

RESEARCH

Open Access

# The spherical-Dirichlet distribution



Jose H. Guardiola 

Correspondence:  
[jose.guardiola@tamucc.edu](mailto:jose.guardiola@tamucc.edu)  
Texas A&M University-Corpus  
Christi, Department of Mathematics  
and Statistics, 6300 Ocean Drive,  
CI-309, 78412 Corpus Christi, TX,  
USA

## Abstract

Today, data mining and gene expressions are at the forefront of modern data analysis. Here we introduce a novel probability distribution that is applicable in these fields. This paper develops the proposed spherical-Dirichlet distribution designed to fit vectors located at the positive orthant of the hypersphere, as it is often the case for data in these fields, avoiding unnecessary probability mass. Basic properties of the proposed distribution, including normalizing constants and moments are developed. Relationships with other distributions are also explored. Estimators based on classical inferential statistics, such as method of moments and maximum likelihood estimators are obtained. Two applications are developed: the first one uses simulated data, and the second uses a real text mining example. Both examples are fitted using the proposed spherical-Dirichlet distribution and their results are discussed.

**Keywords:** Dirichlet distribution, Text mining, Hypersphere, Gene expressions, Positive orthant

## Introduction

In text mining and gene expressions analysis, the collections of texts are represented in a vector-space model, which implies that texts once standardized, are coded as vectors in a sphere of higher dimensions, also called a hypersphere (Suvrit 2016). Many researchers currently model these distributions by means of existing probability density mixtures, however, these approximations waste probability mass in the whole hypersphere, when it is actually only needed at the positive orthant of the hypersphere. This is mainly because of the non-existence of suitable distributions for that subspace. The new proposed distribution fills that void, allowing for an efficient modeling of these vectors.

## Basic properties

In this section we introduce the proposed spherical-Dirichlet distribution, its moments and basic properties.

## Probability density function and normalizing constant

The spherical-Dirichlet distribution is obtained by transforming the Dirichlet distribution on the simplex to the corresponding space on the hypersphere. In this section we derive the density and we compute the normalizing constants. Let  $\mathbf{y}$  have a Dirichlet distribution on the simplex as described by Ingram (Olkin and Rubin 1964).

$$\begin{aligned}
 f_{\text{Dir}}(\mathbf{y}; \alpha) &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m y_i^{\alpha_i-1} \\
 &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} y_i^{\alpha_i-1} \left(1 - \sum_{i=1}^{m-1} y_i\right)^{(\alpha_m-1)}
 \end{aligned} \quad (1)$$

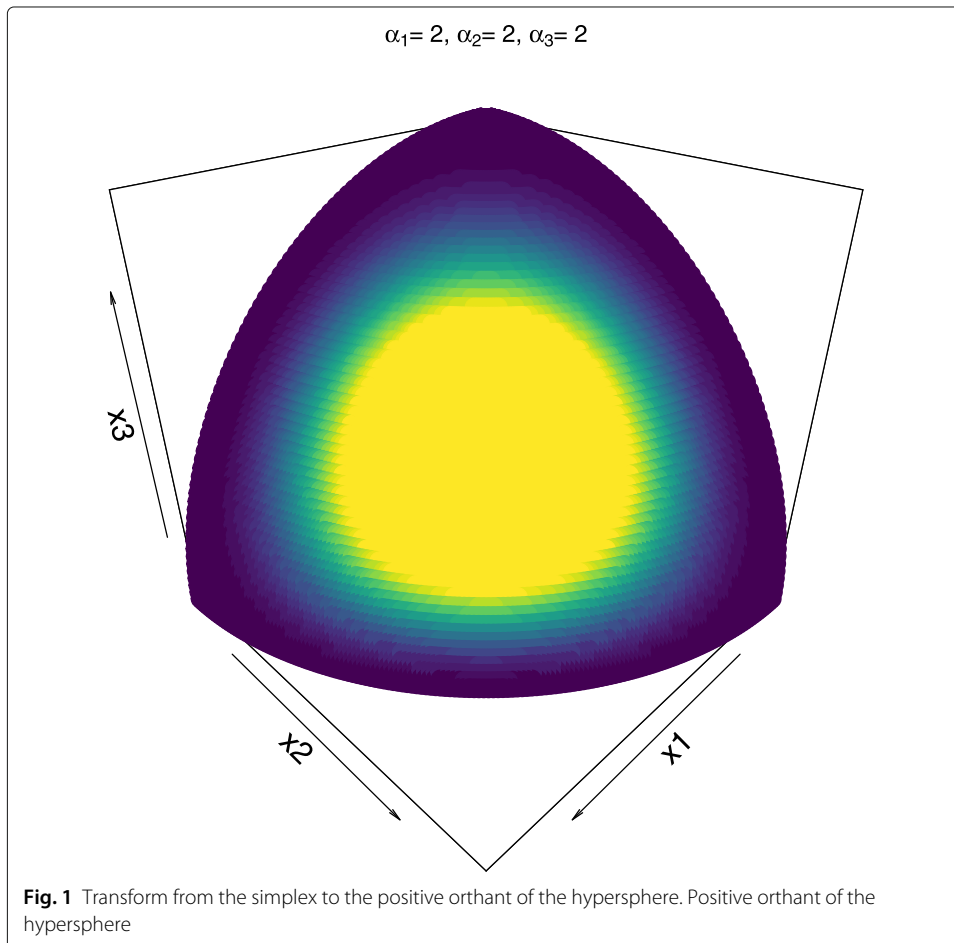
where

$$\alpha_i \in \mathbb{R}^+, \quad \alpha_0 =: \sum_{i=1}^m \alpha_i, \quad 0 \leq y_i \leq 1, \quad \sum_{i=1}^m y_i = 1,$$

