and of the matrix $\boldsymbol{\Upsilon}$ are the same. Observe that the above density is obtained through the transformation technique because it contains the term $D^{-1/2}(\prod_{i=1}^D x_i)^{-1}$, the jacobian of the isometric logratio transformation.

Now it is important to correctly interpret the vector $\mathrm{ilr}(\mathbf{x})$ as the isometric logratio vector or as the coefficients with respect to an orthonormal basis. To avoid possible confusions we denote as $\mathbf{v} = (v_1, v_2, \ldots, v_{D-1})'$ the coefficients with respect to an orthonormal basis of composition $\mathbf{x}$, and as $\mathrm{ilr}(\mathbf{x})$ its isometric logratio transformed vector. Thus, given $\mathbf{x} \sim \mathcal{N}_{\mathcal{S}}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ we can write the probability of any event $A \subseteq \mathcal{S}^D$ as

$$P(A) = \int_{A^*} (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{v} - \boldsymbol{\xi})' \boldsymbol{\Upsilon}^{-1} (\mathbf{v} - \boldsymbol{\xi})\right] dv_1 dv_2 \cdots dv_{D-1}, \quad (12)$$

where A^* represents the coefficients of A with respect to the chosen orthonormal basis.

In both cases, expressions (11) and (12) are standard integrals of a real valued function and we can apply all the standard procedures. In particular we can apply the change of variable theorem. Then, we take the expression (12) and we apply the change $\mathbf{v} = \mathrm{ilr}(\mathbf{x})$, whose jacobian is $D^{-1/2}(\prod_{i=1}^D x_i)^{-1}$. The change of variable theorem assures the equality

$$P(A) = \int_{A^*} (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{v} - \boldsymbol{\xi})' \boldsymbol{\Upsilon}^{-1} (\mathbf{v} - \boldsymbol{\xi})\right] dv_1 dv_2 \cdots dv_{D-1}$$

$$= \int_{\mathrm{ilr}^{-1}(A^*)} \frac{D^{-1/2}\left(\prod_{i=1}^D x_i\right)^{-1}}{(2\pi)^{(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})' \boldsymbol{\Upsilon}^{-1} (\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] dx_1 dx_2 \cdots dx_{D-1}.$$

The coefficients with respect to an orthonormal basis of any element of $\mathcal{S}^D$ are equal to its isometric logratio transformation. Thus, the coefficients of A are also equal to its isometric logratio transformed event, and consequently $\mathrm{ilr}^{-1}(A^*) = A$, where ilr^{-1} denotes the inverse transformation. Observe that the probability of any event $A \subseteq \mathcal{S}^D$ is the same using both models, the normal in $\mathcal{S}^D$ and the logistic normal, and we say that the two laws are equivalent on $\mathcal{S}^D$ in terms of probabilities. But without any doubt, the normal on $\mathcal{S}^D$ and the additive logistic normal models are considerably different, specially from concepts and properties that depend on the geometry of the space.

Now we provide some general properties of the normal in $\mathcal{S}^D$ model that will help us to observe the differences from the additive logistic normal model. To be coherent with the rest of the work, we revert to using the notation $\mathrm{ilr}(\mathbf{x})$ to indicate the coefficients with respect to an orthonormal basis of $\mathcal{S}^D$ of composition $\mathbf{x}$.

Property 1 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{N}_{\mathcal{S}}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ distribution. Let be $\mathbf{a} \in \mathcal{S}^D$ a vector of constants and $b \in \mathbb{R}$ a scalar. Then, the D-part composition $\mathbf{x}^* = \mathbf{a} \oplus (b \otimes \mathbf{x})$ has a $\mathcal{N}_{\mathcal{S}}^D(\mathrm{ilr}(\mathbf{a}) + b\boldsymbol{\xi}, b^2\boldsymbol{\Upsilon})$ distribution.

Proof. The ilr coefficients of composition $\mathbf{x}^*$ are obtained from a linear transformation of the ilr coefficients of composition $\mathbf{x}$ because $\text{ilr}(\mathbf{x}^*) = \text{ilr}(\mathbf{a}) + b\text{ilr}(\mathbf{x})$. We can deal with the density function of the ilr coefficients of $\mathbf{x}$ as a density function in real space. As it is a classical normal density in real space, we use the linear transformation property to obtain the density function of the $\text{ilr}(\mathbf{x}^*)$ vector. Therefore $\mathbf{x}^* \sim \mathcal{N}_S^D(\text{ilr}(\mathbf{a}) + b\boldsymbol{\xi}, b^2\boldsymbol{\Upsilon})$.

Observe that the resulting parameters are $\text{ilr}(\mathbf{a})+b\boldsymbol{\xi}$ and $b^2\boldsymbol{\Upsilon}$. These values are the expected value and covariance matrix of the $\text{ilr}(\mathbf{x}^*)$ coefficients because

$$\begin{aligned}
\text{E}[\text{ilr}(\mathbf{x}^*)] &= \text{E}[\text{ilr}(\mathbf{a}) + b\text{ilr}(\mathbf{x})] = \text{ilr}(\mathbf{a}) + b\text{E}[\text{ilr}(\mathbf{x})] = \text{ilr}(\mathbf{a}) + b\boldsymbol{\xi},\\
\text{var}[\text{ilr}(\mathbf{x}^*)] &= \text{var}[\text{ilr}(\mathbf{a}) + b\text{ilr}(\mathbf{x})] = b^2\text{var}[\text{ilr}(\mathbf{x})] = b^2\boldsymbol{\Upsilon}.
\end{aligned}$$

Property 2 Let be a random composition $\mathbf{x} \sim \mathcal{N}_S^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ and $\mathbf{a} \in \mathcal{S}^D$ a vector of constants. Then $f^*_{\mathbf{a}\oplus\mathbf{x}}(\mathbf{a}\oplus\mathbf{x}) = f^*_{\mathbf{x}}(\mathbf{x})$, where $f^*_{\mathbf{a}\oplus\mathbf{x}}$ and $f^*_{\mathbf{x}}$ represent the density function of random compositions $\mathbf{x}$ and $\mathbf{a}\oplus\mathbf{x}$ respectively.

Proof. Using Property 1 we have that $\mathbf{a}\oplus\mathbf{x} \sim \mathcal{N}_S^D(\text{ilr}(\mathbf{a}) + \boldsymbol{\xi}, \boldsymbol{\Upsilon})$. Therefore,

$$\begin{aligned}
f^*_{\mathbf{a}\oplus\mathbf{x}}(\mathbf{a}\oplus\mathbf{x}) &= (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2}\\
&\times \exp\left[-\frac{1}{2}\left(\text{ilr}(\mathbf{a}\oplus\mathbf{x}) - (\text{ilr}(\mathbf{a}) + \boldsymbol{\xi})\right)' \boldsymbol{\Upsilon}^{-1}\left(\text{ilr}(\mathbf{a}\oplus\mathbf{x}) - (\text{ilr}(\mathbf{a}) + \boldsymbol{\xi})\right)\right]\\
&= (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2}\\
&\times \exp\left[-\frac{1}{2}\left(\text{ilr}(\mathbf{a}) + \text{ilr}(\mathbf{x}) - (\text{ilr}(\mathbf{a}) + \boldsymbol{\xi})\right)' \boldsymbol{\Upsilon}^{-1}\left(\text{ilr}(\mathbf{a}) + \text{ilr}(\mathbf{x}) - (\text{ilr}(\mathbf{a}) + \boldsymbol{\xi})\right)\right]\\
&= (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}\left(\text{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1}\left(\text{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)\right] = f^*_{\mathbf{x}}(\mathbf{x}),
\end{aligned}$$

as indicated in (9).

Note that Property 2 is not hold true for the additive logistic normal distribution.

