Journal of Multivariate Analysis 114 (2013) 412–426

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# A generalization of the Dirichlet distribution

A. Ongaro *, S. Migliorati

*Department of Statistics, University of Milano-Bicocca, Milano, Italy*

ARTICLE INFO

*Article history:*
Received 4 January 2012
Available online 1 August 2012

*AMS subject classifications:*
60E05
62E15
62H05

*Keywords:*
Dirichlet mixture
Subcomposition
Amalgamation
Compositional invariance
Neutrality
Multi-modality

ABSTRACT

A new parametric family of distributions on the unit simplex is proposed and investigated. Such family, called flexible Dirichlet, is obtained by normalizing a correlated basis formed by a mixture of independent gamma random variables. The Dirichlet distribution is included as an inner point. The flexible Dirichlet is shown to exhibit a rich dependence pattern, capable of discriminating among many of the independence concepts relevant for compositional data. At the same time it can model multi-modality. A number of stochastic representations are given, disclosing its remarkable tractability. In particular, it is closed under marginalization, conditioning, subcomposition, amalgamation and permutation.


## 1. Introduction

In many of the problems of interest to scientists, data consist of proportions and thus are subject to non-negativity and unit-sum constraints. Such data originate, for instance, when analyzing rock compositions, household budgets, pollution components and so on. They are called compositional and arise naturally in a great variety of disciplines such as biology, medicine, chemistry, economics, psychology, environmetrics, psephology and many others.

Modeling compositional data requires the choice of a distribution defined on the proper bounded domain: the simplex. In such context a critical issue is the definition of appropriate concepts of independence, as standard ones become unsuitable. To this end a number of new "simplicial" independence forms have been developed entailing different degrees of strength and different important practical interpretations (for a discussion see [1]).

The most widely studied distribution on the simplex is the Dirichlet, whose popularity can be mainly traced back to its several mathematical properties as well as easiness of parameter interpretation. However, its use in applications is quite limited as it implies the most extreme forms of simplicial independence.

In the light of such shortcomings, Aitchison proposed a powerful and far-reaching approach based on mapping the simplex into a standard Euclidean space by means of log-ratio transformations of the original variables (see [4,2,3]). The consequent logistic normal distribution is one of the most practical and widespread for compositional data analysis. Still yet, such model does not contain the Dirichlet and does not allow to incorporate many useful forms of simplicial independence. Moreover, it does not possess other relevant compositional properties such as closure under amalgamation, compositional invariance and symmetry with respect to the components. Aitchison proposed several interesting related models (for an overview see [1]) aimed at dealing with specific inferential problems, but it appears still lacking a unifying family capable of simultaneously accommodating for the just mentioned requirements.

* Corresponding author.
*E-mail address:* andrea.ongaro@unimib.it (A. Ongaro).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.07.007

Alternatively, one can stay on the simplex looking for appropriate generalizations of the Dirichlet distribution as proposed by Connor and Mosimann [6], Gupta and Richards [8], Barndorff-Nielsen and Jorgensen [5], and Favaro et al. [7]. In particular, the Liouville distribution has been widely studied (see [8–12]). All such proposals do not allow for a dependence structure sufficiently richer than the Dirichlet (see Section 8). Generalizations of the Liouville distribution have also been considered in [19,20]: a detailed discussion of such models and their relations to our work is deferred to Section 6. Other distributions including the Dirichlet aim at suiting different purposes, e.g. the grouped Dirichlet [17] and the nested Dirichlet [18] are presented as tools for modeling incomplete categorical data.

Following this second (stay on the simplex) approach, we shall propose a new generalization of the Dirichlet exhibiting a substantially greater flexibility in terms of dependence/independence structure and shape of the density, but maintaining, to a large extent, several mathematical and compositional properties which make the Dirichlet so manageable.

This objective is achieved by extending the basis of gamma independent random variables which generates the Dirichlet distribution via normalization. Specifically, the new basis exhibits correlated components which are obtained by randomly allocating a further independent gamma variable to the components of the original independent basis. This leads to a Dirichlet mixture structure for the corresponding normalized distribution.

The present article is organized as follows. After presenting some preliminary definitions central to compositional data analysis in Section 2, we propose the new basis deriving some of its properties in Section 3. Next, in Section 4, we define and study the new distribution on the simplex: the flexible Dirichlet. In particular, we give stochastic representations, joint moments, marginals, conditionals, covariance of log-ratios, amalgamation, subcomposition and permutation properties. Then, in Section 5, we investigate independence relationships for the flexible Dirichlet and in Section 6, we discuss Liouville type distributions and their connection with the flexible Dirichlet. In Section 7 some inferential aspects are touched upon and an illustrative application is given. We conclude with some final remarks in Section 8. The lengthy proofs are presented in the Appendix.

## 2. Preliminary definitions and properties

In this section, we introduce the basic notation and recall some definitions and properties particularly relevant for compositional data.

A basis is a random vector $\boldsymbol{Y} = (Y_1, \dots, Y_D)$ with positive components. The corresponding composition $\boldsymbol{X} = (X_1, \dots, X_D) = \mathcal{C}(\boldsymbol{Y})$ is obtained by normalizing $\boldsymbol{Y}$, i.e. $X_i = Y_i/Y^+$ where $Y^+ = \sum_{i=1}^{D} Y_i$. Its support is the simplex $\mathcal{S}^D = \left\{\boldsymbol{x} : x_i > 0, i = 1, \dots, D, \sum_{i=1}^{D} x_i = 1\right\}$. Such a definition considers the simplex as a $(D-1)$-dimensional subspace of $\mathcal{R}_+^D$ and has the advantage of treating all the components symmetrically. In some contexts, it is instead mathematically convenient to view the simplex as an open subset of the $(D-1)$-dimensional space $\mathcal{R}_+^{D-1}$. This leads to the asymmetric form $\mathcal{S}_a^D = \left\{\boldsymbol{x} : x_i > 0, i = 1, \dots, D-1, \sum_{i=1}^{D-1} x_i < 1\right\}$. In such a case $1 - x_1 - \cdots - x_{D-1}$ is called the fill-up value.

To treat all components symmetrically, densities on the simplex will be defined on $\mathcal{S}^D$. Formally, they can be interpreted as densities with respect to a suitable uniform measure on $\mathcal{S}^D$. They can be transformed into standard densities, i.e. with respect to the Lebesgue measure on $\mathcal{R}^{D-1}$ restricted to $\mathcal{S}_a^D$, by simply replacing the last component $x_D$ by $1 - x_1 - \cdots - x_{D-1}$.

Often the elements of a composition can be grouped according to some relevant homogeneity criterion. Thus, it is of interest to study the group composition (i.e. the totals of the groups) and the relative magnitudes of the components within each group. This gives rise to the two fundamental concepts of amalgamation and subcomposition. Such concepts can be formally defined as follows. Let $a_0 = 0 < a_1 < \cdots < a_{C-1} < a_C = D$ be integers and let

$$X_1, \dots, X_{a_1} | X_{a_1+1}, \dots, X_{a_2} | \cdots | X_{a_{C-1}+1}, \dots, X_{a_C} \tag{1}$$

be a general partition (of order $C-1$) of the vector $\boldsymbol{X}$ into $C$ subsets. Then the $i$th subcomposition is defined as $\boldsymbol{S}_i = \left(X_{a_{i-1}+1}, \dots, X_{a_i}\right)/X_i^+$ where $X_i^+ = X_{a_{i-1}+1} + \cdots + X_{a_i}$, $(i = 1, \dots, C)$. The amalgamation is the vector of the totals of the $C$ subsets: $\boldsymbol{X}^+ = (X_1^+, \dots, X_C^+)$.

The first property which turns out to be important for compositional data modeling is compositional invariance: a basis $\boldsymbol{Y}$ is compositionally invariant if the corresponding composition $\boldsymbol{X} = \mathcal{C}(\boldsymbol{Y})$ is independent of its size $Y^+$.

Almost all other independence concepts present in the literature can be expressed in terms of subcompositions $\boldsymbol{S}_i$ $(i = 1, \dots, C)$ and amalgamation $\boldsymbol{X}^+$. In particular, the most widespread ones are defined in terms of partition of order 1 $(C = 2)$. In such a case, for simplicity, we shall set $a_1 = k$. Denoting independence by $\perp$ and a set of independent random variables by $A \perp B \perp C$, such concepts can be described as follows: partition independence means that $\boldsymbol{S}_1 \perp \boldsymbol{S}_2 \perp \boldsymbol{X}^+$; subcompositional invariance means that $(\boldsymbol{S}_1, \boldsymbol{S}_2) \perp \boldsymbol{X}^+$; neutrality on the right means that $\boldsymbol{S}_2 \perp \left(\boldsymbol{S}_1, \boldsymbol{X}^+\right)$; neutrality on the left means that $\boldsymbol{S}_1 \perp \left(\boldsymbol{S}_2, \boldsymbol{X}^+\right)$; subcompositional independence means that $\boldsymbol{S}_1 \perp \boldsymbol{S}_2$.

Finally, the previous properties can be extended in the following more complex form: an independence property is complete if it holds for partitions $\boldsymbol{X}_1 = (X_1, \dots, X_k)$ and $\boldsymbol{X}_2 = (X_{k+1}, \dots, X_D)$ of order 1 at all possible levels $k$.

For a discussion and interpretation of such properties see [1].

We end this section recalling a standard construction and some basic properties of the Dirichlet distribution. Let $W_i \sim Ga(\alpha_i)$, $i = 1, \dots, D$, denote $D$ independent gamma distributed random variables with a common arbitrary scale parameter and shape parameter $\alpha_i > 0$. Then, the normalized random vector $\boldsymbol{X} = \mathcal{C}(W_1, \ldots, W_D)$ is Dirichlet distributed denoted by $\boldsymbol{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)$.

The Dirichlet distribution is closed under marginalization, conditioning (after normalization), amalgamation and subcomposition. On the other hand, it implies all the above independence properties, which are a consequence of the following well known result: the random elements $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C, \boldsymbol{X}^+, W^+ = \sum_{i=1}^{D} W_i$ are independent under the Dirichlet model for any choice of the groups of the partition.

