A new distribution on the simplex containing the Dirichlet family

A. Ongaro[1], S. Migliorati[1], and G.S. Monti[1]
[1]Department of Statistics, University of Milano-Bicocca, Milano, Italy;
E-mail for correspondence: andrea.ongaro@unimib.it

Abstract

The Dirichlet family owes its privileged status within simplex distributions to easyness of interpretation and good mathematical properties. In particular, we recall fundamental properties for the analysis of compositional data such as closure under amalgamation and subcomposition. From a probabilistic point of view, it is characterised (uniquely) by a variety of independence relationships which makes it indisputably the reference model for expressing the non trivial idea of substantial independence for compositions.

Indeed, its well known inadequacy as a general model for compositional data stems from such an independence structure together with the poorness of its parametrisation.

In this paper a new class of distributions (called Flexible Dirichlet) capable of handling various dependence structures and containing the Dirichlet as a special case is presented. The new model exhibits a considerably richer parametrisation which, for example, allows to model the means and (part of) the variance-covariance matrix separately. Moreover, such a model preserves some good mathematical properties of the Dirichlet, i.e. closure under amalgamation and subcomposition with new parameters simply related to the parent composition parameters. Furthermore, the joint and conditional distributions of subcompositions and relative totals can be expressed as simple mixtures of two Flexible Dirichlet distributions.

The basis generating the Flexible Dirichlet, though keeping compositional invariance, shows a dependence structure which allows various forms of partitional dependence to be contemplated by the model (e.g. non-neutrality, subcompositional dependence and subcompositional non-invariance), independence cases being identified by suitable parameter configurations. In particular, within this model substantial independence among subsets of components of the composition naturally occurs when the subsets have a Dirichlet distribution.

Key words: Compositional data, Generalized Dirichlet distribution, Compositional invariance, Neutrality, Subcompositional independence.

1 Introduction

Historically the Dirichlet distribution has represented the first tool for modeling compositional data thanks to its simplicity and good mathematical properties. It can be considered as the model entailing the most extreme forms of independence for compositional data. As remarked by Aitchison (2003, p.60) and others, such characteristic feature heavily restricts its potential for applications. In the light of such inadequacies Aitchison (1980, 1982) proposed a powerful methodology based on log ratio transformations of the original variables. In such approach parametric models are built on the unconstrained transformed sample space.

Here we shall instead work directly on the original sample space with the aim of finding distributions which include the Dirichlet but allow to model weaker forms of independence.

Several generalizations of the Dirichlet distribution have been proposed in the literature, e.g. the scaled Dirichlet (Aitchison, 2003, pp.305-306), the Liouville (Rayens and Srinivasan, 1994 and the references therein), Connor and Mosimann's distribution (Connor and Mosimann, 1969). However it is still an open problem to find a tractable parametric class which contains the Dirichlet but also models showing significant departures from its strong independence properties (see Aitchison, 2003, p.305).

This paper proposes a new generalization of the Dirichlet (called Flexible Dirichlet) and gives an in-depth study of the theoretical properties most relevant for the analysis of compositional data. In particular, special emphasis will be given to various independence concepts for proportions developed in the literature.

The present work is organised as follows. Some important properties of the Dirichlet distribution are briefly recalled in Section 2 whereas some of the main forms of independence for compositional data are summarised in Section 3. The Flexible Dirichlet distribution is then defined by introducing its generating basis (Section 4) and its great tractability is shown through a number of useful properties (such as closure under permutation, amalgamation and subcompositions) and simple expressions for moments, marginals and conditionals. A detailed analysis of the independence relations holding within the Flexible Dirichlet model together with simple characterisations in terms of marginal distributions are given in Section 5. Section 6 contains some final remarks and hints for future developments, while, for clarity of exposition, the most involved proofs are reported in the Appendix.

2 The Dirichlet distribution

The Dirichlet distribution has density function

$$f_{\mathcal{D}}\left(\underline{x};\underline{\alpha}\right)=\frac{\Gamma\left(\sum_{i=1}^{D}\alpha_i\right)}{\prod_{i=1}^{D}\Gamma\left(\alpha_i\right)}\prod_{i=1}^{D}x_i^{\alpha_i-1} \tag{1}$$

where $\underline{x}=(x_1,\ldots,x_D)\in S^D=\left\{\underline{x}:x_i\geqslant 0\ ,i=1,...,D \text{ and } \sum_{i=1}^{D}x_i=1\right\}$, i.e. $\underline{x}$ takes values on the unitary simplex and $\underline{\alpha}=(\alpha_1,\ldots,\alpha_D)\in\mathcal{R}_+^{\mathcal{D}}$ is the parameter vector. We shall denote it by $\underline{X}\sim\mathcal{D}^D(\underline{\alpha})$. For our purposes we shall allow some (although not all) α_i's to take the value zero; in such a case the corresponding components must be interpreted as degenerate at zero.

In this section we shall recall some important properties of the Dirichlet family particularly relevant for the analysis of compositional data.

Let us first introduce some useful definitions. Let $0<a_1<\ldots<a_{C-1}<a_C=D$ and let

$$X_1,\ldots,X_{a_1}|X_{a_1+1},\ldots,X_{a_2}|\ldots|X_{a_{C-1}+1},\ldots,X_{a_C} \tag{2}$$

be a general partition (of order $C-1$) of the vector $\underline{X}$ into C subsets. Often it is of interest to study the behaviour of the C groups by analyzing the subcompositions and by comparing the corresponding totals. The i^{th} subcomposition is defined as

$$\underline{S}_i = \frac{(X_{a_{i-1}+1}, \ldots, X_{a_i})}{T_i}$$

where $T_i = X_{a_{i-1}+1} + \ldots + X_{a_i}$, $(i = 1, \ldots, C)$. The amalgamation is the vector of totals $\underline{T} = (T_1, \ldots, T_C)$.

1. Genesis.

   The Dirichlet distribution can be obtained by normalizing a basis of independent, equally scaled Gamma random variables (r.v.s) $W_i \sim Ga(\alpha_i)$. Formally, if $\underline{X} = \mathcal{C}(\underline{W}) \equiv \underline{W}/W_+$ where $W_+ = \sum_{i=1}^{D} W_i$ then $\underline{X} \sim \mathcal{D}^D(\underline{\alpha})$.