Property 3 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{N}_S^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ distribution. Let be $\mathbf{x}_P = \mathbf{P}\mathbf{x}$ the composition $\mathbf{x}$ with the parts reordered by a permutation matrix $\mathbf{P}$. Then $\mathbf{x}_P$ has a $\mathcal{N}_S^D(\boldsymbol{\xi}_P, \boldsymbol{\Upsilon}_P)$ distribution with

$$\boldsymbol{\xi}_P = \mathbf{U}'\mathbf{P}\mathbf{U}\boldsymbol{\xi} \qquad \text{and} \qquad \boldsymbol{\Upsilon}_P = (\mathbf{U}'\mathbf{P}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}'\mathbf{P}\mathbf{U})',$$

where $\mathbf{U}$ is a $D\times(D-1)$ matrix with vectors $\text{clr}(\mathbf{e}_i)$ $(i = 1, 2, \ldots, D-1)$ as columns.

Proof. To obtain the distribution of a random composition $\mathbf{x}_P$ in terms of the distribution of $\mathbf{x}$, it is necessary to find a matrix relationship between the ilr coefficients of both compositions $\mathbf{x}_P$ and $\mathbf{x}$. If we work with the clr coefficients, we have $\text{clr}(\mathbf{x}_P) = \mathbf{P}\text{clr}(\mathbf{x})$. Applying (6) we obtain $\text{ilr}(\mathbf{x}_P) = (\mathbf{U}'\mathbf{P}\mathbf{U})\,\text{ilr}(\mathbf{x})$. As the $\text{ilr}(\mathbf{x})$ vector has a normal distribution, we can apply the change of variable theorem or the linear transformation property of the normal distribution in real space to obtain a $\mathcal{N}_S^D(\mathbf{U}'\mathbf{P}\mathbf{U}\boldsymbol{\xi}, (\mathbf{U}'\mathbf{P}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}'\mathbf{P}\mathbf{U})')$ distribution for the random composition $\mathbf{x}_P$.

Note that the parameters of the model agree with the expected value and the covariance matrix of the $\text{ilr}(\mathbf{x}_P)$ vector.

Property 4 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{N}_S^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ distribution. Let be $\mathbf{s} = \mathcal{C}(\mathbf{S}\mathbf{x})$ a C-part subcomposition obtained from the $C\times D$ selection matrix $\mathbf{S}$. Then $\mathbf{s}$ has a $\mathcal{N}_S^C(\boldsymbol{\xi}_S, \boldsymbol{\Upsilon}_S)$ distribution, with

$$\boldsymbol{\xi}_S = \mathbf{U}^{*\prime}\mathbf{S}\mathbf{U}\boldsymbol{\xi} \qquad \text{and} \qquad \boldsymbol{\Upsilon}_S = (\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U})',$$

where $\mathbf{U}$ is a $D \times (D-1)$ matrix with the clr coefficients of an orthonormal basis of $\mathcal{S}^D$ as columns, and $\mathbf{U}^*$ is a $C \times (C-1)$ matrix with the clr coefficients of an orthonormal basis of $\mathcal{S}^C$ as columns.

Proof. We know that the coefficients $\text{alr}(\mathbf{s})$ and $\text{alr}(\mathbf{x})$ are equal to the respective alr transformed vectors. Aitchison (1986, p. 119) proves that $\text{alr}(\mathbf{s}) = \left(\mathbf{F}_{C-1,C}\mathbf{S}\mathbf{F}^*_{D,D-1}\right)\text{alr}(\mathbf{x})$. Applying (6) we obtain $\text{ilr}(\mathbf{s}) = \left(\mathbf{U}^{*\prime}\mathbf{F}^*_{C,C-1}\mathbf{F}_{C-1,C}\mathbf{S}\mathbf{F}^*_{D,D-1}\mathbf{F}_{D-1,D}\mathbf{U}\right)\text{ilr}(\mathbf{x})$. We can easily check that matrices $\mathbf{F}^*_{C,C-1}\mathbf{F}_{C-1,C}$ and $\mathbf{F}^*_{D,D-1}\mathbf{F}_{D-1,D}$ have the columns of matrices $\mathbf{U}^*$ and $\mathbf{U}$, respectively, as eigenvectors with eigenvalue 1. Consequently we have $\mathbf{U}^{*\prime}\left(\mathbf{F}^*_{C,C-1}\mathbf{F}_{C-1,C}\right) = \mathbf{U}^{*\prime}$ and $\left(\mathbf{F}^*_{D,D-1}\mathbf{F}_{D-1,D}\right)\mathbf{U} = \mathbf{U}$, and the relationship between the ilr coefficients of subcomposition $\mathbf{s}$ and composition $\mathbf{x}$ is $\text{ilr}(\mathbf{s}) = \left(\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U}\right)\text{ilr}(\mathbf{x})$. Given the density of the $\text{ilr}(\mathbf{x})$ vector and applying the change of variable theorem or the linear transformation property of the normal distribution in real space, we obtain the density of the $\text{ilr}(\mathbf{s})$ vector, that corresponds to a $\mathcal{N}^C_S\left(\mathbf{U}^{*\prime}\mathbf{P}\mathbf{U}\boldsymbol{\xi}, (\mathbf{U}^{*\prime}\mathbf{P}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}^{*\prime}\mathbf{P}\mathbf{U})'\right)$ density function.

Observe that properties of the classical normal distribution in real space have allowed us to prove the closeness under perturbation, power transformation, permutation and subcompositions of the normal on $\mathcal{S}^D$ family. Nevertheless, given $\mathbf{x} \sim \mathcal{N}^D_S(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$, it has not seem possible up to now to describe the distribution of any amalgamation in terms of the distribution $\mathbf{x}$. In particular, we have been unable to find a matrix relationship between the ilr coefficients of $\mathbf{x}$ and the corresponding amalgamated composition.

Following the methodology stated by Pawlowsky-Glahn, Egozcue, and Burger (2003), we can define the expected value of a normal on $\mathcal{S}^D$ distributed random composition:

Property 5 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{N}^D_S(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ distribution and let be $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{D-1})'$. Then $\text{E}[\mathbf{x}] = (\xi_1 \otimes \mathbf{e}_1) \oplus (\xi_2 \otimes \mathbf{e}_2) \oplus \ldots \oplus (\xi_{D-1} \otimes \mathbf{e}_{D-1})$, where $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ is an orthonormal basis of $\mathcal{S}^D$.

Proof. The expected value of any random vector is an element of the support space. If we apply the standard definition of the expected value to the coefficients of composition $\mathbf{x}$ with respect to the orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ using density (10), we obtain the coefficients of composition $\text{E}[\mathbf{x}]$ with respect to the considered orthonormal basis. Applying standard integration methods we have $\text{ilr}(\text{E}[\mathbf{x}]) = \boldsymbol{\xi}$. Finally, we obtain the composition $\text{E}[\mathbf{x}]$ through the linear combination $(\xi_1 \otimes \mathbf{e}_1) \oplus (\xi_2 \otimes \mathbf{e}_2) \oplus \ldots \oplus (\xi_{D-1} \otimes \mathbf{e}_{D-1})$.

Recall that the vector $\boldsymbol{\xi}$ denotes the expected value of the $\text{ilr}(\mathbf{x})$ vector. Then we have that $\text{ilr}(\text{E}[\mathbf{x}]) = \text{E}[\text{ilr}(\mathbf{x})]$. Also, equality (8) says that the vector of coefficients with respect to the orthonormal basis of composition $\text{cen}(\mathbf{x})$ is $\text{E}[\text{ilr}(\mathbf{x})]$. Consequently we have that $\text{E}[\mathbf{x}] = \text{cen}[\mathbf{x}]$.

In the additive logistic normal case, we can compute the expected value using the standard procedure. But, as Aitchison (1986) adverts, the integral expressions are not reducible to any simple form and it is necessari to apply Hermitian integration to obtain numerical results. But, certainly, the expected value is not equal to the composition $\text{cen}(\mathbf{x})$ obtained in the normal in $\mathcal{S}^D$ case.

Property 6 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{N}^D_S(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ distribution. Then a dispersion measure around the expected value is $\text{Mvar}[\mathbf{x}] = \text{trace}(\boldsymbol{\Upsilon})$.