For completeness we report its first two moments: $E(X_i) = \alpha_i/\alpha^+$, $\text{Var}(X_i) = E(X_i)(1 - E(X_i))/(\alpha^+ + 1)$ and $\text{Cov}(X_i, X_r) = -E(X_i)E(X_r)/(\alpha^+ + 1)$ $i \neq r$. Note that, once the mean vector is chosen, only one parameter, namely $\alpha^+$, is devoted to modeling the whole variance–covariance structure. In particular, all covariances are proportional to the product of the corresponding means.

## 3. The generating basis: the flexible gamma distribution

Consider the basis $\boldsymbol{Y} = (Y_1, \ldots, Y_D)$ defined as

$$Y_i = W_i + Z_i\, U \quad i = 1, \ldots, D \tag{2}$$

where the random variables $W_i \sim Ga(\alpha_i)$ are independent, $U \sim Ga(\tau)$ is an independent gamma random variable with the same scale parameter as the $W_i$'s. Moreover $\boldsymbol{Z} = (Z_1, \ldots, Z_D) \sim Mu(1, \boldsymbol{p})$ is a multinomial random vector independent of $U$ and of the $W_i$'s which is equal to $\boldsymbol{e}_i$ with probability $p_i$, where $\boldsymbol{e}_i$ is a vector whose elements are all equal to zero except for the $i$th element which is one. Here the vector $\boldsymbol{p} = (p_1, \ldots, p_D)$ is such that $0 \le p_i < 1$ and $\sum_{i=1}^{D} p_i = 1$, $\alpha_i > 0$ and $\tau > 0$.

The basis defined by (2) will be called flexible Gamma (FG) and denoted by $\text{FG}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$. It defines a parametric family of positive and dependent random variables obtained by starting from the Dirichlet basis and randomly allocating to the $i$th component a further independent gamma random variable with probability $p_i$, $(i = 1, \ldots, D)$.

By conditioning on $\boldsymbol{Z}$, a representation of the FG as a finite mixture of random vectors with gamma independent components is immediately obtained. This directly leads to explicit expressions for the density (including the marginal and the conditional one) and joint moments of any order. In particular $\text{Cov}(Y_i, Y_r) = -p_i p_r \tau^2$, $i \neq r$. Thus the random allocation of the gamma variate $U$ introduces flexibility in the variance–covariance structure of the basis, inducing arbitrary non-positive correlation between its components. Although such covariance structure is not completely general, it still allows for rich (in)dependence relationships in the corresponding normalized distribution, as we shall see.

We now give two properties of the FG of particular relevance for our intended application to compositional analysis.

The following amalgamation property is a consequence of the infinite divisibility of the gamma distribution and of the structure of the random allocation scheme defining the FG.

**Proposition 3.1** (*Closure Under Amalgamation*). *Let $\boldsymbol{Y} \sim \text{FG}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ and denote by $\boldsymbol{Y}^+ = (Y_1^+, \ldots, Y_C^+)$, $\boldsymbol{\alpha}^+ = (\alpha_1^+, \ldots, \alpha_C^+)$ and $\boldsymbol{p}^+ = (p_1^+, \ldots, p_C^+)$ the amalgamation vectors where each component is the total within the corresponding group of the partition. Then $\boldsymbol{Y}^+ \sim \text{FG}^C(\boldsymbol{\alpha}^+, \boldsymbol{p}^+, \tau)$.*

**Proposition 3.2** (*Compositional Invariance*). *The basis defined by (2) is compositionally invariant, i.e. the normalized vector $\boldsymbol{X} = \mathcal{C}(\boldsymbol{Y})$ and the size $Y^+$ are independent.*

**Proof.** As, conditionally on $\boldsymbol{Z} = \boldsymbol{e}_i$, $\boldsymbol{X}$ is a $\mathcal{D}^D(\boldsymbol{\alpha}_i)$ where $\boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \tau\boldsymbol{e}_i$ independent of $Y^+ \sim Ga(\alpha^+ + \tau)$, the distribution of $(\boldsymbol{X}, Y^+)$ can be written as

$$F_{\boldsymbol{X},Y^+}(\boldsymbol{x}, y^+) = \sum_{i=1}^{D} p_i F_{\boldsymbol{X},Y^+|\boldsymbol{Z}=\boldsymbol{e}_i}(\boldsymbol{x}, y^+) = Ga(y^+; \alpha^+ + \tau) \sum_{i=1}^{D} p_i \mathcal{D}^D(\boldsymbol{x}; \boldsymbol{\alpha}_i) \tag{3}$$

where $\mathcal{D}^D(\boldsymbol{x}; \boldsymbol{\alpha})$ denotes the distribution function of a Dirichlet random variable. Notice that (3) also gives the marginal distributions of $\boldsymbol{X}$ and $Y^+$. □

Strictly speaking, the FG does not contain the standard Dirichlet basis with independent gamma components. However, it is easy to check that any density obtained by multiplying the density of such basis by a function depending only on the sum of the components generates the Dirichlet distribution. This is the case of the FG density when $\tau = 1$ and $p_i = \alpha_i/\alpha^+$, $i = 1, \ldots, D$.

## 4. The flexible Dirichlet distribution

The new distribution, called flexible Dirichlet (FD), derives from the normalization of a FG basis. Formally, let $\boldsymbol{Y} \sim \text{FG}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$, then the normalized vector $\boldsymbol{X} = \mathcal{C}(\boldsymbol{Y})$ has a FD distribution denoted by $\text{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$.

Some important properties of such distribution are given and discussed below. The following mixture representation of the FD can be easily derived by conditioning on $\boldsymbol{Z}$.

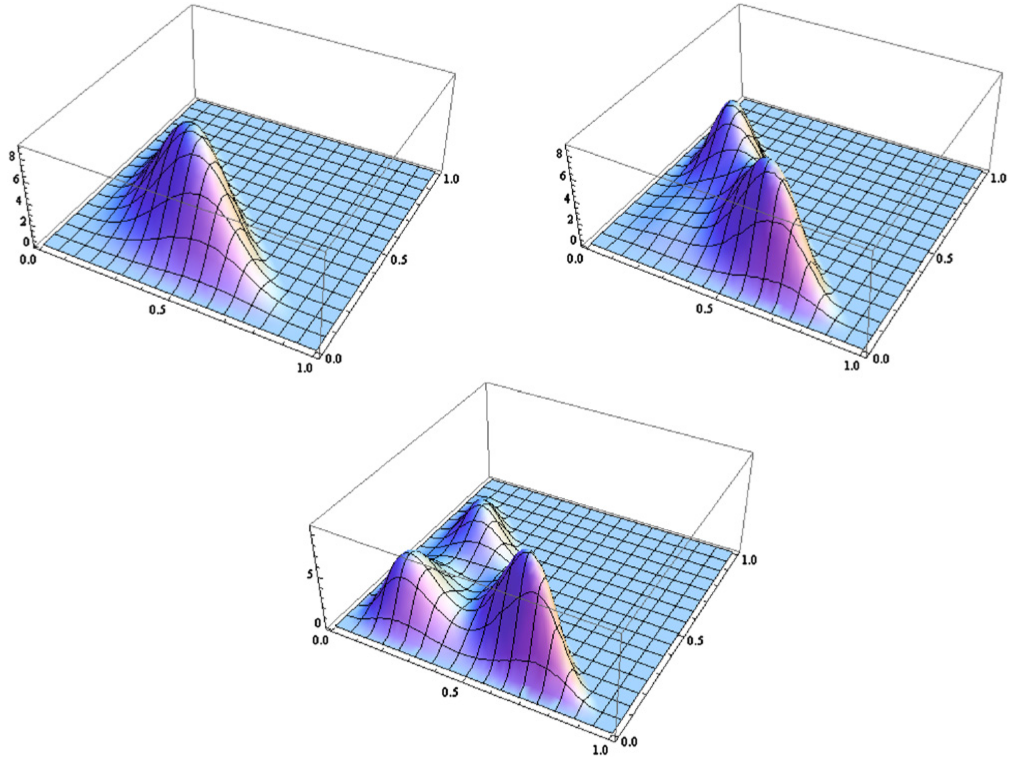


**Fig. 1.** The FD density with $\boldsymbol{\alpha} = (4, 4, 4)$, $\boldsymbol{p} = (0.6, 0.35, 0.05)$ and $\tau = 2$ (top left), $\boldsymbol{\alpha} = (4, 4, 4)$, $\boldsymbol{p} = (0.55, 0.4, 0.05)$ and $\tau = 5$ (top right) and $\boldsymbol{\alpha} = (4, 4, 4)$, $\boldsymbol{p} = (0.5, 0.2, 0.3)$ and $\tau = 7$ (bottom).

**Proposition 4.1** (*Dirichlet Mixture Representation*)**.** *The distribution function of* $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *is a finite mixture of Dirichlet distributions:*

$$\mathrm{FD}^D(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{p}, \tau) = \sum_{i=1}^{D} p_i \mathcal{D}^D(\boldsymbol{x}; \boldsymbol{\alpha} + \tau \boldsymbol{e}_i). \tag{4}$$

Consequently, the density function of the FD can be expressed as

$$f_{\mathrm{FD}}(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{p}, \tau) = \frac{\Gamma\left(\alpha^+ + \tau\right)}{\prod_{r=1}^{D} \Gamma(\alpha_r)} \left(\prod_{r=1}^{D} x_r^{\alpha_r - 1}\right) \sum_{i=1}^{D} p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + \tau)} x_i^{\tau} \tag{5}$$

for $\boldsymbol{x} \in \mathcal{S}^D$.

The FD contains the Dirichlet distribution as the inner point $\tau = 1$ and $p_i = \alpha_i/\alpha^+$, $i = 1, \ldots, D$. That this is the only case can be seen by inspection of (5).

The FD density displays a large variety of shapes as a consequence of its Dirichlet mixture structure. Recall that the Dirichlet density can take many different forms. Specifically, if $\alpha_i < 1$ for all $i$ it is U-shaped, if $\alpha_i = 1$ for all $i$ it is constant and if $\alpha_i > 1$ for all $i$ it is unimodal with mode $(\alpha_i - 1)/(\alpha^+ - D)$ $(i = 1, \ldots, D)$.

It follows, in particular, that the FD allows for multi-modality. By suitably choosing the $\alpha_i$'s and $p_i$'s and by increasing $\tau$ any $k$-modality can be reached for $k \le D$ (see Fig. 1 for a graphical illustration when $D = 3$). Indeed, the larger the value of $\tau$ the more the mode of the $i$th component of the mixture is shifted toward the $i$th vertex of the simplex (i.e. the one with $x_i = 1$).