Transforming the Dirichlet distribution from the simplex to the positive orthant of the hypersphere (Fig. 1)

taking the square root transformation

$$x_i = \sqrt{y_i}, \quad y_i = x_i^2, \quad \frac{\partial y_i}{\partial x_i} = 2x_i, \quad \text{for } i = 1, \dots, (m-1), \quad x_m = \sqrt{y_m}. \quad (2)$$



Computing the Jacobian for all independent variables, it follows that

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} = 2x_1 & \frac{\partial y_1}{\partial x_2} = 0 & 0 & \dots \\ \frac{\partial y_2}{\partial x_1} = 0 & \frac{\partial y_2}{\partial x_2} = 2x_2 & 0 & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{\partial y_{m-1}}{\partial x_{m-1}} = 2x_{m-1} \end{vmatrix} = \prod_{i=1}^{m-1} 2x_i = 2^{m-1} \prod_{i=1}^{m-1} x_i$$

the proposed transformation from (1) results in

$$\begin{aligned} f_{\text{SDir}}(\mathbf{x}; \alpha) &= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} x_i^{2\alpha_i-1} \cdot x_m^{2\alpha_m-2} \\ &= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m x_i^{2\alpha_i-1} \cdot x_m^{-1} \\ &= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} x_i^{2\alpha_i-1} \left(1 - \sum_{i=1}^{m-1} x_i^2\right)^{(\alpha_m-1)} \end{aligned} \quad (3)$$

where

$$\alpha_0 =: \sum_{i=1}^m \alpha_i, \quad \alpha_i \in \mathbb{R}^+, \quad 0 \leq x_i \leq 1, \quad \sum_{i=1}^m x_i^2 = 1.$$

We refer to (3) as the spherical-Dirichlet distribution (SDD) and write  $x \sim \text{SDD}(\alpha_i)$ . We introduce the parameters  $\alpha_i$  as the concentration parameters in a similar manner to the corresponding parameters of the Dirichlet distribution.

### Moments

In this section we compute the first and second order moments, mode, standard deviation, variances and covariances and its corresponding covariance matrix. First, we compute the expected value for one of the variables, for example let us consider the expected value of  $x_1$

$$E(x_1) = \int \dots \int \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} x_1 \left( \prod_{i=1}^m x_i^{2\alpha_i-1} \right) \cdot x_m^{-1} dx_1 \dots dx_m \quad (4)$$

$$= \int \dots \int \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} x_1^{2(\alpha_1+\frac{1}{2})-1} \left( \prod_{i=2}^m x_i^{2\alpha_i-1} \right) \cdot x_m^{-1} dx_1 \dots dx_m, \quad (5)$$

where we recognize the expression inside the integral as the kernel of the proposed SDD with a new first parameter  $\alpha_1 + \frac{1}{2}$ , then we can rewrite immediately this expression as

$$\begin{aligned} E(x_1) &= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \frac{\Gamma(\alpha_1 + \frac{1}{2}) \prod_{i=2}^m \Gamma(\alpha_i)}{2^{m-1} \Gamma(\alpha_0 + \frac{1}{2})} \\ &= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \frac{1}{2})} \frac{\Gamma(\alpha_1 + \frac{1}{2})}{\Gamma(\alpha_1)}, \end{aligned} \quad (6)$$

we define  $\mu_i$  as,

$$\mu_i =: \frac{\Gamma(\alpha_i + \frac{1}{2})}{\Gamma(\alpha_i)}, \quad (7)$$

the expected value from (6) can be rewritten as,

$$E(x_i) = \frac{\mu_i}{\mu_0}. \quad (8)$$

The general solution for the first moment of a vector  $\mathbf{x} = \{x_1, \dots, x_m\}^T$  with a vector of parameters  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_m\}^T$  can be written as

$$E(\mathbf{x}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \frac{1}{2})} \left( \frac{\Gamma(\alpha_1 + \frac{1}{2})}{\Gamma(\alpha_1)}, \dots, \frac{\Gamma(\alpha_{m-1} + \frac{1}{2})}{\Gamma(\alpha_{m-1})} \right) = \frac{1}{\mu_0} \frac{\Gamma(\boldsymbol{\alpha} + \frac{1}{2})}{\Gamma(\boldsymbol{\alpha})}, \quad (9)$$

let

$$\boldsymbol{\mu} =: \frac{\Gamma(\boldsymbol{\alpha} + \frac{1}{2})}{\Gamma(\boldsymbol{\alpha})}, \quad C =: \frac{\|\boldsymbol{\mu}\|}{\mu_0}, \quad \bar{\boldsymbol{\mu}} =: \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}, \quad \bar{\boldsymbol{\mu}} \in \Omega_{m-1}, \quad (10)$$

then, the expected value for a vector  $\mathbf{x}$  can also be written as

$$E(\mathbf{x}) = \frac{\boldsymbol{\mu}}{\mu_0} = \frac{\|\boldsymbol{\mu}\|}{\mu_0} \cdot \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|} = C \cdot \bar{\boldsymbol{\mu}}. \quad (11)$$