Proof. The metric variance is defined as $\text{Mvar}[\mathbf{x}] = \text{E}[d^2_a(\mathbf{x}, \text{cen}[\mathbf{x}])]$. Given $\mathbf{x} \sim \mathcal{N}^D_S(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ we know from property 5 that $\text{cen}[\mathbf{x}] = \text{E}[\mathbf{x}]$. Then, the metric variance is a dispersion measure around the expected value $\text{Mvar}[\mathbf{x}] = \text{E}[d^2_a(\mathbf{x}, \text{E}[\mathbf{x}])]$. The distance d_a between two elements is equal to the Euclidean distance d_{eu} between the corresponding coefficients with respect to an orthonormal basis. Then, we can write $\text{Mvar}[\mathbf{x}] = \text{E}[d^2_{eu}(\text{ilr}(\mathbf{x}), \text{E}[\text{ilr}(\mathbf{x})])]$. This value corresponds to the trace of matrix $\text{var}(\text{ilr}(\mathbf{x}))$. Finally, using the covariance matrix of a normal distribution in real space we obtain $\text{Mvar}[\mathbf{x}] = \text{trace}(\boldsymbol{\Upsilon})$.

As Aitchison (1986) averts, we cannot interpret the crude covariance or correlations. Therefore, we always compute the covariances and correlations of the coefficients with respect to the orthonormal basis. In the case of the normal on $\mathcal{S}^D$ model, the covariances between components are equal to the off-diagonal elements of matrix $\boldsymbol{\Upsilon}$.

3.2 Inferential aspects

Given a compositional data set $\mathbf{X}$, the estimates of parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Upsilon}$ can be calculated through the sample mean and sample covariance matrix of the ilr coefficients of the data set:

$$\hat{\boldsymbol{\xi}} = \overline{\mathrm{ilr}(\mathbf{X})} \qquad \widehat{\boldsymbol{\Upsilon}} = \mathrm{var}(\mathrm{ilr}(\mathbf{X})).$$

With these values we can obtain the estimates of $\mathrm{E}[\mathbf{x}]$ and $\mathrm{Mvar}[\mathbf{x}]$ as

$$\widehat{\mathrm{E}[\mathbf{x}]} = (\hat{\xi}_1 \otimes \mathbf{e}_1) \oplus (\hat{\xi}_2 \otimes \mathbf{e}_2) \oplus \cdots \oplus (\hat{\xi}_{D-1} \otimes \mathbf{e}_{D-1}),$$
$$\widehat{\mathrm{Mvar}[\mathbf{X}]} = \mathrm{trace}(\widehat{\boldsymbol{\Upsilon}}).$$

The estimators $\widehat{\mathrm{E}[\mathbf{x}]}$ and $\widehat{\mathrm{Mvar}[\mathbf{X}]}$ are consistent and minimize the variance. This can be proved using properties of estimators for the normal law in real space.

To validate the normal on $\mathcal{S}^D$ law, we have only to apply a goodness-of-fit test for the multivariate normal distribution to the coefficients with respect to an orthonormal basis of sample $\mathbf{X}$. Unfortunately the most common tests of normality as the Anderson-Darling or the Kolmogorov-Smirnov tests are dependent on the orthonormal basis chosen. But in this particular case, we can reproduce the singular value decomposition and a power-perturbation characterisation of compositional variability of the random composition as proposed by Aitchison et al. (2003).

3.3 Another parametrization

Aitchison (1986) introduces the additive logistic normal distribution using the additive logratio transformation, but equivalent parametrizations using the centred logratio transformation (Aitchison, 1986, p. 116) or the isometric logratio transformation (Mateu Figueras, 2003, p. 78) could be obtained. The normal on $\mathcal{S}^D$ law is defined using the ilr coefficients. Given the similarity among the alr, clr and ilr coefficients and the additive logratio, centred logratio and isometric logratio vectors, it is natural to ask about the expression of the density function in terms of the alr or clr coefficients. To avoid large expressions we will use $\mathbf{v}$ to denote the coefficients of composition $\mathbf{x}$ with respect to an orthonormal basis and $\mathbf{y}$ to denote the coefficients of $\mathbf{x}$ with respect to the B basis introduced earlier in section 1.

Given a $\mathcal{N}_{\mathcal{S}}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon})$ law, the probability of any event $A \subseteq \mathcal{S}^D$ is

$$P(A) = \int_{A^*} (2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{v}-\boldsymbol{\xi})' \boldsymbol{\Upsilon}^{-1} (\mathbf{v}-\boldsymbol{\xi})\right] d\mathbf{v},$$

where A^* denotes the coefficients of A with respect to the orthonormal basis. Matrix $\mathbf{FU}$ is a change of basis matrix from the orthonormal basis to the B basis. Then, the transformation $\mathbf{v} = (\mathbf{FU})^{-1}\mathbf{y}$ makes effective this change of basis and we obtain

$$\begin{aligned} P(A) &= \int_{\mathbf{FU}A^*} \frac{(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[-\frac{1}{2}\left((\mathbf{FU})^{-1}\mathbf{y}-\boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1} \left((\mathbf{FU})^{-1}\mathbf{y}-\boldsymbol{\xi}\right)\right] \frac{1}{|\mathbf{FU}|} d\mathbf{y} \\ &= \int_{\mathbf{FU}A^*} \frac{(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Upsilon} \mid^{1/2} |\mathbf{FU}|} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{FU}\boldsymbol{\xi})' \left((\mathbf{FU})^{-1}\right)' \boldsymbol{\Upsilon}^{-1} (\mathbf{FU})^{-1} (\mathbf{y}-\mathbf{FU}\boldsymbol{\xi})\right] d\mathbf{y} \\ &= \int_{\mathbf{FU}A^*} \frac{(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})\right] d\mathbf{y}, \end{aligned}$$

where $\boldsymbol{\mu} = \mathbf{FU}\boldsymbol{\xi}$ and $\boldsymbol{\Sigma} = \mathbf{FU}\boldsymbol{\Upsilon}(\mathbf{FU})'$. Using (6) it is easy to see that $\boldsymbol{\mu} = \mathrm{E}[\mathbf{y}]$, $\boldsymbol{\Sigma} = \mathrm{var}[\mathbf{y}]$ and $\mid \boldsymbol{\Upsilon} \mid^{1/2} |\mathbf{FU}| = \mid \boldsymbol{\Sigma} \mid^{1/2}$. Observe that $\mathbf{FU}A^*$ represents the coefficients of the event A with respect to the B basis.

In conclusion, the normal on $\mathcal{S}^D$ law using $\mathrm{alr}(\mathbf{x})$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ parametrization can be obtained. But recall that the basis B is not orthonormal, and therefore the Euclidean distance between two alr coefficients is not equal to the Aitchison distance between the corresponding compositions. Thus, it will have no sense to use the density of the alr coefficients in the procedures that use distances or scalar products.

In a similar way, we could work with the density of the clr coefficients. Even though the Euclidean distance between two clr coefficients is equal to the Aitchison distance between the corresponding compositions, we have an additional difficulty: the density function of $\mathrm{clr}(\mathbf{x})$ coefficients is degenerate.

4 The skew-normal distribution on $\mathcal{S}^D$

4.1 Definition and properties

Definition 2 Let be $(\Omega, \mathcal{F}, p)$ a probability space. A random composition $\mathbf{x} : \Omega \longrightarrow \mathcal{S}^D$ is said to have a skew-normal on $\mathcal{S}^D$ distribution with parameters $\boldsymbol{\xi}$, $\boldsymbol{\Upsilon}$ and $\boldsymbol{\varrho}$ (APPENDIX), if the density function of the coefficients with respect to an orthonormal basis of $\mathcal{S}^D$ is

$$
\begin{aligned}
f^*_{\mathbf{x}}(\mathbf{x}) = {} & 2(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1} \left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)\right] \\
& \times \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right],
\end{aligned} \tag{13}
$$

where Φ is the $\mathcal{N}(0,1)$ distribution function and $\boldsymbol{\upsilon}$ is the square root of $\mathrm{diag}(\boldsymbol{\Upsilon})$, where $\mathrm{diag}(\boldsymbol{\Upsilon})$ stands for the matrix obtained putting to zero all the off-diagonal elements of $\boldsymbol{\Upsilon}$.