2. Moments.

   If $\underline{X} \sim \mathcal{D}^D(\underline{\alpha})$ then expressions for the first and the second moments are particularly simple:

$$E(X_i) = \frac{\alpha_i}{\alpha^+}$$

$$Var(X_i) = \frac{\alpha_i(\alpha^+ - \alpha_i)}{(\alpha^+)^2(\alpha^+ + 1)} = \frac{E(X_i)(1 - E(X_i))}{(\alpha^+ + 1)}$$

$$Cov(X_i, X_r) = -\frac{\alpha_i \alpha_r}{(\alpha^+)^2(\alpha^+ + 1)} = -\frac{E(X_i)E(X_r)}{(\alpha^+ + 1)}$$

   where $\alpha^+ = \sum_{i=1}^{D} \alpha_i$. It follows that no constraints on the means of the components are imposed, though only one parameter α^+ is devoted to modeling the whole variance-covariance structure. In particular all covariances are proportional to the product of the corresponding means.

3. Marginals and conditionals.

   The Dirichlet distribution is essentially closed under operations of marginalization and conditioning, the consequent distributions being simply related to the full one. More precisely, given a partition of order 1 $\underline{X} = (X_1, \ldots, X_k | X_{k+1}, \ldots, X_D) = (\underline{X}_1, \underline{X}_2)$ then

$$(\underline{X}_1, 1 - \sum_{i=1}^{k} X_i) \sim \mathcal{D}^{k+1}(\alpha_1, \ldots, \alpha_k, \alpha^+ - \sum_{i=1}^{k} \alpha_i).$$

   Furthermore, the normalized conditional

$$\frac{\underline{X}_1}{1 - x_2^+} \mid \underline{X}_2 = \underline{x}_2 \quad \sim \quad \underline{S}_1 \mid \underline{X}_2 = \underline{x}_2$$

   has distribution $\mathcal{D}^k(\alpha_1, \ldots, \alpha_k)$.

   Notice that the only effect of conditioning is on the support of the r.v., so that after normalization the conditional and the marginal distributions do coincide. This can be interpreted as a strong form of independence for subsets of unit-sum constrained random vectors.

4. Permutation

   The Dirichlet family is closed under permutation and the permuted r.v. has Dirichlet distribution with parameter vector permuted accordingly. This apparently trivial property expresses the fact that the Dirichlet distribution treats in a completely symmetric way all the components, which allows to freely rearrange the order of the components.

5. Amalgamation

   The Dirichlet family is closed under amalgamation, i.e.

$$\underline{T} \sim \mathcal{D}^C(\alpha_1^+, \ldots, \alpha_C^+)$$

   where $\alpha_i^+ = \alpha_{a_{i-1}+1} + \ldots + \alpha_{a_i}$.

6. Subcomposition

   The Dirichlet family is closed under subcomposition. In particular we have

$$\underline{S}_1 = \frac{(X_1, \ldots, X_{a_1})}{T_1} \sim \mathcal{D}^{a_1}(\alpha_1, \ldots, \alpha_{a_1}),$$

   analogous expressions holding for all the other subcompositions.

3 Independence concepts for compositional data

Obviously the components of a r.v. $\underline{X}$ defined on the simplex cannot be independent because of the unit-sum constraint. Therefore a large variety of forms of independence has been developed in the literature. Here we shall briefly recall some particularly relevant ones. We shall mainly follow Aitchison (2003) which the reader is referred to for a deeper discussion and further references on the topic.

A first idea of independence somewhat related to a r.v. $\underline{X}$ defined on the simplex concerns the basis $\underline{W}$ generating the composition. More precisely, a basis is **compositionally invariant** if the corresponding composition $\underline{X} = \mathcal{C}(\underline{W})$ obtained by normalizing $\underline{W}$ is independent of its size $W^+ = \sum_{i=1}^{D} W_i$.

A concept which is often considered as the analogous of independence for unconstrained r.v.s is **complete subcompositional independence**. A r.v. $\underline{X}$ possesses such property if the subcompositions $\underline{S}_i$ formed from any partition of the vector form an independent set.

Many other independence definitions are present in the literature, most of which can be expressed in terms of subcompositions $\underline{S}_i$ $(i = 1, \ldots, C)$ and amalgamation $\underline{T}$. For the sake of simplicity we shall illustrate such ideas focusing on partitions of order 1, i.e. partitions formed by two subsets $(C = 2)$. In such case all the forms of independence involve the three r.v.s $\underline{S}_1$, $\underline{S}_2$ and $\underline{T}$. The following table reports a list of the main independence properties, with $\perp$ standing for independence and $A \perp B \perp C$ denoting a set of independent r.v.s.

Table 1: Independence properties for partitions of order 1.

Property	**Independence**
Partition independence	$\underline{S}_1 \perp \underline{S}_2 \perp \underline{T}$
Subcompositional invariance	$(\underline{S}_1, \underline{S}_2) \perp \underline{T}$
Neutrality on the right	$\underline{S}_2 \perp (\underline{S}_1, \underline{T})$
Neutrality on the left	$\underline{S}_1 \perp (\underline{S}_2, \underline{T})$
Subcompositional independence	$\underline{S}_1 \perp \underline{S}_2$

The Dirichlet distribution can be shown to possess all the above independence properties and therefore it can be properly considered as the model of maximum independence compatible with unit-sum constrained r.v.s.

In particular, its generating Gamma basis is compositionally invariant. Moreover, all other independence properties are a consequence of the following well known result. Suppose $\underline{X} \sim \mathcal{D}^D(\underline{\alpha})$

and let $(\underline{S}_1, \ldots, \underline{S}_C)$ be the subcompositions and $\underline{T}$ the amalgamation derived from an arbitrary partition of $\underline{X}$ into C subsets. Then the $C+1$ r.v.s $\underline{S}_1, \ldots, \underline{S}_C, \underline{T}$ are independent.

4 The Flexible Dirichlet

The new distribution, called Flexible Dirichlet, has been derived by normalizing an appropriate basis of positive (but dependent) r.v.s as follows.

4.1 Generating basis

We shall define a basis through a parametric family of positive and dependent r.v.s which contains Gamma independent variates as a particular case. This is achieved by starting from the usual basis of independent equally scaled Gamma r.v.s $W_i \sim Ga(\alpha_i)$ and randomly adding to one of such D r.v.s an independent Gamma r.v. (with the same scale parameter) $U \sim Ga(\tau)$. The latter variate is allocated to the i^{th} component of the basis with probability p_i, $(i = 1, \ldots, D)$.

Formally, let $\underline{Z} = (Z_1, \ldots, Z_D)$ be a r.v. independent from U and from the W_i's which is equal to $\underline{e}_i$ with probability p_i where $\underline{e}_i$ is a vector whose elements are all equal to zero except for the i^{th} element which is one. Here the vector $\underline{p} = (p_1, \ldots, p_D)$ is such that $0 \leq p_i < 1$ and $\sum_{i=1}^{D} p_i = 1$.

