

# The Dirichlet distribution with respect to the Aitchison measure on the simplex - a first approach

G. Mateu-Figueras and V. Pawlowsky-Glahn

Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona; [gloria.mateu@udg.es](mailto:gloria.mateu@udg.es)

## Abstract

The algebraic-geometric structure of the simplex, known as Aitchison geometry, is used to look at the Dirichlet family of distributions from a new perspective. A classical Dirichlet density function is expressed with respect to the Lebesgue measure on real space. We propose here to change this measure by the Aitchison measure on the simplex, and study some properties and characteristic measures of the resulting density.

**Key words:** Radon-Nykodym derivative, perturbation, expected value.

## 1 Introduction

Given a measurable space,  $E$ , the *Radon-Nikodym derivative* of a probability,  $P$ , with respect to a measure,  $\lambda_E$ , is a measurable and non-negative function,  $f$ , such that the probability of any event,  $A$ , of the corresponding  $\sigma$ -algebra is

$$P(A) = \int_A f(x) d\lambda_E(x), \quad (1)$$

for  $x \in E$ . We also call  $f$  *density function with respect to the measure  $\lambda_E$* .

In general we work with real random vectors, i.e.  $E = \mathbb{R}^D$ , and we use density functions with respect to the Lebesgue measure, which is a natural measure in real space. The Lebesgue measure allows to compute easily any probability using equation (1), because the integral is an ordinary one. The Lebesgue measure plays a fundamental role in real analysis and is compatible with the inner vector space structure of  $\mathbb{R}^D$ .

Sometimes, like in the case of compositional data, we work with random vectors defined in a space  $E$  different from real space, but with an Euclidean space structure. In those cases it might not be suitable to apply the same methods and concepts used in real space, as they might lead to inconsistent results, like the spurious correlation mentioned by Pearson (1897). The underlying reason is that most methods have been developed for the real case, as they are based on the usual Euclidean space structure of  $\mathbb{R}^D$ . This problem can be circumvented defining a probability law on  $E$  using a density function with respect to an appropriate measure,  $\lambda_E$ . If properly defined, the measure  $\lambda_E$  in  $E$  will have comparable properties to the Lebesgue measure in real space, and densities will show a nice behavior. But this has important consequences. For example, it might be difficult to compute any moment, or to make effective the calculation of any probability, because we have not an ordinary integral. Those difficulties can be solved working on coordinates (Eaton, 1983) and, in particular, working on coordinates with respect to an orthonormal basis (Pawlowsky-Glahn, 2003).

The *principle of working on coordinates* is based on the following facts. If  $E$  is an Euclidean vector space with an internal operation,  $\oplus$ , an external operation,  $\odot$ , and an inner product,  $\langle \cdot, \cdot \rangle_E$ , then the general theory of linear algebra proves the existence of a (non unique) orthonormal basis with respect to which the coefficients or coordinates behave like usual elements in real space, satisfying all the standard rules (sum, product, ordinary scalar product, ...). Properties that hold in the space of coordinates transfer directly to the space  $E$ . This allows us to identify statistical analysis

in  $E$  with conventional statistical analysis on real coordinates. Furthermore, concepts like measure or density function can be taken as defined on coordinates and, consequently, it is possible to work with the Lebesgue measure and the classical density in real space.

This idea can be applied to the simplex. Do do so, in the following sections first the Aitchison measure and space structure, as well as the expression of the coefficients of an arbitrary element with respect to an orthonormal basis, are introduced. Then, the expression of the Dirichlet density function with respect to this natural measure on the simplex is provided. Finally, this density is compared with the classical Dirichlet density function with respect to the Lebesgue measure.

We have to insist on the fact that this approach implies using a measure which is different from the usual Lebesgue measure. But the most important aspect is that it opens the door to alternative statistical models depending not only on the assumed distribution, but also on the measure which is considered as appropriate or natural for the studied phenomenon.