**Proposition 4.2** (*Dirichlet Contamination Representation*)**.** *If* $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *then*

$$\boldsymbol{X} \sim \beta \boldsymbol{X}' + (1 - \beta)\boldsymbol{Z} \tag{6}$$

*where* $\boldsymbol{X}' \sim Dir(\boldsymbol{\alpha})$, $\beta \sim Be(\alpha^+, \tau)$ *(i.e. $\beta$ is beta distributed) and* $\boldsymbol{Z} \sim Mu(1, \boldsymbol{p})$ *are all independent.*

**Proof.** The basis (2) can be written as $\boldsymbol{Y} = \boldsymbol{W} + U\boldsymbol{Z}$ where $\boldsymbol{W} = (W_1, \ldots, W_D)$, $U$ and $\boldsymbol{Z}$ are independent. By normalizing it we obtain

$$\boldsymbol{X} = \frac{W^+}{W^+ + U} \frac{\boldsymbol{W}}{W^+} + \frac{U}{W^+ + U} \boldsymbol{Z}.$$

The result then follows as $\boldsymbol{W}/W^+ \sim \mathcal{D}^D(\boldsymbol{\alpha})$ is independent of $W^+$ (because of compositional invariance of $\boldsymbol{W}$) and therefore of $W^+/(W^+ + U) \sim Be(\alpha^+, \tau)$. □

Thus the FD can be viewed as a contamination of the Dirichlet distribution: a random fraction $\beta$ of the unitary total is distributed according to a Dirichlet, the remaining part is randomly allocated to a single component. In agreement with the above remarks on multi-modality, by increasing $\tau$ the mass concentrated around the mode of the Dirichlet is more and more driven toward the $D$ vertices of the simplex, collapsing in the limit to the multinomial.

We now give a description of the distribution of $\boldsymbol{X}$ through the equivalent (one to one) representation $(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{X}^+)$ derived by exploiting the random allocation scheme defining the basis.

**Proposition 4.3** (*Partition Representation*)**.** *If* $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *then the joint distribution of* $(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{X}^+)$ *can be expressed as the following mixture of two components with independent elements:*

$$p_1^+ F_1(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{x}^+) + (1 - p_1^+) F_2(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{x}^+) \tag{7}$$

*where:*

$$F_1(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{x}^+) = \mathrm{FD}^k\left(\boldsymbol{s}_1; \boldsymbol{\alpha}_1, \boldsymbol{p}_1/p_1^+, \tau\right) \mathcal{D}^{D-k}(\boldsymbol{s}_2; \boldsymbol{\alpha}_2)\, Be(\boldsymbol{x}^+; \alpha_1^+ + \tau, \alpha_2^+)$$
$$F_2(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{x}^+) = \mathcal{D}^k(\boldsymbol{s}_1; \boldsymbol{\alpha}_1)\, \mathrm{FD}^{D-k}\left(\boldsymbol{s}_2; \boldsymbol{\alpha}_2, \boldsymbol{p}_2/p_2^+, \tau\right) Be(\boldsymbol{x}^+; \alpha_1^+, \alpha_2^+ + \tau).$$

**Proof.** We shall prove that, conditionally on $Z_1^+ = \sum_{i=1}^k Z_i$, the vectors $\boldsymbol{S}_1, \boldsymbol{S}_2$ and $\boldsymbol{X}^+$ are independent with appropriate distributions. Let us first condition on $Z_1^+ = 1$. Then, $\boldsymbol{Y}_1 = (Y_1, \ldots, Y_k)$ and $\boldsymbol{Y}_2 = (Y_{k+1}, \ldots, Y_D)$ are independent with $\boldsymbol{Y}_1 \sim \mathrm{FG}^k\left(\boldsymbol{\alpha}_1, \boldsymbol{p}_1/p_1^+, \tau\right)$ whereas $\boldsymbol{Y}_2$ has independent gamma components. It follows that $(\mathcal{C}(\boldsymbol{Y}_1), Y_1^+)$ and $(\mathcal{C}(\boldsymbol{Y}_2), Y_2^+)$ are independent and, by compositional invariance of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, that $\mathcal{C}(\boldsymbol{Y}_1), \mathcal{C}(\boldsymbol{Y}_2), Y_1^+$ and $Y_2^+$ are all independent. Therefore, $\boldsymbol{S}_1 = \mathcal{C}(\boldsymbol{Y}_1), \boldsymbol{S}_2 = \mathcal{C}(\boldsymbol{Y}_2)$ and $\boldsymbol{X}^+ = Y_1^+/(Y_1^+ + Y_2^+)$ are independent with distributions respectively $\mathrm{FD}^k\left(\boldsymbol{\alpha}_1, \boldsymbol{p}_1/p_1^+, \tau\right)$, $\mathcal{D}^{D-k}(\boldsymbol{\alpha}_2)$ and $Be(\alpha_1^+ + \tau, \alpha_2^+)$.

Analogously, conditionally on $Z_1^+ = 0$, one has that $\boldsymbol{S}_1 \sim \mathcal{D}^k(\boldsymbol{\alpha}_1), \boldsymbol{S}_2 \sim \mathrm{FD}^{D-k}\left(\boldsymbol{\alpha}_2, \boldsymbol{p}_2/p_2^+, \tau\right)$ and $\boldsymbol{X}^+ \sim Be(\alpha_1^+, \alpha_2^+ + \tau)$ are independent. □

The simple mixture structure of the above representation is very informative. The marginals of $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{X}^+$ as well as of $(\boldsymbol{S}_1, \boldsymbol{S}_2)$, $(\boldsymbol{S}_1, \boldsymbol{X}^+)$ and $(\boldsymbol{S}_2, \boldsymbol{X}^+)$ are immediately available as two component mixtures. Moreover, conditionals and corresponding independence properties can also be easily derived.

Joint moments of the FD can be directly computed from the moments of the FG by exploiting compositional invariance of the basis (Proposition 3.2).

**Proposition 4.4** (*Joint Moments*)**.** *The joint moments of* $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *admit the following explicit form:*

$$E\left[\prod_{i=1}^D X_i^{n_i}\right] = \frac{1}{(\alpha^+ + \tau)^{[n^+]}} \prod_{i=1}^D \alpha_i^{[n_i]} \sum_{i=1}^D \frac{(\alpha_i + \tau)^{[n_i]}}{\alpha_i^{[n_i]}}\, p_i \tag{8}$$

*where* $n^+ = \sum_{i=1}^D n_i$.

Particularly simple expressions are obtained when $\tau$ is an integer. For example, when $\tau = 1$ the sum in (8) becomes $1 + \sum_{i=1}^D p_i\, n_i/\alpha_i$.

The first two moments of the FD take the form

$$\begin{aligned}
E(X_i) &= \frac{\alpha_i + p_i\tau}{\alpha^+ + \tau} = \frac{\alpha_i}{\alpha^+}\left(\frac{\alpha^+}{\alpha^+ + \tau}\right) + p_i\left(\frac{\tau}{\alpha^+ + \tau}\right) \\
\mathrm{Var}(X_i) &= \frac{E(X_i)(1 - E(X_i))}{(\alpha^+ + \tau + 1)} + \frac{\tau^2 p_i(1 - p_i)}{(\alpha^+ + \tau)(\alpha^+ + \tau + 1)} \\
\mathrm{Cov}(X_i, X_r) &= -\frac{E(X_i)E(X_r)}{(\alpha^+ + \tau + 1)} - \frac{\tau^2 p_i p_r}{(\alpha^+ + \tau)(\alpha^+ + \tau + 1)} \quad i \neq r.
\end{aligned} \tag{9}$$

It follows that the presence of 2D parameters allows to model the means and (part of) the variance–covariance matrix separately. In particular, unlike the Dirichlet distribution, the FD accounts for components with the same mean but different variances or for covariances which do not show proportionality with respect to the product of means. Covariances are all negative, but this limitation is not particularly severe as the unit sum constraint naturally introduces negative dependence.

Indeed, log-ratio covariances are more suitable than raw covariances to analyze compositional data dependences (see [1]). The following log-ratio covariances of the FD can be easily computed by conditioning on $\boldsymbol{Z}$ and can be written in terms of mean and variance of the logarithmic transform of a gamma random variable, i.e. the digamma and trigamma functions.

**Proposition 4.5** (*Log-Ratio Covariances*)**.** *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$, then the covariance between any two log-ratios with distinct indices i, j, k and l is*

$$\mathrm{Cov}\left(\log\frac{X_i}{X_j}, \log\frac{X_k}{X_l}\right) = -(c_i - c_j)(c_k - c_l)$$

*where $c_m = p_m\left[\Psi(\alpha_m + \tau) - \Psi(\alpha_m)\right]$, $(m = i, j, k, l)$ and $\Psi(\alpha)$ is the digamma function.*
*If the log-ratios have the same denominator (i.e. $j = l$) then*

$$\mathrm{Cov}\left(\log\frac{X_i}{X_j}, \log\frac{X_k}{X_j}\right) = -c_ic_k + c_jc_k + c_ic_j + p_jg(\alpha_j + \tau) + (1 - p_j)g(\alpha_j) - \left[c_j + \Psi(\alpha_j)\right]^2$$

*where $g(\alpha) = \Psi'(\alpha) + [\Psi(\alpha)]^2$ and $\Psi'$ is the trigamma function.*

Thus the FD log-ratio covariances can take both positive and negative values, as can be easily verified in the case $\tau = 1$. This is in contrast with the Dirichlet and Liouville distributions, which have null covariances when the indices are all distinct and positive ones when the ratios have the same denominator (see [13]). Actually the same restrictions on covariances do hold for all models obtained by normalizing a basis of independent random variables as, for example, the ones proposed in [7].

Because of the amalgamation property of the FG, the FD distribution is closed under marginalization, simple relationships holding between the parameters of the joint and the marginal distributions.

