SORT 37 (1) January-June 2013, 29-56

The normal distribution in some constrained sample spaces

Glòria Mateu-Figueras[1], Vera Pawlowsky-Glahn[1] and Juan José Egozcue[2]

Abstract

Phenomena with a constrained sample space appear frequently in practice. This is the case, for example, with strictly positive data, or with compositional data, such as percentages or proportions. If the natural measure of difference is not the absolute one, simple algebraic properties show that it is more convenient to work with a geometry different from the usual Euclidean geometry in real space, and with a measure different from the usual Lebesgue measure, leading to alternative models that better fit the phenomenon under study. The general approach is presented and illustrated using the normal distribution, both on the positive real line and on the *D*-part simplex. The original ideas of McAlister in his introduction to the lognormal distribution in 1879, are recovered and updated.

MSC: 60A10, 60E10, 62E10.

Keywords: Additive logistic normal distribution, Aitchison measure, Lebesgue measure, lognormal distribution, orthonormal basis, simplex.

1. Introduction

In general, continuous multivariate observations are assumed to be real random vectors which density functions are defined with respect to the Lebesgue measure. The Lebesgue measure is compatible with the inner vector space structure of real space and thus natural in $\mathbb{R}$. When random vectors are defined on a constrained sample space, $E \subset \mathbb{R}^D$, methods and concepts used in real space can lead to nonsensical results. For example, for positive random variables, the usual confidence interval $\bar{x} \pm kS$, where S is the standard deviation, can include negative values. In the case of random compositions, i.e., of random vectors

[1] Dept. d'Informàtica i Matemàtica Aplicada, Campus Montilivi, Universitat de Girona, Spain. Email: gloria.mateu@udg.edu
[2] Dept. de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Barcelona, Spain.
Received: December 2011
Accepted: September 2012

defined on the simplex that represent proportions of some whole, problems appear if correlations between components are used. This is a well-known problem stated by Pearson (1897) and called *spurious correlation*. A way to avoid these problems is to use transformations, such as the logarithm on the positive real line or logratio transformations on the simplex (Aitchison, 1986). There is a long history behind the logarithmic and the logratio transformations. The well-known lognormal and the logistic normal distributions were introduced in $\mathbb{R}_+$ and in the simplex, respectively, through these transformations. In this contribution, we revise those definitions and propose a common and new theory to introduce a normal distribution in constrained sample spaces. In particular, we focus on a general constrained sample space, $E \subset \mathbb{R}^D$, which admits a meaningful Euclidean vector space structure, possibly different from the usual structure of real space. The idea, previously used in Eaton (1983), is that, for any Euclidean vector space E with a one-to-one transformation to $\mathbb{R}^D$, a measure λ_E, compatible with its structure, is obtained from the respective structure of $\mathbb{R}^D$ and its Lebesgue measure. This allows us to define density functions on E considering the measure λ_E or, equivalently, the corresponding density functions of the coordinates.

Every one-to-one transformation between a set E and real space induces a real Euclidean vector space structure in E, with associated measure λ_E. Particularly interesting are those transformations that are meaningful and related to the measure of difference between observations. This idea can be found in Galton (1879), as an introduction to the logarithmic transformation as a means to acknowledge Fechner's law, according to which *perception equals log(stimulus)*. The idea was then formalised by McAlister (1879). This approach has acquired a growing importance in applications, due to the fact that some *constrained sample spaces*, which are subsets of some real space — like $\mathbb{R}_+$ or the simplex — can be structured as Euclidean vector spaces (Pawlowsly-Glahn and Egozcue, 2001). It is important to emphasize that this approach implies using a measure which is different from the usual Lebesgue measure.

The advantage of this approach is that it opens the door to study statistical models using a measure which is considered to be appropriate or natural for the studied phenomenon, instead of the ordinary Lebesgue measure. Here we apply this idea to the normal distribution on the two mentioned constrained sample spaces, the positive real line, $\mathbb{R}_+$, and the simplex, $\mathcal{S}^D$. They are well known as the lognormal distribution and the additive logistic normal distribution when expressed with respect to the Lebesgue measure. We focus on their representation when the reference measure is the measure associated to the Euclidean vector space structure of the sample space. While the probability law is the same, the change of representation produces a change in some characteristic values of the distribution. Also, some invariance properties of normal distributions appear as natural within the structure of the sample space. These properties usually get lost when representing these distributions with respect to the Lebesgue measure. The idea of using not only an interpretable space structure, but also to change the measure, is a powerful tool because it leads to results coherent with the interpretation of the measure of difference, and because they are mathematically more straightforward.

Section 2 describes some technical details in an abstract setting concerning Euclidean vector spaces, their reference measure, and the definition of the normal probability density functions on them. Sections 3 and 4 present the application of these concepts to the positive real line and the simplex, respectively, as well as some examples on normal modelling in these constrained spaces.

2. Probability densities in Euclidean vector spaces

Let $E \subseteq \mathbb{R}^D$ be the sample space for a random vector $\mathbf{X}$, i.e. each realization of $\mathbf{X}$ is in E. Assume there exists a one-to-one, differentiable, mapping $h : E \to \mathbb{R}^d$ with $d \leq D$. In the case of the positive real line, $E = \mathbb{R}_+$ and $d = D = 1$, i.e. $\mathbb{R}$ and $\mathbb{R}_+$ have the same dimension. This is not the case of the simplex $\mathcal{S}^D$, which consists of vectors of D positive components adding up to a fixed constant. Only $d = D-1$ components are required to specify a point in it, i.e. the dimension of $\mathcal{S}^D$ is $d = D-1 < D$. The mapping h allows to define a Euclidean vector structure on E just translating the standard properties of $\mathbb{R}^d$ into E. The existence of the mapping h implies some characteristics of E. An important one is that E must have some border set so that h transforms neighbourhoods of this border into neighbourhoods of infinity in $\mathbb{R}^d$. For instance, a sphere in $\mathbb{R}^3$ with a defined pole can be transformed into $\mathbb{R}^2$, but, if no pole is defined, this is no longer possible.

The vector addition or internal operation $\oplus$ and the scalar multiplication or external operation $\odot$ in E are defined as

$$\mathbf{x} \oplus \mathbf{y} = h^{-1}(h(\mathbf{x}) + h(\mathbf{y})) \,, \quad \alpha \odot \mathbf{x} = h^{-1}(\alpha \cdot h(\mathbf{x})) \,,$$

for $\mathbf{x}, \mathbf{y} \in E$ and $\alpha \in \mathbb{R}$. With these definitions, E is a d-dimensional vector space. The metric structure is induced by the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_E = \langle h(\mathbf{x}), h(\mathbf{y}) \rangle$. It implies the norm, $\|\mathbf{x}\|_E = \|h(\mathbf{x})\|$, and the distance, $\mathrm{d}_E(\mathbf{x}, \mathbf{y}) = \mathrm{d}(h(\mathbf{x}), h(\mathbf{y}))$, thus completing the Euclidean vector space structure of E. This structure is derived from the inner product, norm and distance in $\mathbb{R}^d$, denoted as $\langle \cdot, \cdot \rangle$, $\| \cdot \|$ and $\mathrm{d}(\cdot, \cdot)$, respectively. By construction, $h(\mathbf{x})$ is the vector of coordinates of $\mathbf{x} \in E$. The coordinates correspond to the orthonormal basis in E given by the images of the canonical basis in $\mathbb{R}^d$ by h^{-1}. The origin of the space E is then $h^{-1}(\vec{\mathbf{0}})$ where $\vec{\mathbf{0}}$ is the neutral element of $\mathbb{R}^d$ with respect to the ordinary sum. The Lebesgue measure in $\mathbb{R}^d$, λ_d, induces a measure in E, denoted λ_E, using the fact that h is one-to-one and setting $\lambda_E(h^{-1}(B)) = \lambda_d(B)$, for any Borelian B in $\mathbb{R}^d$ (Eaton, 1983). This idea was used in Pawlowsky-Glahn (2003) to define the Aitchison measure on the simplex.

In order to define probability density functions (pdf's) in E, a reference measure is needed. A pdf is the Radon–Nikodym derivative of a probability measure P with respect to a measure on E. When the reference measure is λ_E, we denote the pdf as $f^E = dP/d\lambda_E$. When E is viewed as a subset of $\mathbb{R}^D$, the pdf with respect to the Lebesgue measure λ_D could be eventually considered. However, if $d < D$, the random vector $\mathbf{X}$

cannot be absolutely continuous with respect to λ_D and the pdf does not exist. Our approach, and a more natural way to define a pdf for $\mathbf{X}$, is to start with a pdf for the (random) coordinates $\mathbf{Y} = h(\mathbf{X})$ in $\mathbb{R}^d$. Assume that $f_{\mathbf{Y}}$ is the pdf of $\mathbf{Y}$ with respect to the Lebesgue measure, λ_d, in $\mathbb{R}^d$, i.e. $\mathbf{Y}$ is absolutely continuous with respect to λ_d and the pdf is the Radon–Nikodym derivative $f_{\mathbf{Y}} = dP/d\lambda_d$. The random vector $\mathbf{X}$ is recovered from $\mathbf{Y}$ as $\mathbf{X} = h^{-1}(\mathbf{Y})$. When $D > d$, $h^{-1}(\mathbf{Y})$ can be expressed using only d of its components. Let h_d^{-1} be such a restriction and $\mathbf{X}_d = h_d^{-1}(\mathbf{Y})$. The inverse mapping is denoted by $h_d(\mathbf{X}_d)$ and it holds that $h_d(\mathbf{X}_d) = h(\mathbf{X})$. This means that more than d components of $\mathbf{X}$ are redundant. When $D = d$, the restriction of h^{-1} reduces to $h_d^{-1} = h^{-1}$. For instance, to recover a vector in the simplex of D components from its representation using $d = D-1$ coordinates, one can recover $d = D-1$ components. The remaining one is obtained from the constant sum of all components. The pdf of $\mathbf{X}_d$ with respect to the Lebesgue measure in $\mathbb{R}^d$ is computed using the Jacobian rule

$$f_{\mathbf{X}_d}(\mathbf{x}_d) = \frac{dP}{d\lambda_d}(\mathbf{x}_d) = f_{\mathbf{Y}}(h_d(\mathbf{x}_d)) \cdot \left| \frac{\partial h_d(\mathbf{x}_d)}{\partial \mathbf{x}_d} \right| , \tag{1}$$

where the last term is the d-dimensional Jacobian of h_d. The next step is to express the pdf with respect to λ_E, the compatible measure in E. The chain rule for Radon–Nikodym derivatives implies

$$f^E_{\mathbf{X}_d}(\mathbf{x}_d) = \frac{dP}{d\lambda_E}(\mathbf{x}_d) = \frac{dP}{d\lambda_d}(\mathbf{x}_d) \cdot \frac{d\lambda_d}{d\lambda_E}(\mathbf{x}_d) , \tag{2}$$

and, due to the inverse function theorem, the last derivative is

$$\frac{d\lambda_d}{d\lambda_E}(\mathbf{x}_d) = \left| \frac{\partial h_d^{-1}(h_d(\mathbf{x}_d))}{\partial \mathbf{y}} \right| = \left| \frac{\partial h_d(\mathbf{x}_d)}{\partial \mathbf{x}_d} \right|^{-1} , \tag{3}$$

Substituting (2) and (3) into (1),

$$f^E_{\mathbf{X}}(\mathbf{x}) = \frac{dP}{d\lambda_E}(\mathbf{x}) = f_{\mathbf{Y}}(h(\mathbf{x})) , \tag{4}$$

where the subscripts d have been suppressed, as they only play a role when computing the Jacobian. Difficulties using $f^E_{\mathbf{X}}$, arising from the fact that the integral $\mathrm{P}(A) = \int_A f^E_{\mathbf{X}}(\mathbf{x}) d\lambda_E(\mathbf{x})$ is not an integral with respect to the Lebesgue measure in $\mathbb{R}^d$ but with respect to the Lebesgue type measure in E, are solved working with coordinates. Particularly, they are solved working with coordinates with respect to an orthonormal basis in E. Using (4) the probability of an event $A \subseteq E$ can be computed as $\mathrm{P}(A) = \int_{h(A)} f_{\mathbf{Y}}(h(\mathbf{x}))\, d\lambda_d(h(\mathbf{x}))$ or, in simpler notation, $\mathrm{P}(A) = \int_{h(A)} f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}$.

