Proceedings of the 4th International Workshop
on Compositional Data Analysis (2011)

The shifted-scaled Dirichlet distribution in the simplex

G.S. Monti[1], G. Mateu-Figueras[2], V. Pawlowsky-Glahn[2] and J.J. Egozcue [3]

[1] Department of Statistics, University of Milan-Bicocca, Italy; gianna.monti@unimib.it;
[2] Department of Computer Science and Applied Mathematics, University of Girona, Spain;
[3] Department of Applied Mathematics III, Technical University of Catalonia, Spain.

Abstract

Perturbation and powering are two operations in the simplex that define a vector-space structure. Perturbation and powering in the simplex play the same role as the sum and product by scalars in real space. A standard Dirichlet random composition can be shifted by perturbation, and scaled powering by a real scalar. The obtained random composition has a *shifted-scaled Dirichlet* distribution. The procedure is analogous to standardization of real random variables. The derived distribution is a generalization of the Dirichlet one, and it is studied from a probabilistic point of view. In the simplex, considered as an Euclidean space, the Aitchison measure is the natural (Lebesgue type) measure, which is compatible with its operations and metrics. Therefore, a natural way of describing the generalized (shifted-scaled) Dirichlet probability distributions is using probability densities with respect to the Aitchison measure. This density representation is compared with the traditional probability density with respect to the Lebesgue measure. In particular, the center and variability for both representations are compared.

1 Introduction

Compositional data, or parts of some whole, have been originally defined as non-negative bounded-sum data. The sample space for this kind of data is the simplex, denoted

$$\mathcal{S}^D = \{\mathbf{x} = (x_1, \ldots, x_D),\ x_i > 0,\ \sum_{i=1}^{D} x_i = \kappa\}\,,$$

where κ is a constant, usually taken to be one or one hundred. Given the fact that compositional data only carry relative information, the requirement of scale invariance is natural, and they can be understood as equivalence classes of vectors passing through the origin (Barceló-Vidal et al., 2001). From this perspective, a composition in the simplex is a representative of an equivalence class.

Historically the prevalent model for compositional data was the Dirichlet distribution (Gupta and Richards, 2001). The Dirichlet distribution provides a handy tool for modeling data restricted to the unit simplex. It has been widely used in geology, biology, and chemistry for handling compositional data such as rates, percentages and proportions (Wong, 1998). For example, one may be interested in modeling the proportions of a given set of mineral components in rocks, the proportions of nucleotides in a DNA sequence, the proportions of time people spend a day in different activities among work, sport and extra work, or the impact of the market share distribution on industry performance. The drawbacks of the Dirichlet distribution are described extensively in Aitchison (1986, p. 58-60). They include, among others, a completely negative correlation structure and a very strong implied independence structure, namely full partition independence. As stated in Aitchison (1986, p. 305) "The realization that the Dirichlet class leans so heavily towards independence has prompted a number of authors (Connor and Mosimann, 1969; Darroch and James, 1974; Mosimann, 1975; James, 1981; James and Mosimann, 1980) to search for generalizations of the Dirichlet class with less independence structure. Their efforts have been met with only limited success ...". One of the generalizations was the *scaled Dirichlet* distribution, which is just a perturbation (shift) in the simplex and thus belongs to the same class of distributions as the Dirichlet one (Monti et al., 2011).

In this paper we present a further generalization of the Dirichlet distribution which naturally includes the scaled Dirichlet, namely the distribution of a random vector obtained after applying the perturbation and powering operations to a Dirichlet random composition. Perturbation and powering are two operations that define a vector-space structure in the simplex (Billheimer et al., 2001; Pawlowsky-Glahn and Egozcue, 2001; Aitchison et al., 2002). They play the same role as sum

Egozcue, J.J., Tolosana-Delgado, R. and Ortego, M.I. (eds.)
ISBN: 978-84-87867-76-7

and product by scalars in real space. When a random variable is standardized, first the distribution is centered on the mean through the sum operation, and second it is divided by the standard error through multiplication. This procedure can be translated into the compositional framework starting with a Dirichlet random composition and considering the Aitchison geometry of the simplex (Pawlowsky-Glahn and Egozcue, 2001). This motivates the name *shifted-scaled* Dirichlet distribution. *Shifted* refers to perturbation in the simplex; it was previously called *scaled* due to the multiplicative character of perturbation; now *scaled* means scaling in the simplex, i.e. powering, and therefore the extended generalization of the distribution is related to this kind of scaling. The derived distribution, which is a generalization of the Dirichlet one, is studied from a probabilistic point of view. In particular, the resulting probability density function is presented both with respect to the Lebesgue measure on real space and to the Aitchison measure on the simplex. This latter measure is compatible with the Euclidean space structure of the simplex (Pawlowsky-Glahn and Egozcue, 2001). The center and variability of the shifted-scaled Dirichlet distribution are studied in both representations.

In Section 2, some background information on compositional data analysis and on working in coordinates in the simplex is introduced, as well as the main ideas on center and variability of random compositions. In Section 3, we introduce the generalization of the Dirichlet distribution within the Aitchison geometry that we call shifted-scaled Dirichlet distribution. The section starts reviewing the Dirichlet (Kotz et al., 2000) and scaled Dirichlet distributions (Monti et al., 2011). The way to construct the new distribution, as well as some of its properties and features, are described in Subsection 3.3.

2 Compositional data in the simplex

2.1 The vector space structure of the simplex

In $\mathcal{S}^D$, an internal operation, $\oplus$, called *perturbation*, and an external operation, $\odot$, called *powering*, are defined (Aitchison, 1986). Given two D-part compositions $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ the perturbation operation $\mathbf{x} \oplus \mathbf{y}$ is defined by

$$\mathbf{x} \oplus \mathbf{y} = \frac{\kappa \cdot (x_1 y_1, \ldots, x_D y_D)}{x_1 y_1 + \ldots + x_D y_D} = \mathcal{C}\,(x_1 y_1, \ldots, x_D y_D) \;, \tag{1}$$

where the *closure* operator $\mathcal{C}$ standardizes the contained vector by dividing each component by the sum of its components and multiplying them by the constant κ so that, after closure, the components sum to κ. The symbol $\oplus$ emphasizes the analogy with vector addition in real space. Perturbation defines a commutative group structure on the simplex with identity $(1/D, \ldots, 1/D)$ and inverse $\mathbf{x}^{-1} = \mathcal{C}\,(1/x_1, \ldots, 1/x_D)$. Given a real number α and a composition $\mathbf{x} \in \mathcal{S}^D$, powering of $\mathbf{x}$ is

$$\alpha \odot \mathbf{x} = \mathcal{C}\,(x_1^{\alpha}, \ldots, x_D^{\alpha}) \;. \tag{2}$$

Powering is analogous to multiplication by scalars in real space. The symbol $\odot$ emphasizes the analogy with multiplication by a scalar in real space.

Furthermore, an inner product $\langle \cdot, \cdot \rangle_a$ can be defined

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \sum_{i=1}^{D} \ln \frac{x_i}{\mathrm{g_m}(\mathbf{x})} \ln \frac{y_i}{\mathrm{g_m}(\mathbf{y})} \;, \tag{3}$$

where $\mathrm{g_m}(\mathbf{x})$ denotes the geometric mean of the components of $\mathbf{x}$ (Billheimer et al., 2001; Pawlowsky-Glahn and Egozcue, 2001).

