

A Bicompositional Dirichlet Distribution

Jakob Bergman

2009:3



LUND
UNIVERSITY

**DEPARTMENT OF
STATISTICS**

S-220 07 LUND
SWEDEN

A Bivariate Dirichlet Distribution

Jakob Bergman^{*†}

December 10, 2009

Abstract

The Simplex \mathcal{S}^D is the sample space of a D -part composition. There are only a few distributions defined on the Simplex and even fewer defined on the Cartesian product $\mathcal{S}^D \times \mathcal{S}^D$. Based on the Dirichlet distribution, defined on \mathcal{S}^D , we propose a new bicompositional Dirichlet distribution defined on $\mathcal{S}^D \times \mathcal{S}^D$, and examine some of its properties, such as moments as well as marginal and conditional distributions. The proposed distribution allows for modelling of covariation between compositions without leaving $\mathcal{S}^D \times \mathcal{S}^D$.

Keywords: Cartesian product, compositional data, Dirichlet distribution, Simplex

1 Introduction

1.1 Compositions

A composition is a vector of non-negative components summing to a constant, usually 1. The components of a composition are what we usually think of as proportions (at least when the vector sums to one). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all three examples of compositions from different scientific areas. For more examples of compositions, see for instance Aitchison (1986) (or the reprint (Aitchison, 2003)).

Compositions differ from other multivariate random vectors on the real space due the summation constraint. Whereas the Cartesian product of two random vectors on the real space \mathcal{R}^p will form a new random vector on the real space \mathcal{R}^{p+p} , this is not the case for compositions.

^{*}Department of Statistics, Lund University, Box 743, SE-220 07 Lund, Sweden

[†]e-mail: jakob.bergman@stat.lu.se

When describing dependency structures, the compositional analysis has been primarily concerned with describing the dependency structures within a composition, i.e. the relation between the components of a composition. Aitchison (1986, Ch. 10) for instance devotes an entire chapter to this, and as a recent example Ongaro et al. (2008) construct a new distribution for modelling such relations. In this paper we will not be considering the relation between the components of a composition, but the relation between two compositions. We will use the term *bicompositional* when referring to two compositions (with same number of components) and the term *unicompositional* when referring to one composition. A composition with two components will be referred to as a *bicomponent* composition, as opposed to a *multicomponent* composition.

1.2 The Simplex

The sample space of a composition is the Simplex. (For simplicity we will always take the summing constant to be 1.) We define the D -dimensional Simplex space \mathcal{S}^D as

$$\mathcal{S}^D = \left\{ (x_1, \dots, x_D)' \in \mathcal{R}_+^D : \sum_{j=1}^D x_j = 1 \right\}$$

where \mathcal{R}_+ is the positive real space. As noted above, it is a sample space that occurs in a wide variety of applications. There are however only a limited number of distributions defined on the Simplex; the two most notable are the Dirichlet distribution and the logistic normal distribution class described by Aitchison and Shen (1980). There are also a number of generalisations of these two distributions.

The sample space of the joint distribution of two compositions is the Cartesian product $\mathcal{S}^D \times \mathcal{S}^D$. This sample space is the subspace of \mathcal{R}_+^{D+D} , where

$$\sum_{j=1}^D x_j = \sum_{j=D+1}^{D+D} x_j = 1.$$

There are almost no distributions defined on $\mathcal{S}^D \times \mathcal{S}^D$. For the case $D = 2$, the bicomponent case, there have been proposed a few bivariate Beta distributions (though usually not in a compositional context). See for instance in recent years Olkin and Liu (2003), Nadarajah and Kotz (2005), Nadarajah (2006) or Nadarajah (2007). For $D > 2$, the multicomponent case, we have not found any distributions at all in the literature.