Then the new basis $\underline{Y} = (Y_1, \ldots, Y_D)$ is defined as

$$Y_i = W_i + Z_i\, U \qquad i = 1, \ldots, D. \tag{3}$$

In the above definition we let $\alpha_i \geq 0$ and $\tau \geq 0$, a $Ga(0)$ r.v. being interpreted as degenerate at zero. We only ask that $\alpha^+ + \tau > 0$ to prevent all components of $\underline{Y}$ from being degenerate at zero.

By conditioning on $\underline{Z}$ one can easily obtain the joint distribution of $\underline{Y} = (Y_1, \ldots, Y_D)$:

$$F(\underline{y}; \underline{\alpha}, \underline{p}, \tau) = \sum_{i=1}^{D} \left\{ Ga(y_i; \alpha_i + \tau) \prod_{r \neq i} Ga(y_r; \alpha_r) \right\} p_i \tag{4}$$

where $Ga(y_i; \lambda)$ denotes the distribution function of a Gamma r.v. with shape parameter λ. Such density is therefore a finite mixture of random vectors with Gamma independent components. In particular the one-dimensional marginal distribution function takes the form:

$$F(y_i; \alpha_i, p_i, \tau) = p_i Ga(y_i; \alpha_i + \tau) + (1 - p_i) Ga(y_i; \alpha_i). \tag{5}$$

Notice that when $\tau = 0$ the basis $\underline{Y}$ has independent Gamma components and coincides with the Dirichlet basis.

Straightforward calculation shows that the first two moments of $\underline{Y}$ are equal to:

$$\begin{aligned} E(Y_i) &= \alpha_i + p_i \tau \\ Var(Y_i) &= \alpha_i + p_i \tau + p_i (1 - p_i) \tau^2 \\ Cov(Y_i, Y_r) &= -p_i p_r \tau^2. \end{aligned} \tag{6}$$

It is noticeable that the random allocation of the Gamma U variate leads to more flexibility in the variance-covariance structure of the basis, inducing, in particular, negative correlation between its components.

4.2 Definition and first properties

Definition Let $\underline{Y}$ be defined as in (3), then the normalized vector

$$\underline{X} = \mathcal{C}(\underline{Y}) = \left(\frac{Y_1}{Y^+}, \ldots, \frac{Y_D}{Y^+}\right) \tag{7}$$

where $Y^+ = \sum_{i=1}^D Y_i$, has a Flexible Dirichlet distribution denoted by $FD^D(\underline{\alpha}, \underline{p}, \tau)$. □

Some important properties of such distribution are given and discussed below. In order to simplify the notation, given a partition of order 1 $\underline{X} = (X_1, \ldots, X_k | X_{k+1}, \ldots, X_D) = (\underline{X}_1, \underline{X}_2)$ we shall adopt the following definitions:

$$X_1^+ = \sum_{i=1}^{k} X_i, \quad X_2^+ = \sum_{i=k+1}^{D} X_i$$

and in an analogous way we shall define the following quantities: $\underline{\alpha}_1$, $\underline{\alpha}_2$, α_1^+, α_2^+, $\underline{p}_1$, $\underline{p}_2$, p_1^+ and p_2^+. The following mixture representation and the moments of the Flexible Dirichlet can be easily derived by conditioning on $\underline{Z}$.

Property 1 (Mixture representation) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ is a finite mixture of Dirichlet distributions:

$$FD^D(\underline{\alpha}, \underline{p}, \tau) = \sum_{i=1}^{D} p_i \mathcal{D}^D(\underline{\alpha} + \tau \underline{e}_i). \tag{8}$$

In particular, the one-dimensional marginals are mixtures of two Betas:

$$X_i \sim p_i Be\left(\alpha_i + \tau, \alpha_{-i}^+\right) + (1 - p_i) Be\left(\alpha_i, \alpha_{-i}^+ + \tau\right) \tag{9}$$

where $\alpha_{-i}^+ = \alpha^+ - \alpha_i$. Consequently, if all $\alpha_i > 0$ the density function of the Flexible Dirichlet can be expressed as

$$f_{FD}(\underline{x}; \underline{\alpha}, \underline{p}, \tau) = \frac{\Gamma(\alpha^+ + \tau)}{\prod_{r=1}^D \Gamma(\alpha_r)} \left(\prod_{r=1}^{D} x_r^{\alpha_r - 1}\right) \sum_{i=1}^{D} p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + \tau)} x_i^\tau \tag{10}$$

for $\underline{x} \in S^D$. □

Property 2 (Moments) The first two moments of the Flexible Dirichlet take the form

$$\begin{aligned}
E(X_i) &= \frac{\alpha_i + p_i\tau}{\alpha^+ + \tau} = \frac{\alpha_i}{\alpha^+}\left(\frac{\alpha^+}{\alpha^+ + \tau}\right) + p_i\left(\frac{\tau}{\alpha^+ + \tau}\right) \\
Var(X_i) &= \frac{\alpha_i(\alpha^+ - \alpha_i) + \tau[\alpha_i(1 - p_i) + (\alpha^+ - \alpha_i)p_i] + \tau^2 p_i(1 - p_i)(\alpha^+ + \tau + 1)}{(\alpha^+ + \tau)^2(\alpha^+ + \tau + 1)} = \\
&= \frac{E(X_i)(1 - E(X_i))}{(\alpha^+ + \tau + 1)} + \frac{\tau^2 p_i(1 - p_i)}{(\alpha^+ + \tau)(\alpha^+ + \tau + 1)} \\
Cov(X_i, X_r) &= -\frac{\alpha_i\alpha_r + \tau(p_r\alpha_i + p_i\alpha_r) + \tau^2 p_i p_r(\alpha^+ + \tau + 1)}{(\alpha^+ + \tau)^2(\alpha^+ + \tau + 1)} = \\
&= -\frac{E(X_i)E(X_r)}{(\alpha^+ + \tau + 1)} - \frac{\tau^2 p_i p_r}{(\alpha^+ + \tau)(\alpha^+ + \tau + 1)}
\end{aligned} \tag{11}$$

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 Flexible Dirichlet 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. □

Property 3 (Marginals) As in the Dirichlet case, the Flexible Dirichlet distribution is closed under marginalization, simple relationships holding between the parameters of the joint and the parameters of the marginal distributions. More precisely

$$(\underline{X}_1, 1 - X_1^+) \sim FD^{k+1}(\underline{\alpha}_1, \alpha^+ - \alpha_1^+, \underline{p}_1, 1 - p_1^+, \tau). \tag{12}$$

This can be easily seen by considering the distribution of $(Y_1, \ldots, Y_k, \sum_{i=k+1}^{D} Y_i)$. □

Property 4 (Conditionals) The (normalized) conditional distributions of a Flexible Dirichlet are mixtures of a Flexible Dirichlet and of a Dirichlet. More precisely, the normalized conditional

$$\frac{\underline{X}_1}{1 - x_2^+} \mid \underline{X}_2 = \underline{x}_2 \quad \sim \quad \underline{S}_1 \mid \underline{X}_2 = \underline{x}_2$$

has distribution

$$\frac{p_1^+}{p_1^+ + q(\underline{x}_2)} FD^k \left(\underline{\alpha}_1, \frac{\underline{p}_1}{p_1^+}, \tau\right) + \frac{q(\underline{x}_2)}{p_1^+ + q(\underline{x}_2)} \mathcal{D}^k(\underline{\alpha}_1) \tag{13}$$

where

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

Notice that in general, unlike the Dirichlet distribution, the normalized conditional does depend on $\underline{X}_2$. Necessary and sufficient conditions for independence will be discussed in the next section where all independence relationships will be analyzed.