With these operations $(\mathcal{S}^D, \oplus, \odot, \langle \cdot, \cdot \rangle_a)$ has a $(D-1)$-dimensional real Euclidean vector space structure. The geometry on the simplex, in which to consider statistical modeling, is called *simplicial* or *Aitchison geometry* (Pawlowsky-Glahn and Egozcue, 2001). To complete the metric vector space structure of the simplex we report the expression for the Aitchison distance between two compositions $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{S}^D$

$$\mathrm{d}_a(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{D} \sum_{i<j} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2} \;. \tag{4}$$

ISBN: 978-84-87867-76-7

The Aitchison distance is compatible with the vector space structure of the simplex. Given three compositions $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{p}$ in $\mathcal{S}^D$ and a scalar α the following properties hold

$$\mathrm{d}_a(\mathbf{x},\mathbf{y}) = \mathrm{d}_a(\mathbf{p}\oplus\mathbf{x},\mathbf{p}\oplus\mathbf{y}), \qquad \mathrm{d}_a(\alpha\odot\mathbf{x},\alpha\odot\mathbf{y}) = |\alpha|\mathrm{d}_a(\mathbf{x},\mathbf{y}). \tag{5}$$

In our discussion we use an orthonormal basis in $\mathcal{S}^D$ for deriving coordinate representations of compositional vectors. Given an orthonormal basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_{D-1}\}$ of $\mathcal{S}^D$, which elements satisfy the conditions $\langle\mathbf{e}_i,\mathbf{e}_i\rangle_a = 1$ and $\langle\mathbf{e}_i,\mathbf{e}_j\rangle_a = 0$, $(i=1,\ldots,D-1,\ i\neq j)$, a composition $\mathbf{x}\in\mathcal{S}^D$ can be expressed as a linear-combination,

$$\mathbf{x} = (\varepsilon_1\odot\mathbf{e}_1)\oplus(\varepsilon_2\odot\mathbf{e}_2)\oplus\ldots\oplus(\varepsilon_{D-1}\odot\mathbf{e}_{D-1}) = \bigoplus_{i=1}^{D-1}(\varepsilon_i\odot\mathbf{e}_i),$$

where the symbol $\bigoplus$ represents repeated perturbation. The coefficients ε_i are the coordinates of the composition $\mathbf{x}\in\mathcal{S}^D$ with respect to the given orthonormal basis, i.e. $\varepsilon_i = \langle\mathbf{x},\mathbf{e}_i\rangle_a$, and the vector $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_{D-1})$ is a vector of $\mathbb{R}^{D-1}$. For a fixed basis they are uniquely determined, given that a composition can always be represented in a unique way by its coordinates with respect to an orthonormal basis. Once an orthonormal basis has been chosen, all standard statistical methods can be applied to coordinates and transferred to the simplex preserving their properties (Mateu-Figueras et al., 2011). The vector of coordinates is obtained applying to a composition $\mathbf{x}$ a function, which has been called *isometric-logratio-transformation* and denoted ilr (Egozcue et al., 2003). This transformation goes from $\mathcal{S}^D$ to $\mathbb{R}^{D-1}$ and is defined by

$$\mathrm{ilr}(\mathbf{x}) = (\langle\mathbf{x},\mathbf{e}_1\rangle_a,\ldots,\langle\mathbf{x},\mathbf{e}_{D-1}\rangle_a). \tag{6}$$

The resulting vector is a $(D-1)$ vector of real coordinates. Other frequent representations of compositions involve transformations based on log-ratios, such as the additive log-ratio (alr) transformation and the centered log-ratio (clr). The alr transformation assigns a vector of coordinates

$$\mathrm{alr}(\mathbf{x}) = (a_1,\ldots,a_{D-1}) = \left(\ln\frac{x_1}{x_D},\ldots,\ln\frac{x_{D-1}}{x_D}\right) \tag{7}$$

to the composition $\mathbf{x}$, but the coordinates obtained correspond to an oblique basis (Egozcue and Pawlowsky-Glahn, 2006). The inverse transformation is written

$$\mathbf{x} = \mathcal{C}\left(e^{a_1},\ldots,e^{a_{D-1}},1\right). \tag{8}$$

On the other hand the clr transformation is defined as

$$\mathrm{clr}(\mathbf{x}) = (c_1,\ldots,c_D) = \left(\ln\frac{x_1}{\mathrm{g_m}(\mathbf{x})},\ldots,\ln\frac{x_D}{\mathrm{g_m}(\mathbf{x})}\right), \tag{9}$$

but its coefficients are coordinates in a generating system, not coordinates with respect to a basis. An inverse transformation also exists in this case and the composition $\mathbf{x}$ can be written in terms of the clr-coefficients as

$$\mathbf{x} = \mathcal{C}\left(e^{c_1},\ldots,e^{c_D}\right). \tag{10}$$

2.2 The Aitchison measure on the simplex

Traditionally, a Lebesgue measure is used in the simplex, in particular when the Dirichlet distribution is involved. A composition $\mathbf{x}$ in $\mathcal{S}^D$ is represented e.g. by $(D-1)$ parts $\mathbf{x}_- = (x_1,x_2,\ldots,x_{D-1})$ living in a $(D-1)$-dimensional space, because the last component of $\mathbf{x}$, x_D, is implicitly equal to one minus the sum of the others, so that all D components sum to one. The vector $\mathbf{x}_-$ is considered as an element of $\mathbb{R}^{D-1}$ and the measure in this space is then taken as the Lebesgue measure λ.

Pawlowsky-Glahn (2003) defined an alternative measure on the simplex, denoted as λ_a and called Aitchison measure, which is compatible with the inner vector space structure of the simplex. The same strategy can be used to define a Lebesgue type measure on any Euclidean space (Eaton, 1983).

A simple way to define the Aitchison measure is to translate the Lebesgue measure in the space of orthonormal coordinates into the simplex. Let $\mathbf{e}_i$, $(i = 1, 2, \ldots, D-1)$ be an orthonormal basis of $\mathcal{S}^D$. If $\mathbf{x} \in \mathcal{S}^D$, then its coordinates are denoted $\mathrm{ilr}_i(\mathbf{x})$, $(i = 1, 2, \ldots, D-1)$. In the space of coordinates, $\mathbb{R}^{D-1}$, a hyper-parallelepiped R is characterized by two points, say $\mathbf{a} = (a_1, a_2, \ldots, a_{D-1})$, $\mathbf{b} = (b_1, b_2, \ldots, b_{D-1})$. The hyper-volume (Lebesgue measure) of the hyper-parallelepiped is the product of the lengths of the edges in the direction of each coordinate, this is

$$\lambda_{D-1}(R) = \prod_{i=1}^{D-1} |b_i - a_i| \ ,$$

where the Lebesgue measure on $\mathbb{R}^{D-1}$ has been subscripted with $D-1$ in order to distinguish it from the previous λ. Their difference is not the space where they are defined but on how the elements of $\mathbb{R}^{D-1}$ are interpreted: $(D-1)$ parts of a composition, $\mathbf{x}_-$, for λ; and $(D-1)$ ilr-coordinates for λ_{D-1}. Using the inverse transformation ilr^{-1}, the Aitchison measure of a subset $S = \mathrm{ilr}^{-1}(R) \subset \mathcal{S}^D$, with $R \subset \mathbb{R}^{D-1}$, is defined as

$$\lambda_a(S) = \lambda_a(\mathrm{ilr}^{-1}(R)) = \lambda_{D-1}(R) \ .$$

The measure for hyper-parallelepipeds is extended to the whole sigma-algebra following the general measure theory (Athreya and Lahiri, 2006; Ash and Doléans-Dade, 2000). Invariance with respect to the choice of the orthonormal basis derives from this property in $\mathbb{R}^{D-1}$. The Aitchison measure is absolutely continuous with respect to λ.

Consider a probability measure P defined on the simplex. If P is absolutely continuous with respect to λ and λ_a, the Radon-Nikodym derivative of P with respect to each of the two measures is a probability density, and it is denoted $dP/d\lambda$, respectively $dP/d\lambda_a$. Their integrals on a measurable set, $S \subseteq \mathcal{S}^D$, are the corresponding probabilities:

$$\mathrm{P}(S) = \int_S \frac{dP}{d\lambda}\, d\lambda = \int_S \frac{dP}{d\lambda_a}\, d\lambda_a \ .$$

The relationship between the two probability densities, $dP/d\lambda_a$ and $dP/d\lambda$, is given by the chain rule for measures

$$\frac{dP}{d\lambda} = \frac{dP}{d\lambda_a} \cdot \frac{d\lambda_a}{d\lambda} \ ,$$

where the Jacobian $d\lambda_a/d\lambda$ describes the relationship between the Lebesgue measure for parts of the simplex and the Aitchison measure. As shown in (Mateu-Figueras, 2003, p. 53) or in the appendix, this Jacobian is

$$\frac{d\lambda_a}{d\lambda} = \frac{1}{\sqrt{D}\ x_1 \cdots x_D} \ . \tag{11}$$

In the next section the Dirichlet, scaled Dirichlet and shifted-scaled Dirichlet distributions are analyzed changing the measure from λ to λ_a.