**Proposition 4.6** (*Marginals*)**.** *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$, then*

$$(\boldsymbol{X}_1, 1 - X_1^+) \sim \mathrm{FD}^{k+1}(\boldsymbol{\alpha}_1, \alpha^+ - \alpha_1^+, \boldsymbol{p}_1, 1 - p_1^+, \tau) \tag{10}$$

*where $\boldsymbol{X}_1 = (X_1, \ldots, X_k)$.*

Consequently, the one-dimensional marginals are mixtures of two betas:

$$X_i \sim p_iBe\left(\alpha_i + \tau, \alpha^+ - \alpha_i\right) + (1 - p_i)Be\left(\alpha_i, \alpha^+ - \alpha_i + \tau\right). \tag{11}$$

Let us now focus on the conditionals. The distribution of $\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2$ coincides, up to a scale transformation, with the distribution of $\boldsymbol{S}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2$ as $\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim c\boldsymbol{S}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2$ where $c = 1 - x_2^+$. Therefore we shall give the latter being more relevant for compositional data.

**Proposition 4.7** (*Conditionals*)**.** *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$. Then, the distribution of $\boldsymbol{S}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2$ is a mixture of a FD and of a Dirichlet:*

$$p(\boldsymbol{x}_2)\mathrm{FD}^k\left(\boldsymbol{\alpha}_1, \frac{\boldsymbol{p}_1}{p_1^+}, \tau\right) + [1 - p(\boldsymbol{x}_2)]\,\mathcal{D}^k(\boldsymbol{\alpha}_1) \tag{12}$$

*where*

$$p(\boldsymbol{x}_2) = \frac{p_1^+}{p_1^+ + q(\boldsymbol{x}_2)} \tag{13}$$

*and*

$$q(\boldsymbol{x}_2) = \frac{\Gamma(\alpha_1^+ + \tau)}{\Gamma(\alpha_1^+)(1 - x_2^+)^\tau}\sum_{i=k+1}^{D} p_i\frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + \tau)}x_i^\tau. \tag{14}$$

Notice that in general, unlike the Dirichlet distribution, the normalized conditional does depend on $\boldsymbol{X}_2$. Independence of $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$ is easily seen to be equivalent to neutrality on the left. Necessary and sufficient conditions for this independence will be given in the next section.

**Proof.** Let us first condition on $Z_1^+ = 1$. From the definition of the basis one immediately sees that $\boldsymbol{S}_1 \sim \mathrm{FD}^k\left(\boldsymbol{\alpha}_1, \frac{\boldsymbol{p}_1}{p_1^+}, \tau\right)$. By compositional invariance of the FD we also have that $\boldsymbol{S}_1$ is independent of $Y_1^+ = W_1^+ + U$. This implies that $\boldsymbol{S}_1$ is independent of $\boldsymbol{X}_2 = \frac{1}{W^+ + U}(W_{k+1}, \ldots, W_D)$. An analogous argument shows that, conditionally on $Z_1^+ = 0$, $\boldsymbol{S}_1$ is again independent of $\boldsymbol{X}_2$ but with distribution $\mathcal{D}^k(\boldsymbol{\alpha}_1)$. Therefore we can write

$$F_{\boldsymbol{S}_1|\boldsymbol{X}_2=\boldsymbol{x}_2}(\boldsymbol{s}_1) = \mathrm{FD}^k\left(\boldsymbol{s}_1; \boldsymbol{\alpha}_1, \frac{\boldsymbol{p}_1}{p_1^+}, \tau\right) Pr\left(Z_1^+ = 1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2\right) + \mathcal{D}^k\left(\boldsymbol{s}_1; \boldsymbol{\alpha}_1\right) Pr\left(Z_1^+ = 0 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2\right).$$

Finally, the weight $Pr\left(Z_1^+ = 1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2\right)$ can be computed as $f_{\boldsymbol{X}_2|Z_1^+=1}(\boldsymbol{x}_2)Pr\left(Z_1^+ = 1\right)/f_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$ by noticing that, conditionally on $Z_1^+ = 1$, $(\boldsymbol{X}_2, 1 - X_2^+)$ is distributed as $\mathcal{D}^{D-k+1}\left(\boldsymbol{\alpha}_2, \alpha^+ + \tau - \alpha_2^+\right)$, whereas unconditionally, by Proposition 4.6, $(\boldsymbol{X}_2, 1 - X_2^+) \sim \mathrm{FD}^{D-k+1}\left(\boldsymbol{\alpha}_2, \alpha^+ - \alpha_2^+, \boldsymbol{p}_2, 1 - p_2^+, \tau\right)$. □

**Proposition 4.8** (*Conditional Moments*). *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ and let $\boldsymbol{S}_1 = (S_{11}, \ldots, S_{1k})$. Then*

$$E\left[\prod_{i=1}^{k} S_{1i}^{n_i} | \boldsymbol{X}_2 = \boldsymbol{x}_2\right] = \prod_{i=1}^{k} \alpha_i^{[n_i]} \left[\frac{p(\boldsymbol{x}_2)}{(\alpha_1^+ + \tau)^{[n^+]}} \sum_{i=1}^{k} \frac{(\alpha_i + \tau)^{[n_i]}}{\alpha_i^{[n_i]}} \frac{p_i}{p_1^+} + \frac{1 - p(\boldsymbol{x}_2)}{(\alpha_1^+)^{[n^+]}}\right] \tag{15}$$

*where $p(\boldsymbol{x}_2)$ is given by (13). In particular, the conditional mean is*

$$E\left(\boldsymbol{S}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2\right) = \frac{\boldsymbol{\alpha}_1}{\alpha_1^+} + w\left(\frac{\boldsymbol{p}_1}{p_1^+} - \frac{\boldsymbol{\alpha}_1}{\alpha_1^+}\right) p(\boldsymbol{x}_2) \tag{16}$$

*where $w = \tau/(\tau + \alpha_1^+)$.*

The conditional mean (16) does not depend on $\boldsymbol{x}_2$ if and only if $p_1^+ = 0$ or $p_2^+ = 0$ or $\boldsymbol{\alpha}_1/\alpha_1^+ = \boldsymbol{p}_1/p_1^+$. Otherwise, such regression function allows for a variety of different dependence forms. For any given component $S_{1i}$, one can model the form of the dependence on $\boldsymbol{x}_2$ by suitably choosing $\alpha_i$ and $p_i$. Specifically, $\frac{\alpha_i}{\alpha_1^+} = \frac{p_i}{p_1^+}$ gives independence, whereas $\frac{\alpha_i}{\alpha_1^+} > \frac{p_i}{p_1^+}$ ($\frac{\alpha_i}{\alpha_1^+} < \frac{p_i}{p_1^+}$) produces positive (negative) dependence, as $p(\boldsymbol{x}_2)$ is non increasing in each $x_j, j = k + 1, \ldots, D$.

Furthermore, $E\left(S_{1i} | \boldsymbol{X}_2 = \boldsymbol{x}_2\right)$ varies from $\frac{\alpha_i}{\alpha_1^+} + w\left(\frac{p_i}{p_1^+} - \frac{\alpha_i}{\alpha_1^+}\right)$ when $x_2^+ \to 0$ to $\frac{\alpha_i}{\alpha_1^+}$ when $x_2^+ \to 1$. Therefore the bigger the difference $\left|\frac{p_i}{p_1^+} - \frac{\alpha_i}{\alpha_1^+}\right|$ the stronger the dependence on $\boldsymbol{x}_2$ and the larger the range of variation.

As a consequence of the definition of the generating basis, the FD is permutation invariant.

**Proposition 4.9** (*Closure Under Permutation*). *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$. Then, any permutation of $\boldsymbol{X}$ is $\mathrm{FD}^D(\boldsymbol{\alpha}', \boldsymbol{p}', \tau)$ where $\boldsymbol{\alpha}'$ and $\boldsymbol{p}'$ are the corresponding permutation of $\boldsymbol{\alpha}$ and $\boldsymbol{p}$, respectively.*

This result fully expresses the symmetric nature of the FD, implying that the particular order chosen to form the composition does not affect statistical analyses. This property is not shared by many other relevant models where typically the fill-up value is given a different status.

**Proposition 4.10** (*Closure Under Amalgamation*). *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$. Then, in the notation of Proposition 3.1:*

$$\boldsymbol{X}^+ \sim \mathrm{FD}^C(\alpha_1^+, \ldots, \alpha_C^+, p_1^+, \ldots, p_C^+, \tau).$$

Notice that the parameters of the amalgamation $\boldsymbol{X}^+$ are simply related with the parameters of $\boldsymbol{X}$, just as in the Dirichlet case. This derives from the analogous property of the basis.

The distribution of subcompositions can be derived from the partition representation after a suitable permutation of the components (see Propositions 4.3 and 4.9).

**Proposition 4.11** (*Subcompositions*). *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$. Then*

$$\boldsymbol{S}_i = \frac{\left(X_{a_{i-1}+1}, \ldots, X_{a_i}\right)}{X_i^+} \sim p_i^+ \mathrm{FD}^{b_i}\left(\boldsymbol{\alpha}_i, \frac{\boldsymbol{p}_i}{p_i^+}, \tau\right) + (1 - p_i^+) \mathcal{D}^{b_i}(\boldsymbol{\alpha}_i)$$

*where $\boldsymbol{\alpha}_i = (\alpha_{a_{i-1}+1}, \ldots, \alpha_{a_i})$, $\boldsymbol{p}_i = (p_{a_{i-1}+1}, \ldots, p_{a_i})$ and $b_i = a_i - a_{i-1}$.*

It is noticeable that subcompositions and (normalized) conditional distributions, being a mixture of a Dirichlet and a FD, may still be described in terms of the original random allocation scheme (2) if we allow the probabilities $p_i$'s to have sum smaller than one.

## 5. Independence relationships for the flexible Dirichlet

The FD exhibits a sophisticated dependence structure which admits various levels of independence corresponding to simple suitable parameter configurations.

Let us focus first on independence properties for partitions of order 1 as described in Section 2. All such properties can be easily derived from the joint distribution of $(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{X}^+)$ given in Proposition 4.3. To avoid trivial independence relations suppose in the following that $\boldsymbol{S}_i$ $(i = 1, 2)$ contains at least two components, i.e. $k \geq 2$ and $D - k \geq 2$.

The following parameter configurations will play a fundamental role in characterizing the various independence concepts:

1.a. $p_1 = \cdots = p_k = 0$;
1.b. $p_{k+1} = \cdots = p_D = 0$;
2.a. $\tau = 1$ and $\frac{p_1}{\alpha_1} = \cdots = \frac{p_k}{\alpha_k}$;
2.b. $\tau = 1$ and $\frac{p_{k+1}}{\alpha_{k+1}} = \cdots = \frac{p_D}{\alpha_D}$.