For the Proof see the Appendix. □

Property 5 (Permutation) The Flexible Dirichlet is closed under permutation. Furthermore, the parameters $\underline{\alpha}$ and $\underline{p}$ of the permuted random vector are obtained by applying the same permutation to the original parameters. As already noticed for the Dirichlet case, this means that all the components are treated symmetrically and that any rearrangement of their order can be straightforwardly dealt with.

The proof directly follows from the definition of the generating basis. □

Property 6 (Amalgamation) The Flexible Dirichlet is closed under amalgamation:

$$\underline{T} \sim FD^C(\alpha_1^+, \ldots, \alpha_C^+, p_1^+, \ldots, p_C^+, \tau)$$

where $\alpha_i^+ = \alpha_{a_{i-1}+1} + \ldots + \alpha_{a_i}$ and $p_i^+ = p_{a_{i-1}+1} + \ldots + p_{a_i}$. Notice that the parameters of the amalgamation $\underline{T}$ are easily obtained by summing up the α_i's and p_i's within each group of the partition (as in the Dirichlet case). This can be shown by proving that the generating basis is closed under amalgamation, the latter property deriving from the infinite divisibility of Gamma r.v.s. and from the structure of the allocation scheme of the $Ga(\tau)$ r.v.. □

Property 7 (Subcompositions) The distribution of subcompositions from a Flexible Dirichlet is simply related to such family being a mixture of a Flexible Dirichlet and of a Dirichlet. Formally

$$\underline{S}_1 = \frac{(X_1, \ldots, X_{a_1})}{T_1} \sim p_1^+ FD^k \left(\underline{\alpha}_1, \frac{\underline{p}_1}{p_1^+}, \tau\right) + (1 - p_1^+) \mathcal{D}^k(\underline{\alpha}_1)$$

where $\underline{\alpha}_1 = (\alpha_1, \ldots, \alpha_{a_1})$, $\underline{p}_1 = (p_1, \ldots, p_{a_1})$ and $p_1^+ = p_1 + \ldots + p_{a_1}$. Clearly, analogous expressions do hold for any other subcomposition.

The proof can be given following the same argument as in the proof of Property 4. □

Notice that Properties 6 and 7 were expressed in terms of partitions of the type defined in (2), which, strictly speaking, are all derived from the same given order of the components X_i. However, as the Flexible Dirichlet family is closed under permutation (with new parameters simply obtained by permutation of the original ones) such properties hold for a completely arbitrary partition.

5 Independence relationships for the Flexible Dirichlet

The Flexible Dirichlet exhibits a sophisticated dependence structure we shall thoroughly study in the present section. Various forms of independence are admitted corresponding to suitable parameter configurations, compositional invariance being the only one (among those listed in Section 3) always holding.

Property 8 (Compositional Invariance) The basis $\underline{Y}$ generating the Flexible Dirichlet distribution is compositionally invariant, i.e. the size of the basis Y^+ and the composition $\underline{X}$ are independent.

Proof

Compositional invariance can be proved by deriving the conditional density function of $\underline{X}$ given Y^+ and further conditioning on $\underline{Z}$:

$$f_{\underline{X}|Y^+=y^+}(\underline{x}) = \sum_{i=1}^{D} f_{\underline{X}|Y^+=y^+,\underline{Z}=\underline{e}_i}(\underline{x}) Pr\left(\underline{Z}=\underline{e}_i|Y^+=y^+\right).$$

In particular, from definition (3) it immediately follows that $Y^+ = \sum_{i=1}^{D} W_i + U$ and $\underline{Z}$ are independent so that $Pr\left(\underline{Z}=\underline{e}_i|Y^+=y^+\right) = p_i$. Besides, conditionally on $\underline{Z}=\underline{e}_i$, $\underline{X}$ has a Dirichlet distribution and, consequently, it is independent of Y^+. It follows that the conditional density $f_{\underline{X}|Y^+=y^+}(\underline{x})$ does not depend on Y^+. □

Before further investigating the independence properties of the Flexible Dirichlet it is necessary to determine all parameter configurations which produce the Dirichlet, the latter being the reference model for complete strong forms of independence.

Proposition 1 The Flexible Dirichlet coincides with the Dirichlet if and only if one of the two following conditions holds:

1. $\tau = 0$
2. $\tau = 1$ and $p_i = \alpha_i/\alpha^+$, $\forall i = 1, \ldots, D$.

In both cases $FD^D(\underline{\alpha}, \underline{p}, \tau) = \mathcal{D}^D(\underline{\alpha})$.

For the proof see the Appendix. □

As a consequence, to make the model identifiable we have to adjust the parameter space so as to exclude one of the two cases. We shall address this problem after having discussed the dependence properties and their relation with the parameters. For the time being we keep both cases.

Let us now focus on independence properties for partitions of order 1 as described in Section 3. All such properties can be derived from the following characterization of the joint distribution of $(\underline{S}_1, \underline{S}_2, T)$.

Proposition 2 If $\underline{X} \sim FD^D(\underline{\alpha}, \underline{p}, \tau)$ then the joint distribution of $(\underline{S}_1, \underline{S}_2, T)$ can be expressed as follows:

$$T \sim FD^2(\alpha_1^+, \alpha_2^+, p_1^+, p_2^+, \tau)$$

and the distribution function $F_{\underline{S}_1,\underline{S}_2|T=t}(\underline{s}_1, \underline{s}_2)$ of $(\underline{S}_1, \underline{S}_2 \mid T = t)$ is

$$FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) \mathcal{D}^{D-k}\left(\underline{s}_2; \underline{\alpha}_2\right) p(t) + \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right) FD^{D-k}\left(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau\right)(1 - p(t)) \quad (15)$$

where $FD^k\left(\underline{w}; \underline{\alpha}, \underline{p}, \tau\right)$ denotes the distribution function of a Flexible Dirichlet and the weight $p(t)$ is given by

$$p(t) = \frac{p_1^+}{p_1^+ + p_2^+ \left(\frac{1-t}{t}\right)^{\tau} \frac{\Gamma(\alpha_1^+ + \tau)\Gamma(\alpha_2^+)}{\Gamma(\alpha_2^+ + \tau)\Gamma(\alpha_1^+)}}.$$

For the proof see the Appendix. □

Let us now examine when the independence properties of Table 1 hold for the Flexible Dirichlet. To avoid trivial independence relations suppose in the following that $\underline{S}_i$ contains at least two components, i.e. $k \geq 2$ and $D-k \geq 2$.