The representation of the distribution of random variables by pdf's defined with respect to the measure λ_E requires a review of the moments and other characteristics of the pdf's. Following Eaton (1983), the expectation and variance of $\mathbf{X}$ are defined as follows:

Definition 2.1 *Let* $\mathbf{X}$ *be a random variable supported on* E *and* $h : E \to \mathbb{R}^d$ *the coordinate function defined on* E*. The expectation in* E *is*

$$\mathrm{E}^E[\mathbf{X}] = h^{-1}\left(\int_{\mathbb{R}^d} \mathbf{y}\, f_{h(\mathbf{X})}(\mathbf{y})\, d\mathbf{y}\right) \tag{5}$$

$$= h^{-1}\left(\mathrm{E}[h(\mathbf{X})]\right), \tag{6}$$

provided the integral in (5) exists in the Lebesgue sense.

Intuitively, the expectation (5) in E consists of representing the elements of E using coordinates and to integrate using the pdf of the coordinates; the result is transformed back into E. Equation (6) summarizes this result using the standard definition of expectation of the coordinates in $\mathbb{R}^d$.

The variance involves only real expectations and can be identified with the variance of coordinates. Special attention deserves the metric or total variance (Aitchison, 1986; Pawlowsly-Glahn and Egozcue, 2001). Assuming the existence of the integrals, the metric variability of $\mathbf{X}$ with respect to a point $\mathbf{z} \in E$ is defined as $\mathrm{Var}^E[\mathbf{X}, \mathbf{z}] = \mathrm{E}[\mathrm{d}_E^2(\mathbf{X}, \mathbf{z})]$. The minimum metric variability is attained for $\mathbf{z} = \mathrm{E}^E[\mathbf{X}]$, thus supporting the definition in (5)–(6). The metric variance is then

$$\mathrm{Var}^E[\mathbf{X}] = \mathrm{E}[\mathrm{d}_E^2(\mathbf{X}, \mathrm{E}^E[\mathbf{X}])]\ . \tag{7}$$

The mode of a pdf is normally defined as its maximum value, although local maxima are also frequently called modes. However, the shape and, particularly, the maximum values depend on the reference measure taken in the Radon-Nikodym derivatives of the density. Since the Lebesgue measure in the coordinate space, $\mathbb{R}^d$, corresponds to the measure λ_E in E, the mode can be defined as

$$\mathrm{Mode}^E[\mathbf{X}] = \underset{\mathbf{x}\in E}{\mathrm{argmax}}\{f_{\mathbf{X}}^E(\mathbf{x})\} = h^{-1}\left(\underset{\mathbf{y}\in\mathbb{R}^d}{\mathrm{argmax}}\{f_{h(\mathbf{X})}(\mathbf{y})\}\right)\ .$$

3. The positive real line

The real line, with the ordinary sum and product by scalars, has a vector space structure. The ordinary inner product and the Euclidean distance are compatible with these

operations, i.e. they satisfy the translation invariance and the homogeneity properties. But this geometry is not suitable for the positive real line. Confront, for example, some meteorologists with two pairs of samples taken at two rain gauges, $\{5;10\}$ and $\{100;105\}$ in mm, and ask for the difference; quite probably, in the first case they will say there was double the total rain in the second gauge compared to the first, while in the second case they will say it rained a lot but approximately the same. They are assuming a relative measure of difference. Consequently, for them the natural measure of difference is not the usual Euclidean one, and the ordinary vector space structure of $\mathbb{R}$ does not behave suitably for the problem. In fact, problems might appear shifting a positive number (vector) by a negative real number (vector); or multiplying a positive number (vector) by an arbitrary real number (positive or negative scalar), because results can be outside $\mathbb{R}_+$.

There are two operations, $\oplus$, $\odot$, which induce a vector space structure in $\mathbb{R}_+$ (Pawlowsly-Glahn and Egozcue, 2001). In fact, given $x, x^* \in \mathbb{R}_+$, the internal operation, which plays an analogous role to addition in $\mathbb{R}$, is the usual product $x \oplus x^* = x \cdot x^*$ and, for $\alpha \in \mathbb{R}$, the external operation, which plays an analogous role to the product by scalars in $\mathbb{R}$, is $\alpha \odot x = x^\alpha$. An inner product, compatible with $\oplus$ and $\odot$ is $\langle x, x^* \rangle_+ = \ln x \cdot \ln x^*$, which induces a norm, $\|x\|_+ = |\ln x|$, and a distance, $\mathrm{d}_+(x, x^*) = |\ln x^* - \ln x|$, and thus the complete Euclidean vector space structure in $\mathbb{R}_+$. Since $\mathbb{R}_+$ is a one-dimensional vector space, there are only two orthonormal bases: the unit-vector (e) and its inverse element with respect to the internal operation (e^{-1}). From now on the first option is considered and it will be denoted by e. Any $x \in \mathbb{R}_+$ can be expressed as $x = \ln x \odot e = e^{\ln x}$ which reveals that $h(x) = \ln x$ is the coordinate of x with respect to the basis e. The measure λ_+ in $\mathbb{R}_+$ can be defined so that, for λ_1 the Lebesgue measure in $\mathbb{R}^1$, and an interval $(a,b) \subset \mathbb{R}_+$, $\lambda_+(a,b) = \lambda_1(\ln a, \ln b) = |\ln b - \ln a|$ and $d\lambda_+ / d\lambda_1 = 1/x$ (Mateu-Figueras, 2003). Following the notation in Section 2, all these definitions can be obtained by setting $E = \mathbb{R}_+$, $D = d = 1$ and $h(x) = \ln x$. The generalization to $E = \mathbb{R}_+^D$ is straightforward: for $\mathbf{x} \in \mathbb{R}_+^D$, the coordinate function can be defined as $h(\mathbf{x}) = \ln(\mathbf{x})$, where the logarithm applies component-wise.

3.1. The normal distribution on $\mathbb{R}_+$

Using the algebraic-geometric structure in $\mathbb{R}_+$ and the measure λ_+, the normal distribution on $\mathbb{R}_+$ was defined in Mateu-Figueras et al. (2002) through the density function of orthonormal coordinates.

Definition 3.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable $X : \Omega \longrightarrow \mathbb{R}_+$ is said to have a normal on $\mathbb{R}_+$ distribution with two parameters μ and σ^2, written $\mathcal{N}_+(\mu, \sigma^2)$, if its density function with respect to λ_+ is*

$$f_X^+(x) = \frac{dP}{d\lambda_+}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(\ln x - \mu)^2}{\sigma^2} \right), \qquad x \in \mathbb{R}_+. \tag{8}$$

The density (8) is the usual normal density applied to coordinates $\ln x$ as implied by (4) and it is a density in $\mathbb{R}_+$ with respect to the λ_+ measure. This density function is completely restricted to $\mathbb{R}_+$ and its expression corresponds to the law of frequency introduced by McAlister (1879). The probability law corresponding to the density (8) is that of the lognormal distribution, denoted Λ, where μ and σ^2 are the logarithmic mean and variance. The continuous line in Figure 1 represents the density function (8) for $\mu = 0$ and $\sigma^2 = 1$. Note that the areas under the log-normal density f_X are proportional to probabilities, whereas areas under f_X^+, as shown in the figure, are not. In the case of f_X^+ a probability is proportional to the ordinate of the curve times the length of dx, i.e. times $\lambda_+(x, x+dx) = |\ln(x+dx) - \ln(x)|$.

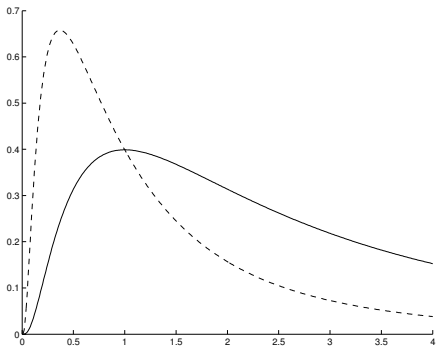


Figure 1: *Density function f_X^+ (——) and f_X (- - - -) with $\mu = 0$ and $\sigma = 1$.*

According to this approach, the normal distribution in $\mathbb{R}_+$ exhibits properties analogous to the normal distribution in $\mathbb{R}$, the most relevant of which are summarized in the following properties. The corresponding proofs are presented in the appendix.

Property 3.1 *Let $X \sim \mathcal{N}_+(\mu, \sigma^2)$, $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$. Then, the random variable $X^* = a \oplus (b \odot X) = a \cdot X^b$ is distributed as $\mathcal{N}_+(\ln a + b\mu, b^2\sigma^2)$.*

Property 3.2 *Let $X \sim \mathcal{N}_+(\mu, \sigma^2)$ and $a \in \mathbb{R}_+$. Then, $f^+_{a \oplus X}(a \oplus x) = f_X^+(x)$, where f_X^+ and $f^+_{a \oplus X}$ represent the probability density functions of the random variables X and $a \oplus X = a \cdot X$, respectively.*

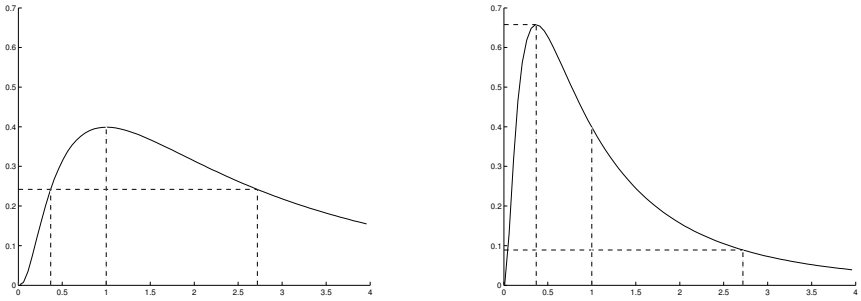
Property 3.3 *If $X \sim \mathcal{N}_+(\mu, \sigma^2)$, then $\mathrm{E}^+[X] = \mathrm{Med}^+[X] = \mathrm{Mode}^+[X] = e^{\mu}$.*

Property 3.4 *If $X \sim \mathcal{N}_+(\mu, \sigma^2)$, then $\mathrm{Var}^+[X] = \sigma^2$.*

Note that property 3.1 implies that the family $\mathcal{N}_+(\mu, \sigma^2)$ is closed under the operations in $\mathbb{R}_+$ and property 3.2 asserts the equivariance under translations in $\mathbb{R}_+$.