We use the notation $\mathbf{x} \sim \mathcal{SN}^D_{\mathcal{S}}(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$. The subscript $\mathcal{S}$ indicates that it is a model on the simplex and the superscript D indicates the number of parts of the composition. As in the normal on $\mathcal{S}^D$ case, we use the notation $\boldsymbol{\xi}$ and $\boldsymbol{\Upsilon}$ to represent the parameters of the model, but in this case neither $\boldsymbol{\xi}$ nor $\boldsymbol{\Upsilon}$ denote the expected value and covariance matrix of the $\mathrm{ilr}(\mathbf{x})$ vector.

The density (13) corresponds to a density function of a skew-normal random vector in $\mathbb{R}^{D-1}$. This is the reason why we call it the skew-normal on $\mathcal{S}^D$ law. We want to insist that (13) is the density of the coefficients of $\mathbf{x}$ with respect to an orthonormal basis on $\mathcal{S}^D$, and therefore it is a Radon-Nikodym derivative with respect to the Lebesgue measure in $\mathbb{R}^{D-1}$, the coefficients space. This allows us to compute the probability of any event $A \subseteq \mathcal{S}^D$ with an ordinary integral as

$$
P(A) = \int_{A^*} 2(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp[\mathbf{M}]\, \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] d\lambda(\mathrm{ilr}(\mathbf{x})),
$$

with

$$
\mathbf{M} = -\frac{1}{2}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1} \left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right),
$$

where A^* represents the coefficients of the event A with respect to the considered orthonomal basis, and λ is the Lebesgue measure in $\mathbb{R}^{D-1}$.

Like for the normal on $\mathcal{S}^D$ law, which is equivalent, on $\mathcal{S}^D$, in terms of probabilities, to the additive logistic normal law, we can prove that the skew-normal on $\mathcal{S}^D$ law is equivalent, on $\mathcal{S}^D$ and in terms of probabilities, to the additive logistic skew-normal law studied in Mateu Figueras (2003). To define the additive logistic skew-normal distribution we consider $x_D = 1 - x_1 - x_2 - \cdots - x_{D-1}$ and $\mathcal{S}^D \subset \mathbb{R}^D$. Its density function is the Radon-Nikodym derivative of the probability with respect to the Lebesgue measure in $\mathbb{R}^{D-1}$, and it is obtained using the additive logratio transformation. In Mateu Figueras (2003, p. 96) the density function using the isometric logratio transformation is provided, and consequently the probability of any event $A \subseteq \mathcal{S}^D$ is

$$P(A) = \int_A \frac{2D^{-1/2}\left(\prod_{i=1}^D x_i\right)^{-1}}{(2\pi)^{(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[\mathbf{M}\right] \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] dx_1 dx_2 \cdots dx_{D-1}, \tag{14}$$

with

$$\mathbf{M} = -\frac{1}{2}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right),$$

where now the vector $\mathrm{ilr}(\mathbf{x})$ denotes the isometric logratio transformation of $\mathbf{x}$. Observe that the above density function is obtained through the transformation technique because it contains the term $D^{-1/2}(\prod_{i=1}^D x_i)^{-1}$, the jacobian of the isometric logratio transformation. We also use the notation $\boldsymbol{\xi}$, $\boldsymbol{\Upsilon}$ and $\boldsymbol{\varrho}$ for the parameters of the model, but in this case neither $\boldsymbol{\xi}$ nor $\boldsymbol{\Upsilon}$ correspond to the expected value and the covariance matrix of the isometric logratio transformed vector.

Now it is important to correctly interpret the $\mathrm{ilr}(\mathbf{x})$ vector as the isometric logratio vector or as the vector of coefficients with respect to an orthonormal basis, both denoted as $\mathrm{ilr}(\mathbf{x})$. Therefore, to avoid possible confusions, we denote as $\mathbf{v} = (v_1, v_2, \ldots, v_{D-1})'$ the coefficients with respect to an orthonormal basis of composition $\mathbf{x}$, and as $\mathrm{ilr}(\mathbf{x})$ its isometric logratio transformed vector. Thus, given $\mathbf{x} \sim \mathcal{SN}_{\mathcal{S}}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$, we can write the probability of any event $A \subseteq \mathcal{S}^D$ as

$$P(A) = \int_{A^*} 2(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[\mathbf{M}^*\right] \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathbf{v} - \boldsymbol{\xi})\right] dv_1 dv_2 \cdots dv_{D-1}, \tag{15}$$

with

$$\mathbf{M}^* = -\frac{1}{2}\left(\mathbf{v} - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1}\left(\mathbf{v} - \boldsymbol{\xi}\right),$$

where A^* represents the coefficients of A with respect to the chosen orthonormal basis.

In both cases, the expressions (14) and (15) are standard integrals of a real valued function and we can apply all the standard procedures. In particular we can apply the change of variable theorem in expression (15) and, taking $\mathbf{v} = \mathrm{ilr}(\mathbf{x})$, whose jacobian is $D^{-1/2}(\prod_{i=1}^D x_i)^{-1}$, we obtain the equality

$$\begin{aligned} P(A) &= \int_{A^*} 2(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[\mathbf{M}^*\right] \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathbf{v} - \boldsymbol{\xi})\right] dv_1 dv_2 \cdots dv_{D-1} \\ &= \int_{\mathrm{ilr}^{-1}(A^*)} \frac{2D^{-1/2}\left(\prod_{i=1}^D x_i\right)^{-1}}{(2\pi)^{(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[\mathbf{M}\right] \Phi\left[\boldsymbol{\varrho}' \boldsymbol{\upsilon}^{-1}(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi})\right] dx_1 dx_2 \cdots dx_{D-1}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{M}^* &= -\frac{1}{2}\left(\mathbf{v} - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1}\left(\mathbf{v} - \boldsymbol{\xi}\right), \\ \mathbf{M} &= -\frac{1}{2}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right)' \boldsymbol{\Upsilon}^{-1}\left(\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\xi}\right). \end{aligned}$$

The second term of this equality agrees with (14) because $\mathrm{ilr}^{-1}(A^*) = A$. Observe that the probability of any event $A \subseteq \mathcal{S}^D$ is the same using both models. In these cases we say that both models are equivalent on $\mathcal{S}^D$ in terms of probabilities. But the skew-normal on $\mathcal{S}^D$ and the additive logistic skew-normal models present essential differences. The two density functions will differ and some properties that depend on the space structure also differ.

Next, we study the principal properties of the skew-normal on $\mathcal{S}^D$ model. To be coherent with the rest of our work, we revert to using the notation $\mathrm{ilr}(\mathbf{x})$ to indicate the coefficients with respect to an orthonormal basis of $\mathcal{S}^D$ of composition $\mathbf{x}$.

Property 7 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{SN}_{S}^{D}(\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ distribution. Let be $\mathbf{a}\in\mathcal{S}^{D}$ a vector of constants and $b\in\mathbb{R}$ a scalar. Then, the D-part composition $\mathbf{x}^{*}=\mathbf{a}\oplus(b\otimes\mathbf{x})$ has a $\mathcal{SN}_{S}^{D}(\mathrm{ilr}(\mathbf{a})+b\boldsymbol{\xi},b^{2}\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ distribution.