Similarly, we compute the expected value for  $x_1^2$  as

$$E(x_1^2) = \int \dots \int \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} x_1^2 \left( \prod_{i=1}^m x_i^{2\alpha_i-1} \right) \cdot x_m^{-1} dx_1 \dots dx_m \quad (12)$$

$$= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \int \dots \int x_1^{2(\alpha_1+1)-1} \left( \prod_{i=2}^m x_i^{2\alpha_i-1} \right) \cdot x_m^{-1} dx_1 \dots dx_m, \quad (13)$$

again, we can recognize the expression inside the integral as the kernel of the proposed SDD with a new first parameter  $\alpha_1 + 1$ , that yields

$$\begin{aligned} E(x_1^2) &= \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \frac{\Gamma(\alpha_1 + 1) \prod_{i=2}^{m-1} \Gamma(\alpha_i)}{2^{m-1} \Gamma(\alpha_0 + 1)} \\ &= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + 1)} \frac{\Gamma(\alpha_1 + 1)}{\Gamma(\alpha_1)} = \frac{\alpha_1}{\alpha_0}, \end{aligned} \quad (14)$$

this result can be generalized to

$$E(x_i^2) = \frac{\alpha_i}{\alpha_0}. \quad (15)$$

Moreover, the variance for any variable  $x_i$  is

$$V(x_i) = \frac{\alpha_i}{\alpha_0} - \frac{\mu_i^2}{\mu_0^2}, \quad (16)$$

and the covariance for  $x_1, x_2$  can be written as

$$E(x_1 \cdot x_2) = \int \dots \int \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} x_1 \cdot x_2 \left( \prod_{i=1}^m x_i^{2\alpha_i-1} \right) \cdot x_m^{-1} dx_1 \dots dx_m, \quad (17)$$

after some arrangements, we can identify the kernel of the proposed SDD with the first two parameters as  $\alpha_1 + \frac{1}{2}$ , and  $\alpha_2 + \frac{1}{2}$ , where we can solve the corresponding integral, and our result takes the form

$$E(x_1 \cdot x_2) = \frac{\Gamma(\alpha_1 + \frac{1}{2}) \Gamma(\alpha_2 + \frac{1}{2})}{\alpha_0 \Gamma(\alpha_1) \Gamma(\alpha_2)} = \frac{\mu_1 \cdot \mu_2}{\alpha_0}. \quad (18)$$

In general, for any pair of variables  $(x_i, x_j)$  we can write

$$E(x_i \cdot x_j) = \delta_{ij} \cdot \frac{\alpha_i}{\alpha_0} + (1 - \delta_{ij}) \cdot \frac{\mu_i \cdot \mu_j}{\alpha_0}, \quad (19)$$

where  $\delta_{ij}$  is the delta Kronecker, and we can also write the covariance for any pair of variables  $(x_i, x_j)$  as

$$COV(x_i, x_j) = \left( \frac{1}{\alpha_0} - \frac{1}{\mu_0^2} \right) \mu_i \cdot \mu_j \text{ for } i \neq j. \quad (20)$$

We can also write the covariance for any pair of variables  $(x_i, x_j)$  as

$$COV(x_i, x_j) = \delta_{ij} \cdot \left( \frac{\alpha_{i=j}}{\alpha_0} - \frac{\mu_i^2}{\mu_0^2} \right) + (1 - \delta_{ij}) \cdot \left( \frac{1}{\alpha_0} - \frac{1}{\mu_0^2} \right) \mu_i \cdot \mu_j, \quad (21)$$

that in matrix notation can also be written as

$$\Sigma = \begin{bmatrix} \frac{\alpha_1}{\alpha_0} - \frac{\mu_1^2}{\mu_0^2} & \left( \frac{1}{\alpha_0} - \frac{1}{\mu_0^2} \right) \mu_1 \cdot \mu_2 & \dots & \dots \\ \left( \frac{1}{\alpha_0} - \frac{1}{\mu_0^2} \right) \mu_2 \cdot \mu_1 & \frac{\alpha_2}{\alpha_0} - \frac{\mu_2^2}{\mu_0^2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \frac{\alpha_m}{\alpha_0} - \frac{\mu_m^2}{\mu_0^2} \end{bmatrix},$$

an equivalent expression is

$$\Sigma = \frac{1}{\alpha_0} \begin{bmatrix} \alpha_1 - \mu_1^2 & 0 & \dots & \dots \\ 0 & \alpha_2 - \mu_2^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \alpha_m - \mu_m^2 \end{bmatrix} - \left( \frac{1}{\mu_0^2} - \frac{1}{\alpha_0} \right) \mu \mu^T,$$

similarly we let

$$\Sigma = \frac{1}{\alpha_0} \text{diag}(\alpha) - \frac{C^2 \mu_0^2}{\alpha_0} \text{diag}(\bar{\mu} \bar{\mu}^T) - C^2 \left( 1 - \frac{\mu_0^2}{\alpha_0} \right) \bar{\mu} \bar{\mu}^T, \quad (22)$$

where

$$C = \frac{\|\mu\|}{\mu_0}, \quad \bar{\mu} = \frac{\mu}{\|\mu\|}, \quad \bar{\mu} \in \Omega_{m-1}, \quad (23)$$

that summarizes our results in a succinct form.

### Mode and relationship with the mean

The mode for the SDD can be determined by finding the values of  $\alpha_i$  that maximize this function, alternatively, we can also maximize the log of this function as it is customary

and usually an easier procedure. First, taking the natural log of the SDD and adding the constraint  $\sum_{i=1}^m x_i^2 = 1$  for the purpose of using Lagrange multipliers we get

$$\ln f_{\text{SDDir}}(\mathbf{x}, \alpha) = \ln \left( \frac{2^{m-1} \Gamma(\alpha_0)}{\prod_{i=1}^m \Gamma(\alpha_i)} \right) + \sum_{i=1}^m (2\alpha_i - 1) \ln x_i - \ln x_m - \lambda \left( \sum_{i=1}^m x_i^2 - 1 \right), \quad (24)$$