The expected value, the median and the mode are elements of the support space $\mathbb{R}_+$, but the variance is only a numerical value which describes the dispersion of X. We are

used to taking the square root of σ^2 as a means to represent intervals centred at the mean and with radius equal to some standard deviations. Such an interval, centred at $\mathrm{E}^+[X] = e^{\mu}$ and with length $2k\sigma$, is $(e^{\mu-k\sigma}, e^{\mu+k\sigma})$, as $\mathrm{d}_+(e^{\mu-k\sigma}, e^{\mu+k\sigma}) = 2k\sigma$. This kind of interval is used in practice (Ahrens, 1954); for instance, under log-normality assumption, predictive intervals in $\mathbb{R}_+$ are computed on log-transformed data, and then back-transformed using exponentiation. In figure 2(a) we represent the interval $(e^{\mu-\sigma}, e^{\mu+\sigma})$ for a $\mathcal{N}_+(\mu, \sigma^2)$ density function with $\mu = 0$ and $\sigma^2 = 1$. It can be shown that it is of minimum length in $\mathbb{R}_+$, and also an isodensity interval, as the distribution is symmetric around e^{μ} in $\mathbb{R}_+$. This symmetry might seem paradoxical, in view of the shape of the density function. But still, it is symmetric within the Euclidean vector space structure of $\mathbb{R}_+$, although certainly not within the space structure of $\mathbb{R}$.



(a) *(b)*

Figure 2: *Dashed line: interval* $(e^{\mu-\sigma}, e^{\mu+\sigma})$*; (a)* f_X^+*, corresponding to* $\mathcal{N}_+(\mu = 0, \sigma^2 = 1)$*, (b)* f_X *corresponding to* $\Lambda(\mu = 0, \sigma^2 = 1)$*.*

An important aspect of this approach is that consistent estimators and exact confidence intervals for the expected value are easy to obtain. It is enough to take exponentials of those obtained from normal theory using log-transformed data, i.e. using the coordinates with respect to the orthonormal basis. Thus, let $x_1, x_2, \ldots, x_n$ be a random sample and $y_i = \ln x_i$ for $i = 1, 2, \ldots, n$. Then, the optimal estimator for the mean of a normal in $\mathbb{R}_+$ population is the geometric mean $(x_1 x_2 \cdots x_n)^{1/n}$, that equals to $e^{\bar{y}}$. An exact $(1-\alpha)100\%$ confidence interval for the mean is $(e^{\bar{y}-t_{\alpha/2}S/\sqrt{n}}, e^{\bar{y}+t_{\alpha/2}S/\sqrt{n}})$, where S denotes the standard deviation of the log transformed sample and $t_{\alpha/2}$ the $(1-\alpha/2)$ t-student ($n-1$ d.f.) quantile.

The normal distribution plays a relevant role in statistics mainly due to its relationship with the central limit theorem. The central limit theorem for the log-normal model is well-known (Aitchison and Brown, 1957). The sums of random variables are replaced by multiplications of positive variables and the limiting distribution is the log-normal. As the central limit theorem concerns the limiting probability law and the multiplication of random variables, it can be translated into terms of the normal in $\mathbb{R}_+$. Let X_1, X_2, ..., X_n be a sequence of random variables in $\mathbb{R}_+$. Define the coordinates of X_i as $Y_i = \ln X_i$ for

$i = 1,2,\ldots$, and assume they are mutually independent and identically distributed with mean μ and variance σ^2. The standardized variable $(1/n)\sum_{i=1}^{n}(Y_i-\mu)/\sigma$ converges in law, for $n \to \infty$, to a random variable with standard normal distribution due to the central limit theorem. Transforming back using $h^{-1} = \exp$, the central limit theorem for the normal in $\mathbb{R}_+$ yields: the random standardized geometric mean

$$Z_n = \frac{1}{n\sigma} \odot \bigoplus_{i=1}^{n} [X_i \ominus \exp(\mu)] = \left[\prod_{i=1}^{n} \left(\frac{X_i}{\exp(\mu)} \right)^{1/\sigma} \right]^{1/n},$$

converges in law, as $n \to \infty$, to a random variable distributed $\mathcal{N}_+(1,0)$.

The role of the operations $\oplus$ and $\odot$ (in $\mathbb{R}_+$ they are the multiplication and powering) in the central limit theorem is remarkable. Its relevance relies on the fact that the operations on random variables involved are interpretable and of frequent use.

3.2. Normal on $\mathbb{R}_+$ versus lognormal

The lognormal distribution has long been recognized as a useful model in the evaluation of random phenomena whose distribution is positive and skew, and specially when dealing with measurements in which the random errors are multiplicative rather than additive. The history of this distribution dates back to 1879, when Galton (1879) observed that the *law of frequency of errors* was incorrect in many groups of phenomena. This observation was based on Fechner's law which, in its approximate and simplest form, is *sensation=log(stimulus)*. According to this law, an error of the same magnitude in excess or in deficiency (in the absolute sense) is not equally probable; therefore, he proposed the geometric mean as a measure of the most probable value instead of the arithmetic mean. This remark was followed by the memoir of McAlister (1879), where a mathematical development concluding with the lognormal distribution was performed. He proposed a practical and easy method for the treatment of a data set grouped around its geometric mean: *convert the observations into logarithms and treat the transformed data set as a series round its arithmetic mean*, and introduced a density function called the *law of frequency* which is the normal density function applied to the log-transformed variable, i.e. the density (8). In order to compute probabilities in given intervals, he introduced also the *law of facility*, nowadays known as the lognormal density function (9).

A unified treatment of lognormal theory is presented in Aitchison and Brown (1957); recent developments are compiled in Crow and Shimizu (1988). A great number of authors use the lognormal model from an applied point of view. Their approach assumes $\mathbb{R}_+$ to be a subset of the real line with the usual Euclidean geometry restricted to it. This is how everybody understands the sentence *an error of the same magnitude in excess or in deficiency* in the same way. One might ask oneself why there is much to say about the lognormal distribution if the data analysis can be referred to the intensively studied normal distribution by taking logarithms. One of the generally accepted reasons is that

parameter estimates are biased if obtained from the inverse transformation. As noted above, the normal on $\mathbb{R}_+$ distribution is well known as the lognormal distribution. But the proposed change of representation produces differences in some properties which are studied below.

Recall that a positive random variable X is said to be lognormally distributed with two parameters μ and σ^2 if $Y = \ln X$ is normally distributed with mean μ and variance σ^2. We write $X \sim \Lambda(\mu, \sigma^2)$. Its probability density function is

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right) & x > 0, \\ 0 & x \leq 0. \end{cases} \tag{9}$$

Comparing (9) with (8), subtle differences can be observed. One is that (9) includes a case for the zero and for the negative values of the random variable. This fact is paradoxical, because the lognormal model is completely restricted to $\mathbb{R}_+$. It is forced by the fact that $\mathbb{R}_+$ is considered as a subset of $\mathbb{R}$ with the restricted structure and, consequently, the variable is assumed to be a real random variable, hence the name *lognormal distribution in* $\mathbb{R}$. Another difference lies in the coefficient $1/x$, the Jacobian, which is necessary to work with real analysis in $\mathbb{R}$. In the lognormal case the Jacobian is necessary because the density is written with respect to the Lebesgue measure, but in the normal in $\mathbb{R}_+$ case the Jacobian is not necessary as the density is expressed with respect to λ_+. More obvious differences are that (9) is not equivariant under translations and is not symmetric around the mean. Note that for the lognormal case, $\mathrm{E}[X] = e^{\mu + \frac{1}{2}\sigma^2}$, the $\mathrm{Med}[X] = e^{\mu}$ and $\mathrm{Mod}[X] = e^{\mu - \sigma^2}$. Using our approach a different expected value and a different mode are obtained, while the value for the median is the same. The dashed line in Figure 1 illustrates the probability density function (9) for $\mu = 0$ and $\sigma^2 = 1$. It clearly differs from the density function (8) plotted in continuous line.

As for the normal in $\mathbb{R}_+$ case, an interval centered at the mean and with radius equal to some standard deviations can be represented for the lognormal in $\mathbb{R}$. Considering $\mathbb{R}_+$ as a subset of $\mathbb{R}$ with an Euclidean structure, such an interval is: $(\mathrm{E}[X] - k\mathrm{Stdev}[X], \mathrm{E}[X] + k\mathrm{Stdev}[X])$. But it has no sense, because the lower bound might take a negative value. For example, for $\mu = 0$ and $\sigma^2 = 1$, the above interval with $k = 1$ is $(-0.512, 3.810)$. This is the reason why sometimes intervals $(e^{\mu - k\sigma}, e^{\mu + k\sigma})$ are used, which are considered to be *non-optimal*, because they are neither isodensity intervals, nor do they have minimum length. In Figure 2(b) we represent the interval $(e^{\mu-\sigma}, e^{\mu+\sigma})$ for $\Lambda(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma^2 = 1$. It is clear that in the bounds of the interval the density function takes different values.

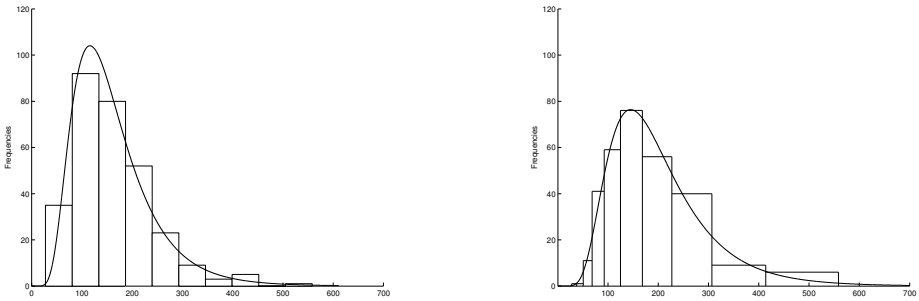
Consistent estimators and exact confidence intervals for the mean and the variance of a lognormal variable are difficult to compute. Early methods are summarised by Aitchison and Brown (1957) and Crow and Shimizu (1988). In the literature an extensive number of procedures and discussions can be found. It is not the objective of this paper

to summarise them all and to provide a complete set of formulas. For the mean, the term $e^{\bar{y}}$ multiplied by a term depending on σ, expressed as an infinite series or tabulated in a set of tables, is obtained in most cases (Aitchison and Brown, 1957; Krige, 1981; Clark and Harper, 2000). For example, Sichel (1996) optimal estimator for the mean of a lognormal population is used by Clark and Harper (2000). This estimator is obtained as $e^{\bar{y}}\gamma$, where γ is a bias correction factor depending on the variance and the size of the data set. It is tabulated in a set of tables. A similar bias correction factor is used to obtain confidence intervals on the population mean (Clark and Harper, 2000). Nevertheless, in practical situations, sometimes the sample geometric mean, $e^{\bar{y}}$, is used to represent the mean and in some cases also to represent the mode of a lognormally distributed variable (Herdan, 1960). But, as adverted by Crow and Shimizu (1988), those affirmations cannot be justified using the lognormal theory. On the contrary, using the normal in $\mathbb{R}_+$ approach, those affirmations are completely justified.