1.3 An Example

A trivial example of a composition is the vector of length two consisting of the proportion Employed and the proportion Not employed, the latter of which of

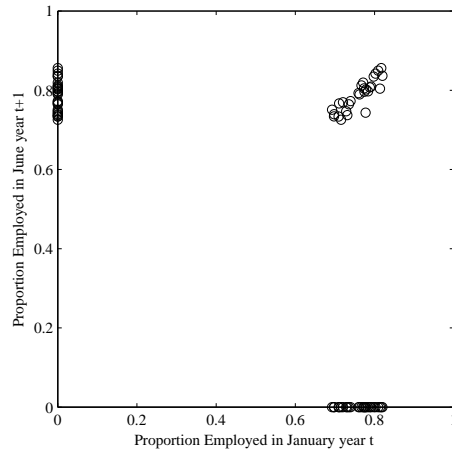


Figure 1. The proportion Employed in Sweden in June plotted versus the corresponding proportion in January the previous year, for the years 1977 through 2004. The marginal proportions are plotted on the axes.

course is one minus the first one. In Figure 1 the composition (Employed, Not employed) in Sweden in June has been plotted versus the same composition for January the previous year, for the years 1977 through 2004. Data come from the Swedish Labour Force Survey (AKU), conducted by Statistics Sweden; see Persson and Henkel (2005) for a description of the Swedish Labour Force Survey. The marginal compositions are plotted on the axes. We clearly see a positive linear relation between the two months. To be able to model this relation we need distributions defined on $\mathcal{S}^2 \times \mathcal{S}^2$.

It would of course also be interesting to examine the three-component composition: (Employed, Unemployed, Not in the labour force). This would however require a four-dimensional plot. We acknowledge though the need for modelling the covariation between two compositions with D components in the space $\mathcal{S}^D \times \mathcal{S}^D$.

As well as studying the joint probability density of two compositions, we could study the probability density for a composition conditioned on the values of another composition. In our example we might study the composition (Employed, Not employed) or (Employed, Unemployed, Not in the labour force) in June, conditioned on the value in January the previous year. Also for this context we need suitable distributions to be able to create models.

2 A Bicompositional Dirichlet Distribution

Following Aitchison (1986), we define $d = D - 1$. We let $\mathbf{x} = (x_1, \dots, x_d, x_D)' \in \mathcal{S}^D$. The well-known (unicompositional) Dirichlet distribution with parameter $\boldsymbol{\alpha} \in \mathcal{R}_+^D$ is defined as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_D)}{\prod_{j=1}^D \Gamma(\alpha_j)} \prod_{j=1}^D x_j^{\alpha_j - 1}$$

where $\Gamma(\cdot)$ is the Gamma function.

Based on the Dirichlet distribution we will now define a bicompositional generalisation for \mathbf{X} and \mathbf{Y} .

Definition 1 (Probability Density Function). Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}^D$ and let

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = A \left(\prod_{j=1}^D x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) (\mathbf{x}' \mathbf{y})^\gamma \quad (1)$$

where $\alpha_j, \beta_j \in \mathcal{R}_+$ for $j = 1, \dots, D$ and γ is a real number.

The parameter γ models the degree of covariation between \mathbf{X} and \mathbf{Y} . If $\gamma = 0$, (1) reduces to the product of two independent Dirichlet probability density functions

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = A \prod_{j=1}^D x_j^{\alpha_j - 1} y_j^{\beta_j - 1}$$

where A is known:

$$A = \frac{\Gamma(\alpha_1 + \dots + \alpha_D) \Gamma(\beta_1 + \dots + \beta_D)}{\prod_{j=1}^D \Gamma(\alpha_j) \Gamma(\beta_j)}.$$

This motivates that (1) is referred to as the probability density function of a bicompositional Dirichlet distribution. We let $\mathcal{D}_1^D(\boldsymbol{\alpha})$ denote the unicompositional Dirichlet distribution with D components and parameter $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_D)'$ and we let $\mathcal{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ denote the bicompositional Dirichlet distribution with D components and parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and γ , where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_D)'$.

Next we examine some of the properties of this distribution: first in the special bicomponent case, and then in the general multicomponent case.