Proof. The ilr coefficients of composition $\mathbf{x}^{*}$ are obtained from a linear transformation of the ilr coefficients of composition $\mathbf{x}$ because $\mathrm{ilr}(\mathbf{x}^{*})=\mathrm{ilr}(\mathbf{a})+b\mathrm{ilr}(\mathbf{x})$. We deal with the density function of the $\mathrm{ilr}(\mathbf{x})$ coefficients as a density function in real space. As this density is a classical skew-normal density, we use the linear transformation property (APPENDIX) to obtain the density function of the $\mathrm{ilr}(\mathbf{x}^{*})$ coefficients. Therefore $\mathbf{x}^{*}\sim\mathcal{SN}_{S}^{D}(\mathrm{ilr}(\mathbf{a})+b\boldsymbol{\xi},b^{2}\boldsymbol{\Upsilon},\boldsymbol{\varrho})$.

Property 8 Let be a random composition $\mathbf{x}\sim\mathcal{SN}_{S}^{D}(\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ and $\mathbf{a}\in\mathcal{S}^{D}$ a vector of constants. Then $f_{\mathbf{a}\oplus\mathbf{x}}^{*}(\mathbf{a}\oplus\mathbf{x})=f_{\mathbf{x}}^{*}(\mathbf{x})$, where $f_{\mathbf{a}\oplus\mathbf{x}}^{*}$ and $f_{\mathbf{x}}^{*}$ represent the density functions of random compositions $\mathbf{x}$ and $\mathbf{a}\oplus\mathbf{x}$ respectively.

Proof. Using Property 7 we have that $\mathbf{a}\oplus\mathbf{x}\sim\mathcal{SN}_{S}^{D}(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$. Therefore,

$$\begin{aligned}
f_{\mathbf{a}\oplus\mathbf{x}}^{*}(\mathbf{a}\oplus\mathbf{x}) &= 2(2\pi)^{-(D-1)/2}\mid\boldsymbol{\Upsilon}\mid^{-1/2}\\
&\quad\times\exp\left[-\frac{1}{2}\left(\mathrm{ilr}(\mathbf{a}\oplus\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi})\right)'\boldsymbol{\Upsilon}^{-1}\left(\mathrm{ilr}(\mathbf{a}\oplus\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi})\right)\right]\\
&\quad\times\Phi\left[\boldsymbol{\varrho}'\boldsymbol{v}^{-1}(\mathrm{ilr}(\mathbf{a}\oplus\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi}))\right]\\
&= 2(2\pi)^{-(D-1)/2}\mid\boldsymbol{\Upsilon}\mid^{-1/2}\\
&\quad\times\exp\left[-\frac{1}{2}\left(\mathrm{ilr}(\mathbf{a})+\mathrm{ilr}(\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi})\right)'\boldsymbol{\Upsilon}^{-1}\left(\mathrm{ilr}(\mathbf{a})+\mathrm{ilr}(\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi})\right)\right]\\
&\quad\times\Phi\left[\boldsymbol{\varrho}'\boldsymbol{v}^{-1}(\mathrm{ilr}(\mathbf{a})+\mathrm{ilr}(\mathbf{x})-(\mathrm{ilr}(\mathbf{a})+\boldsymbol{\xi}))\right]\\
&= 2(2\pi)^{-(D-1)/2}\mid\boldsymbol{\Upsilon}\mid^{-1/2}\exp\left[-\frac{1}{2}\left(\mathrm{ilr}(\mathbf{x})-\boldsymbol{\xi}\right)'\boldsymbol{\Upsilon}^{-1}\left(\mathrm{ilr}(\mathbf{x})-\boldsymbol{\xi}\right)\right]\\
&\quad\times\Phi\left[\boldsymbol{\varrho}'\boldsymbol{v}^{-1}(\mathrm{ilr}(\mathbf{x})-\boldsymbol{\xi})\right]=f_{\mathbf{x}}^{*}(\mathbf{x}),
\end{aligned}$$

as was indicated in (9).

Note that Property 8 is not hold true for the additive logistic skew-normal distribution.

Property 9 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{SN}_{S}^{D}(\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ distribution. Let be $\mathbf{x}_{P}=\mathbf{P}\mathbf{x}$ the composition $\mathbf{x}$ with the parts reordered by a permutation matrix $\mathbf{P}$. Then $\mathbf{x}_{P}$ has a $\mathcal{SN}_{S}^{D}(\boldsymbol{\xi}_{P},\boldsymbol{\Upsilon}_{P},\boldsymbol{\varrho}_{P})$ distribution with

$$\boldsymbol{\xi}_{P}=\mathbf{U}'\mathbf{P}\mathbf{U}\boldsymbol{\xi},\quad\boldsymbol{\Upsilon}_{P}=(\mathbf{U}'\mathbf{P}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}'\mathbf{P}\mathbf{U})',\quad\boldsymbol{\varrho}_{P}=\frac{\boldsymbol{v}_{P}\boldsymbol{\Upsilon}_{P}^{-1}\mathbf{B}'\boldsymbol{\varrho}}{\sqrt{1+\boldsymbol{\varrho}'(\boldsymbol{v}^{-1}\boldsymbol{\Upsilon}\boldsymbol{v}^{-1}-\mathbf{B}\boldsymbol{\Upsilon}_{P}^{-1}\mathbf{B}')\boldsymbol{\varrho}}},$$

where $\mathbf{U}$ is a $D\times(D-1)$ matrix with vectors $\mathrm{clr}(\mathbf{e}_{i})$ $(i=1,2,\ldots,D-1)$ as columns, $\mathbf{B}=\boldsymbol{v}^{-1}\boldsymbol{\Upsilon}(\mathbf{U}'\mathbf{P}'\mathbf{U})$, and $\boldsymbol{v}$ and $\boldsymbol{v}_{P}$ are the square roots of $\mathrm{diag}(\boldsymbol{\Upsilon})$ and $\mathrm{diag}(\boldsymbol{\Upsilon}_{P})$, respectively.

Proof. In property 3 we have seen that $\mathrm{ilr}(\mathbf{x}_{P})=(\mathbf{U}'\mathbf{P}\mathbf{U})\,\mathrm{ilr}(\mathbf{x})$. Applying the change of variable theorem or the linear transformation property of the skew-normal distribution in real space, we obtain a $\mathcal{SN}_{S}^{D}(\boldsymbol{\xi}_{P},\boldsymbol{\Upsilon}_{P},\boldsymbol{\varrho}_{P})$ distribution for the random composition $\mathbf{x}_{P}$.

Property 10 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{SN}_{S}^{D}(\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ distribution. Let be $\mathbf{s}=\mathcal{C}(\mathbf{S}\mathbf{x})$ a C-part subcomposition obtained from the $C\times D$ selection matrix $\mathbf{S}$. Then $\mathbf{s}$ has a $\mathcal{SN}_{S}^{C}(\boldsymbol{\xi}_{S},\boldsymbol{\Upsilon}_{S},\boldsymbol{\varrho}_{S})$ distribution, with

$$\boldsymbol{\xi}_{S}=\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U}\boldsymbol{\xi},\quad\boldsymbol{\Upsilon}_{S}=(\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U})\boldsymbol{\Upsilon}(\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U})',\quad\boldsymbol{\varrho}_{S}=\frac{\boldsymbol{v}_{S}\boldsymbol{\Upsilon}_{S}^{-1}\mathbf{B}'\boldsymbol{\varrho}}{\sqrt{1+\boldsymbol{\varrho}'(\boldsymbol{v}^{-1}\boldsymbol{\Upsilon}\boldsymbol{v}^{-1}-\mathbf{B}\boldsymbol{\Upsilon}_{S}^{-1}\mathbf{B}')\boldsymbol{\varrho}}},$$

where $\mathbf{U}$ is a $D \times (D-1)$ matrix with the clr coefficients of an orthonormal basis of $\mathcal{S}^D$ as columns, $\mathbf{U}^*$ is a $C \times (C-1)$ matrix with the clr coefficients of an orthonormal basis of $\mathcal{S}^C$ as columns, $\mathbf{B} = \boldsymbol{\upsilon}^{-1}\boldsymbol{\Upsilon}(\mathbf{U}'\mathbf{S}'\mathbf{U}^*)$, and $\boldsymbol{\upsilon}$ and $\boldsymbol{\upsilon}_S$ are the square roots of $\mathrm{diag}(\boldsymbol{\Upsilon})$ and $\mathrm{diag}(\boldsymbol{\Upsilon}_S)$, respectively.