## 2 A relative measure on $\mathcal{S}^D$

In the 80's, Aitchison (1982, 1986) showed that the standard operations we use in real space make no sense from a compositional point of view, and introduced perturbation,  $\oplus$ , powering,  $\odot$  and a distance,  $d_a$ . Later Billheimer et al. (2001) and Pawlowsky-Glahn and Egozcue (2001) introduced independently an inner product,  $\langle \cdot, \cdot \rangle_a$ , which is compatible with these operations, and showed that  $(\mathcal{S}^D, \oplus, \odot)$  has an Euclidean vector space structure of dimension  $D - 1$ . A proof can be found in Pawlowsky-Glahn and Egozcue (2002). The resulting geometry is known as the Aitchison geometry of the simplex.

In face of a compositional problem, we have to focus on the relative magnitudes of the parts of compositions, that is, the sizes are irrelevant (Aitchison, 1997). This argument is used to define operations, functions, or methods, on  $\mathcal{S}^D$ . The same argument can be used for the measure. Pawlowsky-Glahn (2003) defines a measure on the simplex, denoted as  $\lambda_a$  and called Aitchison measure. This measure is relative and compatible with the inner vector space structure of the simplex. It is also absolutely continuous with respect to the Lebesgue measure,  $\lambda$ , on real space. The relationship between them is given by the jacobian

$$\frac{\lambda_a}{\lambda} = \frac{1}{\sqrt{D}x_1x_2 \cdots x_D}. \quad (2)$$

A proper statistical analysis is expected to produce meaningful results. In this case, we would like to have statistical techniques compatible with this measure; e.g. a measure of variability expressed in terms of the natural measure of difference, a mean which minimizes in some sense the natural variability of the data, or a density function with respect to this natural measure.

As mentioned, the simplex  $\mathcal{S}^D$ , with the operations  $\oplus$  and  $\odot$  and the inner product  $\langle \cdot, \cdot \rangle_a$ , has an Euclidean vector space structure of dimension  $D - 1$ . The general theory of linear algebra guarantees the existence of a (non unique) orthonormal basis that we denote by  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$ . The coefficients of any composition  $\mathbf{x} \in \mathcal{S}^D$  with respect to this orthonormal basis are

$$\text{ilr}(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \langle \mathbf{x}, \mathbf{e}_2 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a)',$$

which is a  $D - 1$  vector of real coordinates. We will use the notation  $\text{ilr}(\mathbf{x})$  to emphasize the similarity with the vector obtained applying the isometric logratio transformation to composition  $\mathbf{x}$ , a transformation from  $\mathcal{S}^D$  to  $\mathbb{R}^{D-1}$  defined by Egozcue et al. (2003).

We can apply standard real analysis to the ilr coordinates. It is easy to see that operations  $\oplus$  and  $\odot$  are equivalent to the sum and the scalar product of the respective coordinates. We can apply also the standard inner product, the ordinary Euclidean distance and the Lebesgue measure in  $\mathbb{R}^{D-1}$  to the ilr coordinates .

We know that, like in every inner product space, the orthonormal basis is not unique. But in this case it is not straightforward to determine which one is the most appropriate to solve a specific problem. Nevertheless, the important point is that, once we have chosen an appropriate orthonormal basis, all standard statistical methods can be applied to the coefficients, and results can be transferred to the simplex preserving their properties.

### 3 The Dirichlet distribution

As stated in Aitchison(1986, p. 58), a random composition  $\mathbf{x} \in \mathcal{S}^D$  is said to have a Dirichlet distribution with parameter  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_D) \in \mathbb{R}_+^D$  if its density function is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_D)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_D)} x_1^{\alpha_1-1} \dots x_D^{\alpha_D-1}, \quad (3)$$

where  $\Gamma$  is the Gamma function. Every Dirichlet random composition is formed from the closure of  $D$  independent and equally scaled gamma distributed random variables. When  $D = 2$  this density is known as the Beta density.