2.3 Center and variability

As shown by Tolosana-Delgado (2006), characterization of the variability of a random composition $\mathbf{X}$ following the development by Eaton (1983) can be summarized as follows. Consider two log-contrasts of $\mathbf{X}$, with coefficients $z_1^{(j)}, z_2^{(j)}, \ldots, z_D^{(j)}$ such that

$$Z^{(j)}(\mathbf{X}) = \sum_{i=1}^{D} z_i^{(j)} \log X_i \ , \ \sum_{i=1}^{D} z_i^{(j)} = 0 \ , \ j = 1, 2.$$

The variability (second order moment) of a random composition $\mathbf{X}$ is a bilinear form which assigns a covariance to each couple of log-contrasts $Z^{(1)}(\mathbf{X})$, $Z^{(2)}(\mathbf{X})$. Depending on how $Z^{(1)}(\mathbf{X})$, $Z^{(2)}(\mathbf{X})$ are represented, the bilinear form takes also different forms, although all of them are related. For instance, log-contrasts can be expressed as

$$Z^{(j)}(\mathbf{X}) = \sum_{i=1}^{D} z_i^{(j)} \log \frac{X_i}{\mathrm{g_m}(\mathbf{X})} = \sum_{i=1}^{D} z_i^{(j)} \mathrm{clr}_i(\mathbf{X}) \ , \ j = 1, 2,$$

and therefore,

$$\mathrm{Cov}\left(Z^{(1)}, Z^{(2)}\right) = \mathbf{z^{(1)}}'\mathbf{\Gamma}\mathbf{z^{(2)}} \;,\; [\mathbf{\Gamma}]_{ij} = \mathrm{Cov}(\mathrm{clr}_i(\mathbf{X}), \mathrm{clr}_j(\mathbf{X})) \;,\; i,j = 1,2,\ldots,D \;.$$

Alternatively, the log-contrasts can be represented as linear combinations of the $(D-1)$ coordinates $\mathrm{ilr}(\mathbf{X})$ of $\mathbf{X}$ with respect to an orthonormal basis of the simplex $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}$. Then, the $(D-1, D-1)$-matrix $\mathbf{\Upsilon} = \mathrm{Cov}(\mathrm{ilr}(\mathbf{X}))$, represents the second order moment of $\mathbf{X}$. Similarly, the variability can be represented by the $(D-1, D-1)$-matrix $\mathbf{\Sigma}$, when the log contrasts are expressed as a linear combination of alr-coordinates, with $\mathbf{\Sigma} = \mathrm{Cov}(\mathrm{alr}(\mathbf{X}))$. The relationship between the covariance matrices $\mathbf{\Sigma}$, $\mathbf{\Gamma}$ and $\mathbf{\Upsilon}$, is (Mateu-Figueras, 2003; Aitchison, 1986)

$$\begin{aligned} \mathbf{\Gamma} &= \mathbf{F}^{\star}\mathbf{\Sigma}\mathbf{F}^{\star\prime} \,, \\ \mathbf{\Upsilon} &= \left(\mathbf{U}'\mathbf{F}^{\star}\right)\mathbf{\Sigma}\left(\mathbf{U}'\mathbf{F}^{\star}\right)' = \mathbf{U}'\mathbf{\Gamma}\mathbf{U} \,, \end{aligned} \tag{12}$$

where $\mathbf{U}$ is a $(D, D-1)$-matrix with column vectors $\mathbf{u_i} = \mathrm{clr}\left(\mathbf{e_i}\right)$, $i = 1, \ldots, (D-1)$, and $\mathbf{F}^{\star}$ is a $(D, D-1)$-matrix

$$\mathbf{F}^{\star} = \frac{1}{D}\begin{bmatrix} D-1 & -1 & \ldots & -1 \\ -1 & D-1 & \ldots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \ldots & D-1 \\ -1 & -1 & \ldots & -1 \end{bmatrix} ,$$

A global measure of variability with respect to the Aitchison measure λ_a is the metric variance (Pawlowsky-Glahn and Egozcue, 2001, 2002). It is equivalent to the concept of total variance (Aitchison, 1997, 2002), and describes the dispersion around a given point. Consider a random composition $\mathbf{X}$ with sample space $\mathcal{S}^D$. The dispersion or metric variance around $\xi \in \mathcal{S}^D$ is the expected value of the squared distance between $\mathbf{X}$ and ξ

$$\mathrm{Mvar}(\mathbf{X}, \xi) = \mathrm{E}\left(\mathrm{d}_a^2\left(\mathbf{X}, \xi\right)\right) \,, \tag{13}$$

where d_a is the Aitchison distance defined above. The metric variance can be computed directly using coordinates. Assuming the metric variance of $\mathbf{X}$ exists, the center of the distribution of $\mathbf{X}$ is that element $\xi \in \mathcal{S}^D$ which minimizes $\mathrm{Mvar}(\mathbf{X}, \xi)$. It is usually denoted by $\mathrm{Cen}(\mathbf{X})$ (Aitchison, 1997, 2002; Pawlowsky-Glahn and Egozcue, 2001, 2002). Therefore, the metric variance around the center $\mathrm{Cen}(\mathbf{X})$ of the distribution of $\mathbf{X}$ is given by $\mathrm{Mvar}(\mathbf{X}, \mathrm{Cen}(\mathbf{X}))$ or $\mathrm{Mvar}(\mathbf{X})$ for short. The metric variance is the trace of the covariance matrices $\mathbf{\Gamma}$, and $\mathbf{\Upsilon}$,

$$\mathrm{Mvar}(\mathbf{X}) = \mathrm{trace}\left(\mathbf{\Gamma}\right) = \mathrm{trace}\left(\mathbf{\Upsilon}\right) \,. \tag{14}$$

Aitchison (1997, 2002) defines the total variance as the trace of matrix $\mathbf{\Gamma}$. Mateu-Figueras (2003), using the relationship (12), proves that $\mathrm{Mvar}(\mathbf{X}) = \mathrm{trace}\left(\mathbf{\Gamma}\right)$. Also, Aitchison (1986) proves that the trace of matrices $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ are not equal. In particular, it can be shown that $\mathrm{trace}\left(\mathbf{\Gamma}\right) = \mathrm{trace}\left(\mathbf{\Sigma}\right) - D^{-1}\mathbf{j}'\mathbf{\Sigma}\mathbf{j}$ where $\mathbf{j}$ represents a column vector of units (Aitchison, 1986, p. 103).

3 The shifted-scaled Dirichlet distribution

This section has two main goals: to revise the Dirichlet and scaled Dirichlet distribution, and to introduce a new generalization of the latter distribution called shifted-scaled Dirichlet distribution.

3.1 Dirichlet distribution

Definition 3.1 (Dirichlet distribution (Wilks, 1962; Kotz et al., 2000)) *A random vector $\mathbf{X} \in \mathcal{S}^D$ has a D-variate Dirichlet distribution with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D) \in \mathbb{R}_+^D$ if it has density function*

$$f(\mathbf{x}) = \frac{dP}{d\lambda}(\mathbf{x}) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{D}\Gamma(\alpha_i)}\prod_{i=1}^{D} x_i^{\alpha_i - 1} \,, \tag{15}$$

where P is the Dirichlet probability measure, $\alpha_+ = \sum_{i=1}^{D}\alpha_i$, and Γ denotes the Euler gamma function.

This distribution will be denoted $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$. The number of its parameters is D. Aitchison (1986) pointed out that the variables in a Dirichlet random vector exhibit strong conditional independence relationships. However, the Dirichlet distribution is still popular for analyzing compositional data because of its conjugate property with the multinomial likelihood in Bayesian analysis and its computational efficiency.

Equation (15) corresponds to a classical density, i.e. it is a Radon-Nikodym derivative with respect to the Lebesgue measure in the space of $(D-1)$ parts. Using (11) we can change the measure and express this density with respect to the measure λ_a (Mateu-Figueras and Pawlowsky-Glahn, 2005). The resulting expression for the Dirichlet density function is

$$f_a(\mathbf{x}) = \frac{dP}{d\lambda_a}(\mathbf{x}) = \frac{\Gamma(\alpha_+)\sqrt{D}}{\prod_{i=1}^{D}\Gamma(\alpha_i)} \prod_{i=1}^{D} x_i^{\alpha_i} \,. \tag{16}$$

Property 3.1 (Genesis) *The Dirichlet distribution can be obtained by normalizing a set of independent, equally scaled gamma random variables (r.v.s)* $W_i \sim Ga(\alpha_i, 1)$, $i = 1, \ldots, D$. *Formally, if* $\mathbf{X} = \mathcal{C}(\mathbf{W}) \equiv \mathbf{W}/W_+$, *where* $W_+ = \sum_{i=1}^{D} W_i \sim Ga(\alpha_+, 1)$, *then* $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$.