Proof. In property 4 we have seen that $\mathrm{ilr}(\mathbf{s}) = \left(\mathbf{U}^{*\prime}\mathbf{S}\mathbf{U}\right)\mathrm{ilr}(\mathbf{x})$. Given the density of the $\mathrm{ilr}(\mathbf{x})$ coefficients and applying the change of variable theorem or the linear transformation property of the skew-normal distribution in real space, we obtain the density of the $\mathrm{ilr}(\mathbf{s})$ coefficients that corresponds to a $\mathcal{SN}_{\mathcal{S}}^{C}(\boldsymbol{\xi}_S, \boldsymbol{\Upsilon}_S, \boldsymbol{\varrho}_S)$ density function.

We have seen that the skew-normal on $\mathcal{S}^D$ family is closed under perturbation, power transformation, permutation and subcompositions. Again, given $\mathbf{x} \sim \mathcal{SN}_{\mathcal{S}}^{D}(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$, it has not been possible up to now to describe the distribution of any amalgamation in terms of the distribution of $\mathbf{x}$ because we have not a matricial relationship between both compositions.

We can also compute the expected value of a skew-normal in $\mathcal{S}^D$ distributed random composition.

Property 11 Let be $\mathbf{x}$ a D-part composition with a $\mathcal{SN}_{\mathcal{S}}^{D}(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$ distribution. Then, $\mathrm{E}[\mathbf{x}] = (\beta_1 \otimes \mathbf{e}_1) \oplus (\beta_2 \otimes \mathbf{e}_2) \oplus \ldots \oplus (\beta_{D-1} \otimes \mathbf{e}_{D-1})$, with $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ an orthonormal basis of $\mathcal{S}^D$ and $\boldsymbol{\beta} = \boldsymbol{\xi} + \boldsymbol{\upsilon}\boldsymbol{\delta}\sqrt{2/\pi}$, where $\boldsymbol{\delta}$ is a parameter related with $\boldsymbol{\varrho}$ following the equality (17), and $\boldsymbol{\upsilon}$ is the square root of $\mathrm{diag}(\boldsymbol{\Upsilon})$.

Proof. The expected value of any random composition is an element of the support space. From the coefficients of $\mathbf{x}$ with respect to the orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ and from the density (13), we obtain the coefficients of the vector $\mathrm{E}[\mathbf{x}]$ with respect to the same orthonormal basis. Using the expected value of the skew-normal distribution in real space we have that $\mathrm{E}[\mathrm{ilr}(\mathbf{x})] = \boldsymbol{\xi} + \boldsymbol{\upsilon}\boldsymbol{\delta}\sqrt{2/\pi}$ denoted as $\boldsymbol{\beta}$. Finally, we obtain composition $\mathrm{E}[\mathbf{x}]$ through the linear combination $(\beta_1 \otimes \mathbf{e}_1) \oplus (\beta_2 \otimes \mathbf{e}_2) \oplus \ldots \oplus (\beta_{D-1} \otimes \mathbf{e}_{D-1})$.

In this case we have obtained $\mathrm{E}[\mathrm{ilr}(\mathbf{x})] = \boldsymbol{\beta}$. Using (8) we also conclude that $\mathrm{ilr}(\mathrm{cen}[\mathbf{x}]) = \boldsymbol{\beta}$ and consequently we have $\mathrm{cen}[\mathbf{x}] = \mathrm{E}[\mathbf{x}] = (\beta_1 \otimes \mathbf{e}_1) \oplus (\beta_2 \otimes \mathbf{e}_2) \oplus \ldots \oplus (\beta_{D-1} \otimes \mathbf{e}_{D-1})$. This is an essential difference between the skew-normal on $\mathcal{S}^D$ law and the additive logistic skew-normal law. As it is observed in Mateu Figueras (2003), we have not the equality between $\mathrm{cen}[\mathbf{x}]$ and the expected value of an additive logistic skew-normal model.

Property 12 Let be $\mathbf{x}$ a D-part random composition with a $\mathcal{SN}_{\mathcal{S}}^{D}(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$ distribution. A dispersion measure around the expected value is $\mathrm{Mvar}[\mathbf{x}] = \mathrm{trace}\left(\boldsymbol{\Upsilon} - (2/\pi)\boldsymbol{\upsilon}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\upsilon}\right)$, where $\boldsymbol{\upsilon}$ is the square root of $\mathrm{diag}(\boldsymbol{\Upsilon})$, and $\boldsymbol{\delta}$ is the parameter related with $\boldsymbol{\varrho}$ following the equality (17).

Proof. The metric variance is defined as $\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_a^2(\mathbf{x}, \mathrm{cen}[\mathbf{x}])]$. We know from Property 11 that $\mathrm{cen}[\mathbf{x}] = \mathrm{E}[\mathbf{x}]$. Then the metric variance is a dispersion measure around the expected value $\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_a^2(\mathbf{x}, \mathrm{E}[\mathbf{x}])]$. The distance d_a between two compositions is equal to the Euclidean distance d_{eu} between the corresponding ilr coefficients. Therefore, we can write $\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_{eu}^2(\mathrm{ilr}(\mathbf{x}), \mathrm{E}[\mathrm{ilr}(\mathbf{x})])]$. This value corresponds to the trace of matrix $\mathrm{var}(\mathrm{ilr}(\mathbf{x}))$. Finally, using the covariance matrix of a skew-normal distribution in real space we obtain $\mathrm{Mvar}[\mathbf{x}] = \mathrm{trace}(\boldsymbol{\Upsilon} - (2/\pi)\boldsymbol{\upsilon}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\upsilon})$.

4.2 Inferential aspects

Given a compositional data set $\mathbf{X}$, the estimates of parameters $\boldsymbol{\xi}$, $\boldsymbol{\Upsilon}$ and $\boldsymbol{\varrho}$ can be calculated applying the maximum likelihood procedure to the ilr coefficients of the data set. These estimates cannot be expressed in analytic terms and we have to use numerical methods to compute an approximation from the sample.

The estimated values $\hat{\boldsymbol{\xi}}, \widehat{\boldsymbol{\Upsilon}}$ and $\hat{\boldsymbol{\varrho}}$ allow us to compute the estimates of the expected value and metric variance of composition $\mathbf{x}$:

$$\widehat{\mathrm{E}[\mathbf{x}]} = (\hat{\beta}_1 \otimes \mathbf{e}_1) \oplus (\hat{\beta}_2 \otimes \mathbf{e}_2) \oplus \cdots \oplus (\hat{\beta}_{D-1} \otimes \mathbf{e}_{D-1}),$$

$$\widehat{\mathrm{Mvar}[\mathbf{X}]} = \mathrm{trace}\left(\widehat{\boldsymbol{\Upsilon}} - \frac{2}{\pi}\hat{\boldsymbol{\upsilon}}\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\upsilon}}\right),$$

where $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\upsilon}}\hat{\boldsymbol{\delta}}\sqrt{2/\pi}$ and $\hat{\boldsymbol{\upsilon}}$ is the square root of $\mathrm{diag}(\widehat{\boldsymbol{\Upsilon}})$.

The normal model in $\mathcal{S}^D$ is a particular case of the skew-normal model in $\mathcal{S}^D$ because it corresponds to the case $\boldsymbol{\varrho} = \mathbf{0}$. Thus, to decide if a skew-normal on $\mathcal{S}^D$ model is better than a normal on $\mathcal{S}^D$ model, it suffices to test the null hypothesis $\mathrm{H}_0 : \boldsymbol{\varrho} = \mathbf{0}$ versus the hypothesis $\mathrm{H}_1 : \boldsymbol{\varrho} \neq \mathbf{0}$ applying a likelihood ratio test to the ilr coefficients of the sample.