taking derivatives with respect to  $x_i$  and setting them to zero we have

$$\frac{\partial \ln f_{\text{SDDir}}}{\partial x_i} = (2\alpha_i - 1) \frac{1}{x_i} - 2x_i \lambda = 0 \text{ for } i < m, \quad (25)$$

solving for  $x_i^2$ , it yields

$$x_i^2 = \frac{2\alpha_i - 1}{2\lambda} \text{ for } i < m, \quad (26)$$

similarly, taking derivatives with respect to  $x_m$

$$\frac{\partial \ln f_{\text{SDDir}}}{\partial x_m} = (2\alpha_m - 1) \frac{1}{x_m} - \frac{1}{x_m} - 2x_m \lambda = 0 \text{ for } i = m, \quad (27)$$

and solving for  $x_m$ , we have

$$x_m^2 = \frac{\alpha_m - 1}{\lambda} \text{ for } i = m, \quad (28)$$

substituting these results at the constraint, we can solve for  $\lambda$  as

$$\lambda = \frac{1}{2}(2\alpha_0 - m - 1), \quad (29)$$

where we can obtain the mode for  $x_i$  as

$$(\text{mode})x_i = \sqrt{\frac{2\alpha_i - 1}{2\alpha_0 - m - 1}} \text{ for } i < m, \quad (30)$$

and for  $x_m$

$$(\text{mode})x_m = \sqrt{\frac{2(\alpha_m - 1)}{2\alpha_0 - m - 1}} \text{ for } i = m. \quad (31)$$

Considering the special case of a symmetric SDD, we set up  $\alpha_i = \alpha$  for  $i < m$ , and  $\alpha_m = \alpha + \frac{1}{2}$ , both (30) and (31) yield

$$(\text{mode})x_i = \sqrt{\frac{2\alpha - 1}{m \cdot (2\alpha - 1)}} = \frac{1}{\sqrt{m}} \text{ for } \alpha \neq \frac{1}{2} \text{ for } i \leq m, \quad (32)$$

the mean for a symmetric SDD for  $\alpha_i = \alpha$  for  $i < m$ , and  $\alpha_m = \alpha + \frac{1}{2}$ , yields

$$E(x_i) = \frac{\mu_i}{\mu_0} = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \frac{1}{2})} = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} \cdot \frac{\Gamma(m\alpha + \frac{1}{2})}{\Gamma(m\alpha + 1)}, \quad (33)$$

where we can see that the mode does not match the expected value for a symmetric SDD, however, we can still find an asymptotic relationship using the expression developed by Frame (Frame 1949),

$$\lim_{x \rightarrow \infty} f(x) = \frac{\Gamma(x+a)}{\Gamma(x)} = x^a, \quad (34)$$

using this approximation it yields

$$\lim_{\alpha \rightarrow \infty} E(x_i) = \left(\alpha^{\frac{1}{2}}\right) \cdot \frac{1}{(m\alpha)^{\frac{1}{2}}} = \frac{1}{\sqrt{m}}, \quad (35)$$

where

$$\alpha_i = \alpha \text{ for } i < m, \quad \alpha_m = \alpha + \frac{1}{2}, \text{ and } \alpha \neq \frac{1}{2},$$

that in the limit matches the mode at (32).

### Relationships of the SDD with other distributions

In this section we explore the relationships, or lack thereof, between the SDD and other popular distributions such as the uniform, von Mises and its particular case of the Fisher Bingham distribution. We consider limiting cases for different values of the concentration parameters  $\alpha_i$ .

#### Limiting case symmetric distribution for large $\alpha$

Assuming a symmetric SDD with  $\alpha_i = \alpha$ , for  $\forall \alpha_i$  we can write

$$f_{\text{SDD}}(\mathbf{x}; \alpha) = 2^{m-1} \frac{\Gamma(m\alpha)}{\Gamma(\alpha)^m} \prod_{i=1}^m x_i^{2\alpha-1} \cdot x_m^{-1}, \quad (36)$$

subject to the restrictions

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^m x_i^2 = 1, \quad \alpha \in \mathbb{R}^+,$$

in this case the covariance matrix can be reduced to

$$\Sigma = \frac{1}{m} \left(1 - \frac{\mu_\alpha^2}{\alpha}\right) \mathbf{I} - \left(\frac{\mu_\alpha}{\mu_0}\right)^2 \left(1 - \frac{\mu_0^2}{m\alpha}\right) \mathbf{1}\mathbf{1}^T, \quad (37)$$

where

$$\mu_\alpha = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)}, \quad \mu_0 = \frac{\Gamma(\alpha_0 + \frac{1}{2})}{\Gamma(\alpha_0)},$$

in an attempt to write the SDD as a rotational distribution of the type shown by Mardia (Mardia and Jupp 2000), the latter expression can be rewritten as

$$\Sigma = \left(1 - \frac{\mu_\alpha^2}{\alpha}\right) \left(\frac{1}{m} \mathbf{I} - \bar{\mu} \bar{\mu}^T\right) + \left(1 - m \frac{\mu_\alpha^2}{\mu_0^2}\right) \bar{\mu} \bar{\mu}^T, \quad (38)$$

or equivalently

$$\Sigma = \text{var}(x)m\bar{\mu}\bar{\mu}^T + \left(1 - \frac{\mu_\alpha^2}{m}\right) \left(I - m\bar{\mu}\bar{\mu}^T\right), \quad (39)$$

we can't determine an equivalence to the von Mises or similar rotationally symmetric distributions, however, using the expression developed by Frame (Frame 1949), we can see that in the limiting case for  $\alpha \rightarrow \infty$  and consequently  $\alpha_0 \rightarrow \infty$  we have

$$\lim_{\alpha \rightarrow \infty} \mu_\alpha = \lim_{\alpha \rightarrow \infty} \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} = \alpha^{\frac{1}{2}},$$