3 The Bicomponent Case

For two bicomponent compositions, $\mathbf{X} = (X, 1 - X)'$ and $\mathbf{Y} = (Y, 1 - Y)'$, the probability density function is a function of $\mathbf{x} = (x, 1 - x)'$ and $\mathbf{y} = (y, 1 - y)'$,

but as it is completely determined by (x, y) we will for simplicity treat it as function of (x, y) . Hence the probability density function (1) is reduced to

$$f_{X,Y}(x, y) = Ax^{\alpha_1-1}(1-x)^{\alpha_2-1}y^{\beta_1-1}(1-y)^{\beta_2-1} (xy + (1-x)(1-y))^\gamma. \quad (2)$$

This distribution could of course also be considered a bivariate Beta distribution, if we regard it as a function of (x, y) .

We begin our investigation of the distribution by stating for what values of γ the distribution exists.

Theorem 1. *The distribution $\mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$.*

Proof. If γ is a non-negative then the last factor in

$$x^{\alpha_1-1}(1-x)^{\alpha_2-1}y^{\beta_1-1}(1-y)^{\beta_2-1} (xy + (1-x)(1-y))^\gamma \quad (3)$$

is bounded for $x, y \in [0, 1]$ and the integral of (3) exists. If γ is negative then the last factor is unbounded.

Let γ be negative and $g = -\gamma$. Then the integral of (3) may be written as

$$\int_0^1 \int_0^1 \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}y^{\beta_1-1}(1-y)^{\beta_2-1}}{(xy + (1-x)(1-y))^g} dx dy. \quad (4)$$

The denominator is close to 0 only when (x, y) is close to either $(0, 1)$ or $(1, 0)$. For the rest of the integration area the denominator is bounded away from 0. In the triangle $(0, .5)$, $(0, 1)$, $(.5, 1)$, i.e. the top left triangle, $(xy + (1-x)(1-y))$ may estimated from below by $(x + 1 - y)/2$ and from above by $(x + 1 - y)$. In the bottom right triangle, $(xy + (1-x)(1-y))$ may estimated from below by $(y + 1 - x)/2$ and from above by $(y + 1 - x)$. The integration area, with the top left and bottom right triangles shaded, is shown in Figure 2.

The integral (4) thus exists if and only if the integral

$$\int_{\frac{1}{2}}^1 \left(\int_0^{y-\frac{1}{2}} \frac{x^{\alpha_1-1}(1-y)^{\beta_2-1}}{(x+1-y)^g} dx \right) dy \quad (5)$$

and the corresponding integral over the bottom triangle both exist. Introducing $u = x$ and $v = x + 1 - y$, (5) turns into

$$\int_0^{\frac{1}{2}} \left(\int_0^v \frac{u^{\alpha_1-1}(v-u)^{\beta_2-1}}{v^g} du \right) dv. \quad (6)$$

If we replace u by vt , (6) may be written as

$$\int_0^{\frac{1}{2}} v^{\alpha_1+\beta_2-g-1} dv \int_0^1 t^{\alpha_1-1}(1-t)^{\beta_2-1} dt.$$

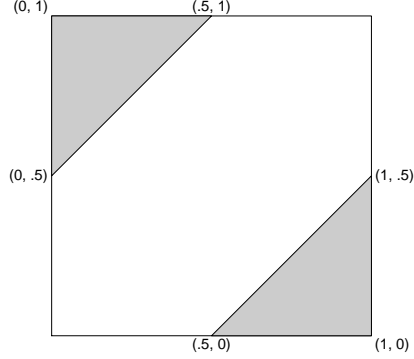


Figure 2. The integration area of (4). The denominator is close to 0 only when (x, y) is close to either $(0, 1)$ or $(1, 0)$.

The second integral of this product always exists, but the first one exists if and only if $\alpha_1 + \beta_2 > g$. With an analogous argument for the integral over the bottom right triangle, we can show that this integral exists if and only if $\alpha_2 + \beta_1 > g$. Hence the density (2) exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. \square

We next determine the normalisation constant A .