**Proposition 5.1.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has* subcompositional independence, *i.e.* $\boldsymbol{S}_1 \perp \boldsymbol{S}_2$, *if and only if at least one of the four conditions* 1.a, 1.b, 2.a, 2.b *is satisfied.*

**Proposition 5.2.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *is* neutral on the left, *i.e.* $\boldsymbol{S}_1 \perp (\boldsymbol{S}_2, \boldsymbol{X}^+)$, *if and only if at least one among the three conditions* 1.a, 1.b, 2.a *is satisfied.*

As neutrality on the left is equivalent to independence between $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$ (being $\boldsymbol{X}_2$ in a one to one correspondence with $(\boldsymbol{S}_2, \boldsymbol{X}^+)$), conditions in Proposition 5.2 guarantee independence between $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$ as well. Complete analogous conditions can be obtained for right neutrality.

**Proposition 5.3.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has* subcompositional invariance, *i.e.* $(\boldsymbol{S}_1, \boldsymbol{S}_2) \perp \boldsymbol{X}^+$, *if and only if at least one of the three conditions* 1.a, 1.b, *both* 2.a *and* 2.b *is satisfied.*

**Proposition 5.4.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has* partition independence, *i.e.* $\boldsymbol{S}_1 \perp \boldsymbol{S}_2 \perp \boldsymbol{X}^+$, *if and only if it has subcompositional invariance.*

The above properties show that the FD model is capable of distinguishing several types of independences relevant for compositional data. In particular, it succeeds in discriminating among subcompositional independence, left neutrality, right neutrality and partition independence, the only two undistinguishable forms of independence being subcompositional invariance and partition independence.

To underline the different implications of the two sets of conditions, namely type 1 (equality to zero of a group of $p_i$'s) and type 2 ($\tau = 1$ and proportionality of a group of $\alpha_i$'s to the corresponding $p_i$'s), we shall consider the general and central independence concept of neutrality. Such concept, which includes both left neutrality and right neutrality, was first introduced by Connor and Mosimann [6]. A certain group of components is said neutral if it has no influence on the remaining ones which are the primary interest of the analysis. Formally, this is defined as independence between such a group and the relative proportions of the remaining variables. Given the permutation invariance property of the FD, we can always take $\boldsymbol{X}_1$ as the variables of interest so that neutrality can be reduced to left neutrality, i.e. independence between $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$.

In order to model such neutrality, type 1 conditions can be imposed either on the first $k$ variables or on the remaining ones. In both cases they give rise to a stronger and symmetric form of independence, i.e. neutrality of both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. However, if interest of the analysis is on the components in $\boldsymbol{X}_1$, it may be preferable to assume type 1 conditions on $\boldsymbol{X}_2$, as this allows a completely general FD distribution for $\boldsymbol{S}_1$ instead of a Dirichlet one.

Type 2 conditions lead to neutrality of $\boldsymbol{X}_2$ only if they are fulfilled by the first $k$ variables. This does not imply neutrality of $\boldsymbol{X}_1$, but it forces severe constraints on the distribution of the variables of interest, which can only be Dirichlet.

Let us now consider complete independence properties. The following results can be proved by intersection of the conditions assuring the corresponding properties for a given level of a partition.

**Proposition 5.5.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has* complete left neutrality *if and only if at least one of the following conditions is satisfied*

(i) $\tau = 1$ *and* $\frac{p_1}{\alpha_1} = \cdots = \frac{p_{D-1}}{\alpha_{D-1}}$;
(ii) $\tau = 1$, $\frac{p_1}{\alpha_1} = \cdots = \frac{p_{k-1}}{\alpha_{k-1}}$ *and* $p_{k+1} = \cdots = p_D = 0$, *where k can take any value in* $\{3, \ldots, D-1\}$;
(iii) $p_3 = \cdots = p_D = 0$.

Analogous conditions can be obtained for complete right neutrality.

**Proposition 5.6.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has* complete subcompositional independence *if and only if at least one of the following conditions is satisfied*

(i) *it has complete right neutrality;*
(ii) *it has complete left neutrality;*
(iii) $\tau = 1$ *and* $\frac{p_1}{\alpha_1} = \cdots = \frac{p_{D-2}}{\alpha_{D-2}}$;
(iv) $\tau = 1$ *and* $\frac{p_3}{\alpha_3} = \cdots = \frac{p_D}{\alpha_D}$;
(v) $\tau = 1$, $\frac{p_1}{\alpha_1} = \cdots = \frac{p_{k-1}}{\alpha_{k-1}}$ *and* $\frac{p_{k+1}}{\alpha_{k+1}} = \cdots = \frac{p_D}{\alpha_D}$ *where k can take any value in* $\{3, \ldots, D-2\}$.

The above complete properties are referred to a given order of the components in agreement with the literature (see [1,6]). The permutation invariance of the FD model allows to identify immediately analogous conditions for any reordering of the variables.

Clearly, independence issues are relevant for higher order partitions as well, i.e. when more than two groups are considered. Here, actually, a larger variety of independence concepts can be defined. The treatment of such concepts does not give rise to particular difficulties within the FD model as they can be analyzed resorting to a generalization of the partition representation (see Proposition 4.3).

**Proposition 5.7.** *Let $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ be partitioned into C subsets as defined in (1). Then, the joint distribution function of the corresponding subcompositions and amalgamation $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C, \boldsymbol{X}^+)$ can be written as*

$$\sum_{i=1}^{C} G_i(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_C, \boldsymbol{x}^+)\, p_i^+ \tag{17}$$

*where* $G_i(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_C, \boldsymbol{x}^+) = \mathcal{D}^C(\boldsymbol{x}^+; \boldsymbol{\alpha}^+ + \tau \boldsymbol{e}_i)\, \mathrm{FD}^{b_i}(\boldsymbol{s}_i; \boldsymbol{\alpha}_i, \boldsymbol{p}_i/p_i^+, \tau) \prod_{j \neq i} \mathcal{D}^{b_j}(\boldsymbol{s}_j; \boldsymbol{\alpha}_j)$ *and* $b_j = a_j - a_{j-1}$.

Such expression can be proved along the same lines of Proposition 4.3. It allows a straightforward derivation of all marginal and conditional distributions of subcompositions and totals.

Furthermore, from (17) conditions for any independence among the components of $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C, \boldsymbol{X}^+)$ can be easily established and expressed through simple parameter configurations. As an illustrative example, consider subcompositional independence. To avoid trivial independences suppose each group contains at least two components.

**Proposition 5.8.** $\boldsymbol{X} \sim \mathrm{FD}^D(\boldsymbol{\alpha}, \boldsymbol{p}, \tau)$ *has subcompositional independence, i.e.* $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C$ *are independent, if and only if at least one of the following conditions holds.*

1. *The $p_i^+$'s are all equal to zero except one.*
2. $\tau = 1$, *$m$ $p_i^+$'s are equal to zero and in at least $C - m - 1$ of the remaining groups the $\alpha_i$'s are proportional to the $p_i$'s, with $m$ ranging from 0 to $C - 2$.*

**Proof.** By integrating $\boldsymbol{X}^+$ out in (17) one immediately obtains a mixture representation of the distribution of $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C)$. Suppose there are $m$ $p_i^+$'s equal to zero, $m = 0, 1, \ldots, C-2$. Then $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_C)$ are independent if and only if at least $C-m-1$ of the remaining $C - m$ groups have identical (Dirichlet) distribution in all terms of the mixture representation. But this is possible if and only if $\tau = 1$ and the $\alpha_i$'s are proportional to the $p_i$'s within the $C - m - 1$ groups. If only one $p_i^+$ is positive, then by (17) the $\boldsymbol{S}_i$'s are always independent. □

## 6. Flexible Dirichlet and generalized Liouville distributions

The FG and FD distributions are connected with the semiparametric family of Liouville distributions and especially with some of its generalizations. The Liouville distribution has density proportional to

$$\prod_{i=1}^{D} x_i^{\alpha_i - 1} f\left(\sum_{i=1}^{D} x_i\right), \quad \alpha_i > 0, i = 1, \ldots, D$$

where $f$ is positive, continuous and satisfies proper integrability conditions. In particular, if the support is $\mathcal{R}_+^D$ then $\boldsymbol{X}$ has a I kind Liouville distribution while it is of II kind if the support is the $D$-dimensional simplex $\mathcal{S}^{D+1}$. Its properties have been extensively studied (see [8–12]). The only distribution on the simplex generated as normalization (conditionally on the size or unconditionally) of the I kind Liouville is the Dirichlet. The II kind Liouville has a dependence structure not sufficiently richer than the Dirichlet one (see [13,14]). For example all subcompositions are Dirichlet distributed. Moreover, it is an asymmetric model as the fill-up value exhibits a special behavior.

In [19,20] a number of generalizations of the Liouville are proposed. The most attractive from a compositional viewpoint is the conditional generalized Liouville (CGL) (see [20]). Such distribution can be thought of as obtained by normalization, conditionally on the size, of the I kind generalized Liouville. Both the I kind generalized Liouville and the CGL have density proportional to

$$\prod_{i=1}^{D} x_i^{\alpha_i - 1} f\left[\sum_{i=1}^{D} (x_i/q_i)^{\tau_i}\right], \quad \alpha_i, q_i, \tau_i > 0, i = 1, \ldots, D \tag{18}$$

respectively defined on $\mathcal{R}_+^D$ and on $\mathcal{S}^D$.

The CGL inherits the same flexibility in modeling independence concepts as the normalized I kind generalized Liouville, but it possesses an explicit density up to the normalizing constant. The evaluation of such constant is problematic though (see comments in [13, p. 132]).

Let us now investigate the role of the FD within the generalized Liouville model.

If $f$ is chosen as the identity function, the FD differs from CGL in two respects. On the one side, it allows some of the coefficients of the power terms $x_i^{\tau_i}$ to be equal to zero. This is a useful extension as it introduces a richer spectrum of possibilities to model independence concepts as discussed in Section 5. On the other side, the CGL allows for different values of the $\tau_i$'s. Such an extension loses some important properties such as closure under amalgamation and marginalization as well as simple expressions for conditionals and moments. At the same time it does not produce gains in terms of range of independence relationships which can be modeled.