3.3. Example

The differences between using a density with respect to the Lebesgue measure λ or a density with respect to the measure λ_+ can be best appreciated in practice. In order to compare the classical lognormal estimators with those obtained by the normal in $\mathbb{R}_+$ approach, we have simulated 300 samples representing sizes of oil fields in thousands of barrels, a geological variable often lognormally modelled (Davis, 1986). The objective with this simple example is to estimate a suitable location parameter and a corresponding confidence interval and to compare the results obtained using the lognormal approach with the results obtained using the proposed approach. Using the classical lognormal procedures and Table A2 provided by Aitchison and Brown (1957) we obtain 161.96 as an estimate for the mean. Afterwards, using Tables 1,2 and 3 given by Krige (1981), we obtain 162.00 and $(150.31, 176.78)$ as an estimate and approximate 90% confidence interval for the mean. Also, using Tables 7, 8(b) and 8(e) provided by Clark and Harper (2000), we could apply Sichel's bias correction to obtain 161.86 and $(144.07, 188.39)$ as the optimal estimator and confidence interval for the mean in the context of the lognormal approach.

Using the normal in $\mathbb{R}_+$ approach we obtain 145.04 as the estimate for the mean and $(138.70, 151.68)$ as the exact 90% confidence interval for the mean. Logically, different values are obtained, as different methodologies are used. The mean is obtained as $e^{\bar{y}}$ and it is not necessary to apply any bias correction, as unbiasedness is in this case equivalent to unbiasedness in coordinates. The confidence interval is obtained as $(e^{\bar{y}-t_{0.05}S/\sqrt{n}}, e^{\bar{y}+t_{0.05}S/\sqrt{n}})$ where S denotes the sample standard deviation of the log transformed sample. Note that only exponentials of the mean and the 90% confidence interval obtained from normal theory using log-transformed data are taken. As can be observed, the differences to those obtained using the lognormal approach are important. With the normal in $\mathbb{R}_+$ a much more conservative result is obtained, although it is consistent with the assumed geometry of $\mathbb{R}_+$.



(a) *(b)*

Figure 3: *Simulated sample, $n = 300$. (a) Histogram and fitted lognormal density; the Lebesgue-lengths of the bins are equal. (b) Display analogous to a histogram and fitted normal in $\mathbb{R}_+$ density; the λ_+-lengths of the bins are equal.*

In order to compare graphically the normal in $\mathbb{R}_+$ and the lognormal approaches we can represent the histogram with the corresponding fitted densities. In Figures 3(a) and 3(b) the histogram with the fitted lognormal and normal in $\mathbb{R}_+$ densities are shown. Note that the intervals of the histogram are of equal length in both cases, as the absolute Euclidean distance is used in (a) and the relative distance in $\mathbb{R}_+$ is used in (b) to compute them. Thus, (b) is a display analogous to a histogram, based on the structure defined in Section 3. Finally, in Figure 4 the histogram of the log-transformed data or, equivalently, of the coordinates with respect to the orthonormal basis, with the fitted normal density, is provided. This last figure is adequate using both methodologies, but in this case we have chosen exactly the same intervals as in Figure 3(b). This is only possible using the normal on $\mathbb{R}_+$ approach, as the intervals on the positive real line have corresponding intervals in the space of coordinates.

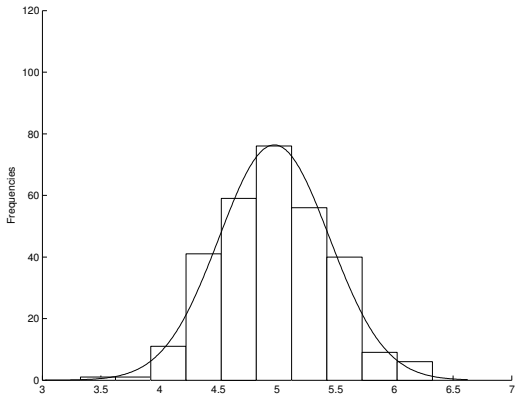


Figure 4: *Simulated sample, $n = 300$. Histogram of the log-transformed sample with the fitted normal density. The bins of the histogram are the log-images of the bins shown in Figure 3(b).*

The normal on $\mathbb{R}_+$ density model and its properties have been applied in a spatial context, and the results have been compared with those obtained with the classical lognormal kriging approach (Tolosana-Delgado and Pawlowsky-Glahn, 2007). Using this approach, problems of non-optimality, robustness and preservation of distribution disappear.

4. The simplex

Compositional data are parts of some whole which carry only relative information. Typical examples are parts per unit, percentages, ppm, or moles per liter. When constrained to sum to a constant, their sample space is the D-part simplex, $\mathcal{S}^D = \{\mathbf{x} = (x_1, x_2, \ldots, x_D)^\mathsf{T} : x_1 > 0, x_2 > 0, \ldots, x_D > 0; \sum_{i=1}^D x_i = \kappa\}$, where $^\mathsf{T}$ stands for transpose and κ is a constant, set to 1 for simplicity in (Aitchison, 1982).

The simplex $\mathcal{S}^D$ has a $(D-1)$-dimensional Euclidean vector space structure (Billheimer et al., 2001; Pawlowsly-Glahn and Egozcue, 2001) with the following operations. Let $\mathcal{C}(\cdot)$ denote the closure operation which normalises any vector $\mathbf{x}$ to a constant sum (Aitchison, 1982), $\mathbf{x}, \mathbf{x}^* \in \mathcal{S}^D$, and $\alpha \in \mathbb{R}$. The internal operation, called *perturbation*, is defined as $\mathbf{x} \oplus \mathbf{x}^* = \mathcal{C}(x_1 x_1^*, x_2 x_2^*, \ldots, x_D x_D^*)^\mathsf{T}$, with inverse $\mathbf{x} \ominus \mathbf{x}^* = \mathcal{C}(x_1/x_1^*, x_2/x_2^*, \ldots, x_D/x_D^*)^\mathsf{T}$. The external operation, called *powering*, is defined as $\alpha \odot \mathbf{x} = \mathcal{C}(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha)^\mathsf{T}$, and the inner product as

$$\langle \mathbf{x}, \mathbf{x}^* \rangle_a = \frac{1}{D} \sum_{i<j} \ln \frac{x_i}{x_j} \ln \frac{x_i^*}{x_j^*} \,. \tag{10}$$

The associated squared distance

$$\mathrm{d}_a^2(\mathbf{x}, \mathbf{x}^*) = \frac{1}{D} \sum_{i<j} \left(\ln \frac{x_i}{x_j} - \ln \frac{x_i^*}{x_j^*} \right)^2 ,$$

is relative and satisfies standard properties of a distance (Martín-Fernández et al., 1998), i.e. $\mathrm{d}_a(\mathbf{x}, \mathbf{x}^*) = \mathrm{d}_a(\mathbf{a} \oplus \mathbf{x}, \mathbf{a} \oplus \mathbf{x}^*)$ and $\mathrm{d}_a(\alpha \odot \mathbf{x}, \alpha \odot \mathbf{x}^*) = \mid \alpha \mid \mathrm{d}_a(\mathbf{x}, \mathbf{x}^*)$. The corresponding geometry is known as *Aitchison geometry*, and therefore the subindex a is used.

The inner product (10) and its associated norm, $\|\mathbf{x}\|_a = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a}$, ensure the existence of orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{D-1}\}$ which, for a given basis, lead to a unique expression of a composition $\mathbf{x}$ as a linear combination,

$$\mathbf{x} = (\langle \mathbf{x}, \mathbf{e}_1 \rangle_a \odot \mathbf{e}_1) \oplus (\langle \mathbf{x}, \mathbf{e}_2 \rangle_a \odot \mathbf{e}_2) \oplus \ldots \oplus (\langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a \odot \mathbf{e}_{D-1}).$$

In inner product spaces, an orthonormal basis is not uniquely determined. It is not straightforward to determine which one is the most appropriate to solve a specific problem, but a promising strategy, based on binary partitions, has been developed in

(Egozcue and Pawlowsky-Glahn, 2005). Here, whenever a specific basis is needed, the basis given in (Egozcue et al., 2003) is used. In this basis, the coordinates of $\mathbf{x} \in \mathcal{S}^D$ are

$$y_i = \frac{1}{\sqrt{i(i+1)}} \ln\left(\frac{x_1 x_2 \cdots x_i}{x_{i+1}^i}\right), \qquad i = 1,2,\ldots,D-1\,. \tag{11}$$

The coordinates in this particular basis are denoted $\mathrm{ilr}(\mathbf{x})$ to emphasise the fact that this coordinate transformation is an isometric mapping from $\mathcal{S}^D$ to $\mathbb{R}^{D-1}$ and that the coordinates are log-ratios (Egozcue et al., 2003). The important point is that, once an orthonormal basis has been chosen, all standard statistical methods can be applied to the coordinates and transferred to the simplex preserving their properties. This is what we call the *principle of working on coordinates* (Mateu-Figueras et al., 2011). As stated in Section 2., the Lebesgue measure in the space of coordinates induces a measure in $\mathcal{S}^D$, denoted here as λ_a and called Aitchison measure on $\mathcal{S}^D$. This measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_{D-1}$, and the relationship between them is $|d\lambda_a/d\lambda_{D-1}| = (\sqrt{D}\; x_1 x_2 \cdots x_D)^{-1}$ (Pawlowsky-Glahn, 2003). Following the notation in Section 2., all these definitions can be obtained by setting $E = \mathcal{S}^D$ and $d = D-1$.

For later use, the concept of subcomposition is required. For $1 < C < D$, a C-part subcomposition, $\mathbf{x}_S$, from a D-part composition, $\mathbf{x}$, can be obtained as $\mathbf{x}_S = \mathcal{C}(\mathbf{S}\mathbf{x})$, 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 (Aitchison, 1986). A subcomposition can be regarded as a composition in a simplex with fewer parts, and thus as an element of a space of lower dimension.