If we remove the requirement of equal scale parameter of the gamma variables, we obtain a generalization of the Dirichlet distribution called the scaled Dirichlet distribution (Aitchison, 1986, p.305). The expression of its density function is

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_D)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_D)} \frac{\beta_1^{\alpha_1} x_1^{\alpha_1-1} \dots \beta_D^{\alpha_D} x_D^{\alpha_D-1}}{(\beta_1 x_1 + \dots + \beta_D x_D)^{\alpha_1 + \dots + \alpha_D}}, \quad (4)$$

where  $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}_+^D$ . Observe that when  $\boldsymbol{\beta}$  is the vector of ones, then the Dirichlet distribution (3) is obtained. Densities (3) and (4) are classical densities, that is, they are Radon-Nikodym derivatives with respect to the Lebesgue measure in real space. Using (2) we can easily change the measure and express both densities with respect to the measure  $\lambda_a$ . The resulting expressions of the Dirichlet and scaled Dirichlet density functions are, respectively,

$$f_{\mathbf{x}}^*(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_D) \sqrt{D}}{\Gamma(\alpha_1) \dots \Gamma(\alpha_D)} x_1^{\alpha_1} \dots x_D^{\alpha_D}, \quad (5)$$

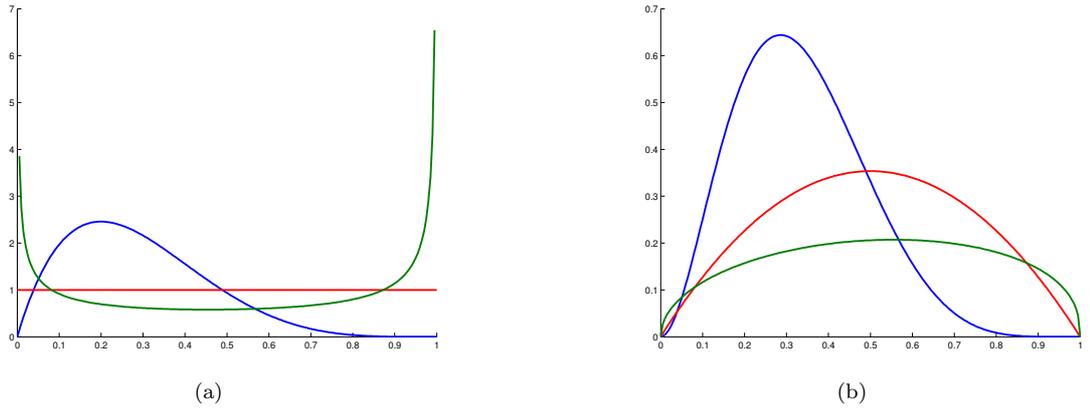
$$f_{\mathbf{x}}^{**}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_D) \sqrt{D}}{\Gamma(\alpha_1) \dots \Gamma(\alpha_D)} \frac{\beta_1^{\alpha_1} x_1^{\alpha_1} \dots \beta_D^{\alpha_D} x_D^{\alpha_D}}{(\beta_1 x_1 + \dots + \beta_D x_D)^{\alpha_1 + \dots + \alpha_D}}. \quad (6)$$

It is also possible to express densities (5) and (6) in terms of the coordinates with respect to an orthonormal basis. This would be useful to compute the expected value of any moment. We do not provide here the expression because it is quite long and complicated. Nevertheless, the important thing is that we could do it, and we could use those densities as classical densities in real space.