The choice of the function $f$ in Liouville models is a non trivial and still open problem. Our results on the FD suggest that the identity function is a simple and fruitful solution as it leads to many important compositional properties, which are not shared in general by generalized Liouville distributions. Furthermore, such properties are obtained without losing flexibility in differentiating the various forms of independence, as can be seen by comparing the CGL and the FD independence properties.

To fully explore the existence of other useful parametric subfamilies of the generalized Liouville class, an analysis of the consequences of any given choice of the nonparametric part $f$ on the various properties of interest would be required. Here, to illustrate, we shall perform such analysis focusing on compositional invariance, a general investigation going beyond the scope of the present paper. In particular, we shall provide necessary and sufficient conditions for compositional invariance to hold for an arbitrary given choice of $f$ in the I kind generalized Liouville. This is in contrast with the approach used in [20] where conditions are given if the property has to hold simultaneously for all possible functions $f$.

Actually, we shall give our result starting from a convenient generalization of the I kind generalized Liouville. The new density is taken to be proportional to

$$g\left(\sum_{i=1}^{D} y_i\right)\prod_{i=1}^{D} y_i^{\alpha_i-1} f\left(\sum_{i=1}^{D}\psi_i y_i^{\tau_i}\right), \quad y_i > 0, i = 1, \dots, D \tag{19}$$

where the $\tau_i$'s and $\alpha_i$'s are all strictly positive and the $\psi_i$'s are non-negative. In fact, if $\psi_i = 0$ the corresponding parameter $\tau_i$ is not defined, therefore we shall assume that the parameter space includes only the $\tau_i$'s corresponding to positive $\psi_i$'s. The function $g$ is assumed strictly positive.

The new factor $g$ allows to enlarge the possible choices of the function $f$ in (19), i.e. the ones which make (19) integrable. Furthermore, it allows general models for the distribution of the size $Y_+ = \sum_{i=1}^{D} Y_i$ without affecting compositional invariance of the I kind generalized Liouville.

**Theorem 1.** *Let* $\boldsymbol{Y} = (Y_1, \dots, Y_D)$ *be a vector of non-negative random variables with density proportional to* (19). *If* $f$ *is continuous then* $\boldsymbol{Y}$ *is compositionally invariant if and only if at least one of the following two cases holds.*

(i) $\boldsymbol{Y}$ *has a Liouville distribution, i.e.* (i$_1$) $f$ *is constant or* (i$_2$) $\psi_1 = \dots = \psi_D > 0$ *and* $\tau_1 = \dots = \tau_D = 1$ *or* (i$_3$) $\psi_1 = \dots = \psi_D = 0$.
(ii) *The* $\tau_i$*'s corresponding to positive* $\psi_i$*'s are all equal and* $f(x) = cx^{\alpha}$, $c > 0$, $\alpha \in \mathcal{R}$.

The proof of Theorem 1, together with a preliminary lemma, is given in the Appendix.

Under subclass (i), the only distribution for the normalized (conditional or unconditional to the size) simplex vector is the Dirichlet. In particular, (i$_2$) and (i$_3$) are necessary conditions if we ask the class to be compositionally invariant for any $f$, in accordance with Theorem 5.2 in [20]. On the contrary, condition (ii) identifies a specific parametric power form for $f$ which gives rise to a potentially useful generalization of the FG basis. The latter corresponds to the case $\alpha = 1$ and $g(x) = e^{-x}$. Notice however that such generalization does not admit a mixture representation and, therefore, it is bound to be less tractable than the FG.

## 7. Inferential issues and an application

Let us give some general indications on inferential aspects so as to highlight FD tractability. We intend to take up a detailed analysis in future work.

Standard methods for likelihood maximization are unsuitable in the present setup. Yet, the finite mixture structure of the model allows to treat the estimation problem as an incomplete data one (see [16]); thus the $E$–$M$ algorithm can be suitably adapted.

More precisely, given $n$ independent observations the true log-likelihood can be thought of as originated from the following complete log-likelihood:

$$\sum_{j=1}^{n}\sum_{i=1}^{D} z_{ji}\left\{\log p_i + \log f_{\mathcal{D}}(\boldsymbol{x}_j; \boldsymbol{\alpha} + \tau \boldsymbol{e}_i)\right\}$$

where the $z_{ji}$'s represent the missing data and take the value zero or one, according to whether $\boldsymbol{x}_j$ has arisen from the $i$th component of the mixture or not. Here $f_{\mathcal{D}}(\boldsymbol{x}; \boldsymbol{\alpha})$ denotes the Dirichlet density with parameter $\boldsymbol{\alpha}$.

The choice of the starting values for the $E$–$M$ algorithm, which is a critical one, can be successfully dealt with combining the $k$-means clustering algorithm for estimating the $p_i$'s and a two-step method of moments for $\tau$ and $\boldsymbol{\alpha}$.

Let us now focus on testing issues and, in particular, on conditions 1.a, 1.b, 2.a or 2.b of Section 5, which give rise to the various kinds of independence described in Section 2. All such hypotheses can be tested through suitable likelihood ratio tests. The constrained maximization required by such tests is typically best achieved by constructing suitable profile likelihoods which exploit properties of marginal, amalgamation and subcomposition distributions.

As an example, consider the more involved type 2 conditions, e.g. $H_0 : \tau = 1, p_1/\alpha_1 = \dots = p_k/\alpha_k$. Under such hypothesis the distribution of $\boldsymbol{X}$ can be conveniently represented through the distribution of $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$, which are independent with $\boldsymbol{S}_1 \sim \mathcal{D}^k(\boldsymbol{\alpha}_1)$ and $\boldsymbol{X}_2 \sim \mathrm{FD}^{D-k}(\dot{\boldsymbol{\alpha}}, \dot{\boldsymbol{p}}, \tau = 1)$ where $\dot{\boldsymbol{\alpha}} = (\alpha_{k+1}, \dots, \alpha_D, \alpha_1^+)$ and $\dot{\boldsymbol{p}} = (p_{k+1}, \dots, p_D, p_1^+)$. This formulation has the advantage of automatically incorporating the null hypothesis constraints, leading to unconstrained

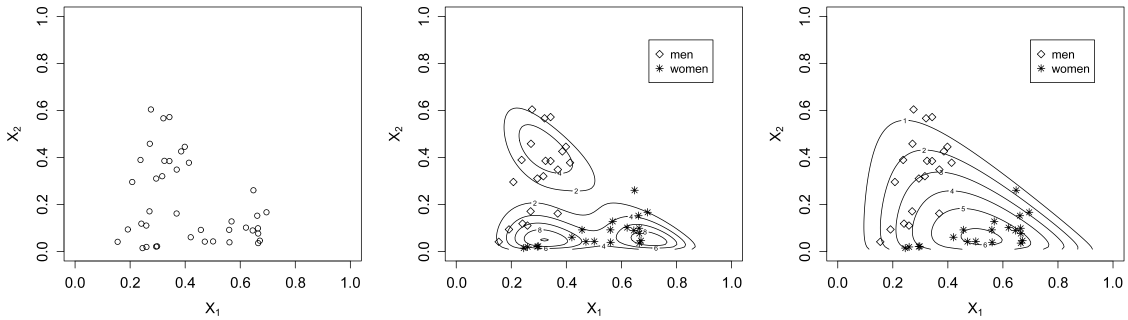


**Fig. 2.** Plot of $X_1$ and $X_2$ (left), estimated contours for the FD model (center) and estimated contours for the Dirichlet model (right).

**Table 1**
Sample and estimated means and variances.

| | $E(X_1)$ | $E(X_2)$ | $E(X_3)$ | $\text{Var}(X_1)$ | $\text{Var}(X_2)$ | $\text{Var}(X_3)$ |
|---|---|---|---|---|---|---|
| Sample | 0.414 | 0.205 | 0.381 | 0.027 | 0.031 | 0.040 |
| Dirichlet | 0.430 | 0.192 | 0.378 | 0.035 | 0.022 | 0.033 |
| FD | 0.421 | 0.206 | 0.373 | 0.033 | 0.031 | 0.034 |

maximization. Furthermore, such maximization can be dealt with by constructing the profile likelihood for $\alpha_1^+$. This is because, being $\alpha_1^+$ the only parameter shared by the two distributions, such profile likelihood can be easily derived by separate maximization of the likelihoods relative to $\boldsymbol{S}_1$ and $\boldsymbol{X}_2$.

Combinations of conditions 1.a, 1.b, 2.a and 2.b can be tested through similar strategies.

To illustrate the potential of the FD in terms of real-world applications, let us consider an example relative to the analysis of household budget surveys aimed at studying consumer demand. A relevant aspect of such analysis is the investigation of the proportions (compositions) of total household expenditures allocated to various types of commodities. In particular, we shall consider a data set made available by Aitchison [1, p. 362] where household expenditures on four commodity groups of a sample of 40 individuals (20 single men and 20 single women living alone) are reported. The variables under study are the proportions spent on housing ($X_1$), including fuel and light, foodstuffs ($X_2$), including alcohol and tobacco, services ($X_3$), including transport and vehicles, and other goods ($X_4$), including clothing, footwear and durable goods.

To make the results of the analysis simpler to be represented and conveyed, we reduce the original sample space to three variables only. This can be achieved, without loosing the main features of the problem, by amalgamating services and other goods. Indeed such commodity groups can be viewed as rather homogeneous corresponding to more luxury goods. Therefore, from now on, $X_3$ will indicate the new amalgamated variable.

A plot of the pair $(X_1, X_2)$ is shown in Fig. 2 (left panel). The correlation coefficients $\rho_{12} = -0.304$, $\rho_{13} = -0.552$ and $\rho_{23} = -0.626$ highlight the (significantly) negative correlation between all components. Means and variances are reported in Table 1.