and

$$\lim_{\alpha \rightarrow \infty} \mu_0 = \lim_{\alpha \rightarrow \infty} \frac{\Gamma(m\alpha + \frac{1}{2})}{\Gamma(m\alpha)} = (m\alpha)^{\frac{1}{2}},$$

which in the limit it yields

$$\lim_{\alpha \rightarrow \infty} \Sigma = \lim_{\alpha \rightarrow \infty} \left(1 - \frac{\mu_\alpha^2}{\alpha}\right) \left(\frac{1}{m}I - \bar{\mu}\bar{\mu}^T\right) + \left(1 - m\frac{\mu_\alpha^2}{\mu_0^2}\right) \bar{\mu}\bar{\mu}^T = 0,$$

we conclude that for large values of  $\alpha$  the covariance matrix tends to zero, consequently, the SDD tends to be concentrated as a vector with no variation.

#### Limiting case uniform distribution

We now consider the case where  $\alpha_i = \frac{1}{2}$ , for  $i < m$  and  $\alpha_m = 1$ , the SDD takes the form

$$f_{\text{SDD}}(\mathbf{x}; \alpha) = \frac{2^{m-1} \Gamma(\frac{m-1}{2} + 1)}{\prod_{i=1}^{m-1} \Gamma(\frac{1}{2}) \Gamma(1)} \prod_{i=1}^{m-1} x_i^{2\frac{1}{2}-1} \cdot x_m^{2(1)-2} = \frac{2^{m-1} \Gamma(\frac{m+1}{2})}{\pi^{\frac{m-1}{2}}}, \quad (40)$$

which is a constant thickness independent of the values of  $x_i$ , then the SDD becomes the uniform distribution over the positive orthant of the hypersphere.

#### Similarities and differences of the SDD with the von Mises and Fisher-Bingham distributions

The von Mises distribution is usually considered the analogue of the normal distribution in the circle as described by Mardia in (Mardia 1975). The von Mises distribution and its particular case for the three dimensional sphere, the Fisher-Bingham distribution, both tend to converge to a multivariate and bivariate normal distribution respectively for large values of  $\kappa$  as shown by Kent (Kent 1982).

The proposed SDD doesn't seem to converge to the von Mises distribution or to a multivariate normal distribution for large values of  $\alpha_i$ , but rather it tends to be concentrated as a vector as it was established at the end of the corresponding section for the limiting cases for the SDD.

Moreover, both the von Mises and the Fisher-Bingham distribution converge to the uniform distribution for very small values of  $\kappa$ , in a similar way as the SDD becomes the uniform distribution for the values of the parameters described at the end of the previous subsection.



### Inference for the spherical-Dirichlet distribution

We now consider estimation of the parameters of the SDD. Our main interest is to develop suitable procedures to estimate the set of parameters  $\alpha_i$ , given a sample of random vectors located at the positive orthant of the hypersphere. We first derive estimators for  $\alpha_i$  using the method of moments (MOM), and next we develop estimators for the same set of parameters using the method of maximum likelihood estimation (MLE).

#### Method of moments (MOM)

Using a similar procedure as the one developed by Narayanan (Narayanan 1992) to estimate the parameters of the Dirichlet distribution, suppose we have a random sample with  $n$  random vectors  $X_1, X_2, \dots, X_n$  such that  $X_i \in \mathcal{R}^m = \left[ X_j | j = 1, \dots, m; X_j > 0, \sum_{j=1}^m x_j^2 = 1 \right]$  that are i.i.d., then

$$E(x_i) = \frac{\Gamma(\alpha_i + \frac{1}{2})}{\Gamma(\alpha_i)} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \frac{1}{2})} = \frac{\mu_i}{\mu_0} \text{ for } \forall i, \quad (41)$$

and

$$E(x_i^2) = \frac{\alpha_i}{\alpha_0} \text{ for } \forall i. \quad (42)$$

We define the sample moments as

$$X'_{1j} = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad j = 1, \dots, m, \quad (43)$$

and

$$X'_{2j} = \frac{1}{n} \sum_{i=1}^n x_{ij}^2 \quad j = 1, \dots, m. \quad (44)$$

We have  $m-1$  first order moment equations and  $m-1$  second order moment equations to solve for  $m$  unknowns  $\alpha_i$ . To avoid linear dependency and for the sake of simplicity we choose one of the first order moments, and  $m-1$  of the second order moment equations

$$\frac{\Gamma(\alpha_1 + \frac{1}{2})}{\Gamma(\alpha_1)} \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \frac{1}{2})} = \frac{1}{n} \sum_{i=1}^n x_{i1} = X'_{11}, \quad (45)$$

then, the remaining  $m-1$  second order moment equations are

$$\frac{\alpha_i}{\alpha_0} = \frac{1}{n} \sum_{i=1}^n x_{ij}^2 = X'_{2j} \quad j = 2, \dots, (m-1). \quad (46)$$

There is no closed form solution for  $\alpha_i$  in solving simultaneously (45) and (46), so we must solve numerically to obtain the corresponding method of moments estimators for  $\alpha_i$ . Results from MOM can be used as initial values for the MLE that usually exhibit better statistical properties.