### 4 Comparison

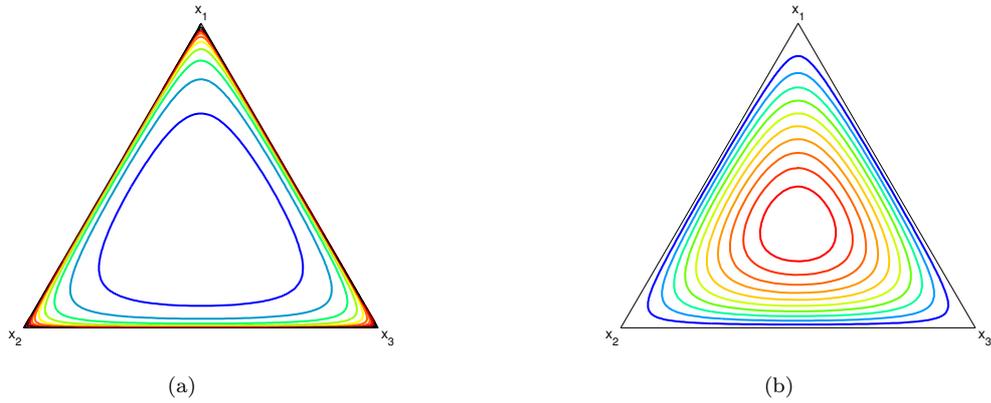
The objective of this section is to compare both approaches and to see some first consequences of changing the measure. We start with a graphical comparison when  $D = 2$ . In Figure 1(a) we represent densities with respect to the Lebesgue measure. In Figure 1(b) we represent densities with respect to the  $\lambda_a$  measure. Obviously we observe differences. Using the density with respect to  $\lambda_a$  there is always a mode. This is not the case using the classical beta densities, because for  $\boldsymbol{\alpha} = (1, 1)$  a constant density function is obtained, and for  $\boldsymbol{\alpha} = (0.5, 0.4)$  it is a density with vertical asymptotes at 0 and 1.

In Figure 2 the isodensity contour plots of a Dirichlet density with  $D = 3$  in the ternary diagram are represented. In both cases the red curves represent the highest values and the blue ones the



**Figure 1:** Beta density curves with respect to (a) the Lebesgue measure  $\lambda$  and (b) the Aitchison measure  $\lambda_a$  with parameters  $\alpha = (2, 5)$  (—);  $\alpha = (1, 1)$  (—) and  $\alpha = (0.5, 0.4)$  (—).

lowest. Important differences can be observed. Using the classical approach (the density with respect to the Lebesgue measure Fig.2(a)) we observe that the density increases when we tend to the boundary of the ternary diagram. Using the proposed approach (Fig.2(b)) we obtain the opposite behavior.



**Figure 2:** Isodensity contour plots of a Dirichlet density with parameter  $\alpha = (0.9, 0.9, 0.9)$  with respect to (a) the Lebesgue measure  $\lambda$  and (b) the Aitchison measure  $\lambda_a$ .

Differences can also be found in the properties and characteristic measures. The mode of a Dirichlet density with respect to the Lebesgue measure is

$$\text{mode}_{\mathbb{R}}(\mathbf{x}) = \left( \frac{\alpha_1 - 1}{\alpha_0 - D}, \frac{\alpha_2 - 1}{\alpha_0 - D}, \dots, \frac{\alpha_D - 1}{\alpha_0 - D} \right),$$

whereas the mode of a Dirichlet density with respect to the natural measure on the simplex is

$$\text{mode}_S(\mathbf{x}) = \left( \frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_0}, \dots, \frac{\alpha_D}{\alpha_0} \right),$$

where  $\alpha_0 = \alpha_1 + \alpha_2 + \dots + \alpha_D$  in both cases.

The expected value is also different. Using the classical approach  $E_{\mathbb{R}}(\mathbf{x})$  is computed using the

standard definition, i.e.

$$E_{\mathbb{R}}(\mathbf{x}) = \left( \frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_0}, \dots, \frac{\alpha_D}{\alpha_0} \right).$$

There are some difficulties to compute the expected value using the density with respect to the  $\lambda_a$  measure. One easy way to obtain it is expressing the Dirichlet density function in terms of the coordinates with respect to an orthonormal basis, and then using the standard definition of the expected value applied to the  $\text{ilr}(\mathbf{x})$  vector. The result are the coordinates with respect to the same orthonormal basis of the composition  $E_{\mathcal{S}}(\mathbf{x})$ . Then, using a linear combination we obtain the expected composition as

$$E_{\mathcal{S}}(\mathbf{x}) = \mathcal{C} \left( e^{\Psi(\alpha_1)}, e^{\Psi(\alpha_2)}, \dots, e^{\Psi(\alpha_D)} \right),$$

where  $\Psi$  represents the Digamma function. The differences between  $E_{\mathbb{R}}(\mathbf{x})$  and  $E_{\mathcal{S}}(\mathbf{x})$  are obvious. In particular, observe that  $E_{\mathbb{R}}(\mathbf{x})$  is equal to the  $\text{mode}_{\mathcal{S}}(\mathbf{x})$ .