The maximum likelihood estimates of the FD parameters obtained via implementation in $R$ language of the above described adaptation of the $E$–$M$ algorithm are $\hat{\boldsymbol{\alpha}} = (6.60, 1.88, 5.31)$, $\hat{\boldsymbol{p}} = (0.305, 0.347, 0.348)$ and $\hat{\tau} = 6.78$, whereas the estimates of the Dirichlet model parameters are $\hat{\boldsymbol{\alpha}} = (2.63, 1.18, 2.31)$. The contours of the two estimated models are shown in Fig. 2 (center and right panels). Unlike the Dirichlet model, the FD produces a fairly accurate fit. In particular, its ability of identifying the presence of three clusters of roughly similar sizes is remarkable. Two clusters are originated by attitudes toward the household expenditures typically associated with gender: the first one, including only men, is characterized by high values of $X_2$ and the second one, including only women, by high values of $X_1$. A third perhaps less predictable cluster presents high values of $X_3$. An inspection of the absolute expenditures (contained in the original data set) highlights that such group corresponds to individuals with particularly high income levels whatever the gender.

The superiority of the FD model is strongly confirmed by AIC (as well as by BIC) values (−96.10 (−85.96) for the FD and −84.16 (−79.10) for the Dirichlet). A better behavior of the FD is also supported by estimates of means and variances reported in Table 1. In particular, notice that the Dirichlet model, as a consequence of the relations it imposes between means and variances, is forced to underestimate the variance of $X_2$.

Interesting considerations can be drawn from the analysis of the one-dimensional marginal plots as well. They all suggest the presence of bimodality which, again, is captured fairly well by the FD model in contrast with the Dirichlet (beta in this case) one (see Fig. 3).

In this example no components seem to be neutral (as implied by the Dirichlet model), although associations between each subcomposition and the remaining variable are rather low. The highest and perhaps most interesting one concerns the pair $S_{12} = X_1/(X_1 + X_2)$ and $X_3$. The corresponding plot is reported in Fig. 4 (left panel) together with the estimate of the FD conditional mean (16) and, as a benchmark, a normal kernel nonparametric regression. Such conditional mean is consistent with the pattern of the data implying that increasing levels of $X_3$ lead to a higher relative importance of $X_1$ (housing) with respect to $X_2$ (food).

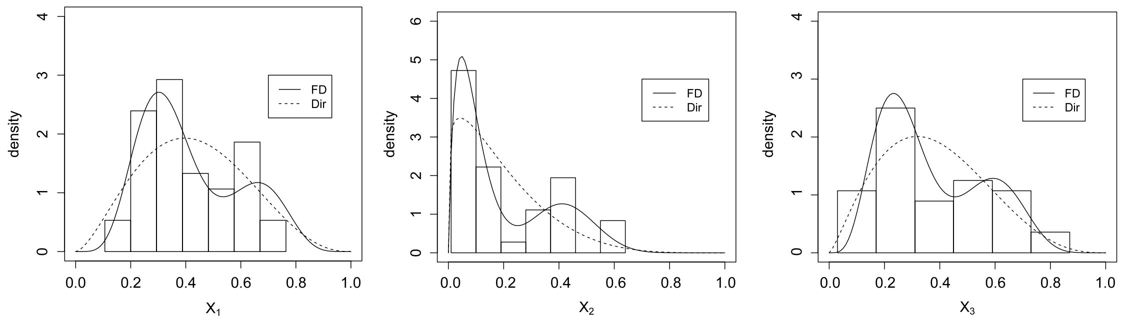


**Fig. 3.** Histogram and estimated densities of $X_1$ (left), $X_2$ (center) and $X_3$ (right).

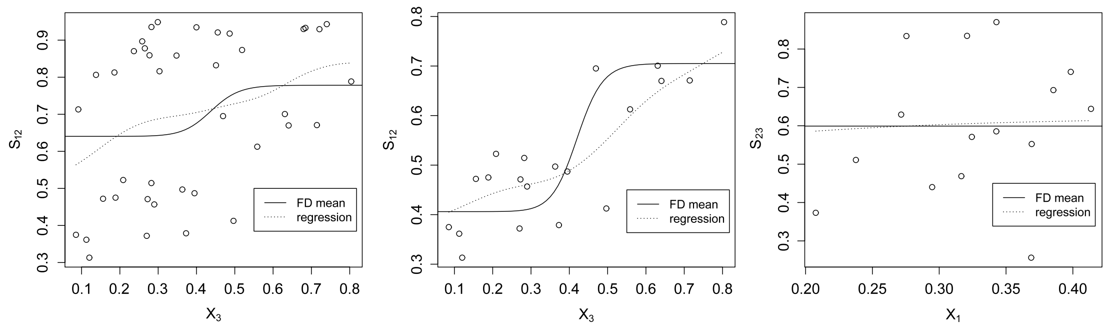


**Fig. 4.** Conditional mean and nonparametric regression for all individuals (left panel), for all men (center panel) and for the first cluster of men only (right panel).

Quite stronger associations are obtained if the analysis is performed distinguishing individuals on the basis of gender. As an example we report the same variables for the group of men (see center panel of Fig. 4). If attention is focused on the cluster level, then neutral components are detected by the FD model, coherently with data evidence. For example $X_1$ is neutral, i.e. it is independent of $S_{23} = X_2/(X_2 + X_3)$, for the first cluster consisting of men only (see Fig. 4, right panel).

In general, it seems that the conditional mean implied by the FD model is very often sufficiently flexible to capture the main pattern inherent in the data.

## 8. Conclusions

The FD incorporates a structure of independences/dependences, which, although not completely general, represents a substantial improvement on the Dirichlet model. This is witnessed by the generality of (log-ratio) covariances and conditional mean expression as well as by the wide range of simplicial independences it admits. In addition, it allows for great flexibility in the density shape including multi-modality.

The FD model succeeds in combining the above described flexibility with a considerable tractability. The latter derives from the special structure of the random allocation scheme defining the basis FG which gives rise to some very useful alternative representations. Moreover, it implies some general properties such as symmetry, simple and explicit expressions for marginals, conditionals and moments, as well as properties of specific relevance for compositional data such as compositional invariance and closure under amalgamation and subcomposition.

Such combination of flexibility and tractability makes the FD model distinctive when compared with log-ratio transformation models (as already underlined in Section 1) as well as with other generalizations of the Dirichlet present in the literature. Indeed, the FD constitutes a privileged case of the semiparametric class of generalized Liouville distributions, sharing its quite elaborate dependence patterns but displaying a special tractability. On the other hand, it allows less restrictive dependence forms than other relevant distributions on the simplex. In particular, although some models in [5] show interesting statistical properties (including chi-squared decompositions, closure under amalgamation, marginalization, permutation and conditioning), they exhibit a poor dependence structure, which is determined by a single precision parameter, as in the Dirichlet case. Furthermore, Connor and Mosimann's generalization of the Dirichlet [6] only models completely right neutral compositions whereas the general class proposed in [7], because of the independence of the generating basis, implies severe restrictions on the relations between ratios (see comments below Proposition 4.6).

We end by pointing to some further developments of the FD.

A limitation common to most compositional models is the assumption that all components are strictly positive. This is in contrast with frequent situations in applications where one or more elements of the composition may be absent (essential zeros). The FD allows to model such elements by equating to zero the corresponding $\alpha_i$'s. If $\alpha_i = 0$ then $p_i$ assumes the meaning of probability that the $i$th component is strictly positive.

An analysis of the FD from a Bayesian point of view, that is as prior for categorical data models, is also worth exploring as the FD is easily seen to be conjugate with respect to the multinomial sampling scheme. We plan to tackle these issues in future work.

**Acknowledgments**

We are grateful to the referees and to the editor for their constructive comments, which helped to improve the paper. This research was supported by grants from University of Milano-Bicocca.

## Appendix

The following lemma gives all solutions of a functional equation appearing in the proof of Theorem 1. As pointed out by a referee, Eq. (20) below belongs to a sub-family of the general class of Levi-Civita functional equations, known as Pexider equations. Solutions of such equations are generally given assuming their validity for domains and codomains different from the ones needed in our case (cfr. [15]). For this reason and for completeness an ad hoc self-contained proof of the lemma is reported.

**Lemma 1.** *Let $f$ defined on $\mathcal{R}_+$ be a non-negative continuous function not identically equal to zero. The function $f$ factors as*

$$f(xy) = k(y)h(x) \quad \forall x > 0, y \in (a, b) \tag{20}$$

*with $0 < a < b$, for some arbitrary functions $k$ and $h$ if and only if $f(x)$ is of the type $f(x) = cx^{\alpha}$, $c > 0$, $\alpha \in \mathcal{R}$.*

**Proof.** Clearly if $f(x) = cx^{\alpha}$ then (20) is satisfied for any choice of real $\alpha$ and positive $c$, for example by choosing $k(x) = h(x) = \sqrt{c}x^{\alpha}$.

Suppose now that (20) holds. First notice that $k$ must be strictly positive on $(a, b)$, otherwise $f$ would be identically equal to zero. Furthermore, we can assume without loss of generality that $a < 1 < b$. This is because for any $u > 0$ one can write $f(xy) = k'(y)h'(x)$ where $k'(y) = k(yu)$ and $h'(x) = h(x/u)$ which holds for any $x > 0$ and $y \in (a' = a/u, b' = b/u)$. By choosing, for example, $u = (a + b)/2$ one has $a' < 1 < b'$.