Property 3.2 (Marginals and Conditionals) *Let* $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$. *The following properties hold:*

- *For* $D = 2$, *the Dirichlet reduces to the beta distribution, as can be derived from Equation (15), which can be regarded as a multivariate beta distribution.*
- *The marginal* $\left(X_1, \ldots, X_k, 1 - \sum_{i=1}^{k} X_i\right)$ *is* $\mathcal{D}^{k+1}(\alpha_1, \ldots, \alpha_k, \alpha_+ - \sum_{i=1}^{k} \alpha_i)$. *It follows that the marginal distribution of each* X_i *is beta with parameters* $(\alpha_i, \alpha_+ - \alpha_i)$.
- *The conditional distribution* $\left(X_1, \ldots, X_k | X_{k+1}, \ldots, X_D, 1 - \sum_{j=k+1}^{D} X_j\right)$ *is* $\mathcal{D}^k(\alpha_1, \ldots, \alpha_k)$.

See the appendix for a proof.

Property 3.3 *The Dirichlet density has complete permutation symmetry.*

Proof. Any X_i and X_j $(i, j = 1, \ldots, D)$ may be interchanged, as long as the corresponding α_i and α_j are interchanged at the same time. As can be seen in Eqs. (15) and (16), the density is the same, independently of the measure of reference. □

Property 3.4 (Location) *Let* $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$. *The mode and the mean of a Dirichlet r.v., with respect to the measures* λ *and* λ_a, *are:*

$$\begin{aligned} \mathrm{mode}(\mathbf{X}) &= \left(\frac{\alpha_1 - 1}{\alpha_+ - D}, \ldots, \frac{\alpha_D - 1}{\alpha_+ - D}\right), \\ \mathrm{mode}_a(\mathbf{X}) &= \left(\frac{\alpha_1}{\alpha_+}, \ldots, \frac{\alpha_D}{\alpha_+}\right), \end{aligned} \tag{17}$$

$$\begin{aligned} \mathrm{E}(\mathbf{X}) &= \left(\frac{\alpha_1}{\alpha_+}, \ldots, \frac{\alpha_D}{\alpha_+}\right), \\ \mathrm{E}(\mathbf{X})_a &= \mathcal{C}\left(e^{\psi(\alpha_1)}, \ldots, e^{\psi(\alpha_D)}\right), \end{aligned} \tag{18}$$

where $\psi(t) = \frac{\partial \ln \Gamma(t)}{\partial t}$ *is the digamma function (Abramovitz and Stegun, 1965) and* $\mathcal{C}$ *is the closure operator.*

Proof. To find the mode of the Dirichlet distribution we have to maximize the log density subject to the unit sum constraint $\sum_i x_i = 1$; this can be accomplished using Lagrange multipliers (Monti et al., 2011). □

Note that the mean $\mathrm{E}(\mathbf{X})$ of the Dirichlet distribution coincides with the normalized parameter vector $\boldsymbol{\alpha}$, the same as the mode of $\mathbf{X}$ with respect to the Aitchison measure. Also, α_+ may be regarded as a precision parameter, as when it increases, the distributions become more tightly concentrated

ISBN: 978-84-87867-76-7

around the mean. With respect to the Lebesgue measure, when the components of $\boldsymbol{\alpha}$ are all greater than 1, the density has a single mode; otherwise when the components of $\boldsymbol{\alpha}$ are all less than 1, the density is unbounded when approaching the edges and corners of the simplex. As $\alpha_i \to 1$ $(i = 1, \ldots, D)$ their sum tends to D and the Dirichlet probability function becomes nearly uniform.

Property 3.5 (Measures of dispersion) *Let $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$, the variance and covariance with respect to the measure λ are*

$$\begin{aligned} \mathrm{Var}(X_i) &= \frac{\alpha_i(\alpha_+ - \alpha_i)}{(\alpha_+)^2(\alpha_+ + 1)} = \frac{\mathrm{E}(X_i)(1-\mathrm{E}(X_i))}{(\alpha_+ + 1)} , \\ \mathrm{Cov}(X_i, X_r) &= -\frac{\alpha_i \alpha_r}{(\alpha_+)^2(\alpha_+ + 1)} = -\frac{\mathrm{E}(X_i)\mathrm{E}(X_r)}{(\alpha_+ + 1)} , \end{aligned} \tag{19}$$

which reveals that all pairwise correlations are negative.

Property 3.6 (Metric variance) *Given a random vector $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ defined in the unit simplex $\mathcal{S}^D$ the metric variance of $\mathbf{X}$ is*

$$\mathrm{Mvar}(\mathbf{X}) = \frac{D-1}{D} \left(\psi'(\alpha_1) + \ldots + \psi'(\alpha_D)\right) , \tag{20}$$

where $\psi'(t)$, $(t > 0)$ is the trigamma function (Abramovitz and Stegun, 1965).

Proof. To prove this result, recall the expression for the covariance matrix of $\mathrm{alr}(\mathbf{X})$ (Eq. 7) provided by Aitchison (1986, p. 60)

$$\mathrm{Var}\left(\ln \frac{X_i}{X_j}\right) = \psi'(\alpha_i) + \psi'(\alpha_j) , \quad \mathrm{Cov}\left(\ln \frac{X_i}{X_k}, \ln \frac{X_j}{X_k}\right) = \psi'(\alpha_k) . \tag{21}$$

The two Equations in (21) are the elements of the variance and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathrm{alr}(\mathbf{X}))$. Using the relationship (12) between the matrix representation $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ it is easy to find the trace $(\boldsymbol{\Gamma})$. For example, when $D = 3$, the matrix $\boldsymbol{\Gamma}$ assumes the form

$$\boldsymbol{\Gamma} = \frac{1}{9} \begin{bmatrix} 4\psi'(\alpha_1) + \psi'(\alpha_2) + \psi'(\alpha_3) & \psi'(\alpha_3) - 2\psi'(\alpha_1) - 2\psi'(\alpha_2) & \psi'(\alpha_2) - 2\psi'(\alpha_1) - 2\psi'(\alpha_3) \\ \psi'(\alpha_3) - 2\psi'(\alpha_1) - 2\psi'(\alpha_2) & 4\psi'(\alpha_2) + \psi'(\alpha_1) + \psi'(\alpha_3) & \psi'(\alpha_1) - 2\psi'(\alpha_2) - 2\psi'(\alpha_3) \\ \psi'(\alpha_2) - 2\psi'(\alpha_1) - 2\psi'(\alpha_3) & \psi'(\alpha_1) - 2\psi'(\alpha_2) - 2\psi'(\alpha_3) & 4\psi'(\alpha_3) + \psi'(\alpha_1) + \psi'(\alpha_2) \end{bmatrix} ,$$

whose trace is equal to $\frac{2}{3}\left(\psi'(\alpha_1) + \psi'(\alpha_2) + \psi'(\alpha_3)\right)$. □

Note that, in general, when the number of components of $\mathbf{X}$ is D, the elements of the variance and covariance matrix $\boldsymbol{\Gamma}$ are:

$$\begin{aligned} \mathrm{Var}\left(\ln \frac{X_i}{\mathrm{g_m}(\mathbf{X})}\right) &= \left(\frac{D-1}{D}\right)^2 \psi'(\alpha_i) + \frac{1}{D^2} \sum_{j=1,\, j\neq i}^{D} \psi'(\alpha_j) , \\ \mathrm{Cov}\left(\ln \frac{X_i}{\mathrm{g_m}(\mathbf{X})}, \ln \frac{X_j}{\mathrm{g_m}(\mathbf{X})}\right) &= -\left(\frac{D-1}{D^2}\right)\left(\psi'(\alpha_i) + \psi'(\alpha_j)\right) + \frac{1}{D^2} \sum_{k=1,\, k\neq i,j}^{D} \psi'(\alpha_k) . \end{aligned}$$

3.2 Scaled Dirichlet distribution

In a Bayesian analysis of a multinomial situation it is necessary to choose a suitable prior from the set of all probability measures on the simplex. Conjugate analysis leads to the Dirichlet-family as a natural choice. But while this way is mathematically convenient, it is not the most appropriate solution; e.g. it does not take into account relative positions between categories or multinomial cells (Lochner, 1975). In this framework Savage, in a personal and unpublished work, presented a transformed Dirichlet density (Dichey, 1968) as an alternative to the Dirichlet density, taking

$$\tilde{x}_i = \frac{p_i x_i}{\sum_{i=1}^{D} p_i x_i} , \qquad x_i = \frac{\tilde{x}_i / p_i}{\sum_{i=1}^{D} (\tilde{x}_i / p_i)} ,$$

where $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ and each $p_i > 0$. This kind of generalization permits some more flexibility in the choice of a prior in multinomial problems. The probability density function of $\tilde{\mathbf{X}}$ is the scaled Dirichlet distribution (Monti et al., 2011).