Property 9 (Subcompositional independence) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ has subcompositional independence, i.e. $\underline{S}_1 \perp \underline{S}_2$, if and only if at least one of the following conditions is satisfied

1. $\underline{X} \sim \mathcal{D}^D(\alpha)$;

2.a $p_1^+ = 0$;

2.b $p_2^+ = 0$;

3.a $\tau = 1$ and $\frac{p_i}{p_1^+} = \frac{\alpha_i}{\alpha_1^+}$, $(i = 1, \ldots k)$;

3.b $\tau = 1$ and $\frac{p_i}{p_2^+} = \frac{\alpha_i}{\alpha_2^+}$, $(i = k+1, \ldots D)$.

Proof

By marginalization of the distribution given in Proposition 2, the joint distribution function of $(\underline{S}_1, \underline{S}_2)$ can be written in the form

$$FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) \mathcal{D}^{D-k}\left(\underline{s}_2; \underline{\alpha}_2\right) p_1^+ + \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right) FD^{D-k}\left(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau\right)(1-p_1^+).$$

It follows that $\underline{S}_1 \perp \underline{S}_2$ if and only if at least one of the following four conditions is satisfied: $FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) = \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right)$, $\mathcal{D}^{D-k}\left(\underline{s}_2; \underline{\alpha}_2\right) = FD^{D-k}\left(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau\right)$, $p_1^+ = 0$, $p_1^+ = 1$. The last two conditions coincide with cases 2.a and 2.b. Thus suppose now $p_1^+ > 0$ and $p_2^+ > 0$. Then by Proposition 1 we have $FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) = \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right)$ if and only if $\tau = 0$ or condition 3.a is satisfied. Similarly, $\mathcal{D}^{D-k}\left(\underline{s}_2; \underline{\alpha}_2\right) = FD^{D-k}\left(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau\right)$ if and only if $\tau = 0$ or condition 3.b holds. □

Property 10 (Neutrality on the left) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ is neutral on the left, i.e. $\underline{S}_1 \perp (\underline{S}_2, T)$, if and only if at least one among conditions **1.**, **2.a**, **2.b** and **3.a** of Property 9 is satisfied.

Proof

First notice that we can have left neutrality if and only if the conditional distribution of $(\underline{S}_1, \underline{S}_2 \mid T = t)$ factorizes as

$$F_{\underline{S}_1, \underline{S}_2 | T=t}(\underline{s}_1, \underline{s}_2) = F_{\underline{S}_1}(\underline{s}_1) F_{\underline{S}_2 | T=t}(\underline{s}_2).$$

Inspection of expression (15) shows that this can happen if and only if $FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) = \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right)$ or $p(t) = 0$ or $p(t) = 1$ $\forall t \in (0,1)$. The result then follows from the proof of Property 9 as $p(t)$ is identically equal to zero if and only if $p_1^+ = 0$ and it is identically equal to one if and only if $p_1^+ = 1$. □

Notice that Property 10 allows to derive conditions for independence between $\underline{X}_2$ and the normalized version of $\underline{X}_1$ (see the conditional distribution given in Property 4). More precisely, such independence is equivalent to neutrality on the left as there is a one-to-one correspondence between $\underline{X}_2$ and $(\underline{S}_2, T)$.

In complete analogy conditions for right neutrality can be obtained.

Property 11 (Neutrality on the right) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ is neutral on the right, i.e. $\underline{S}_2 \perp (\underline{S}_1, T)$, if at least one among conditions **1.**, **2.a**, **2.b** and **3.b** of Property 9 is satisfied. □

Property 12 (Subcompositional invariance) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ has subcompositional invariance, i.e. $(\underline{S}_1, \underline{S}_2) \perp T$, if at least one of the following conditions is satisfied

1. $\underline{X} \sim \mathcal{D}^D(\alpha)$;

2.a $p_1^+ = 0$;

2.b $p_2^+ = 0$;

3.a $\tau = 1$ and $\frac{p_i}{p_1^+} = \frac{\alpha_i}{\alpha_1^+}$, $(i = 1, \ldots k)$ and **3.b** $\tau = 1$ and $\frac{p_i}{p_2^+} = \frac{\alpha_i}{\alpha_2^+}$, $(i = k+1, \ldots D)$.

Proof

Expression (15) does not depend on t if and only if $p(t)$ is constant (for all $t \in (0,1)$) or $FD^k\left(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau\right) = \mathcal{D}^k\left(\underline{s}_1; \underline{\alpha}_1\right)$ and $\mathcal{D}^{D-k}\left(\underline{s}_2; \underline{\alpha}_2\right) = FD^{D-k}\left(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau\right)$. The former condition is satisfied if and only if either $\tau = 0$ or $p_1^+ = 0$ or $p_2^+ = 0$. The latter condition is fulfilled if and only if $\tau = 0$ or both 3.a and 3.b hold. □

Property 13 (Partition independence) The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ has partition independence, i.e. $\underline{S}_1 \perp \underline{S}_2 \perp T$, if and only if it has subcompositional invariance.