4.1. Some basic statistical concepts in the simplex

A random composition $\mathbf{X}$ is a random vector with $\mathcal{S}^D$ as sample space. In the literature laws of probability over $\mathcal{S}^D$ using the Lebesgue measure can be found. Consequently, the probabilities or any moment are computed using the classical definition. But some usual elements appear to be of little use when working with real situations. One typical example is the expected value, in the sense that frequently it does not lie within the bulk of the sample. As an alternative, the geometric interpretation of the expected value has been used to define the centre, $\mathrm{cen}[\mathbf{X}]$, of a random composition. It is defined as the composition that minimises the expression $\mathrm{E}[\mathrm{d}_a^2(\mathbf{X}, \mathrm{cen}[\mathbf{X}])]$ (Aitchison, 1997; Pawlowsly-Glahn and Egozcue, 2001). The result is $\mathrm{cen}[\mathbf{X}] = \mathcal{C}(\exp(\mathrm{E}[\ln \mathbf{X}]))$, which can be rewritten as (Egozcue et al., 2003) $\mathrm{cen}[\mathbf{X}] = \mathrm{ilr}^{-1}(\mathrm{E}[\mathrm{ilr}(\mathbf{X})])$, or, in general terms, as

$$\mathrm{cen}[\mathbf{X}] = h^{-1}\left(\mathrm{E}\left[h(\mathbf{X})\right]\right).$$

The centre of a random composition is equal to the expectation in $\mathcal{S}^D$ defined in Equation (5) in Section 2. This is an important result because if a law of probability on $\mathcal{S}^D$ is defined using the classical approach, this equality does not hold.

As already mentioned, traditionally the simplex has been considered as a subset of real space and, consequently, the laws of probability have been defined using the standard approach. This is the case for families of distributions like the Dirichlet (Monti et al., 2011), the additive logistic normal (Aitchison, 1982), the additive logistic skew-normal (Mateu-Figueras et al., 2005), or those defined using the Box-Cox family of transformations (Barceló-Vidal, 1996). Except for the Dirichlet, these laws of probability are defined using transformations from the simplex to real space. Two of these transformations will appear later herein, the additive log-ratio (alr) and the centred log-ratio (clr),

$$\operatorname{alr}(\mathbf{x}) = \left(\ln\left(\frac{x_1}{x_D}\right), \ldots, \ln\left(\frac{x_{D-1}}{x_D}\right)\right)^{\mathsf{T}}, \tag{12}$$

$$\operatorname{clr}(\mathbf{x}) = \left(\ln\left(\frac{x_1}{g(\mathbf{x})}\right), \ldots, \ln\left(\frac{x_D}{g(\mathbf{x})}\right)\right)^{\mathsf{T}}, \tag{13}$$

where $g(\mathbf{x})$ is the geometric mean of the components of $\mathbf{x}$. The relationship between the alr and the clr transformations is provided by Aitchison (1986, p.92). The relationships between the alr, clr and ilr transformations are provided by Egozcue et al. (2003).

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

Aitchison (1986) introduced the additive logistic normal (aln) distribution. A random variable on the simplex is aln distributed if the alr transformed random composition (12) has a multivariate normal distribution. The alr transformation is a representation of a composition using coordinates with respect to an oblique basis of the simplex (Egozcue et al., 2011). An equivalent definition can be formulated using orthonormal coordinates. Additionally, the Aitchison measure on the simplex is used to obtain the corresponding density function. Although the following definition is formally different from that of the aln, it corresponds to the same probability law with a different parametrisation.

Consider an orthonormal basis on $\mathcal{S}^D$ and let $h(\cdot)$ be the corresponding orthonormal coordinates.

Definition 4.1 *Let $(\Omega, \mathcal{F}, p)$ be 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, if its moment generating function is*

$$M(\mathbf{t}) = \mathrm{E}[\exp(\mathbf{t}^{\mathsf{T}} h(\mathbf{X}))] = \exp\left(\boldsymbol{\mu}^{\mathsf{T}}\mathbf{t} + \frac{1}{2}\mathbf{t}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{t}\right),$$

where $\mathbf{t}$ *is a* $D-1$ *real vector. The* $D-1$ *vector* $\boldsymbol{\mu}$ *is the mean expressed in coordinates and the* $(D-1)\times(D-1)$ *matrix* $\boldsymbol{\Sigma}$ *is the covariance matrix of the coordinates.*

The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ depend on the specific orthonormal basis selected. If the covariance matrix $\boldsymbol{\Sigma}$ is non singular, it can be inverted and the regular normal distribution can be defined as follows:

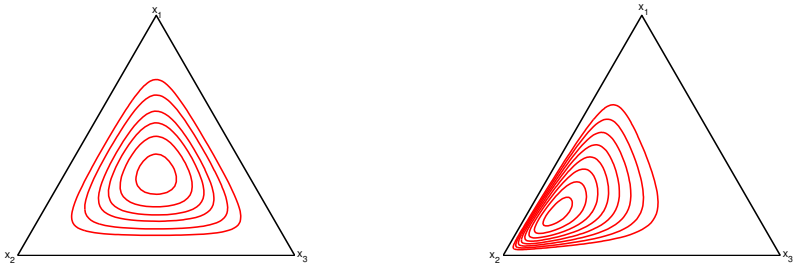
Definition 4.2 *Let* $(\Omega,\mathcal{F},p)$ *be a probability space. A random composition* $\mathbf{X}:\Omega\longrightarrow\mathcal{S}^D$ *is said to have a regular normal on* $\mathcal{S}^D$ *distribution, with parameters* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$*, if its density function is*

$$f_{\mathbf{X}}^{\mathcal{S}}(\mathbf{x})=(2\pi)^{-(D-1)/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}\left(h(\mathbf{x})-\boldsymbol{\mu}\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(h(\mathbf{x})-\mu\right)\right). \qquad (14)$$

The notation $\mathbf{X}\sim\mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is used. The subscript $\mathcal{S}$ indicates that it is a density on the simplex, i.e. a Radon-Nykodym derivative with respect to the Aitchison measure on $\mathcal{S}^D$; the superscript D indicates the number of parts of the composition. Figure 5 shows the isodensity curves of two normal densities on $\mathcal{S}^3$ taking the particular basis given by Egozcue et al. (2003) and using a ternary diagram as a convenient and simple way for representing 3-part compositions (see Aitchison, 1986, p.6). The isodensity curves are not equidistant, the levels are only chosen in order to clearly show the shape of the density function. To understand Figure 5, it should be remarked that the areas on the ternary diagram are computed according λ_a, which significantly differs from the usual Lebesgue area intuitively assigned to the triangle interior. The differences of assigned areas are specially dramatic near the edges of the triangle.

The density (14) is the usual normal density applied to coordinates $h(\mathbf{x})$ as implied by (4). It is a density in $\mathcal{S}^D$ with respect to the λ_a measure. The same strategy is used by Mateu-Figueras and Pawlowsky-Glahn (2007) to define the skew-normal in $\mathcal{S}^D$ law.

The main properties of this model follow. A complete proof of each property can be found in the appendix. The proofs are straightforward for a reader familiar with compositional data analysis.



(a) *(b)*

Figure 5: *Isodensity plots of two* $\mathcal{N}_{\mathcal{S}}^{3}(\mu,\boldsymbol{\Sigma})$ *with (a)* $\mu=(0,0)$*, (b)* $\boldsymbol{\mu}=(-1,1)$ *and* $\boldsymbol{\Sigma}=Id$*.*

Property 4.1 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{a} \in \mathcal{S}^D$ *and* $b \in \mathbb{R}$. *Then, the D-part random composition* $\mathbf{X}^* = \mathbf{a} \oplus (b \odot \mathbf{x})$ *has a* $\mathcal{N}_{\mathcal{S}}^{D}(h(\mathbf{a}) + b\boldsymbol{\mu}, b^2\boldsymbol{\Sigma})$ *distribution.*

Property 4.2 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\mathbf{a} \in \mathcal{S}^D$. *Then* $f^{\mathcal{S}}_{\mathbf{a}\oplus\mathbf{X}}(\mathbf{a} \oplus \mathbf{x}) = f^{\mathcal{S}}_{\mathbf{X}}(\mathbf{x})$, *where* $f^{\mathcal{S}}_{\mathbf{X}}$ *and* $f^{\mathcal{S}}_{\mathbf{a}\oplus\mathbf{X}}$ *represent the density functions of the random compositions* $\mathbf{X}$ *and* $\mathbf{a} \oplus \mathbf{X}$, *respectively.*

Property 4.3 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\mathbf{X}_P = \mathbf{P}\,\mathbf{X}$, *the random composition* $\mathbf{X}$ *with the parts reordered by a permutation matrix* $\mathbf{P}$. *Then* $\mathbf{X}_P \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu}_P,\boldsymbol{\Sigma}_P)$ *with* $\boldsymbol{\mu}_P = \mathbf{U}^{\mathsf{T}}\mathbf{P}\mathbf{U}\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_P = (\mathbf{U}^{\mathsf{T}}\mathbf{P}\mathbf{U})\boldsymbol{\Sigma}(\mathbf{U}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}}\mathbf{U})$, *where* $\mathbf{U}$ *is a* $D \times (D-1)$ *matrix with the* clr *transformation of an orthonormal basis of* $\mathcal{S}^D$ *as columns.*

Property 4.4 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\mathbf{X}_S = \mathcal{C}(\mathbf{S}\mathbf{X})$, *the C-part random subcomposition obtained from the* $C \times D$ *selection matrix* $\mathbf{S}$. *Then* $\mathbf{X}_S \sim \mathcal{N}_{\mathcal{S}}^{C}(\boldsymbol{\mu}_S,\boldsymbol{\Sigma}_S)$, *with* $\mu_S = \mathbf{U}^{*\mathsf{T}}\mathbf{S}\mathbf{U}\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_S = (\mathbf{U}^{*\mathsf{T}}\mathbf{S}\mathbf{U})\boldsymbol{\Sigma}(\mathbf{U}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{*})$, *where* $\mathbf{U}$ *is a* $D \times (D-1)$ *matrix with the* clr *transformation of an orthonormal basis of* $\mathcal{S}^D$ *as columns and* $\mathbf{U}^*$ *is a* $C \times (C-1)$ *matrix with the* clr *transformation of an orthonormal basis of* $\mathcal{S}^C$ *as columns.*

Property 4.5 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. *Then, the expected value in* $\mathcal{S}^D$ *is*

$$\operatorname{cen}[\mathbf{X}] = \mathrm{E}^a[\mathbf{X}] = h^{-1}(\boldsymbol{\mu}) \,,$$

independently of the orthonormal basis of $\mathcal{S}^D$ *for which the coordinate mapping* h *is defined.*

Property 4.6 *Let* $\mathbf{X} \sim \mathcal{N}_{\mathcal{S}}^{D}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. *The metric variance of* $\mathbf{X}$ *is* $\operatorname{Var}^a[\mathbf{X}] = \operatorname{trace}(\boldsymbol{\Sigma})$.

Property 4.1 states that the normal on $\mathcal{S}^D$ law is closed under perturbation and powering. Property 4.2 states that it is equivariant under perturbation. This is important, because when working with compositional data the centring operation (Martín-Fernández et al., 1999), a perturbation using the inverse of the centre of the data set, is often applied in practice to better visualise and interpret the pattern of variability (von Eynatten et al., 2002). Properties 4.3 and 4.4 show that the normal on $\mathcal{S}^D$ family is closed under permutation and subcompositions.