Let us now show that $f$ must be everywhere strictly positive. There must exist $x'$ such that $h(x') > 0$. By writing $f(x'y) = h(x')k(y)$ one has that $f$ must be strictly positive on $(x'a, x'b)$. Observe now that $f(xy) = f(x)k(y)/k(1)$. Then, by induction, it follows that $f$ is strictly positive on $(a^n x', b^n x')$. Therefore $f$ must be everywhere strictly positive and consequently $h$ as well. This implies that $f$ must satisfy the functional equation

$$f(xy) = f(x)f(y)/f(1), \quad x > 0, y \in (a, b). \tag{21}$$

By setting $\bar{f}(z) = f(z)/f(1)$ we obtain the equation

$$\bar{f}(xy) = \bar{f}(x)\bar{f}(y) \quad \forall x > 0, y \in (a, b). \tag{22}$$

Furthermore, any solution $f$ of (21) is of the form $f = c\bar{f}$ where $c$ is some positive constant and $\bar{f}$ is a solution of (22). By induction one has that, for any non-negative integer $n$,

$$\bar{f}(x^n) = \bar{f}^n(x) \quad \forall x \in (a, b). \tag{23}$$

The change of variable $y = x^n$ gives, for any strictly positive integer $n$,

$$\bar{f}^{1/n}(y) = \bar{f}(y^{1/n}) \quad \forall y \in (a^n, b^n). \tag{24}$$

Notice that the latter equation holds simultaneously for all $n$ when $y \in (a, b)$. By combining (23) and (24) we obtain, for any positive integer $m$ and non-negative $n$, $\bar{f}^{m/n}(x) = \bar{f}(x^{m/n})$ $\forall x \in (a, b)$. Finally, as $\bar{f}(x^{-1}) = \bar{f}^{-1}(x)$ $\forall x \in (a, b)$, one has that for any (positive or negative) rational number $q \in \mathbb{Q}$ it holds

$$\bar{f}(x^q) = \bar{f}^q(x) \quad \forall x \in (1/d, d) \tag{25}$$

where $d = \min\{1/a, b\}$. Fix now $x_0 \in (1/d, d)$ with $x_0 \neq 1$ and make the change of variable $z = x_0^q$. Then Eq. (25) implies

$$\bar{f}(z) = z^p \quad \forall z \in x_0^{\mathbb{Q}} \tag{26}$$

where $p = \log f(x_0)/\log x_0$ and $x_0^{\mathbb{Q}} = \{z : z = x_0^q \text{ for some } q \in \mathbb{Q}\}$. As the set $x_0^{\mathbb{Q}}$ is dense in $\mathcal{R}_+$, then the only continuous solutions of (22) are of type (26). □

**Proof of Theorem 1.** Let $\boldsymbol{Y}$ have density proportional to (19). Consider the one to one transformation of $\boldsymbol{Y}$ defined by $(\boldsymbol{X}, T)$ where $X_i = Y_i/Y^+$, $i = 1, \ldots, D-1$ and $T = Y^+$. The Jacobian of such transformation is $t^{D-1}$ so that the joint density of

$(\boldsymbol{X}, T)$ is proportional to

$$g(t)\, t^{\alpha_+-1} \prod_{i=1}^{D} x_i^{\alpha_i-1} f\left(\sum_{i=1}^{D} \psi_i(tx_i)^{\tau_i}\right), \quad \boldsymbol{x} \in \mathcal{S}_a^D, t > 0 \tag{27}$$

where $\mathcal{S}_a^D = \{\boldsymbol{x} : x_i > 0, i = 1, \ldots, D-1, \sum_{i=1}^{D-1} x_i < 1\}$ and $x_D = 1 - x_1 - \cdots - x_{D-1}$.

Clearly $\boldsymbol{X}$ and $T$ are independent if and only if (27) factors properly. This is equivalent to the factorization of $f\left(\sum_{i=1}^{D} \psi_i(tx_i)^{\tau_i}\right)$, i.e.

$$f\left(\sum_{i=1}^{D} \psi_i(tx_i)^{\tau_i}\right) = k(\boldsymbol{x})h(t) \quad \text{a.s.} \tag{28}$$

It is trivial to check that if either condition (i) or condition (ii) is satisfied then $f$ factors.

Let us now prove the only if implication.

For brevity's sake, set $u(\boldsymbol{x}, t) \equiv \sum_{i=1}^{D} \psi_i(tx_i)^{\tau_i}$. We shall first prove that Eq. (28) can be taken to hold everywhere. Let $Q_1$ be the set where $h(t)$ is strictly positive and notice that such set must have positive Lebesgue measure. Consider then the following subset of $Q_1$: $Q_2 = \{t \in Q_1 : f(u(\boldsymbol{x}, t)) = k(\boldsymbol{x})h(t)$ a.s. with respect to $\boldsymbol{x}\}$. Such a set must also have positive Lebesgue measure, given that the factorization in (28) holds a.s. There must then exist a $t' > 0$ such that $k(\boldsymbol{x})$ is a.s. proportional to $f(u(\boldsymbol{x}, t'))$. Therefore $k(\boldsymbol{x})$ is a.s. equal to a continuous function and can be assumed to be continuous. Analogously one can show that $h(t)$ can be taken continuous as well. It follows that factorization (28) can be assumed to hold for any $\boldsymbol{x} \in \mathcal{S}_a^{D-1}$ and $t > 0$, as a continuous function a.s. equal to 0 (i.e. $f(u(\boldsymbol{x}, t)) - k(\boldsymbol{x})h(t)$) must be identically null.

Let us now show that if neither condition (i) nor condition (ii) is satisfied then there are no continuous $f$ satisfying (28).

Suppose first that (i) does not hold and the $\tau_i$'s corresponding to positive $\psi_i$'s are not all equal. Then there must exist at least two strictly positive $\psi_i$'s, say $\psi_i$ and $\psi_j$, with $\tau_i \neq \tau_j$. Let $\boldsymbol{e}_i$ ($\boldsymbol{e}_j$) denote a vector with all zero components except for the $i$th ($j$th) equal to one. By letting $\boldsymbol{x}$ tend to $\boldsymbol{e}_i$ in (28), because of continuity of $f$, we obtain $f(\psi_i t^{\tau_i}) = k'h(t)\ \forall t > 0$. Analogously, by letting $\boldsymbol{x}$ tend to $\boldsymbol{e}_j$, we have $f\left(\psi_j t^{\tau_j}\right) = k''h(t)\ \forall t > 0$. Here $k'$ ($k''$) is the limit of $k(\boldsymbol{x})$ when $\boldsymbol{x}$ tends to $\boldsymbol{e}_i$ ($\boldsymbol{e}_j$). Notice that both $k'$ and $k''$ must be strictly positive, otherwise $f$ would be identically null. Therefore, we can write $f(\psi_i t^{\tau_i}) = wf\left(\psi_j t^{\tau_j}\right)$, where $w = k'/k''$. Without loss of generality, suppose now that $\tau_j > \tau_i$. The change of variable $y = \psi_j t^{\tau_j}$ leads to

$$f(y) = wf(qy^c), \quad \forall y > 0 \tag{29}$$

where $q = \psi_i \psi_j^{-c}$ and $c = \tau_i/\tau_j < 1$. By using (29) inductively, we can write $\forall n \geq 1$

$$f(y) = w^n f\left(q^{1+c+c^2+\cdots+c^{n-1}} y^{c^n}\right), \quad \forall y > 0.$$

If we let $n$ tend to $\infty$, as $(q^{1+c+c^2+\cdots+c^{n-1}} y^{c^n})$ converges to $q^{1/(1-c)}$, we obtain that the only continuous solutions $f$ of (28) are constant functions (actually equal to zero if $w \neq 1$), thus contradicting the hypothesis.

Suppose now that (i) does not hold, the $\tau_i$'s corresponding to positive $\psi_i$'s are all equal (say to $\tau$) and let us show that the only function $f(x)$ satisfying (28) has the form $f(x) = cx^\alpha$, $c > 0$, $\alpha \in \mathcal{R}$. In this case $u(\boldsymbol{x}, t) = t^\tau \sum_{i=1}^{D} \psi_i x_i^\tau$ and therefore $f$ must factor as $f(u(\boldsymbol{x}, t)) = k(\sum_{i=1}^{D} \psi_i x_i^\tau)h(t)$, $\forall \boldsymbol{x} \in \mathcal{S}_a^{D-1}$ and $t > 0$. The result is then a direct consequence of Lemma 1. □

## References


[1] J. Aitchison, The Statistical Analysis of Compositional Data, The Blackburn Press, London, 2003.
[2] J. Aitchison, The statistical analysis of compositional data (with discussion), Journal of the Royal Statistical Society: Series B 44 (1982) 139–177.
[3] J. Aitchison, A general class of distributions on the simplex, Journal of the Royal Statistical Society: Series B 47 (1985) 136–146.
[4] J. Aitchison, S.M. Shen, Logistic-normal distributions: some properties and uses, Biometrika 67 (1980) 261–272.
[5] O.E. Barndorff-Nielsen, B. Jorgensen, Some parametric models on the simplex, Journal of Multivariate Analysis 39 (1991) 106–116.
[6] J.R. Connor, J.E. Mosimann, Concepts of independence for proportions with a generalization of the Dirichlet distribution, Journal of the American Statistical Association 64 (1969) 194–206.
[7] S. Favaro, G. Hadjicharalambous, I. Prunster, On a class of distributions on the simplex, Journal of Statistical Planning and Inference 141 (2011) 2987–3004.
[8] R.D. Gupta, D.St.P. Richards, Multivariate Liouville distributions, Journal of Multivariate Analysis 23 (1987) 233–256.
[9] R.D. Gupta, D.St.P. Richards, Multivariate Liouville distributions, II, Probability and Mathematical Statistics 12 (1991) 291–309.
[10] R.D. Gupta, D.St.P. Richards, Multivariate Liouville distributions, III, Journal of Multivariate Analysis 43 (1992) 29–57.
[11] R.D. Gupta, D.St.P. Richards, Multivariate Liouville distributions, IV, Journal of Multivariate Analysis 54 (1995) 1–17.
[12] R.D. Gupta, D.St.P. Richards, Multivariate Liouville distributions, V, in: N.L. Johnson, N. Balakrishnan (Eds.), Advances in the Theory and Practice of Statistics: A Volume in Honour of Samuel Kotz, Wiley, New York, 1997, pp. 377–396.
[13] R.D. Gupta, D.St.P. Richards, The covariance structure of the multivariate Liouville distributions, Contemporary Mathematics 287 (2001) 125–138.
[14] R.D. Gupta, D.St.P. Richards, The history of the Dirichlet and Liouville distributions, International Statistical Review 69 (2001) 433–446.
[15] M. Kuczma, An Introduction to the Theory of Functional Equations and Inequalities, Birkhauser Verlag, Basel, 2009.
[16] G.J. McLachlan, D. Peel, Finite Mixture Distributions, John Wiley & Sons, New York, 2000.

[17] K.W. Ng, M.L. Tang, M. Tan, G.L. Tian, Grouped Dirichlet distribution: a new tool for incomplete categorical data analysis, Journal of Multivariate Analysis 99 (2008) 490–509.
[18] K.W. Ng, M.L. Tang, G.L. Tian, M. Tan, The nested Dirichlet distribution and incomplete categorical data analysis, Statistica Sinica 19 (2009) 251–271.
[19] W.S. Rayens, C. Srinivasan, Dependence properties of generalized Liouville distributions on the simplex, Journal of the American Statistical Association 89 (1994) 1465–1470.
[20] B. Smith, W.S. Rayens, Conditional generalized Liouville distributions on the simplex, Statistics 36 (2002) 185–194.