To validate the distributional assumption of skew-normality in $\mathcal{S}^D$, we have only to apply some goodness-of-fit tests of multivariate skew-normal distribution to the ilr coefficients of the sample data set. Mateu-Figueras et al. (2003) and Dalla-Valle (2001) have recently developed some tests for the skew-normal distribution. Unfortunately, these tests are dependent on the orthonormal basis chosen.

4.3 Another parametrization

We have defined the skew-normal on $\mathcal{S}^D$ law using the ilr coefficients of composition $\mathbf{x}$. Nevertheless we can obtain the expression of the density function using the alr coefficients. To avoid large expressions we denote as $\mathbf{v}$ and $\mathbf{y}$ the coefficients of composition $\mathbf{x}$ with respect to an orthonormal basis and to the B basis introduced in section 1 respectively.

In terms of the $\mathbf{v}$ coefficients, the probability of any event $A \subseteq \mathcal{S}^D$ is

$$P(A) = \int_{A^*} 2(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{v}-\boldsymbol{\xi})'\boldsymbol{\Upsilon}^{-1}(\mathbf{v}-\boldsymbol{\xi})\right]\Phi\left[\boldsymbol{\varrho}'\boldsymbol{\upsilon}^{-1}(\mathbf{v}-\boldsymbol{\xi})\right]d\mathbf{v},$$

where A^* represents the coefficients of A with respect to the orthonormal basis. Matrix $\mathbf{FU}$ is a change of basis matrix from the orthonormal basis to the B basis. Then, the transformation $\mathbf{v} = (\mathbf{FU})^{-1}\mathbf{y}$ makes effective this change of basis and we obtain

$$\begin{aligned}
P(A) &= \int_{\mathbf{FU}A^*} \frac{2(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Upsilon} \mid^{1/2}} \exp\left[\mathbf{M}\right]\Phi\left[\boldsymbol{\varrho}'\boldsymbol{\upsilon}^{-1}((\mathbf{FU})^{-1}\mathbf{y}-\boldsymbol{\xi})\right]\frac{1}{|\mathbf{FU}|}d\mathbf{y},\\
&= \int_{\mathbf{FU}A^*} \frac{2(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Upsilon} \mid^{1/2}|\mathbf{FU}|} \exp\left[\mathbf{M}^*\right]\Phi\left[\boldsymbol{\varrho}'\boldsymbol{\upsilon}^{-1}(\mathbf{FU})^{-1}(\mathbf{y}-\mathbf{FU}\boldsymbol{\xi})\right]d\mathbf{y},\\
&= \int_{\mathbf{FU}A^*} \frac{2(2\pi)^{-(D-1)/2}}{\mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]\Phi\left[\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]d\mathbf{y},
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{M} &= -\frac{1}{2}\left((\mathbf{FU})^{-1}\mathbf{y}-\boldsymbol{\xi}\right)'\boldsymbol{\Upsilon}^{-1}\left((\mathbf{FU})^{-1}\mathbf{y}-\boldsymbol{\xi}\right),\\
\mathbf{M}^* &= -\frac{1}{2}(\mathbf{y}-(\mathbf{FU})\boldsymbol{\xi})'((\mathbf{FU})^{-1})'\boldsymbol{\Upsilon}^{-1}(\mathbf{FU})^{-1}(\mathbf{y}-(\mathbf{FU})\boldsymbol{\xi}),
\end{aligned}$$

where $\boldsymbol{\mu} = \mathbf{FU}\boldsymbol{\xi}$, $\boldsymbol{\Sigma} = (\mathbf{FU})\boldsymbol{\Upsilon}(\mathbf{FU})'$, $\boldsymbol{\alpha} = \boldsymbol{\omega}((\mathbf{FU})^{-1})'\boldsymbol{\upsilon}^{-1}\boldsymbol{\varrho}$, and $\boldsymbol{\omega}$ is the square root of $\mathrm{diag}(\boldsymbol{\Sigma})$. Observe that $\mathbf{FU}A^*$ represents the coefficients of A with respect to the B basis. In this case neither $\boldsymbol{\mu}$ nor $\boldsymbol{\Sigma}$ represent the expected value and the covariance matrix of the coefficients $\mathbf{y}$.

In conclusion, an equivalent expression of the the skew-normal on $\mathcal{S}^D$ law using $\mathrm{alr}(\mathbf{x})$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ parametrization can be obtained and used in practice. But recall that the basis B is not orthonormal, and therefore the Euclidean distance between two alr coefficients is not equal to the Aitchison distance between the corresponding compositions. Thus, it will have no sense to use the density of the alr coefficients in the procedures that use distances or scalar products.

We could also define the skew-normal on $\mathcal{S}^D$ law using the coefficients with respect to the generating set B^* but we will obtain a degenerate density function. At the moment we have not the definition of the degenerate skew-normal model in real space.

5 Conclusions

The vector space structure of the simplex allows us to define parametric models instead of using transformations to real space keeping the usual Lebesgue measure. We have defined the normal model on $\mathcal{S}^D$ and the skew-normal model on $\mathcal{S}^D$ through their density over the coefficients with respect to an orthonormal basis. In terms of probabilities of subsets of $\mathcal{S}^D$, the normal on $\mathcal{S}^D$ and the skew-normal on $\mathcal{S}^D$ laws are identical to the additive logistic normal and the additive logistic skew-normal distribution. Nevertheless their density functions and some properties are different. For example, for the models defined on $\mathcal{S}^D$ using the coefficients, the expected value is equal to the center of the random composition. An additional reason, to prove properties using the coefficients with respect to an orthonormal basis, is that we can apply standard real analysis to them.

Acknowledgements

This research has received financial support from the *Dirección General de Enseñanza Superior e Investigación Científica (DGESIC)* of the Spanish Ministry of Education and Culture through the project BFM2000-0540 and from the *Direcció General de Recerca* of the *Departament d'Universitats, Recerca i Societat de la Informació* of the *Generalitat de Catalunya* through the project 2001XT 00057.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* London (UK): Chapman and Hall.

Aitchison, J. (1992). On criteria for measures of compositional difference. *Mathematical Geology* vol. *24* (no. 4), pp. 365–379.

Aitchison, J. (1997). The one-hour course in compositional data analysis or compositional data analysis is simple. In V. Pawlowsky-Glahn (Ed.), *Proceedings of IAMG'97 — The third annual conference of the International Association for Mathematical Geology*, Volume I, II and addendum, pp. 3–35. International Center for Numerical Methods in Engineering, Barcelona (E).

Aitchison, J. (2002). Simplicial inference. In M. A. G. Viana and D. S. P. Richards (Eds.), *Algebraic Methods in Statistics and Probability*, Volume 287 of *Contemporary Mathematics Series*, pp. 1–22. American Mathematical Society, Providence, Rhode Island (USA).

Aitchison, J., G. Mateu-Figueras, and K. Ng (2003). Characterisation of distributional forms for compositional data and associated distributional tests. *Submitted to Mathematical Geology*.

Aitchison, J. and S. Shen (1980). Logistic-normal distributions: some properties and uses. *Biometrika* vol. *67* (no. 2), pp. 261–272.

Azzalini, A. and A. Capitanio (1999). Statistical applications of the multivariate skew-normal distribution. *J.R. Statist. Soc. B* vol. *61* (no. 3), pp. 579–602.

Dalla-Valle, A. (2001). A test for the hypothesis of skew normality in a population. *Submitted to Journal of Statistical Computation and Simulation.*

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelò-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology 35*(3), pp. 279–300.

Martín-Fernández, J. A. (2001). *Medidas de diferencia y clasificación no paramétrica de datos composicionales.* Ph. D. thesis, Universitat Politècnica de Catalunya, Barcelona (E).

Mateu Figueras, G. (2003). *Models de distribució sobre el símplex.* Ph. D. thesis, Universitat Politècnica de Catalunya, Barcelona (E).