Theorem 2. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the normalisation constant A is determined by*

$$\frac{1}{A} = \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j) \right) \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \quad (7)$$

where $B(p, q)$ is the Beta function:

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Proof. Let ξ and η be defined by $x = \frac{1}{2}(1 + \xi)$ and $y = \frac{1}{2}(1 + \eta)$. Then $1 - x = \frac{1}{2}(1 - \xi)$, $1 - y = \frac{1}{2}(1 - \eta)$ and $xy + (1 - x)(1 - y) = \frac{1}{2}(1 + \xi\eta)$ and the Binomial expansion yields

$$(xy + (1 - x)(1 - y))^\gamma = \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} (\xi\eta)^i. \quad (8)$$

We note that, since $0 < x < 1$ and $0 < y < 1$, and thus $-1 < \xi < 1$ and $-1 < \eta < 1$, the series on the right-hand side of (8) converges. The normalisation constant is then determined by

$$\begin{aligned}
\frac{1}{A} &= \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \iint x^{\alpha_1-1} (1-x)^{\alpha_2-1} y^{\beta_1-1} (1-y)^{\beta_2-1} (\xi\eta)^i dx dy \\
&= \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\int x^{\alpha_1-1} (1-x)^{\alpha_2-1} \xi^i dx \right) \left(\int y^{\beta_1-1} (1-y)^{\beta_2-1} \eta^i dy \right) \\
&= \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\int x^{\alpha_1-1} (1-x)^{\alpha_2-1} [x - (1-x)]^i dx \right) \\
&\quad \cdot \left(\int y^{\beta_1-1} (1-y)^{\beta_2-1} [y - (1-y)]^i dy \right) \\
&= \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j) \right) \\
&\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right)
\end{aligned}$$

where all integrals are over the unit interval. \square

Empirical trials show that the series (7) converges quickly for most examples. If all the parameters are close to 0, the series may converge very slowly. We have also found that convergence can be slow when γ is negative and close to $-\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. These situations however mean that the probability is concentrated near the edges of the sample space and hence perhaps of little practical importance.

If γ is a non-negative integer the results are simplified as shown next.

Theorem 3. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the normalisation constant A is determined by*

$$\frac{1}{A} = \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \quad (9)$$

where $B(\cdot, \cdot)$ is the Beta function.

Proof. The result follows from expanding the last factor of (2) using the Binomial theorem and then integrating using the definition of the Beta function. \square

We proceed by also stating the cumulative distribution function.

Theorem 4. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the cumulative distribution function is

$$F_{X,Y}(x, y) = \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B_x(\alpha_1 + j, \alpha_2 + i - j) \right) \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B_y(\beta_1 + k, \beta_2 + i - k) \right) \quad (10)$$

where A is the constant given in Theorem 2 and $B_x(p, q)$ is the incomplete Beta function defined as

$$B_x(p, q) = \int_0^x t^{p-1} (1-t)^{q-1} dt.$$

Proof. The proof is similar to the proof of Theorem 2, but uses the definition of the incomplete Beta function instead of the Beta function. \square

If γ is a non-negative integer the cumulative distribution function may be expressed more simply.

Theorem 5. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the cumulative distribution function is

$$F_{X,Y}(x, y) = A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B_x(\alpha_1 + j, \alpha_2 + \gamma - j) B_y(\beta_1 + j, \beta_2 + \gamma - j) \quad (11)$$

where A is the constant given in Theorem 3 and $B_x(\cdot, \cdot)$ is the incomplete Beta function.

Proof. The result follows by integrating (2) with the use of the Binomial theorem and the definition of the incomplete Beta function. \square

Next we give the product moments of the distribution.

Theorem 6. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the product moment $E(X^n Y^m)$ is

$$\begin{aligned} E(X^n Y^m) &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) \right) \\ &\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + m + k, \beta_2 + i - k) \right) \\ &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) \left(\prod_{l=0}^{n-1} \frac{\alpha_1 + j + l}{\alpha_1 + \alpha_2 + \gamma + l} \right) \right) \\ &\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + m + k, \beta_2 + i - k) \left(\prod_{l=0}^{m-1} \frac{\beta_1 + k + l}{\beta_1 + \beta_2 + \gamma + l} \right) \right) \end{aligned}$$

where A is the constant given in Theorem 2.

Proof. The proof of the first equality is similar to that of Theorem 2. The second equality follows from repeated use of the identity

$$B(p+1, q) = \frac{p}{p+q} B(p, q). \quad (12)$$

□

As before, if γ is a non-negative integer, the calculations are simplified.