Proof

Clearly partition independence always implies subcompositional invariance. By using formula (15) it is then easy to check that if any one of the four conditions in Property 12 holds, then $\underline{S}_1$, $\underline{S}_2$ and T are independent. □

The above independence properties can be analyzed and restated in a more compact form by examining the marginal distributions of $\underline{S}_1$, $\underline{S}_2$ and T. More precisely, from Property 6 (amalgamation) and Property 7 (subcomposition) such distributions can be directly specified for the various cases listed above:

1. $\underline{S}_1 \sim \mathcal{D}^k(\underline{\alpha}_1)$, $\underline{S}_2 \sim \mathcal{D}^{D-k}(\underline{\alpha}_2)$ and $T \sim \mathcal{D}^2(\alpha_1^+, \alpha_2^+)$;

2.a $\underline{S}_1 \sim \mathcal{D}^k(\underline{\alpha}_1)$, $\underline{S}_2 \sim FD^{D-k}(\underline{\alpha}_2, \underline{p}_2/p_2^+, \tau)$ and $T \sim \mathcal{D}^2(\alpha_1^+, \alpha_2^+ + \tau)$;

2.b $\underline{S}_1 \sim FD^k(\underline{\alpha}_1, \underline{p}_1/p_1^+, \tau)$, $\underline{S}_2 \sim \mathcal{D}^{D-k}(\underline{\alpha}_2)$ and $T \sim \mathcal{D}^2(\alpha_1^+ + \tau, \alpha_2^+)$;

3.a $\underline{S}_1 \sim \mathcal{D}^k(\underline{\alpha}_1)$, $\underline{S}_2 \sim p_2^+ FD^{D-k}(\underline{\alpha}_2, \underline{p}_2/p_2^+, 1) + (1-p_2^+)\mathcal{D}^{D-k}(\underline{\alpha}_2)$ and $T \sim FD^2(\alpha_1^+, \alpha_2^+, p_1^+, p_2^+, 1)$;

3.b $\underline{S}_1 \sim p_1^+ FD^k(\underline{\alpha}_1, \underline{p}_1/p_1^+, 1) + (1-p_1^+)\mathcal{D}^k(\underline{\alpha}_1)$, $\underline{S}_2 \sim \mathcal{D}^{D-k}(\underline{\alpha}_2)$ and $T \sim FD^2(\alpha_1^+, \alpha_2^+, p_1^+, p_2^+, 1)$.

Thus, careful inspection of the above distributions leads to the following concise characterization of the independence properties.

Corollary Consider the three marginal distributions of $\underline{S}_1$, $\underline{S}_2$ and T. The Flexible Dirichlet $FD^D(\underline{\alpha}, \underline{p}, \tau)$ has:

• partition independence (or subcompositional invariance) if and only if at least two such marginals are Dirichlet;

• neutrality on the left if and only if at least two such marginals are Dirichlet or $\underline{S}_1$ is a Dirichlet;

• neutrality on the right if and only if at least two such marginals are Dirichlet or $\underline{S}_2$ is a Dirichlet;

• subcompositional independence if and only if at least one of the two marginals $\underline{S}_1$ and $\underline{S}_2$ are Dirichlet. □

Let us consider the implications of such independence results in terms of choices of the parameters of the model.

Setting $\tau = 1$ and the α_i's proportional to the p_i's for a given group (say A) of components gives rise to a first type of asymmetric independence between the group A and the group B of the remaining variables: the subcomposition relative to group A is independent of the vector of the components of group B.

Equating to zero the p_i's of group A leads to a stronger and symmetric independence between the two groups: not only do we have independence between the subcomposition relative to group A and the remaining components, but conversely we also get independence between the subcomposition formed by the components of group B and the components of group A.

To shed some light on independence relations for higher order partitions, let us consider the case of subcompositional invariance for a partition of order 2, case which can be easily extended to higher order partitions.

Property 14 Let $\underline{X} \sim FD^D(\underline{\alpha}, \underline{p}, \tau)$ be partitioned into three subsets as $X_1, \dots, X_{a_1} | X_{a_1+1}, \dots, X_{a_2} | X_{a_2+1}, \dots, X_D$ and denote by $\underline{S}_i$ $(i = 1, 2, 3)$ the corresponding subcompositions (see Section 2). To avoid trivial independences suppose each subset contains at least two components (i.e. $a_1 \geq 2$, $a_2 - a_1 \geq 2$, $D - a_2 \geq 2$). Then $\underline{X}$ is subcompositional invariant if and only if at least one of the following conditions holds

1. $\underline{X} \sim \mathcal{D}^D(\underline{\alpha})$;
2. two out of the three p_i^+'s are equal to zero;
3. $\tau = 1$ and in at least two of the three subsets the α_i's are proportional to the p_i's;
4. $\tau = 1$, one of the three subsets has all p_i's equal to zero and one of the other two has proportionality among the α_i's and the p_i's.

Equivalently, $\underline{X}$ is subcompositional invariant if and only if at least two of the $\underline{S}_i$'s have a Dirichlet distribution.

Proof

Clearly if $\tau = 0$ we have joint independence. Thus suppose $\tau > 0$. Let $\underline{Z}_1$, $\underline{Z}_2$ and $\underline{Z}_3$ be defined according to the above partition and denote by Z_i^+ the corresponding partial sums $(i = 1, 2, 3)$. By conditioning on the Z_i^+'s we obtain

$$F_{\underline{S}_1,\underline{S}_2,\underline{S}_3}(\underline{s}_1, \underline{s}_2, \underline{s}_3) = \sum_{i=1}^{3} F_{\underline{S}_1,\underline{S}_2,\underline{S}_3|Z_i^+=1}(\underline{s}_1, \underline{s}_2, \underline{s}_3) Pr(Z_i^+ = 1) =$$

$$= FD^{a_1}(\underline{s}_1; \underline{\alpha}_1, \underline{p}_1/p_1^+, \tau)\mathcal{D}^{a_2-a_1}(\underline{s}_2; \underline{\alpha}_2)\mathcal{D}^{D-a_2}(\underline{s}_3; \underline{\alpha}_3)p_1^+ +$$

$$+\mathcal{D}^{a_1}(\underline{s}_1; \underline{\alpha}_1)FD^{a_2-a_1}(\underline{s}_2; \underline{\alpha}_2, \underline{p}_2/p_2^+, \tau)\mathcal{D}^{D-a_2}(\underline{s}_3; \underline{\alpha}_3)p_2^+ +$$

$$+\mathcal{D}^{a_1}(\underline{s}_1; \underline{\alpha}_1)\mathcal{D}^{a_2-a_1}(\underline{s}_2; \underline{\alpha}_2)FD^{D-a_2}(\underline{s}_3; \underline{\alpha}_3, \underline{p}_3/p_3^+, \tau)p_3^+.$$

Suppose first no p_i^+'s are equal to zero. Then $\underline{S}_1$, $\underline{S}_2$ and $\underline{S}_3$ are independent if and only if for at least a couple of them, say $\underline{S}_j$ and $\underline{S}_{j'}$, the conditional distributions given Z_i^+'s $(i = 1, 2, 3)$ are the same (Dirichlet). But this is possible if and only if $\tau = 1$ and the α_i's are proportional to the p_i's within the subsets j and j'. Furthermore in this case $\underline{S}_1 \perp \underline{S}_2 \perp \underline{S}_3$ whatever the p_i^+'s may be. Suppose now one of the p_i^+'s, say p_j^+, is equal to zero. Then we have independence if and only if for at least one of the other two subsets the conditional distributions are identical (Dirichlet), i.e. if and only if $\tau = 1$ and at least one of the other two subsets has the proportionality property. Finally, suppose only one p_i^+ is positive. Then $\underline{S}_1 \perp \underline{S}_2 \perp \underline{S}_3$ always holds. □