Definition 3.2 (Scaled Dirichlet distribution (Dichey, 1968).) *A random vector $\mathbf{X} \in \mathcal{S}^D$ has a scaled Dirichlet distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_D) \in \mathbb{R}_+^D$ if its density function is*

$$f_s(\mathbf{x}) = \frac{dP_s}{d\lambda}(\mathbf{x}) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^D \Gamma(\alpha_i)} \frac{\prod_{i=1}^D \beta_i^{\alpha_i} x_i^{\alpha_i - 1}}{(\sum_{i=1}^D \beta_i x_i)^{\alpha_+}}, \tag{22}$$

where P_s is the scaled Dirichlet probability measure, $\alpha_+ = \sum_{i=1}^D \alpha_i$, and Γ denotes the gamma function. This distribution will be denoted $\mathbf{X} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

The number of parameters is $2D$. If we fix the parameter $\boldsymbol{\beta}$ equal to $(1, 1, \ldots, 1)$, $\mathcal{C}(1, 1, \ldots, 1)$, or $\mathcal{C}(\beta, \beta, \ldots, \beta)$, we obtain a Dirichlet model. We can express Equation (22) with respect to the measure λ_a in the same way we did for the Dirichlet distribution,

$$f_{sa}(\mathbf{x}) = \frac{dP_s}{d\lambda_a}(\mathbf{x}) = \frac{\Gamma(\alpha_+)\sqrt{D}}{\prod_{i=1}^D \Gamma(\alpha_i)} \frac{\prod_{i=1}^D (\beta_i x_i)^{\alpha_i}}{(\sum_{i=1}^D \beta_i x_i)^{\alpha_+}}. \tag{23}$$

Property 3.7 (Genesis) *The scaled Dirichlet distribution can be obtained by normalizing a vector of D independent, scaled, gamma r.v.s $W_i \sim Ga(\alpha_i, \beta_i)$, $i = 1, 2, \ldots, D$. Formally, if $\mathbf{X} = \mathcal{C}(\mathbf{W})$ then $\mathbf{X} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$.*

Note that the scaled Dirichlet distribution is obtained removing the requirement of equal scaled parameters for the gamma r.v.s. in Property 3.1. The scaled Dirichlet distribution is one of the generalizations of the Dirichlet model we can find in the literature. Nevertheless, the scaled Dirichlet distribution can also be obtained starting from a perturbed random composition with a Dirichlet density (Monti et al., 2011). In fact, let $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ be a random composition defined in $\mathcal{S}^D$, and let $\mathbf{p} \in \mathcal{S}^D$ be a composition. The random composition $\widetilde{\mathbf{X}} = \mathbf{p} \oplus \mathbf{X}$ has distribution $\mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta} = \boldsymbol{p}^{-1})$. Note that in this case the $\boldsymbol{\beta}$ parameter is a composition and the number of parameters of the model is $2D - 1$. The consequence is that the parameter space $\boldsymbol{\beta} \in \mathbb{R}_+^D$ specified in Definition 3.2 should be substituted by $\boldsymbol{\beta} \in \mathcal{S}^D$, because proportional $\boldsymbol{\beta}$'s lead to equal distributions. Thus, taking into account the algebraic-geometric structure of the simplex, the scaled Dirichlet density is just a translation (shift) of a Dirichlet density in the simplex. It follows that the Dirichlet and the scaled Dirichlet belong to the same class of distributions and not to two different classes.

Property 3.8 (Marginals) *Let $\widetilde{\mathbf{X}} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$. When D=2 the densities defined in (22) and (23) correspond respectively to*

$$f_s(\mathbf{x}) = \frac{dP_s}{d\lambda}(\mathbf{x}) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{\beta_1^{\alpha_1} x^{\alpha_1 - 1} \beta_2^{\alpha_2} (1-x)^{\alpha_2 - 1}}{(\beta_1 x + \beta_2(1-x))^{\alpha_1 + \alpha_2}}, \tag{24}$$

and

$$f_{sa}(\mathbf{x}) = \frac{dP_s}{d\lambda_a}(\mathbf{x}) = \frac{\sqrt{2}}{B(\alpha_1, \alpha_2)} \frac{(\beta_1 x)^{\alpha_1} (\beta_2(1-x))^{\alpha_2}}{(\beta_1 x + \beta_2(1-x))^{\alpha_1 + \alpha_2}}. \tag{25}$$

Property 3.9 *The scaled Dirichlet density has complete permutation symmetry.*

Proof. Any X_i and X_j $(i, j = 1, \ldots, D)$ may be interchanged, as long as the corresponding α_i, α_j, and β_i, β_j, are interchanged at the same time. As can be seen in Eqs. (22) and (23), the density is the same, independently of the measure of reference. □

Property 3.10 (Location) *The mode and the expected value of $\widetilde{\mathbf{X}} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to the measure λ_a are, respectively,*

$$\begin{aligned} \mathrm{mode}_a(\widetilde{\mathbf{X}}) &= (\ominus\boldsymbol{\beta}) \oplus \mathrm{mode}_a(\mathbf{X}), \\ \mathrm{E}_a(\widetilde{\mathbf{X}}) &= (\ominus\boldsymbol{\beta}) \oplus \mathrm{E}_a(\mathbf{X}), \end{aligned} \tag{26}$$

ISBN: 978-84-87867-76-7

where $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ *and* $\ominus$ *is the inverse operation of the perturbation, by analogy with standard operations in real space.*

For a proof see Monti et al. (2011).

For a composition $\widetilde{\mathbf{X}} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$, there is no closed form for mode($\widetilde{\mathbf{X}}$) and E($\widetilde{\mathbf{X}}$) with respect to the Lebesgue measure λ in real space. It is necessary to use numerical integration to obtain them.

Equation (26) shows how $\boldsymbol{\beta}$ operates in a simplicial way. The vector of parameters $\boldsymbol{\beta}$ acts on the location of $\widetilde{\mathbf{X}}$, and not on the scale; i.e. it is not related to the measure of scale. Moreover, for $\boldsymbol{\alpha}$ a vector of constants ($\boldsymbol{\alpha} = (\alpha, \alpha, \ldots, \alpha)$), the mean and the mode with respect to the Aitchison measure coincide and are the neutral element in the simplex.

Property 3.11 (Measures of dispersion) *The metric variance for* $\widetilde{\mathbf{X}} \sim \mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *coincides with the metric variance for* $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ *given in Equation (20).*

Proof. The metric variance is invariant under perturbation (see Equation(5)). This property is equivalent to invariance under translation in real space. □

There is no closed form for the covariance matrix of a scaled Dirichlet with respect to the Lebesgue measure λ for $(D-1)$ parts of the composition. It is necessary to use numerical integration to obtain them.

3.3 The shifted-scaled Dirichlet distribution

The aim of this section is to study the distribution of a random composition obtained after applying the perturbation and powering operations to a Dirichlet random composition, i.e, the density of the composition $\widetilde{\mathbf{X}} = \mathbf{p} \oplus (a \odot \mathbf{X})$, with $\boldsymbol{p} \in \mathcal{S}^D$, $a \in \mathbb{R}_+$ and $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$.

Definition 3.3 (Shifted-scaled Dirichlet distribution) *A random vector* $\mathbf{X} \in \mathcal{S}^D$ *has a shifted-scaled Dirichlet distribution with parameters* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D) \in \mathbb{R}_+^D$, $\boldsymbol{p} = (p_1, \ldots, p_D) \in \mathcal{S}^D$ *and* $a \in \mathbb{R}_+$ *if its density function is*

$$f_{ps}(\mathbf{x}) = \frac{dP_{ps}}{d\lambda}(\mathbf{x}) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^D \Gamma(\alpha_i)} \frac{1}{a^{D-1}} \frac{\prod_{i=1}^D p_i^{-(\alpha_i/a)} \; x_i^{(\alpha_i/a)-1}}{\left(\sum_{i=1}^D (x_i/p_i)^{(1/a)}\right)^{\alpha_+}}, \tag{27}$$

where P_{ps} *is the shifted-scaled Dirichlet probability measure,* $\alpha_+ = \sum_{i=1}^D \alpha_i$, *and* Γ *is the gamma function. This distribution will be denoted* $\mathbf{X} \sim p\mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, a)$.