There are also some coincidences, because both densities assign exactly the same probability to any subset of  $\mathcal{S}^D$ , as both models define the same law of probability over  $\mathcal{S}^D$ .

Another coincidence is that, using both approaches, the class of scaled Dirichlet distributions is closed under perturbation. If  $\mathbf{x}$  has a scaled Dirichlet distribution with parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ , and  $\mathbf{p}$  is a constant composition, then the random composition  $\mathbf{x}^* = \mathbf{p} \oplus \mathbf{x}$  has a scaled Dirichlet distribution with parameters  $\boldsymbol{\alpha}$  and  $\mathbf{p}^{-1}\boldsymbol{\beta}$ . As a consequence we have that any random composition  $\mathbf{x}^*$  scaled Dirichlet distributed with parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  can be obtained as  $\mathbf{x}^* = \boldsymbol{\beta}^{-1} \oplus \mathbf{x}$  where  $\mathbf{x}$  has a Dirichlet distribution with parameter  $\boldsymbol{\alpha}$ .

Related to this last property there is another difference: using only the densities with respect to the  $\lambda_a$  measure we have the equality

$$f_{\mathbf{p} \oplus \mathbf{x}}^*(\mathbf{p} \oplus \mathbf{x}) = f_{\mathbf{x}}^*(\mathbf{x}).$$

This equality means that the Dirichlet density with respect to the Aitchison measure on the simplex is invariant under the internal operation. Consequently we can interpret the scaled Dirichlet as the result of applying a perturbation to a Dirichlet density. This is not correct using the classical approach. Observe that this equality has important consequences, because when working with compositional data often the centering operation (Martín-Fernández et al., 1999) is applied, which is a perturbation using the inverse of the center of the data set. One interesting aspect is that the mode and the expected composition of a scaled Dirichlet can be simply obtained as the perturbed mode and expected composition of a Dirichlet.

## 5 Conclusion

The Dirichlet and scaled Dirichlet density functions can be expressed with respect to the Aitchison measure on the simplex. In terms of probabilities of subsets of  $\mathcal{S}^D$ , the laws of probability are identical to the classical Dirichlet and scaled Dirichlet density functions, but their properties and characteristic measures are different. More research is necessary to study the consequences of these differences.

## Acknowledgements

This research has been supported by the Dirección General de Enseñanza Superior (DGES) of the Spanish Ministry for Education and Culture through the project BFM2000-0540.

## References

- Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 44(2), 139–177.
- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (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.
- Billheimer, D., P. Guttorp, and W. Fagan (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association* 96(456), 1205–1214.
- Eaton, M. L. (1983). *Multivariate Statistics. A Vector Space Approach*. John Wiley & Sons.
- Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35(3), 279–300.
- Martín-Fernández, J. A., M. Bren, C. Barceló-Vidal, and V. Pawlowsky-Glahn (1999). A measure of difference for compositional data based on measures of divergence. In S. J. Lippard, A. Næss, and R. Sinding-Larsen (Eds.), *Proceedings of IAMG'99 — The fifth annual conference of the International Association for Mathematical Geology*, Volume I and II, pp. 211–216. Tapir, Trondheim (N), 784 p.
- 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*. Universitat de Girona, ISBN 84-8458-111-X, <http://ima.udg.es/Activitats/CoDaWork03/>.
- Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA)* 15(5), 384–398.
- Pawlowsky-Glahn, V. and J. J. Egozcue (2002). BLU estimators and compositional data. *Mathematical Geology* 34(3), 259–274.
- 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.