After having examined the various forms of independence, let us come back to the identifiability issue raised after Proposition 1. In order to make the Flexible Dirichlet model identifiable we need to exclude either the case $\tau = 0$ or the case $\tau = 1$ and complete proportionality among the α_i's and the p_i's. The exclusion of the first case, namely $\tau = 0$, seems to be preferred as this is a boundary point of the parameter space, thus leading to less convenient inferential procedures. Moreover, the case $\tau = 1$ and complete proportionality can be interpreted as an extreme case of various types of partial proportionality which give rise to (partial) forms of independence as shown above. Notice that if we exclude the value $\tau = 0$ we can drop case 1. in all the above properties as the situation where the Flexible Dirichlet coincides with the Dirichlet distribution is included both in condition 3.a and condition 3.b.

6 Conclusions

The Flexible Dirichlet model has been originated by normalizing a correlated basis of D Gamma mixtures depending on $2D$ parameters. It contains the Dirichlet distribution as an inner point of the parameter space, but it also allows for a more general variance-covariance structure and various forms of independence/dependence for compositional data. Therefore the Flexible Dirichlet may accommodate both for the most extreme form of independence for unit-sum constrained r.v.s (Dirichlet case) and for various weaker independence relations. It keeps several good mathematical properties of the Dirichlet distribution such as “essential” closure under marginalization, conditioning, permutation, amalgamation and subcomposition, the derived distributions displaying quite simple expressions.

Suitable simple parameter configurations allow to model subcompositional independence, neutrality as well as partition independence. Such forms of partial independence can be characterized through the marginal distributions of subcompositions and amalgamation. In particular they quite naturally occur when one or more of the marginals are Dirichlet.

The Flexible Dirichlet allows to account for other special situations such as essential zeroes and multimodality. Let us just briefly comment on these issues. The former can be modeled by equating to zero some α_i's. For example, if we wish to allow the i^{th} component not to be present we can set $\alpha_i = 0$. In such a case $p_i > 0$ assumes the meaning of probability that $X_i > 0$. Multimodality derives from the general mixture representation (8). In particular bimodality can be easily obtained for the generic one dimensional marginal X_i as its distribution is a mixture of two Betas (see (9)). More generally, by Property 3 one has that k-dimensional marginals can display $k+1$-modality.

As far as inferential properties of the model are concerned, we investigated the estimation issue (Migliorati et al., 2008). In particular, the maximum likelihood estimates of the parameters can be found by means of a suitable adaptation of the E-M algorithm which considers as missing data the values of the Z_i's (see expression (3)). The choice of the starting values for the algorithm deserves special attention and a solution which combines the k-means clustering algorithm for estimating the p_i's and a two-step method of moments for τ and $\underline{\alpha}$ has proved fruitful. Moreover distributions and properties of the estimators together with their variance estimators have been analyzed through simulation studies.

Further analysis of inferential aspects is needed, in particular about hypothesis testing relative to the various forms of independence. Moreover, an analysis of the Flexible Dirichlet from a Bayesian point of view is also desirable and promising as it is easily seen to be conjugate with respect to the Multinomial model. We plan to tackle these issues in future work.

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data (with discussion). *Journal of the Royal Statistical Society B*(44), pp. 139–177.

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. London: The Blackburn Press.

Aitchison, J. and Shen, S.M. (1980). Logistic-Normal Distributions: some Properties and Uses. *Biometrika*(67), pp. 261–272.

Connor, J.R. and Mosimann, J.E. (1969). Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution. *Journal of the American Statistical Association*(64), pp. 194–206.

Migliorati, S., Monti, G.S., Ongaro, A. (2008). E–M Algorithm: an Application to a Mixture Model for Compositional Data. *Proceedings of the 44^{th} Scientific Meeting of the Italian Statistical Society*.

Rayens, W.S., Srinivasan, C. (1994). Dependence Properties of Generalized Liouville Distributions on the Simplex. *Journal of the American Statistical Association 89*(428), pp. 1465–1470.

Appendix

Proof of Property 4

Expression (13) can be proved by partitioning the vector $\underline{Z} = (Z_1, \ldots, Z_k | Z_{k+1}, \ldots, Z_D) = (\underline{Z}_1, \underline{Z}_2)$ which enters definition (3) of the basis and conditioning on $\sum_{i=1}^{k} Z_i = Z_1^+$. More precisely, we write the conditional distribution of $\underline{S}_1 \mid \underline{X}_2 = \underline{x}_2$ as

$$F_{\underline{S}_1|\underline{X}_2=\underline{x}_2}(\underline{s}_1) = F_{\underline{S}_1|\underline{X}_2=\underline{x}_2, Z_1^+=1}(\underline{s}_1) Pr\left(Z_1^+ = 1 \mid \underline{X}_2 = \underline{x}_2\right) +$$

$$+ F_{\underline{S}_1|\underline{X}_2=\underline{x}_2, Z_1^+=0}(\underline{s}_1) Pr\left(Z_1^+ = 0 \mid \underline{X}_2 = \underline{x}_2\right)$$

where $F_{Z|W=w}$ denotes the conditional distribution function of Z given $W = w$. It is easy to check that, conditionally on $Z_1^+ = 1$, $\underline{X}$ is distributed as $FD^D(\underline{\alpha}, (\underline{p}_1/p_1^+, \underline{0}), \tau)$. From this one can directly compute the distribution of $\underline{X}_1 \mid \left(\underline{X}_2 = \underline{x}_2, Z_1^+ = 1\right)$ and then normalize it by dividing each component by $1 - X_2^+$ to obtain

$$\frac{\underline{X}_1}{1 - x_2^+} \mid \left(\underline{X}_2 = \underline{x}_2, Z_1^+ = 1\right) \sim \underline{S}_1 \mid \left(\underline{X}_2 = \underline{x}_2, Z_1^+ = 1\right) \sim FD^k\left(\underline{\alpha}_1, \frac{\underline{p}_1}{p_1^+}, \tau\right).$$

The factor $Pr\left(Z_1^+ = 1 \mid \underline{X}_2 = \underline{x}_2\right)$ can be computed as

$$\frac{f_{\underline{X}_2|Z_1^+=1}(\underline{x}_2) Pr\left(Z_1^+ = 1\right)}{f_{\underline{X}_2}(\underline{x}_2)}$$

where $f_{Z|W=w}$ denotes the conditional distribution density function of Z given $W = w$. It is then straightforward to see that, conditionally on $Z_1^+ = 1$, $(\underline{X}_2, 1 - X_2^+)$ is distributed as $\mathcal{D}^{D-k+1}\left(\underline{\alpha}_2, \alpha^+ + \tau - \alpha_2^+\right)$ whereas unconditionally by the marginal property 3 $(\underline{X}_2, 1 - X_2^+) \sim FD^{D-k+1}\left(\underline{\alpha}_2, \alpha^+ - \alpha_2^+, \underline{p}_2, 1 - p_2^+, \tau\right)$. Some algebraic computation leads then to expression (14).