Given a compositional data set the estimates of parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be computed applying the maximum likelihood procedure to the coordinates. The estimated values $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ allow to compute the estimates of the centre (expected value in $\mathcal{S}^D$) and metric variance of the random composition $\mathbf{X}$, as

$$\begin{aligned} \widehat{\mathrm{E}^a[\mathbf{X}]} &= (\hat{\mu}_1 \odot \mathbf{e}_1) \oplus \cdots \oplus (\hat{\mu}_{D-1} \odot \mathbf{e}_{D-1}) = h^{-1}(\hat{\boldsymbol{\mu}}) \,, \\ \widehat{\operatorname{Var}^a[\mathbf{X}]} &= \operatorname{trace}\left(\widehat{\boldsymbol{\Sigma}}\right) . \end{aligned}$$

To validate the distributional assumption of normality on $\mathcal{S}^D$, some goodness-of-fit tests of the multivariate normal distribution have to be applied to the coordinates of the sample data set. There is a large battery of possible tests but, as suggested by Aitchison (1986), it is reasonable to start testing the normality of each marginal using empirical distribution function tests. Unfortunately, the univariate normality of each component is a necessary but not sufficient condition for the normality of the whole vector. Also, these univariate tests depend on the orthonormal basis chosen. This difficulty does not depend on the proposed methodology, as the same problem appears when working with laws of probability defined using transformations and the Lebesgue measure in $\mathcal{S}^D$ (Aitchison et al., 2003). The multivariate normal model can also be validated considering the Mahalanobis distance $(h(\mathbf{X})-\hat{\boldsymbol{\mu}})^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}^{-1}(h(\mathbf{X})-\hat{\boldsymbol{\mu}})$, which is sampled from a χ^2_{D-1}-distribution if the fitted model is appropriate. In this case, the dependence on the chosen orthonormal basis disappears (Stevens, 1986). The use of empirical distribution function tests is also suggested in (Aitchison, 1986).

As mentioned at the beginning of this section, the parametrisation used depends on the selected orthonormal basis. In fact, the vector $\boldsymbol{\mu}$ and the matrix $\boldsymbol{\Sigma}$ are the mean and the covariance matrix of the coordinates random vector $h(\mathbf{X})$. Nevertheless, the distribution can also be defined using object parameters (Tolosana-Delgado, 2005; Eaton, 1983). The idea under the object parametrisation is to define the model independently of the coordinates used for representation. The mean vector of the coordinates $\boldsymbol{\mu}$ is the coordinate representation of a composition in $\mathcal{S}^D$, $\boldsymbol{\mu}_{\mathcal{S}} = h^{-1}(\boldsymbol{\mu})$, that does not depend on the selected basis. The covariance matrix $\boldsymbol{\Sigma}$ can be interpreted as the representation of a symmetric positive semidefinite endomorphism $\Sigma_{\mathcal{S}}$ on $\mathcal{S}^D$. For each choice of basis in $\mathcal{S}^D$ the endomorphism has a different matrix representation $\boldsymbol{\Sigma}$, but the endomorphism itself remains the same. Then, as proposed by Tolosana-Delgado (2005), the normal on $\mathcal{S}^D$ distribution can be defined as

Definition 4.3 *A random composition* $\mathbf{X}$ *is said to follow a normal on* $\mathcal{S}^D$ *distribution with a given mean vector* $\boldsymbol{\mu}_{\mathcal{S}}$ *and a positive semidefinite symmetric endomorphism* $\Sigma_{\mathcal{S}}$ *on* $\mathcal{S}^D$*, if for any testing vector* $\mathbf{x}$*, the projection* $\langle \mathbf{x}, \mathbf{X}\rangle_a$ *follows a univariate normal distribution on* $\mathbb{R}$ *with expectation* $\langle \mathbf{x}, \boldsymbol{\mu}_{\mathcal{S}}\rangle_a$ *and variance* $\langle \mathbf{x}, \Sigma_{\mathcal{S}}\mathbf{x}\rangle_a$*.*

If $\Sigma_{\mathcal{S}}$ is positive definite, the density with respect to λ_a is

$$f^{\mathcal{S}}_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-(D-1)/2}|\Sigma_{\mathcal{S}}|^{-1/2}\exp\left(-\frac{1}{2}\langle \mathbf{x}\ominus\boldsymbol{\mu}_{\mathcal{S}}, \Sigma_{\mathcal{S}}^{-1}(\mathbf{x}\ominus\boldsymbol{\mu}_{\mathcal{S}})\rangle_a\right),$$

where $|\Sigma_{\mathcal{S}}|$ is the determinant of the endomorphism $\Sigma_{\mathcal{S}}$.

As noted by Tolosana-Delgado (2005), given a basis, object definitions may be identified with coordinate ones, proving that the coordinate approach gives the same results whichever basis is used.

4.3. The central limit theorem in S^D

The relevant role of the normal distribution for real vectors is due to the central limit theorem and related properties. The normal distribution in the simplex satisfies a central limit theorem in S^D, as stated in (Aitchison, 1986) to characterize the logistic normal distribution. In the present context, the multivariate central limit theorem (Kocherlakota and Kocherlakota, 1982) holds for coordinates. Let $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ be a sequence of random compositions in $\mathcal{S}^D$. Consider their coordinates with respect to an arbitrary orthonormal basis $\mathbf{Y}_i = h(\mathbf{X}_i) \in \mathbb{R}^{D-1}$, $i = 1,2,\ldots,n$. Assume that the coordinate vectors $\mathbf{Y}_i$ are mutually independent and identically distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; then, being $\overline{\mathbf{Y}}_n = n^{-1}\sum_{i=1}^n \mathbf{Y}_i$, the random vector $\sqrt{n}\left(\overline{\mathbf{Y}}_n - \boldsymbol{\mu}\right)$ converges in distribution to the multivariate normal $\mathcal{N}(0,\boldsymbol{\Sigma})$ as $n \to \infty$. These random vectors are coordinates of the random compositions

$$\sqrt{n} \odot \left(\overline{\mathbf{X}}_n \ominus h^{-1}(\mu)\right) \ , \ \overline{\mathbf{X}}_n = \frac{1}{n} \odot \bigoplus_{i=1}^{n} \mathbf{X}_i = \mathcal{C}\left(\exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln \mathbf{X}_i\right)\right) \ . \tag{15}$$

The random compositions (15) converge in distribution to $\mathcal{N}_{\mathcal{S}}^D(0,\boldsymbol{\Sigma})$, and the multivariate central limit theorem holds in the simplex. The reference operation is the perturbation and the corresponding average equals to the closed geometric mean composition, that is, the geometric mean vector followed by the closure operation. This result justifies the name of normal in the simplex for the studied distribution. However, the relevance of a central limit theorem in this context relays on the interpretation of the average $\overline{\mathbf{X}}_n$ or just the perturbation of random variables in the simplex. Perturbation has many intuitive interpretations depending on the applied context. The following example of concentrations may be illustrative. Suppose that $\mathbf{Z}$ contains the concentrations of D pollutants in a sample of water. The sample is filtered using a permeable membrane with transfer function $\mathbf{X}$, i.e. the components are multiplicative factors applied to the concentrations in $\mathbf{Z}$. The perturbation $\mathbf{Z} \oplus \mathbf{X}$ expresses the output concentrations after applying the filter $\mathbf{X}$. As the filtering membrane is replaced by another similar one after each filtering process, we can assume that $\mathbf{X}$ is random. In order to express the random effect of a filtering membrane we perform a sequence of filtering experiments with similar but different transfer functions. The mean transfer function, say $\overline{\mathbf{X}}_n$, will be approximately distributed as a normal in the simplex as predicted by the central limit theorem.

4.4. The normal on S^D vs the additive logistic normal

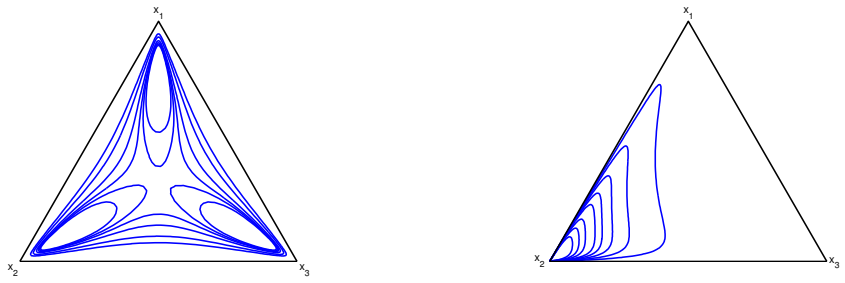
The normal on the simplex is well known as the logistic normal distribution. Nevertheless, the proposed change of representation produces differences in some properties. In this section we study these changes.

The approach used in (Aitchison, 1982) to define the additive logistic normal law on the simplex is standard: transform the random composition from the simplex to real

space, define the density function of the transformed vector, return to the simplex using the change of variable theorem. The result is a density function for the initial random composition with respect to the Lebesgue measure. Therefore, a random composition is said to have an additive logistic normal distribution (aln) when the additive log-ratio (alr) transformed vector – see Eq. (12) – has a normal distribution. Note that this definition does not explicitly state that the change of variable theorem is used. But this is the principal difference between the approach based on working with transformations, and the new approach, based on working with coordinates.

The aln model was initially defined using the additive log-ratio transformation. Using the matrix relationship between log-ratio transformations (Egozcue et al., 2003) the density function in terms of an isometric log-ratio transformation is obtained. Consequently, we can define the logistic normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, with density function:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{(2\pi)^{-(D-1)/2} \mid \boldsymbol{\Sigma} \mid^{-1/2}}{\sqrt{D} x_1 x_2 \cdots x_D} \exp\left(-\frac{1}{2} (\mathrm{ilr}(\mathbf{x}) - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathrm{ilr}(\mathbf{x}) - \mu) \right). \qquad (16)$$



(a) *(b)*

Figure 6: *Isodensity plots of two logistic normal densities with (a)* $\boldsymbol{\mu} = (0,0)$*, (b)* $\boldsymbol{\mu} = (-1,1)$ *and* $\boldsymbol{\Sigma} = Id$*.*

To easily compare both approaches we will use the normal density on the simplex taking the basis given by Egozcue et al. (2003) and consequently the ilr vector stated in (11). Nevertheless, any orthonormal basis could be considered, as the vector $\mathrm{ilr}(\mathbf{x})$ can be obtained from $h(\mathbf{x})$ and the corresponding change of basis matrix. The only difference between expressions (14) and (16) is the term $(\sqrt{D} x_1 x_2 \cdots x_D)^{-1}$, the Jacobian of the isometric log-ratio transformation that reflects the change of the measure on $\mathcal{S}^D$. The influence of this term can be observed in the isodensity curves in Figure 6, where areas on the triangle are computed using the ordinary Lebesgue measure. These curves can be compared with the curves in Figure 5, where areas where computed using λ_a. The differences between Figures 5 and 6 are obvious; in particular the tri-modality in Figure 6(a). This behaviour is not exclusive of the logistic normal density, bimodality is also present with Beta or Dirichlet densities when their parameters are close to 0 and the Lebesgue measure is considered. In Figure 6(b) a single mode can be observed;

nevertheless, the position and the shape of the curves are not the same as in Figure 5(b), the corresponding normal on $\mathcal{S}^3$.