Theorem 7. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the product moment $E(X^n Y^m)$ is*

$$\begin{aligned} E(X^n Y^m) &= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + n + j, \alpha_2 + \gamma - j) B(\beta_1 + m + j, \beta_2 + \gamma - j) \\ &= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \\ &\quad \cdot \left(\prod_{k=0}^{n-1} \frac{\alpha_1 + j + k}{\alpha_1 + \alpha_2 + \gamma + k} \right) \left(\prod_{l=0}^{m-1} \frac{\beta_1 + j + l}{\beta_1 + \beta_2 + \gamma + l} \right) \end{aligned}$$

where A is the constant given in Theorem 3.

Proof. The proof of the first equality is similar to the proof of Theorem 3. The second equality follows from repeated use of the identity (12). □

3.1 Marginal Distributions

In the example in Section 1.3 we noted that not only the joint distribution, but also the conditional distributions may be of interest when modelling bi-compositional data. In order to determine the properties of the conditional distributions, we first need to derive some of the properties of the marginal distributions.

Theorem 8. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the marginal probability density function of X is*

$$\begin{aligned} f_X(x) &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} x^{\alpha_1-1} (1-x)^{\alpha_2-1} (x - (1-x))^i \\ &\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \end{aligned} \quad (13)$$

where A is the constant given in Theorem 2.

Proof. The result follows by expanding (2) similarly to the proof of Theorem 7 and then integrating. \square

As previously the results are simplified for non-negative integer values on γ .

Theorem 9. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the marginal probability density function of X is*

$$f_X(x) = A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\beta_1 + j, \beta_2 + \gamma - j) x^{\alpha_1 + j - 1} (1 - x)^{\alpha_2 + \gamma - j - 1} \quad (14)$$

where A is the constant given in Theorem 3.

Proof. The result follows directly by integrating (2). \square

When $\gamma = 0$, we see that (14) is reduced to the common unicompositional Dirichlet distribution.

For completeness we also state the moments of the marginal distributions.

Theorem 10. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the n^{th} moment of X is*

$$\begin{aligned} E(X^n) &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) \right) \\ &\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \\ &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^i \binom{i}{j} (-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j) \right) \\ &\quad \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \\ &\quad \cdot \left(\prod_{l=0}^{n-1} \frac{\alpha_1 + i + l}{\alpha_1 + \alpha_2 + \gamma + l} \right) \end{aligned}$$

where A is the constant given in Theorem 2.

Proof. The result follows by expanding $(x - (1 - x))^i$, integrating, and repeatedly using the identity (12). \square

Again, the calculations are simplified for integer values.

Theorem 11. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the n^{th} moment of X is

$$\begin{aligned} E(X^n) &= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + n + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \\ &= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \left(\prod_{k=0}^{n-1} \frac{\alpha_1 + j + k}{\alpha_1 + \alpha_2 + \gamma + k} \right) \end{aligned}$$

where A is the constant given in Theorem 3.

Proof. The result follows from direct calculation and repeated use of the identity (12). \square

We note that if $\gamma = 0$, then $E(X) = \frac{\alpha_1}{\alpha_1 + \alpha_2}$, i.e. precisely the expectation of a Dirichlet distribution, as one would expect.

Due to the symmetry of the Bicompositional Dirichlet distribution, all of the above results of course also apply to Y (with the appropriate changes of α_i to β_i and vice versa).

3.2 Conditional Distributions

We now proceed with the bicomponent conditional distributions, first stating the conditional probability density function.