The number of parameters is $2D$. Furthermore, for $a = 1$ a scaled Dirichlet model with parameters $\boldsymbol{\alpha}$ and $\ominus \boldsymbol{p}$ is obtained. If $a = 1$ and the vector $\boldsymbol{p} = \mathcal{C}(1, 1, \ldots, 1)$, or $\mathcal{C}(p, p, \ldots, p)$ for some constant p, we obtain a Dirichlet model because they correspond to the neutral elements with respect to the corresponding operations.

We can express Equation (27) with respect to the measure λ_a, in the same way we did for the scaled Dirichlet density,

$$f_{psa}(\mathbf{x}) = \frac{dP_{ps}}{d\lambda_a}(\mathbf{x}) = \frac{\sqrt{D}\Gamma(\alpha_+)}{\prod_{i=1}^D \Gamma(\alpha_i)} \frac{1}{a^{D-1}} \frac{\prod_{i=1}^D (x_i/p_i)^{(\alpha_i/a)}}{\left(\sum_{i=1}^D (x_i/p_i)^{(1/a)}\right)^{\alpha_+}}. \tag{28}$$

Property 3.12 (Marginals) *When D=2 the densities defined in (27) and (28) correspond respectively to*

$$f_{ps}(\mathbf{x}) = \frac{dP_{ps}}{d\lambda}(\mathbf{x}) = \frac{1}{aB(\alpha_1, \alpha_2)} \frac{(1/p_1)^{\alpha_1/a} \, x^{(\alpha_1/a)-1} \, (1/p_2)^{\alpha_2/a} \, (1-x)^{(\alpha_2/a)-1}}{\left((x_1/p_1)^{1/a} + ((1-x)/p_2)^{1/a}\right)^{\alpha_+}}; \tag{29}$$

$$f_{psa}(\mathbf{x}) = \frac{dP_{ps}}{d\lambda_a}(\mathbf{x}) = \frac{\sqrt{2}}{aB(\alpha_1, \alpha_2)} \frac{(x/p_1)^{\alpha_1/a} \, ((1-x)/p_2)^{\alpha_2/a}}{\left((x_1/p_1)^{1/a} + ((1-x)/p_2)^{1/a}\right)^{\alpha_+}}. \tag{30}$$

PROPERTY 3.13 *The shifted-scaled Dirichlet density has complete permutation symmetry.*

PROOF. Any X_i and X_j $(i, j = 1, \ldots, D)$ may be interchanged, as long as the corresponding α_i, α_j, and p_i, p_j, are interchanged at the same time. As can be seen in Eqs. (27) and (28), the density is the same, independently of the measure of reference. □

Figure 1 provides a graphical comparison between the classical Lebesgue and the Aitchison measure

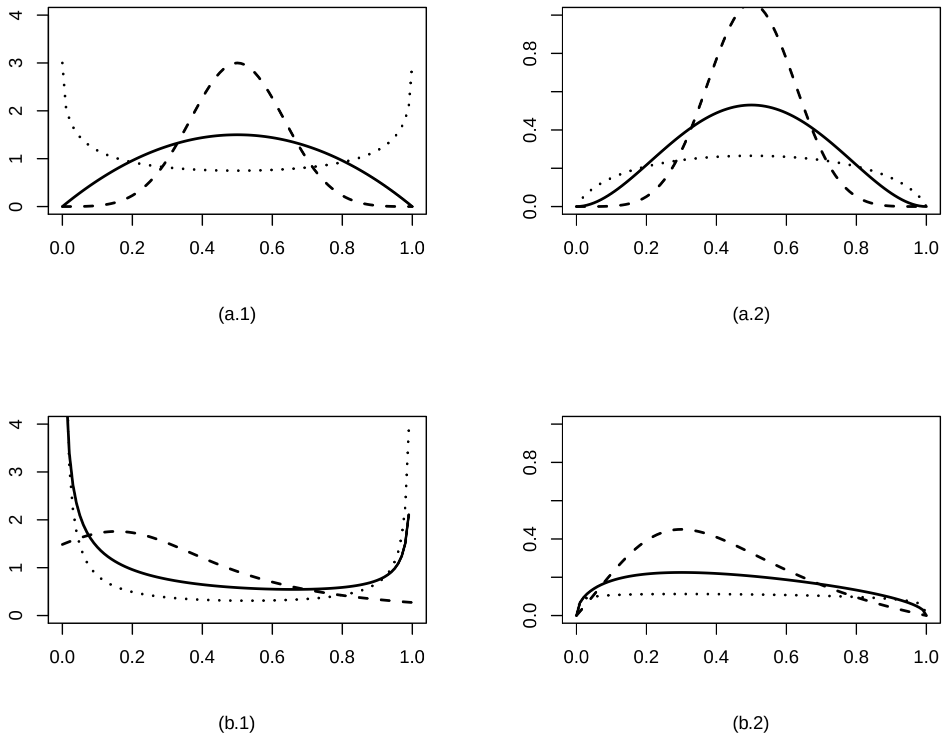

Figure 1: Shifted-scaled Dirichlet density curves when $D = 2$. Figures (a.1) and (b.1) with respect to the restriction of the Lebesgue measure λ on the $[0, 1]$ interval; (a.2) and (b.2) with respect to the Aitchison measure λ_a on $\mathcal{S}^2$. Parameter configurations: (a) $\boldsymbol{\alpha} = (2, 2)$, $\mathbf{p} = (1, 1)$; (b) $\boldsymbol{\alpha} = (0.5, 0.5)$, $\mathbf{p} = (0.3, 0.7)$ for different values of a (dashed curves: $a = 0.5$, solid curves: $a = 1$, dotted curves: $a = 2$) .

when the number of components is $D = 2$. Using the density with respect to the measure λ_a there is always a single mode. This is not the case using the density with respect to the measure λ . For three different values of a, we have considered two different parameterizations: when $\boldsymbol{\alpha} = (2, 2)$, $\mathbf{p} = (1, 1)$ for the (a.1) and (a.2) subfigures and $\boldsymbol{\alpha} = (0.5, 0.5)$, $\mathbf{p} = (0.3, 0.7)$ for the (b.1) and (b.2) subfigures. We can observe that, with respect the Aitchison measure, the a parameter plays the role of a scale parameter; when it increases the distribution becomes more concentrated around the mean.

When the number of components is $D = 3$, the usual representation for a composition is the ternary diagram. Figure 2 shows isodensity contour plots in the ternary diagram of three shifted-scaled Dirichlet densities using the Aitchison measure λ_a. Figure 3 shows the corresponding contour plots in the space of coordinates with respect to an orthonormal basis. The first one, in black, corresponds to a density with $a = 1$, $\boldsymbol{p} = \mathcal{C}(1, 1, 1)$ and $\boldsymbol{\alpha} = (2, 2, 2)$. Thus, in this case we represent a Dirichlet distribution with mean and mode in the center of the ternary diagram. The second one, in red, is obtained with $a = 0.3$, $\boldsymbol{p} = \mathcal{C}(1, 1, \ldots, 1)$ and $\boldsymbol{\alpha} = (2, 2, 2)$. It is the result of applying only a power transformation to the previous Dirichlet random composition. Consequently, in this particular case, this transformation only changes the measure of dispersion around the mean as can be observed in the space of coordinates. Finally, in blue, we represent a shifted-scaled Dirichlet with $a = 0.3$, $\boldsymbol{p} = (0.75, 0.15, 0.12)$ and $\boldsymbol{\alpha} = (2, 2, 2)$. In this case the density is obtained applying a perturbation to the second one. In the space of coordinates, we can observe that the resulting density is a translation of the second one.

ISBN: 978-84-87867-76-7

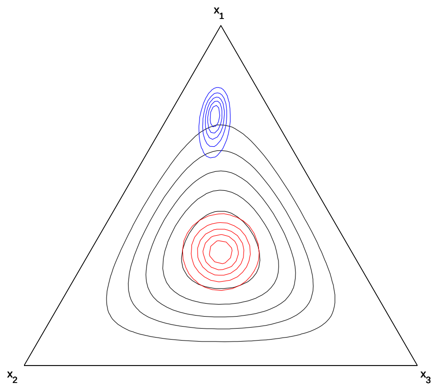

Figure 2: Shifted-scaled Dirichlet density curves when $D = 3$ with respect to the Aitchison measure λ_a in the simplex.