Another difference is the moments of any order. The expression of the density function plays a fundamental role when any moment is computed. The density (16) is a classical density, consequently moments are computed using the standard definition. Obviously, the results are not the same as in the normal on $\mathcal{S}^D$ case. For example, the expected value of an aln density, denoted as $\mathrm{E}[\mathbf{X}]$, exists but numerical procedures have to be applied (see Aitchison, 1986, p.116) to find it and the result is not the same as in property 4.5. Using our approach, the centre of a random composition, denoted as $\mathrm{cen}[\mathbf{X}]$ (Aitchison, 1997), is obtained when the expected value $\mathrm{E}^a[\mathbf{X}]$ is computed. Consequently, it is not necessary to define new characteristic parameters. Using the classical definition, e.g. the expected value, a representative location parameter is obtained. Remember that the centre of a random composition was introduced by Aitchison (1997) because he perceived that the usual expected value $\mathrm{E}[\mathbf{X}]$ is not a representative location parameter. This discrepancy appears because Aitchison (1982) adopts perturbation and powering as operations in the sample space, but uses the density function with respect to the Lebesgue measure, thus assuming for the density a measure not compatible with the operations.

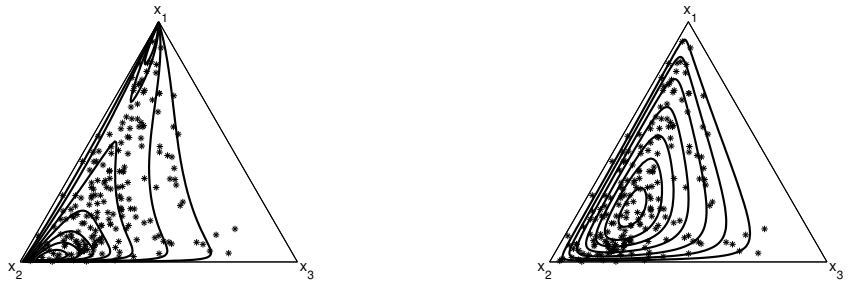
Some coincidences can be found as well. The closure under perturbation, powering, permutation and subcompositions of the logistic normal density model is proven by Aitchison (1986), and stated in Properties 4.1,4.3 and 4.4 for the normal on $\mathcal{S}^D$ density model. Nevertheless, the logistic normal class is not equivariant under perturbation, i.e. $f_{\mathrm{a}\oplus\mathbf{X}}(\mathrm{a}\oplus\mathbf{x}) \neq f_{\mathbf{X}}(\mathbf{x})$.

In summary, the essential differences between both approaches are the shape of the probability density function, in some cases leading to multi-modality for the standard approach; the moments which characterise the density, particularly important in practice for the expected value and the variance; and equivariance under perturbation.

4.5. Example

To illustrate the differences between a density with respect to the Lebesgue measure λ and a density with respect to the measure λ_a in $\mathcal{S}^D$, a GDP data set will be used. The data set used is taken from the National Accounts Statistics database and is available on the United Nations Statistic Division web page http://unstats.un.org/unsd/snaama/dnllist.asp. We use the information corresponding to the year 2009 for 208 countries. The GDP data set is based on the international standard industrial classification (ISIC) of all economic activities. The original data contains the percentages of each economic activity for all countries divided in six categories.

The goal is to compare some characteristics corresponding to the logistic normal and the normal densities on the simplex. In order to provide some useful comprehensive figures a three-part compositional data set is preferred. For this reason the three-part subcomposition (x_1, x_2, x_3) is used, where



(a) *(b)*

Figure 7: *GDP data with isodensity curves of the fitted (a) logistic normal and (b) normal on $\mathcal{S}^3$ densities.*

- $x_1 =$ agriculture, hunting, forestry, fishing (ISIC A-B),
- $x_2 =$ mining, manufacturing, utilities (ISIC C-E),
- $x_3 =$ construction (ISIC F).

Following the suggestions by Aitchison (1986), a battery of 12 tests of goodness-of-fit are used. They are based on the Anderson-Darling, Cramér-von Mises and Watson statistics, applied to the coordinates of the three-part sample data set. In particular, the tests are applied to the marginal distributions, to the bivariate angle distribution and to the radial distribution. Taking a 5% significance level, no significant departure from normality is obtained by any of these tests.

Parameters of the two density models, the normal on $\mathcal{S}^3$ and the logistic normal, are equal. This is a direct consequence of the definition of densities and hence likelihoods. In this case, after taking a suitable ilr transformation (eq. 11 was used), the maximum likelihood estimates for both density models are:

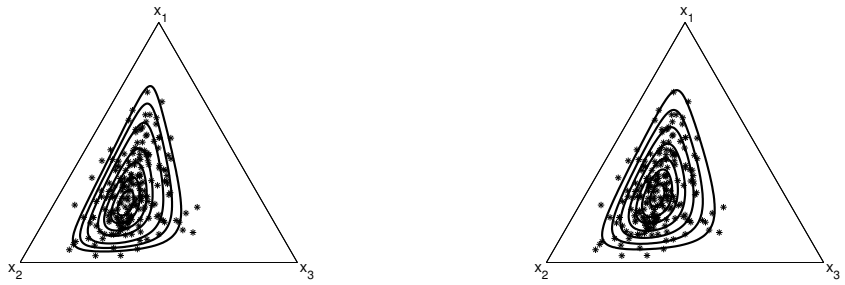
$$\hat{\boldsymbol{\mu}} = (-0.715, 0.521)^{\mathsf{T}}, \qquad \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1.303 & 0.452 \\ 0.452 & 0.680 \end{pmatrix}.$$

Figures 7(a) and 7(b) show the sample in a ternary diagram and the isodensity curves of the fitted logistic normal density and the normal in $\mathcal{S}^3$ density. Different features are observed. The logistic normal density shows two modes whereas the normal in $\mathcal{S}^3$ exhibits a single mode. When contours and sample are plotted in the coordinate space (Figure 9(a)) differences disappear, as the probability density in ilr-coordinates, a bivariate normal in $\mathbb{R}^2$, is equal for the two density models.

After plotting the contours of the density with respect to the Lebesgue measure (Figure 7(a)) showing two modes, one might think about the existence of two subpopulations that could explain the bimodality of the logistic normal density. However, using the available information on the data set concerning geography or development of countries, no coherent reasons were found for the observed bimodality shown in Figure

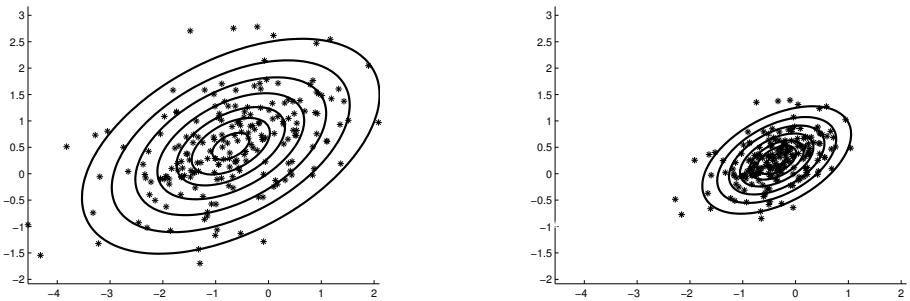
7(a). The bimodality in this case is only due to the measure of reference chosen in the simplex.

For illustration purposes, changes of density contours under powering are shown in Figure 8 for both density models. Let $\mathbf{X}$ denote the three-part random variable of the example with the estimated parameters $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ shown above, and consider $\mathbf{X}_\alpha = \alpha \odot \mathbf{X}$ for $\alpha = 1/2$. In Figures 8(a) and 8(b) the isodensity contours with respect to the Lebesgue measure and the Aitchison measure in the ternary diagram are represented. As can be observed, in the logistic normal case, the bimodality disappears. In other words, the power transformation, which should only move the centre of the density and modify the variability, can eliminate or in other cases generate arbitrary modes if the Lebesgue measure is considered (Mateu-Figueras and Pawlowsky-Glahn, 2008). This undesirable behaviour of modes and isodensity contours prevents the use of the logistic normal density on the simplex and all statistics depending on it, e.g. expectation and covariance with respect to the Lebesgue measure on the simplex, predictive regions, etc.



(a) *(b)*

Figure 8: *Power transformed GDP data with isodensity curves of the fitted (a) logistic normal and (b) normal on $\mathcal{S}^3$ densities.*



(a) *(b)*

Figure 9: ilr *coordinates of the (a) GDP data set and (b) the power transformed data set with the corresponding fitted normal densities.*

5. Conclusions

A particular Euclidean vector space structure of the positive real line and of the simplex, together with the associated measure, allow us to define parametric models with desirable properties. Normal density models on $\mathbb{R}_+$ and on $\mathcal{S}^D$ have been defined through their densities over the coordinates with respect to an orthonormal basis and their main algebraic properties have been studied. From a probabilistic point of view, those laws of probability are identical to the lognormal and to the additive logistic normal distribution defined using the Lebesgue measure and the standard methodology based on transformations. Nevertheless, some differences are obtained in the moments and in the shape of the density function. In particular, the expected value with respect to the new measure differs from what would be obtained with the Lebesgue measure for the lognormal and additive logistic normal distributions, but leads to the parameters that are used for these models. It thus yields directly a suitable characterization of these models. In the normal in $\mathbb{R}_+$ case, a consistent estimator and confidence intervals for the mean are easily obtained directly from the log-transformed data, while in the lognormal case, i.e., keeping the Lebesgue measure and therefore aiming at the corresponding, common expected value, a bias correction is necessary. In the normal in $\mathcal{S}^D$ case we show important differences in the shape of the density. The normal in $\mathcal{S}^D$ always appears unimodal, whereas bimodal and trimodal densities could be obtained using the standard approach.

Acknowledgements

The authors thank two anonymous referees for their suggestions which greatly improved the article. This work has been supported by the Spanish Ministry of Education and Science under project MTM2009-13272 and by the Agència de Gestió d'Ajuts Universitaris i de Recerca of the Generalitat de Catalunya under project Ref: 2009SGR424.

Appendix

This appendix contains the proofs of properties contained in Section 3.1 and Section 4.2. They use the expected value, the covariance matrix, the linear transformation property of the multivariate normal distribution and some matrix relationships among vectors of coordinates and among log-ratio transformations.

Proof of property 3.1. The coordinates of the random variable X^* are obtained from the coordinates of the variable X as $\ln(X^*) = \ln(a) + b\ln(X)$. The density function of $\ln(X)$ is the classical normal density on the real line; thus, the linear transformation property can be used to obtain the density function of the $\ln(X^*)$ random variable. Therefore, $X^* \sim \mathcal{N}_+(\ln a + b\mu, b^2\sigma^2)$.