Theorem 12. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the conditional probability density function for Y conditioned on $X = x$ is

$$f_{Y|X=x}(y) = C y^{\beta_1-1} (1-y)^{\beta_2-1} (xy + (1-x)(1-y))^\gamma \quad (15)$$

where

$$\frac{1}{C} = \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} (x - (1-x))^i \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right).$$

Proof. The result follows directly from (2) and Theorem 8. \square

Theorem 13. If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the conditional probability density function for Y conditioned on $X = x$ is

$$f_{Y|X=x}(y) = D \sum_{k=0}^{\gamma} \binom{\gamma}{k} x^k (1-x)^{\gamma-k} y^{\beta_1+k-1} (1-y)^{\beta_2+\gamma-k-1} \quad (16)$$

where

$$\frac{1}{D} = \sum_{j=0}^{\gamma} \binom{\gamma}{j} x^j (1-x)^{\gamma-j} B(\beta_1 + j, \beta_2 + \gamma - j).$$

Proof. The result follows directly from (2) and Theorem 9. \square

We note that when $\gamma = 0$, (16) simplifies to the probability density function of the unicompositional Dirichlet distribution with parameters β_1 and β_2 .

We also derive the moments for the conditional distributions.

Theorem 14. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the n^{th} moment of $Y|X = x$ is*

$$E(Y^n|X = x) = \frac{C}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} (x - (1-x))^i \cdot \left(\sum_{k=0}^i \binom{i}{k} (-1)^{i-k} B(\beta_1 + n + k, \beta_2 + i - k) \right) \quad (17)$$

where C was given in Theorem 12.

Proof. The proof is similar to the proof of Theorem 10. \square

Theorem 15. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the n^{th} moment of $Y|X = x$ is*

$$E(Y^n|X = x) = D \sum_{j=0}^{\gamma} \binom{\gamma}{j} x^j (1-x)^{\gamma-j} B(\beta_1 + n + j, \beta_2 + \gamma - j) \quad (18)$$

where D was given in Theorem 13.

Proof. The proof is similar to the proof of Theorem 11. \square

In both (17) and (18), the Beta function may be expanded to a product similar to the ones in Theorems 10 and 11.

4 The Multicomponent Case

We now turn to the case when there are $D > 2$ components. Unfortunately, due to the increased complexity, the results for this case are less elaborate than for the bicomponent case.

The distribution exists if γ is a non-negative real number. If γ furthermore is a non-negative integer, the normalisation constant may easily be determined. To simplify the expressions, we let $\alpha. = \alpha_1 + \dots + \alpha_D$, $\beta. = \beta_1 + \dots + \beta_D$ and $\mathbf{k} = (k_1, \dots, k_D)$.

Theorem 16. *If $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\alpha., \beta.; \gamma)$ and γ is a non-negative integer, then the normalisation constant is determined by*

$$\frac{1}{A} = \sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\alpha. + \gamma)} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta. + \gamma)} \quad (19)$$

where $\binom{\gamma}{\mathbf{k}}$ is the multinomial coefficient $\gamma!/(k_1! \cdots k_D!)$.

Proof. The result follows from expanding the factor $(\mathbf{x}'\mathbf{y})^\gamma$ by using the Multinomial theorem and from the properties of the Dirichlet integral. \square

For computational purposes, we note that (19) can for instance be written as

$$\frac{1}{A} = \sum_{k_1=0}^{\gamma} \sum_{k_2=0}^{\gamma-k_1} \cdots \sum_{k_d=0}^{\gamma-\sum_{i=1}^{d-1} k_i} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\boldsymbol{\alpha} + \boldsymbol{\gamma})} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \boldsymbol{\gamma})}$$

if we define $k_D = \gamma - k_1 - \cdots - k_d$.

4.1 Marginal Distributions

Before we examine the conditional distributions we give the marginal distributions for the multicomponent case.

First we give the probability density function.

Theorem 17. *If $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ and γ is a non-negative integer, then the marginal probability density function for \mathbf{X} is*

$$f_{\mathbf{X}}(\mathbf{x}) = A \sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{\alpha_1 + k_1 - 1} \cdots x_D^{\alpha_D + k_D - 1} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \boldsymbol{\gamma})}. \quad (20)$$

where A is the constant given in Theorem 16.

Proof. The proof is similar to that of Theorem 16. \square

Just as for the bicomponent case, we note that when $\gamma = 0$, (20) simplifies to a unicompositional Dirichlet distribution with parameter $\boldsymbol{\alpha}$. When γ is positive integer, (20) just like (1) is a mixture of unicompositional Dirichlet distributions.