**Maximum likelihood estimation (MLE)**

Suppose that we have a random sample of vectors on the positive orthant of the hypersphere,  $X_1, X_2, \dots, X_n$ , where  $X_i \in \mathbb{R}^m$  from an SDD with pdf defined at (3). Then, the log-likelihood is

$$\ln L(\alpha) = \ln \prod_{i=1}^n \frac{2^{m-1} \Gamma\left(\sum_{j=1}^m \alpha_j\right)}{\prod_{j=1}^m \Gamma(\alpha_j)} \prod_{j=1}^m x_{ij}^{2\alpha_j-1} \cdot x_{im}^{-1}. \quad (47)$$

The parameters for an SDD can be estimated by maximizing the log-likelihood function of the data, in a similar procedure as the one used by Minka for the Dirichlet distribution described at (Minka 2000). We can group all the constant terms as  $K$ , and we can rewrite all the products and sums as

$$\begin{aligned} \ln L(\alpha) &= K + n \ln \Gamma\left(\sum_{j=1}^m \alpha_j\right) - n \sum_{j=1}^m \ln \Gamma(\alpha_j) + \sum_{i=1}^n \sum_{j=1}^m (2\alpha_j - 1) \ln x_{ij} - \sum_{i=1}^n \ln x_{im}, \\ &= K + n \left( \ln \Gamma\left(\sum_{j=1}^m \alpha_j\right) - \sum_{j=1}^m \ln \Gamma(\alpha_j) + \sum_{j=1}^m (2\alpha_j - 1) \frac{1}{n} \sum_{i=1}^n \ln x_{ij} - \frac{1}{n} \sum_{i=1}^n \ln x_{im} \right), \end{aligned}$$

where the function that needs to be optimized after removing unnecessary constants is

$$F(\alpha) = \ln \Gamma\left(\sum_{j=1}^m \alpha_j\right) - \sum_{j=1}^m \ln \Gamma(\alpha_j) + \sum_{j=1}^m (2\alpha_j - 1) \left( \frac{1}{n} \sum_{i=1}^n \ln x_{ij} \right) - \frac{1}{n} \sum_{i=1}^n \ln x_{im}.$$

The gradient of the objective function can be obtained by differentiating the log-likelihood  $\ln F(\alpha)$  with respect to  $\alpha_k$  as

$$\nabla(F)_k = \frac{\partial F}{\partial \alpha_k} = \Psi\left(\sum_{j=1}^m \alpha_j\right) - \Psi(\alpha_k) + 2 \left( \frac{1}{n} \sum_{i=1}^n \ln x_{ik} \right), \quad (48)$$

where  $\Psi =: \frac{d \ln \Gamma(x)}{dx}$  is the digamma function. The optimization is subject to the constraints  $\alpha_i \geq 0$ . The SDD is a member of the exponential family and therefore it is a convex function, and the observed sufficient statistic is equal to the expected sufficient statistic, where the latter is

$$E(x_k) = \frac{1}{2} \Psi(\alpha_k) - \frac{1}{2} \Psi\left(\sum_{j=1}^m \alpha_j\right), \quad (49)$$

and the observed sufficient statistic is

$$\frac{1}{n} \sum_{i=1}^n \ln x_{ij}, \quad (50)$$

that leads to the following iterative procedure

$$\Psi(\alpha_k^{new}) = \Psi\left(\sum_{j=1}^m \alpha_j^{old}\right) + 2 \left( \frac{1}{n} \sum_{i=1}^n \ln x_{ik} \right). \quad (51)$$

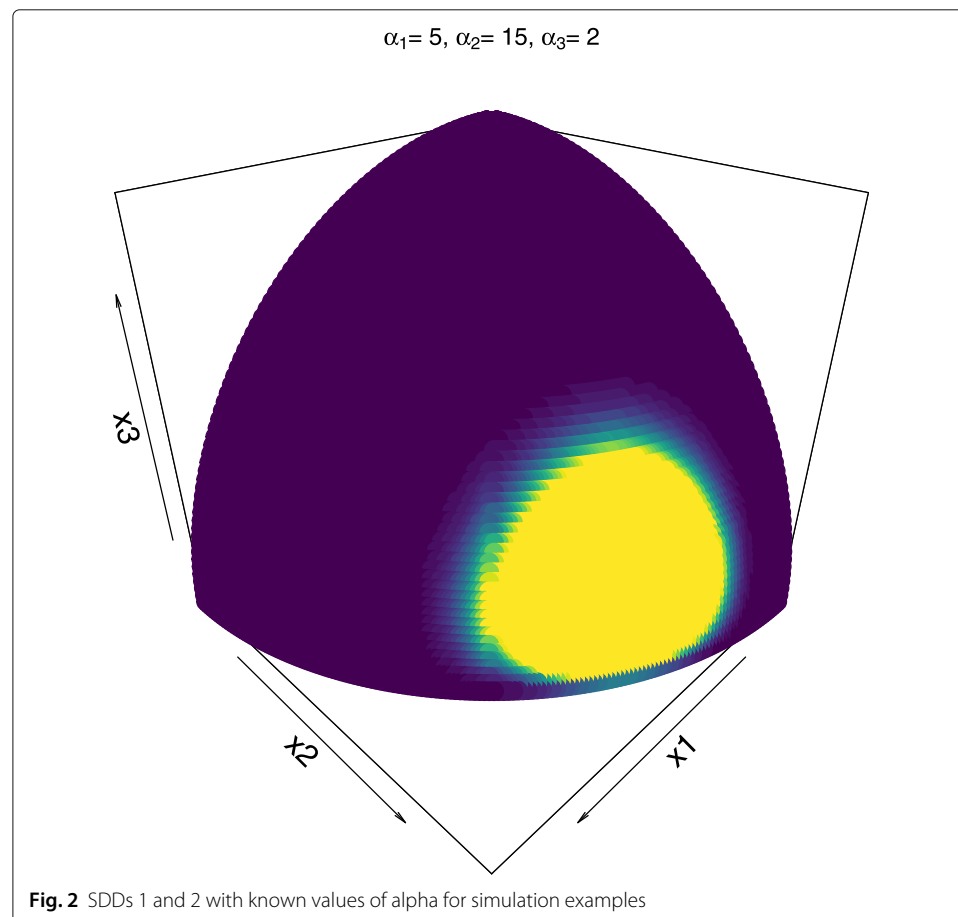
Although the proposed procedure does not guarantee in general reaching a global maximum, updating successively (51) provides reasonable results, and convergence is typically fast.