Proof of property 3.2. From property 3.1 we know that $a \oplus X = a \cdot X \sim \mathcal{N}_+(\ln a + \mu, \sigma^2)$. From (8) we get

$$f^+_{a\oplus X}(a \oplus x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(\ln(ax) - (\ln a + \mu))^2}{\sigma^2}\right) = f^+_X(x).$$

Proof of property 3.3. From (6) we know that $\mathrm{E}^E[X] = \exp(\mathrm{E}[\ln X])$ because the orthonormal coordinates on the positive real line are obtained with the logarithmic transformation. Given some coordinate, the exponential function provides the element on $\mathbb{R}_+$. The density function of $\ln X$ is the normal distribution, as stated in Definition 3.1. Thus, the expected value is the μ parameter and, consequently, $\mathrm{E}^E[X] = \exp(\mu)$. The same result is obtained for the median and the mode, as the normal distribution is symmetric around its expected value μ.

Proof of property 3.4. From (7) we know that the variance can be understood as the expected value of the squared distance around its expected value, i.e. $\mathrm{Var}^+[X] = \mathrm{E}[\mathrm{d}^2_+(X, \mathrm{E}^+[X])]$. Working on coordinates and using the density function of $\ln X$ we obtain $\mathrm{Var}^+[X] = \mathrm{E}[\mathrm{d}^2(\ln X, \mathrm{E}[\ln X])] = \mathrm{Var}[\ln X] = \sigma^2$.

Proof of property 4.1. The orthonormal coordinates of the random composition $\mathbf{X}^*$ are obtained from the orthonormal coordinates of the composition $\mathbf{X}$ via $h(\mathbf{X}^*) = h(\mathbf{a}) + bh(\mathbf{X})$. The density function of $h(\mathbf{X})$ is the classical normal density in real space; thus, the linear transformation property can be used to obtain the density function of $h(\mathbf{X}^*)$. Therefore, $\mathbf{X}^* \sim \mathcal{N}^D_S(h(\mathbf{a}) + b\mu, b^2\mathbf{\Sigma})$.

Proof of property 4.2. Using property 4.1, $\mathbf{a} \oplus \mathbf{X} \sim \mathcal{N}^D_S(h(\mathbf{a}) + \boldsymbol{\mu}, \mathbf{\Sigma})$. We know that $h(\mathbf{a} \oplus \mathbf{x}) = h(\mathbf{a}) + h(\mathbf{x})$, therefore,

$$\begin{aligned} f_{\mathbf{a}\oplus\mathbf{X}}(\mathbf{a} \oplus \mathbf{x}) &= (2\pi)^{-(D-1)/2} \mid \mathbf{\Sigma} \mid^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2}\left(h(\mathbf{a} \oplus \mathbf{x}) - (h(\mathbf{a}) + \boldsymbol{\mu})\right)^{\mathsf{T}} \mathbf{\Sigma}^{-1} \left(h(\mathbf{a} \oplus \mathbf{x}) - (h(\mathbf{a}) + \boldsymbol{\mu})\right)\right] \\ &= (2\pi)^{-(D-1)/2} \mid \mathbf{\Sigma} \mid^{-1/2} \exp\left[-\frac{1}{2}\left(h(\mathbf{x}) - \boldsymbol{\mu}\right)^{\mathsf{T}} \mathbf{\Sigma}^{-1} \left(h(\mathbf{x}) - \boldsymbol{\mu}\right)\right] \\ &= f_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

Proof of property 4.3. For a centered log-ratio transformed vector it is straightforward to see that $\mathrm{clr}(\mathbf{X}_P) = \mathbf{P}\mathrm{clr}(\mathbf{X})$ (Aitchison, 1986, p. 94). Using the matrix relationship between the centered and the isometric log-ratio vectors (Egozcue et al., 2003) we conclude that $h(\mathbf{X}_P) = (\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{U})h(\mathbf{X})$. Given the density of $h(\mathbf{X})$, and applying the linear transformation property of the normal distribution in real space, a $\mathcal{N}_S^D(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P)$ distribution is obtained for the random composition $\mathbf{X}_P$.

Proof of property 4.4. (Aitchison, 1986, p. 119) gives the matrix relationship between $\mathrm{alr}(\mathbf{X}_S)$ and $\mathrm{alr}(\mathbf{X})$. Using the matrix relationships between the additive, centered and isometric log-ratio vectors (Egozcue et al., 2003), we conclude that $h(\mathbf{X}_S) = (\mathbf{U}^{*\prime}\,\mathbf{S}\mathbf{U})h(\mathbf{X})$. Given the density of $h(\mathbf{X})$, and applying the linear transformation property of the normal distribution in real space, the density of $h(\mathbf{X}_S)$ is obtained as that of the $\mathcal{N}_S^C(\boldsymbol{\mu}_S, \boldsymbol{\Sigma}_S)$ distribution.

Proof of property 4.5. From (6) we know that $\mathrm{E}^a[\mathbf{X}] = h^{-1}(\mathrm{E}[h(\mathbf{X})])$, and from (14) we know that the density function of $h(\mathbf{X})$ is the multivariate normal distribution; thus $\mathrm{E}[h(\mathbf{X})] = \boldsymbol{\mu}$. Finally, the composition $\mathrm{E}_a[\mathbf{X}]$ is obtained applying h^{-1}.

Proof of property 4.6. From (7) we know that the variance can be understood as the expected value of the squared distance around its expected value, i.e. $\mathrm{Var}^a[\mathbf{X}] = \mathrm{E}[d_a^2(\mathbf{X}, \mathrm{E}_a[\mathbf{X}])]$. Working on coordinates and using the density function of $h(\mathbf{X})$ we obtain $\mathrm{Var}^a[\mathbf{X}] = \mathrm{trace}(\boldsymbol{\Sigma})$.

References

Ahrens, L. (1954). The lognormal distribution of the elements. *Geochimica et Cosmochimica Acta*, 5, 49–73.

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society: Series B*, 44(2), 139–177.

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

Aitchison, J. (1997). The one-hour course in compositional data analysis or compositional data analysis is simple. *Proceedings of IAMG'97, the third annual conference of the International Association for Mathematical Geology*, V. Pawlowsky-Glahn ed.,International Center for Numerical Methods in Engineering (CIMNE), Barcelona, vol. 1, 3–35.

Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press, Cambridge.

Aitchison, J., Mateu-Figueras, G. and Ng, K. (2003). Characterization of distributional forms for compositional data and associated distributional tests. *Mathematical Geology*, 35(6), 667–680.

Barceló-Vidal, C. (1996). *Mixturas de Datos Composicionales*. Ph.D. Diss., Universitat Politècnica de Catalunya.

Billheimer, D., Guttorp, P. and Fagan, W. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96(456), 1205–1214.

Clark, I. and Harper, W. V. (2000). *Practical Geostatistics 2000*. Ecosse North America Llc., Columbus Ohio.

Crow, E. L. and Shimizu, K. (1988). *Lognormal Distributions. Theory and Applications*. Marcel Dekker, Inc. New York.

Davis, J. C. (1986). *Statistics and Data Analysis in Geology. 2nd ed*. John Wiley & Sons. New York.

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

Egozcue, J. J., Barceló-Vidal, C., Martín-Fernández, J. A., Jarauta-Bragulat, E., Díaz-Barrero, J. L. and Mateu-Figueras, G. (2011). Elements of Simplicial Linear Algebra and Geometry. *Compositional Data Analysis: Theory and Applications*, V. Pawlowsky-Glahn & A. Buccianti eds., John Wiley & Sons, Chichester, 141–157.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37(7), 795–828.

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

Galton, F. (1879). The geometric mean, in vital and social statistics. *Proceedings of the Royal Society of London*, 29, 365–366.

Herdan, G. (1960). *Small Particle Statistics*. Butterwoths, London.

Kocherlakota, S. and Kocherlakota, K. (1982). *Multinormal Distribution. Encyclopedia of statistical sciences*. S. Kotz and N.L. Johnson eds., John Wiley & Sons, New York, vol. 5, 668–677.

Krige, D. G. (1981). *Lognormal-de Wijsian Geostatistics for Ore Evaluation*. South African Inmstitute of Mining and Metallurgy, Johannesburg.

Martín-Fernández, J. A., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (1998). A critical approach to non-parametric classification of compositional data. *Advances in Data Science and Classification, Proceedings of the 6th Conference of the International Federation of Classification Societies, IFCS'98*, A. Rizzi, M. Vichi & H. H. Bock eds., Springer-Verlag, Berlin, 49–56.

Martín-Fernández, J. A., Bren, M., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (1999). A measure of difference for compositional data based on measures of divergence. *Proceedings of IAMG'99, the fifth annual conference of the International Association for Mathematical Geology*, S. J. Lippard, A. Næss & R. Sinding-Larsen eds., Tapir, Trondheim, 211–216.

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

Mateu-Figueras, G. and Pawlowsky-Glahn, V. (2007). The skew-normal distribution on the simplex. *Communications in Statistics-Theory and Methods, Special Issue Skew-elliptical Distributions and their Application*, 36(9), 1787–1802.

Mateu-Figueras, G. and Pawlowsky-Glahn, V. (2008). A critical approach to probability laws in geochemistry. *Mathematical Geosciences*, 40(5), 489–502.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Barceló-Vidal, C. (2005). The additive logistic skew-normal distribution on the simplex.*Stochastic Environmental Research and Risk Assessment*, 19, 205-214.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Egozcue, J. J. (2011). The principle of working on coordinates. *Compositional Data Analysis: Theory and Applications*, V. Pawlowsky-Glahn & A. Buccianti eds., John Wiley & Sons, Chichester, 31–42.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Martín-Fernández, J. A. (2002). Normal in R+ vs lognormal in R. *Terra Nostra*, 3, 305–310.

McAlister, D. (1879). The law of geometric mean. Proceedings of the Royal Society of London, 29, 367–376.

Monti, G. S., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (2011). Notes on the Scaled Dirichlet Distribution. *Compositional Data Analysis: Theory and Applications*, V. Pawlowsky-Glahn & A. Buccianti eds., John Wiley & Sons, Chichester, 128–138.

Pawlowsky-Glahn, V. (2003). Statistical modelling on coordinates. *Compositional Data Analysis Workshop – CoDaWork'03 Proceedings*, S. Thió-Henestrosa & J. A. Martín-Fernández eds., Universitat de Girona.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5), 384–398.

Pearson, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, LX, 489–502.

Stevens, J. (1986). *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates, Hillsdale, USA.

Tolosana-Delgado, R. (2005). *Geostatistics for Constrained Data: Positive Data, Compositions and Probabilities. Application to Environmental Hazard Monitoring.* Ph.D. Diss., Universitat de Girona.

Tolosana-Delgado, R. and Pawlowsky-Glahn, V. (2007). Kriging regionalized positive variables revisited: sample space and scale considerations. *Mathematical Geology*, 39, 529-558.

von Eynatten, H., Pawlowsky-Glahn, V. and Egozcue, J. J. (2002). Undestanding perturbation on the simplex: a simple method to better visualise and interpret compositional data in ternary diagrams. *Mathematical Geology*, 34, 249–257.