Next we determine the moments of the components of the marginal distributions.

Theorem 18. *If $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ and γ is a non-negative integer, then*

$$E(X_j^n) = A \sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\boldsymbol{\alpha} + \boldsymbol{\gamma})} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \boldsymbol{\gamma})} \prod_{l=0}^{n-1} \frac{\alpha_j + k_j + l}{\alpha_j + \gamma + l} \quad (21)$$

for $j = 1, \dots, D$ and $n = 1, 2, \dots$

Proof. The result follows by using the Dirichlet integral and the identity $\Gamma(x+1) = x\Gamma(x)$ repeatedly. \square

4.2 Conditional Distributions

Having determined the marginal distributions we continue by determining the multicomponent conditional distributions. We begin with the conditional probability density function of the composition \mathbf{Y} conditioned on the composition $\mathbf{X} = \mathbf{x}$.

Theorem 19. *If $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ and γ is a non-negative integer, then the conditional probability density function for \mathbf{Y} given $\mathbf{X} = \mathbf{x}$ is*

$$f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{y_1^{\beta_1-1} \cdots y_D^{\beta_D-1} (x_1 y_1 + \cdots + x_D y_D)^\gamma}{\sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \gamma)}}. \quad (22)$$

Proof. The result follows directly from (2) and Theorem 20. \square

As before, we note that if $\gamma = 0$, (22) reduces to a unicompositional Dirichlet distribution. Using (22) we next determine the moments of the multicomponent conditional distribution.

Theorem 20. *If $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ and γ is a non-negative integer, then*

$$E(Y_j^n | \mathbf{X} = \mathbf{x}) = B \sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \gamma)} \prod_{l=0}^{n-1} \frac{\beta_j + k_j + l}{\beta_j + \gamma + l} \quad (23)$$

for $j = 1, \dots, D$ and $n = 1, 2, \dots$; here

$$\frac{1}{B} = \sum_{\substack{k_1, \dots, k_D \geq 0 \\ k_1 + \dots + k_D = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\boldsymbol{\beta} + \gamma)}.$$

Proof. The proof is similar to the proof of Theorem 18. \square

5 Discussion

There are not many distributions defined on the sample space consisting of the Cartesian product of two Simplices. In this paper we have proposed a bicompositional generalisation of the unicompositional Dirichlet distribution. The proposed bicompositional Dirichlet distribution allows for modelling of covariation between compositions without leaving $\mathcal{S}^D \times \mathcal{S}^D$. We have determined in the bicomponent case for what parameter values the distribution exists. We have also derived the marginal and conditional distributions and their moments. For the bicomponent case we have also derived the cumulative distribution function and the product moment.

The proposed distribution is meant to be a first suggestion for modelling bicompositional data and it is developed primarily to possess properties that will be exploited in future work. Among these properties, apart from simplicity, are the ability to model dependence and independence between compositions and also the fact that the distributions constitute an exponential family of distributions.

6 Acknowledgements

The author wishes to thank Professor Jan Lanke for all his time and effort in helping improve this article.

References

- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. London: Chapman and Hall.
- Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.
- Aitchison, J. and S. M. Shen (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* 67(2), 261–272.
- Nadarajah, S. (2006). The bivariate F_3 -beta distribution. *Commun. Korean. Math. Soc.* 21(2), 363–374.
- Nadarajah, S. (2007). A new bivariate beta distribution with application to drought data. *Metron* LXV(2), 153–174.
- Nadarajah, S. and S. Kotz (2005). Some bivariate beta distributions. *Statistics* 39(5), 457–466.
- Olkin, I. and R. Liu (2003). A bivariate beta distribution. *Statistics and Probability Letters* 62(4), 407–412.
- Ongaro, A., S. Migliorati, and G. Monti (2008). A new distribution on the simplex containing the dirichlet family. In J. Daunis i Estadella and J. Martín-Fernández (Eds.), *Proceedings of CODAWORK'08, CD-ROM*. Universitat de Girona.
- Persson, E. and M. Henkel (2005, 10). Arbetskraftsundersökningarna 2004. *Statistiska Meddelanden AM 12 SM 0510*.