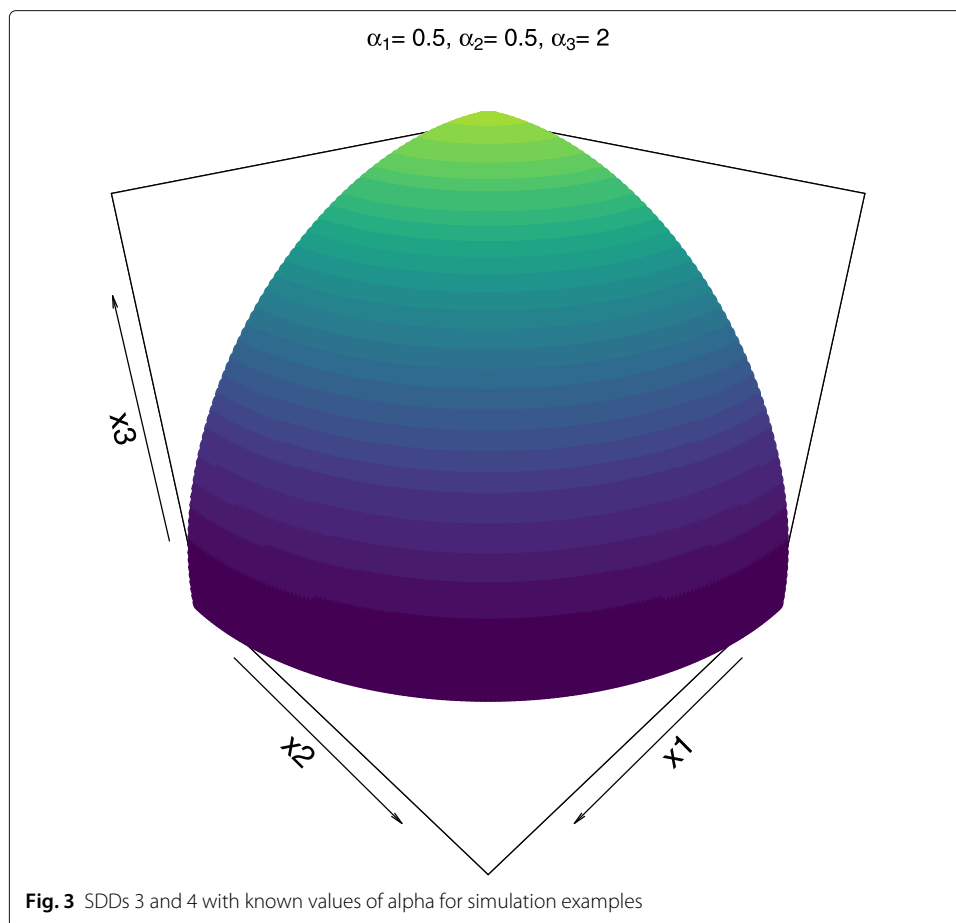
### Applications to data

Let's now consider estimation of the parameters of the SDD. We first developed an example using simulated data generated from the proposed SDD with parameters we assumed to be unknown for the purpose of this estimation. Next, a second example was developed using a text mining example, with data obtained from a publicly available data set. Both examples were solved using MOM and MLE, applying the proposed techniques described at the corresponding sections for the method of moments and maximum likelihood estimation, and results obtained from both methods were compared.

### Simulation example

Four different simulations were performed each with 1,000 randomly generated values from an SDD in a three-dimensional hypersphere, with known values of the parameters  $\alpha_1, \alpha_2$  and  $\alpha_3$ . Plots for the corresponding values of these parameters are shown at Figs. 2 and 3. Inferences to estimate the values of these parameters, assumed to be unknown, were performed using the MOM and MLE procedures developed in the corresponding sections. Graphs for the SDD corresponding to the proposed four different sets of parameters are shown at the following figures:





First, an estimation was performed using MOM and iterating between (45) and (46). These values were updated in each cycle until convergence was achieved within a pre-set tolerance limit. The estimated values of the parameters using MOM were used as the initial values for the iterative process using MLE. For the latter method expression (51) is updated successively until the values of the parameters were stable within a pre-set tolerance level. Results for the estimation of both methods and the true values of the parameters are shown at Table 1.

Note the close agreement between the MLEs and MOMs at the results shown at Table 1.

#### Text mining example

A text mining example was developed using a publicly available data set assembled by Lang (Lang). An example of email messages regarding several interest groups are available, the “auto” topic was selected and summarized using standard data mining techniques. A collection of randomly selected sample of 160 documents (emails) were extracted and summarized as vectors at the positive orthant of the hypersphere. Common terms such as “from” or “subject” were excluded as they did not provide any discriminant power and could potentially bias the analysis. Vocabulary reduction for synonymous and stemming were performed, and the ten most common terms were extracted by obtaining their raw frequencies. The frequencies for these terms were expressed as the components of vectors in a ten-dimensional space. A small fraction of the data set can be seen at Table 2.