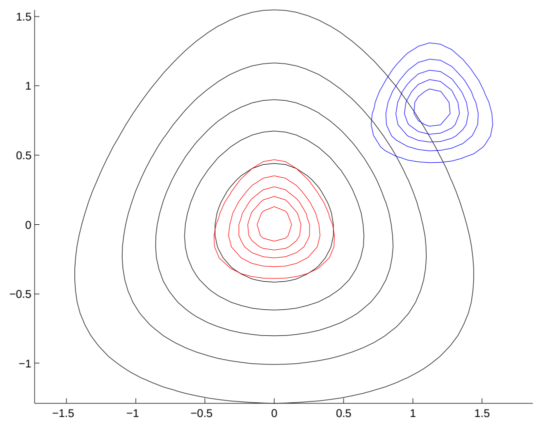

Figure 3: Shifted-scaled Dirichlet density curves when $D = 3$, with respect to the Lebesgue measure λ in real space $\mathbb{R}^2$.

PROPERTY 3.14 (LOCATION) *The mode and the expected value of $\widetilde{\mathbf{X}} \sim p\mathcal{SD}^D(\boldsymbol{\alpha}, \mathbf{p}, a)$ with respect to the measure λ_a are, respectively, equal to*

$$\begin{aligned} \mathrm{mode}_a(\widetilde{\mathbf{X}}) =& \mathbf{p} \oplus (a \odot \mathrm{mode}_a(\mathbf{X}))\ , \\ \mathrm{E}_a(\widetilde{\mathbf{X}}) =& \mathbf{p} \oplus (a \odot \mathrm{E}_a(\mathbf{X}))\ , \end{aligned} \tag{31}$$

where $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$.

PROOF. The proof is straightforward using the properties of perturbation, powering, and vector space structure of the simplex. □

There is no closed form for $\mathrm{mode}(\widetilde{\mathbf{X}})$ or $\mathrm{E}(\widetilde{\mathbf{X}})$ of a composition $\widetilde{\mathbf{X}} \sim p\mathcal{SD}^D(\boldsymbol{\alpha}, \boldsymbol{\beta}, a)$ with respect to the Lebesgue measure λ in the space of $D-1$ parts; in this case, numerical integration is required.

To calculate the metric variance recall that for a random composition $\mathbf{X} \in \mathcal{S}^D$, a perturbation $\mathbf{p} \in \mathcal{S}^D$, and a scalar $a \in \mathbb{R}$,

$$\mathrm{Mvar}(a \odot (\mathbf{p} \oplus \mathbf{X})) = a^2 \mathrm{Mvar}(\mathbf{X})\,. \tag{32}$$

PROPERTY 3.15 (MEASURES OF DISPERSION) *For a random composition $\widetilde{\mathbf{X}} \sim p\mathcal{SD}^D(\boldsymbol{\alpha}, \mathbf{p}, a)$ it holds*

$$\mathrm{Mvar}(\widetilde{\mathbf{X}}) = a^2 \mathrm{Mvar}(\mathbf{X})\,, \tag{33}$$

where $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ and $\mathrm{Mvar}(\mathbf{X})$ is defined by Equation (20).

From Equation (31) for the mean, and from Equation (33) for the metric variance of a shifted-scaled Dirichlet distribution, both with respect to the Aitchison measure, we can see that the perturbation **p** can be interpreted as a location element, as it simply shifts the distribution, while the parameter a can be viewed as a scale parameter, as it stretches or shrinks the distribution.

4 Conclusions

The shifted-scaled Dirichlet model is introduced as a natural generalization of the classical Dirichlet model, i.e, the model obtained after applying perturbation and powering to a Dirichlet random composition. Using the density expressed with respect to Aitchison measure, the mode and the expected value are easily obtained from those of the standard Dirichlet probability density. The Dirichlet and the scaled Dirichlet models are revised and the corresponding metric variance is provided. The corresponding density function with respect to the traditional Lebesgue measure on the simplex is given and important differences in the principal characteristic measures (mean, mode and variability) are obtained.

References

Abramovitz, M. and I. Stegun (1965). *Handbook of mathematical functions.* New Yowk: Dover. 1046 p.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Monographs on Statistics and Applied Probability. London, Chapman & Hall Ltd. (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.

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 (CIMNE), Barcelona (E), 1100 p.

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), 340 p.

Aitchison, J., C. Barceló-Vidal, J. J. Egozcue, and V. Pawlowsky-Glahn (2002). A concise guide for the algebraic-geometric structure of the simplex, the sample space for compositional data analysis. In U. Bayer, H. Burger, and W. Skala (Eds.), *Proceedings of IAMG'02 — The eigth annual conference of the International Association for Mathematical Geology*, Volume I and II, pp. 387–392. Selbstverlag der Alfred-Wegener-Stiftung, Berlin, 1106 p.

Ash, R. B. and C. A. Doléans-Dade (2000). *Probability and Measure Theory, Second Edition.* Academic Press, San Diego, CA (USA). 516 p.

Athreya, K. B. and S. N. Lahiri (2006). *Measure Theory and Probability Theory.* Springer, New York. 625 p.

Barceló-Vidal, C., J. A. Martín-Fernández, and V. Pawlowsky-Glahn (2001). Mathematical foundations of compositional data analysis. In G. Ross (Ed.), *Proceedings of IAMG'01 — The Sixth Annual Conference of the International Association for Mathematical Geology.* CD-ROM.

Billheimer, D., P. Guttorp, and W. Fagan (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association 96*, 1205–1214.

Connor, R. J. and J. E. Mosimann (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association 64*, 194–206.

Darroch, J. and I. James (1974). F-independence and null correlations of bounded-sum, positive variables. *Journal of the Royal Statistical Society, Series B (Methodological) 36*, 467–483.

Dichey, J. (1968). Three multidimensional-integral identities with bayesian applications. *Annals of Mathematical Statistics 39*, 1615–1627.

Eaton, M. L. (1983). *Multivariate Statistics. A Vector Space Approach*. New York, John Wiley & Sons. 512 p.

Egozcue, J. and V. Pawlowsky-Glahn (2006). Simplicial geometry for compositional data. Volume 264 of *Special Publications*. Geological Society, London.

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

Gupta, R. and D. Richards (2001). The history of the Dirichlet and Liouville distributions. *International Statistical Review 69*, 433–446.

James, I. (1981). Distributions associated with neutrality properties for random proportions. In C. Taillie, G. P. Patil, and B. A. Baldessari (Eds.), *Statistical Distributions in Scientific Work—Models, Structures, and Characterizations*, pp. 125–136. D. Reidel Publishing Co., Dordrecht (NL), 455 p.

James, I. and I. Mosimann (1980). A new characterisation of the Dirichlet distribution through neutrality. *The Annals of Statistics 8*, 183–189.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2000). *Continuous Multivariate Distributions:* Volume 1, Models and Applications. New York, John Wiley & Sons. xxii+722 p.

Lochner, R. H. (1975). A Generalized Dirichlet distribution in Bayesian Life Testing. *Journal of the Royal Statistical Society. Series B (Methodological) 37*, 103–113.

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

Mateu-Figueras, G. and V. Pawlowsky-Glahn (2005). The Dirichlet distribution with respect to the Aitchison measure on the simplex — a first approach. In G. Mateu-Figueras and C. Barceló-Vidal (Eds.), *Compositional Data Analysis Workshop – CoDaWork'05, Proceedings*. Universitat de Girona, ISBN 84-8458-222-1, http://ima.udg.es/Activitats/CoDaWork05/. CD-ROM.

Mateu-Figueras, G., V. Pawlowsky-Glahn, and J. J. Egozcue (2011). The principle of working on coordinates. See Pawlowsky-Glahn and Buccianti (2011). *In press.*

Monti, G. S., G. Mateu-Figueras, and V. Pawlowsky-Glahn (2011). Notes on the Scaled Dirichlet distribution. See Pawlowsky-Glahn and Buccianti (2011). *In press.*

Mosimann, I. (1975). Statistical problems of size and shape. II. Characterizations of the lognormal, gamma and Dirichlet distributions. In G. Patil, S. Kotz, and J. e. Ord (Eds.), *Statistical Distributions in Scientific Work*, Volume 2, pp. 219–239. D. Reidel Publishing Company, Dordrecht (NL).

Pawlowsky-Glahn, V. (2003). Statistical modelling on coordinates. In S. Thió-Henestrosa and J. A. Martín-Fernández (Eds.), *Compositional Data Analysis Workshop – CoDaWork'03, Proceedings*. University of Girona, Girona (Spain), ISBN 84-8458-111-X, http://ima.udg.es/Activitats/CoDaWork03/. CD-ROM.