Mateu-Figueras, G., C. Barceló-Vidal, and V. Pawlowsky-Glahn (1998). Modeling compositional data with multivariate skew-normal distributions. In A. Buccianti, G. Nardi, and R. Potenza (Eds.), *Proceedings of IAMG'98 — The fourth annual conference of the International Association for Mathematical Geology*, Volume I and II, pp. 532–537. De Frede Editore, Napoli (I).

Mateu-Figueras, G. and V. Pawlowsky-Glahn (2003). Una alternativa a la distribución lognormal. In J. Saralegui and E. Ripoll (Eds.), *Actas del XXVII Congreso Nacional de la Sociedad de Estadística e Investigación Operativa (SEIO)*, pp. 1849–1858. Sociedad de Estadística e Investigación Operativa, Lleida (E).

Mateu-Figueras, G., P. Puig, and C. Barceló-Vidal (2003). Goodness-of-fit test for the skew-normal distribution with unknown parameters. *(In preparation).*

Pawlowsky-Glahn, V., J. Egozcue, and H. Burger (2003). Proceedings of iamg'03 — the annual conference of the international association for mathematical geology. Portsmouth (UK). International Association for Mathematical Geology. CD-ROM.

Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *SERRA* vol. *15*(no. 5), pp. 384–398.

Pawlowsky-Glahn, V. and J. J. Egozcue (2002). BLU estimators and compositional data. *Mathematical Geology* vol. *34*(no. 3), pp. 259–274.

APPENDIX. The skew-normal distribution

The multivariate skew-normal distribution was studied in detail by Azzalini and Capitanio (1999). According to them, a D-variate random vector $\mathbf{y}$ is said to have a multivariate skew-normal distribution if it is continuous with density function

$$2(2\pi)^{-D/2} \mid \boldsymbol{\Upsilon} \mid^{-1/2} \exp\left[-\frac{1}{2}\,(\mathbf{v}-\boldsymbol{\xi})'\,\boldsymbol{\Upsilon}^{-1}\,(\mathbf{v}-\boldsymbol{\xi})\right]\Phi(\boldsymbol{\varrho}'\boldsymbol{v}^{-1}(\mathbf{y}-\boldsymbol{\xi})), \tag{16}$$

where Φ is the $\mathcal{N}(0,1)$ distribution function and $\boldsymbol{v}$ is the square root of diag$(\boldsymbol{\Upsilon})$. The $\boldsymbol{\varrho}$ parameter is a D-variate vector which regulates the shape of the distribution and indicates the direction of maximum skewness. When $\boldsymbol{\varrho}=\mathbf{0}$ the random vector $\mathbf{y}$ reduces to a $\mathcal{N}^D(\boldsymbol{\xi},\boldsymbol{\Upsilon})$ distributed vector. We will use the notation $\mathbf{y}\sim\mathcal{SN}^D(\boldsymbol{\xi},\boldsymbol{\Upsilon},\boldsymbol{\varrho})$ to symbolize a random vector with a density function given by (16).

We know that each component of $\mathbf{y}$ is univariate skew-normal distributed. Its marginal skewness index can be computed using the parameter $\boldsymbol{\varrho}$ and it varies only in the interval $(-0.995,+0.995)$. In the multivariate case, we can also consider a multivariate index of skewness. This multivariate index is also bounded according to the scalar case. Consequently the skew-normal family allows densities with a moderate skewness.

Given $\mathbf{y} \sim \mathcal{SN}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$, the expected value is $\mathrm{E}(\mathbf{y}) = \boldsymbol{\xi} + \boldsymbol{\upsilon}\boldsymbol{\delta}\sqrt{2/\pi}$ and the covariance matrix is $\mathrm{var}(\mathbf{y}) = \boldsymbol{\Upsilon} - (2/\pi)\boldsymbol{\upsilon}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\upsilon}$, where $\boldsymbol{\delta}$ is a D-variate vector related to $\boldsymbol{\varrho}$ parameter as

$$\boldsymbol{\delta} = \frac{1}{\sqrt{1 + \boldsymbol{\varrho}'\boldsymbol{\upsilon}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\upsilon}^{-1}\boldsymbol{\varrho}}}\boldsymbol{\upsilon}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\upsilon}^{-1}\boldsymbol{\varrho}. \tag{17}$$

Azzalini and Capitanio (1999) provide a wide range of properties for the multivariate skew-normal distribution, most of them similar to the properties of the multivariate normal distribution.

Linear transformation property. If $\mathbf{y} \sim \mathcal{SN}^D(\boldsymbol{\xi}, \boldsymbol{\Upsilon}, \boldsymbol{\varrho})$ and $\mathbf{A}$ is a $D \times H$ matrix of constants, then $\mathbf{y}^* = \mathbf{A}'\mathbf{y} \sim \mathcal{SN}^H(\boldsymbol{\xi}^*, \boldsymbol{\Upsilon}^*, \boldsymbol{\varrho}^*)$, with

$$\boldsymbol{\xi}^* = \mathbf{A}'\boldsymbol{\xi}, \qquad \boldsymbol{\Upsilon}^* = \mathbf{A}'\boldsymbol{\Upsilon}\mathbf{A}, \qquad \boldsymbol{\varrho}^* = \frac{\boldsymbol{\upsilon}^*(\boldsymbol{\Upsilon}^*)^{-1}\mathbf{B}'\boldsymbol{\varrho}}{\sqrt{1 + \boldsymbol{\varrho}'(\boldsymbol{\upsilon}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\upsilon}^{-1} - \mathbf{B}(\boldsymbol{\Upsilon}^*)^{-1}\mathbf{B}')\boldsymbol{\varrho}}},$$

where $\mathbf{B} = \boldsymbol{\upsilon}^{-1}\boldsymbol{\Upsilon}\mathbf{A}$ and $\boldsymbol{\upsilon}$ and $\boldsymbol{\upsilon}^*$ are, respectively, the square root of $\mathrm{diag}(\boldsymbol{\Upsilon})$ and $\mathrm{diag}(\boldsymbol{\Upsilon}^*)$. In particular, if $\mathbf{A}$ is a non singular and square matrix, then $\boldsymbol{\varrho}^* = \boldsymbol{\upsilon}^*\mathbf{A}^{-1}\boldsymbol{\upsilon}^{-1}\boldsymbol{\varrho}$.

Given a sample, to find the estimates $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\Upsilon}}$ and $\hat{\boldsymbol{\varrho}}$ of the parameters $\boldsymbol{\xi}, \boldsymbol{\Upsilon}$ and $\boldsymbol{\varrho}$, we apply the maximum likelihood procedure. But the estimates cannot be expressed in analytic terms and we have to use numerical methods (e.g. Newton-Raphson or the generalized gradient method) to compute an approximation from a sample. There are however some problems. In the univariate case, for example, there is always an inflection point at $\varrho = 0$ of the profile log-likelihood and the shape of this function could be problematic and slows the convergence down. Azzalini and Capitanio (1999, p. 591) suggest to substitute the parameter $\boldsymbol{\varrho}$ by a new parameter $\boldsymbol{\beta} = \boldsymbol{\upsilon}^{-1}\boldsymbol{\varrho}$ in the likelihood function. Then the parameters $\boldsymbol{\Upsilon}$ and $\boldsymbol{\beta}$ appear well separated in two factors in the loglikelihood function and we can exploit the factorization property which makes the computation of estimates easier. Initial estimates of the parameters necessary to start the iterative numerical procedure can be calculated from the sample by the method of moments, using the relations of the mean vector, covariance matrix and the skewness vector with the parameters $\boldsymbol{\xi}, \boldsymbol{\Upsilon}$ and $\boldsymbol{\varrho}$ (respectively $\boldsymbol{\beta}$).

But there are still cases where the behaviour of the maximum likelihood estimates appears unsatisfactory because, with nothing pathological in the data pattern, the shape parameter tends to its maximum value. In these cases we have to adopt a temporary solution suggested by Azzalini and Capitanio (1999, p. 591): when the maximum of $\boldsymbol{\varrho}$ occurs on the frontier, re-start the maximization procedure and stop it when it reaches a loglikelihood value not significantly lower than the maximum.