**Table 1** Simulation results

Method	# Iterations	$\alpha_1 = 2$	$\alpha_2 = 2$	$\alpha_3 = 2$	% Error
MOM	176	2.0557	2.0983	2.0764	3.84
MLE	51	2.0412	2.0798	2.0479	2.81
Method	# Iterations	$\alpha_1 = 5$	$\alpha_2 = 15$	$\alpha_3 = 2$	% Error
MOM	589	5.0351	14.7148	1.9684	1.40
MLE	147	5.1932	15.1998	2.0496	2.56
Method	# Iterations	$\alpha_1 = 0.5$	$\alpha_2 = 0.5$	$\alpha_3 = 2$	% Error
MOM	28	0.4964	0.4639	1.9212	3.96
MLE	23	0.4903	0.4821	1.9503	2.67
Method	# Iterations	$\alpha_1 = 2$	$\alpha_2 = 2$	$\alpha_3 = 10$	% Error
MOM	385	1.9349	2.0153	10.1145	1.72
MLE	85	1.9745	2.0713	10.3528	2.79

An appropriate transformation for these vectors was applied to reduce extreme values and eliminate zeros. The transformation used here was  $x_{transf} = \ln(1.10 + x)$ . These vectors were standardized to a unit length at the positive quadrant of the hypersphere and they were fitted using the proposed multivariate SDD for ten dimensions. The estimation for the corresponding  $\alpha_i$ 's for the proposed distribution was done using both MOM and MLE, and their corresponding estimated values are shown Table 3

The number of iterations needed to fit the SDD for the MOM procedure within a preset tolerance level were 271 iterations. The final results of the MOM estimators were used as the initial values for the MLE procedure, and a new model was fitted using 19 additional iterations. Although the MLE procedure in general does not guarantee finding a global maximum, the proposed method provided reasonable results and the convergence was fast enough.

## Conclusions

The proposed SDD constitutes a superior alternative to other competing methods for fitting unit vectors at the positive orthant of the hypersphere. The SDD avoids wasting probability mass or using distribution mixtures that are not suitable for the positive orthant of the hypersphere. Inference results for MOM and MLE were in close agreement for simulated data, and reasonably close for the real text mining example. The simulated data were randomly generated from the proposed SDD while the text mining data were obtained from a real text mining problem. The SDD is flexible and shows a rich variety of shapes suitable to fit a wide range of data, in a similar way that the beta distribution does for a one-dimensional space. Under an appropriate transformation it can also accommodate zeros for some coordinates of the hyper-vectors. Future research may be aimed to enhance the capability of handling zero-value components, avoiding further need of transforming data.

**Table 2** Terms frequency

Doc ID	ntoken	auto	write	articl	engin	don	good	time	drive	road
103092	0	2	1	1	0	0	0	0	0	0
101671	7	0	2	2	0	2	2	0	0	0
.....	...	...	...	...	...	...	...	...	...	...
101582	6	8	3	2	0	0	0	0	0	0
103050	0	3	1	0	0	0	0	0	0	0

**Table 3** Text mining results

Parameter	MOM	MLE
$\alpha_1$	0.7799	1.1787
$\alpha_2$	0.6545	1.0013
$\alpha_3$	0.2151	0.4755
$\alpha_4$	0.1790	0.4276
$\alpha_5$	0.1182	0.2825
$\alpha_6$	0.1268	0.3224
$\alpha_7$	0.0923	0.2857
$\alpha_8$	0.1054	0.3004
$\alpha_9$	0.0833	0.2796
$\alpha_{10}$	0.0591	0.2481

**Acknowledgements**

The author is grateful to the invaluable help of Eduardo García-Portugués from the Department of Statistics, Carlos III University of Madrid (Spain).

**Authors' contributions**

JHG is the sole author of this article. The author(s) read and approved the final manuscript.

**Funding**

Travel funding for the completion of this project was received by a Research Enhancement grant from the Texas A&M University-Corpus Christi Division of Research and Innovation.

**Availability of data and materials**

The data for the text mining example was obtained from the publicly available data set assembled by Lang (Lang). The particular sample analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests**

The author declares that he has no competing interests.

Received: 27 January 2020 Accepted: 31 July 2020

Published online: 05 September 2020

**References**

- Frame, J. S.: An approximation to the quotient of gamma function. *Am. Math. Mon.* **56**(8), 529–535 (1949)
- Kent, J. T.: The Fisher-Bingham distribution on the sphere. *J. R. Stat. Soc. Ser. B Methodol.* **44**(1), 71–80 (1982). <https://doi.org/10.1111/j.2517-6161.1982.tb01189.x>. <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1982.tb01189.x>
- Lang, K.: CMU Text Learning Group Data Archives. <https://www.cs.cmu.edu/afs/cs/project/theo-20/www/data/news20.html>. Accessed 1 Sep 2019
- Mardia, K. V.: Statistics of directional data. *J. R. Stat. Soc. Ser. B Methodol.* **37**(3), 349–393 (1975)
- Mardia, K. V., Jupp, P. E.: Directional Statistics, 2nd edn. Wiley series in probability and statistics. Wiley, Chichester (2000)
- Minka, T. P.: Estimating a Dirichlet distribution. Technical Report, MIT, (2000). <https://www.microsoft.com/en-us/research/publication/estimating-dirichlet-distribution/>
- Narayanan, A.: A note on parameter estimation in the multivariate beta distribution. *Comput. Math. Appl.* **24**(10), 11–17 (1992). [https://doi.org/10.1016/0898-1221\(92\)90016-B](https://doi.org/10.1016/0898-1221(92)90016-B)
- Olkin, I., Rubin, H.: Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Stat.* **35**(1), 261–269 (1964)
- Suvrit, S.: Directional statistics in machine learning: a brief review. arXiv e-prints, 1605–00316 (2016). [1605.00316](https://arxiv.org/abs/1605.00316)

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.