Pawlowsky-Glahn, V. and A. Buccianti (Eds.) (2011). *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons. *In press.*

Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment 15*, 384–398.

Pawlowsky-Glahn, V. and J. J. Egozcue (2002). BLU estimators and compositional data. *Mathematical Geology 34*(3), 259–274.

Tolosana-Delgado, R. (2006). *Geostatistics for constrained variables: positive data, compositions and probabilities. Applications to environmental hazard monitoring.* Ph. D. thesis, University of Girona, Girona (E). 198 p.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York. 644 p.

Wong, T.-T. (1998). Generalized Dirichlet distribution in Bayesian analysis. *Applied Mathematics and Computation 97*, 165–181.

Appendix

Proof of Equation 11. The Jacobian of the alr transformation is $(x_1 \cdots x_D)^{-1}$ (Aitchison, 1986, p.115). The matrix relationship between the alr and the ilr vectors is $\text{ilr}(\mathbf{x}) = (\mathbf{FU})^{-1}\text{alr}(\mathbf{x})$ (Mateu-Figueras, 2003; Egozcue et al., 2003). Consequently, the Jacobian of the ilr transformation is $(|\mathbf{FU}|x_1 \cdots x_D)^{-1}$.

The product $\mathbf{UU}'$ is the (D, D)-matrix

$$\frac{1}{D}\begin{bmatrix} D-1 & -1 & \dots & -1 \\ -1 & D-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & D-1 \end{bmatrix},$$

that has two eigenvalues: 0 with multiplicity 1 and 1 with multiplicity $D-1$. The eigenspace with eigenvalue 1 is the clr hyperplane, that is, the $(D-1)$-dimensional linear subspace $V = \{\mathbf{z} \in \mathbb{R}^D; \sum_{i=1}^{D} z_i = 0\}$. As the rows of matrix $\mathbf{F}$ belong to V, we obtain that $\mathbf{FUU}'\mathbf{F}' = \mathbf{FF}'$. We known that $\mathbf{FF}'$ equals the $(D-1, D-1)$-matrix $\mathbf{H}$ (Aitchison, 1986, p. 343) that has two eigenvalues: 1 with multiplicity $(D-2)$ and D with multiplicity 1. Consequently, $|\mathbf{FUU}'\mathbf{F}'| = |\mathbf{FF}'| = |\mathbf{H}| = D$. Also, as $\mathbf{FU}$ is a square matrix, it holds that $|\mathbf{FUU}'\mathbf{F}'| = |\mathbf{FU}||(\mathbf{FU})'| = |\mathbf{FU}|^2$ and thus $|\mathbf{FU}| = \sqrt{D}$. □

Proof of Property 3.2. Let $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$. To show that the block of marginals $(X_1, \dots, X_k, 1 - \sum_{i=1}^{k} X_i)$ has a Dirichlet distribution with parameter $(\alpha_1, \dots, \alpha_k, \alpha_+ - \sum_{i=1}^{k} \alpha_i)$ it is convenient to integrate out the variable X_{D-1}, starting from the joint probability density function of $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$, in its domain of integration. We obtain the joint probability density function of $(X_1, \dots, X_{D-2}, 1 - \sum_{i=1}^{D-2} X_i)$

$$f(x_1, \dots, x_{D-2}, 1 - \sum_{i=1}^{D-2} x_i) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{D} \Gamma(\alpha_i)} \prod_{i=1}^{D-2} x_i^{\alpha_i - 1} \int_0^{1-\sum_{j=1}^{D-2} x_j} x_{D-1}^{\alpha_{D-1}-1} \left(1 - \sum_{j=1}^{D-2} x_j\right)^{\alpha_D - 1} dx_{D-1}\,.$$

A change of variable allows to rewrite the last integral in the interval $[0, 1]$, i.e. $y = \frac{x_{D-1}}{1-\sum_{j=1}^{D-2} x_j}$, to obtain

$$\begin{aligned} &f(x_1, \dots, x_{D-2}, 1 - \sum_{i=1}^{D-2} x_i) = \\ &= \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{D} \Gamma(\alpha_i)} \prod_{i=1}^{D-2} x_i^{\alpha_i - 1} \left(1 - \sum_{j=1}^{D-2} x_j\right)^{\alpha_{D-1}+\alpha_D - 1} \int_0^1 y^{\alpha_{D-1}-1} (1-y)^{\alpha_D - 1} dy \\ &= \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{D-2} \Gamma(\alpha_i)\Gamma(\alpha_{D-1} + \alpha_D)} \prod_{i=1}^{D-2} x_i^{\alpha_i - 1} \left(1 - \sum_{j=1}^{D-2} x_j\right)^{\alpha_{D-1}+\alpha_D - 1}, \end{aligned}$$

which corresponds to the probability density function of a Dirichlet distribution with parameter $(\alpha_1, \ldots, \alpha_{D-2}, \alpha_{D-1} + \alpha_D)$. Repeating iteratively the process of integration for the other variables $X_{D-2}, X_{D-3}, \ldots, X_k$ we obtain the stated result.

Now we can prove the third statement: the conditional distribution of any subset of the X_i's given any other subset is also (nonstandard) Dirichlet. Let's consider the conditional probability density function of $\left(X_1, \ldots, X_k | X_{k+1}, \ldots, X_D, 1 - \sum_{j=k+1}^{D} X_j\right)$ starting from the definition

$$f\left(x_1, \ldots, x_k | x_{k+1}, \ldots, x_D, 1 - \sum_{j=k+1}^{D} x_j\right) =$$

$$= \frac{\frac{\Gamma(\alpha_+)}{\prod_{i=1}^{D}\Gamma(\alpha_i)} x_1^{\alpha_1-1} \cdots x_{k-1}^{\alpha_{k-1}-1} \left(1 - \sum_{i=1, i\neq k}^{D} x_i\right)^{\alpha_k - 1} \prod_{j=k+1}^{D} x_j^{\alpha_j - 1}}{\frac{\Gamma(\alpha_+)}{\Gamma(\sum_{r=1}^{k}\alpha_r)\prod_{j=k+1}^{D}\Gamma(\alpha_j)} \prod_{j=k+1}^{D} x_j^{\alpha_j-1} (1 - \sum_{j=k+1}^{D} x_j)^{\sum_{r=1}^{k}\alpha_r - 1}} =$$

$$= \frac{\Gamma(\sum_{r=1}^{k}\alpha_r)}{\prod_{r=1}^{k}\Gamma(\alpha_r)} \prod_{r=1}^{k-1} \left(\frac{x_r}{1 - \sum_{j=k+1}^{D} x_j}\right)^{\alpha_r - 1} \left(1 - \frac{\sum_{r=1}^{k-1} x_r}{1 - \sum_{j=k+1}^{D} x_j}\right)^{\alpha_k - 1} \frac{1}{\left(1 - \sum_{j=k+1}^{D} x_j\right)^{D-k-1}} .$$

It is convenient to use the following change of variables

$$z_r = \frac{x_r}{1 - \sum_{j=k+1}^{D} x_j} \quad \text{so that} \quad x_r = z_r \left(1 - \sum_{j=k+1}^{D} x_j\right) , \quad r = 1, \ldots, k-1 .$$

The Jacobian can be calculated easily

$$\left|\frac{\partial(z_1, \ldots, z_{k-1})}{\partial(x_1, \ldots, x_{k-1})}\right| = \left|\det \begin{bmatrix} 1 - \sum_{j=k+1}^{D} x_j & 0 & \ldots & 0 \\ 0 & 1 - \sum_{j=k+1}^{D} x_j & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 - \sum_{j=k+1}^{D} x_j \end{bmatrix}\right| =$$

$$= \left(1 - \sum_{j=k+1}^{D} x_j\right)^{D-k-1} .$$

We then obtain the following formula for the joint probability density function of variables $Z_1, \ldots, Z_{k-1}$

$$f(z_1, \ldots, z_{k-1}) = \frac{\Gamma(\sum_{r=1}^{k}\alpha_r)}{\prod_{r=1}^{k}\Gamma(\alpha_r)} \prod_{r=1}^{k-1} z_r^{\alpha_r - 1} \left(1 - \sum_{r=1}^{k-1} z_r\right)^{\alpha_k - 1} ,$$

which corresponds to a probability density function of a Dirichlet distribution with parameter vector $(\alpha_1, \ldots, \alpha_k)$.

□