Finally, by analogous arguments one can prove that $\underline{S}_1 \mid \left(\underline{X}_2 = \underline{x}_2, Z_1^+ = 0\right) \sim \mathcal{D}^k\left(\underline{\alpha}_1\right)$. □

Proof of Proposition 1

Sufficiency

Condition 1. is obvious by the definition of the Flexible Dirichlet and condition 2. can be easily checked by examining (10) if all $\alpha_i > 0$. If some α_i's (and therefore some p_i's) are equal to zero then the corresponding random components are degenerate at zero. The remaining components are easily seen to be Dirichlet distributed. It follows that the whole vector is Dirichlet with some degenerate components.

Necessity

We shall prove that if neither condition 1. nor condition 2. is satisfied, then the Flexible Dirichlet is not a Dirichlet. So we can assume $\tau > 0$.

Suppose first that the α_i's are all strictly positive. Careful inspection of (10) shows that the Flexible Dirichlet coincides with the Dirichlet if and only if $\sum_{i=1}^{D} p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+\tau)} x_i^{\tau}$ is constant. This can happen only if $\tau = 1$ and $p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+\tau)} = k$ ($k \in \Re^+$), which is equivalent to $\tau = 1$ and $\frac{p_i}{\alpha_i} = k$. Notice that $\sum_{i=1}^{D} p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+\tau)} x_i^{\tau}$ can not be proportional to x_i^{τ} for some i as no p_i's is allowed to be equal to one (see Section 4.1).

Suppose now some α_i's are equal to zero, say $\alpha_{i_1}, \ldots, \alpha_{i_k}$. If even a single one of the corresponding p_i is positive (say p_{i_1}) then the component Y_{i_1} is zero with probability p_{i_1}. It follows that in this case $\underline{X}$ can not have a Dirichlet distribution. Therefore $p_{i_1}, \ldots, p_{i_k}$ must be all equal to zero, the corresponding components being degenerate at zero. Notice that $k \leq D-2$ since the p_i's have sum one and each p_i is strictly less than one. The remaining $D-k \geq 2$ components have a density of the form (10) which is a Dirichlet density only if $\tau = 1$ and $\frac{p_i}{\alpha_i} = k$. □

Proof of Proposition 2

The distribution of T derives from Property 6.

The conditional distribution (15) is derived by conditioning on $Z_1^+ = \sum_{i=1}^{k} Z_i$, $\underline{Z}$ being the Multinomial vector entering the generating basis $\underline{Y}$ (see (3)). Thus we obtain

$$F_{\underline{S}_1,\underline{S}_2|T=t}(\underline{s}_1,\underline{s}_2) = F_{\underline{S}_1,\underline{S}_2|T=t,Z_1^+=1}(\underline{s}_1,\underline{s}_2) Pr\left(Z_1^+ = 1 \mid T = t\right) +$$

$$+F_{\underline{S}_1,\underline{S}_2|T=t,Z_1^+=0}(\underline{s}_1,\underline{s}_2) Pr\left(Z_1^+ = 0 \mid T = t\right).$$

Let us consider $F_{\underline{S}_1,\underline{S}_2|T=t,Z_1^+=1}(\underline{s}_1,\underline{s}_2)$: we shall prove that $\underline{S}_1$, $\underline{S}_2$ and T are independent conditionally on $Z_1^+ = 1$.

Thus in the following let us suppose $Z_1^+ = 1$ so that all the distributions are to be interpreted conditionally on it. By definition $\underline{Y}_1 = (Y_1, \ldots, Y_k) \perp \underline{Y}_2 = (Y_{k+1}, \ldots, Y_D)$. Define $\underline{\tilde{Y}}_1 = \underline{Y}_1 / Y_1^+$ and $\underline{\tilde{Y}}_2 = \underline{Y}_2 / Y_2^+$ and notice that $\underline{\tilde{Y}}_1 = \underline{S}_1 \sim FD^k(\underline{\alpha}_1, \underline{p}_1/p_1^+, \tau)$ and $\underline{\tilde{Y}}_2 = \underline{S}_2 \sim \mathcal{D}^{D-k}(\underline{\alpha}_2)$. As $\underline{\tilde{Y}}_1 \perp Y_1^+$ (by compositional invariance of the Flexible Dirichlet) and clearly $(\underline{\tilde{Y}}_1, Y_1^+) \perp \underline{Y}_2$, it follows that $\underline{\tilde{Y}}_1 \perp Y_1^+ \perp \underline{Y}_2$. This, in particular, implies that

$$\underline{\tilde{Y}}_1 \perp (\underline{\tilde{Y}}_2, Y_2^+, Y_1^+). \tag{16}$$

In an analogous way, as $\underline{\tilde{Y}}_2 \perp Y_2^+$ (by compositional invariance of the Dirichlet), one can prove that

$$\underline{\tilde{Y}}_2 \perp (\underline{\tilde{Y}}_1, Y_1^+, Y_2^+). \tag{17}$$

Independence relations (16) and (17) imply that $\underline{\tilde{Y}}_1 \perp \underline{\tilde{Y}}_2 \perp (Y_1^+, Y_2^+)$ and therefore that

$$\underline{\tilde{Y}}_1 \perp \underline{\tilde{Y}}_2 \perp \frac{Y_1^+}{(Y_1^+, Y_2^+)}$$

which coincides with $\underline{S}_1, \underline{S}_2, T$. An analogous argument shows that, conditionally on $Z_1^+ = 0$, $\underline{S}_1 \perp \underline{S}_2 \perp T$ with $\underline{\tilde{Y}}_1 = \underline{S}_1 \sim \mathcal{D}^k(\underline{\alpha}_1)$ and $\underline{\tilde{Y}}_2 = \underline{S}_2 \sim FD^{D-k}(\underline{\alpha}_2, \underline{p}_2/p_2^+, \tau)$.

Finally, the weight $p(t) = Pr\left(Z_1^+ = 1 \mid T = t\right)$ can be computed by applying Bayes formula and noting that $T \mid Z_1^+ = 1 \sim \mathcal{D}^2(\alpha_1^+ + \tau, \alpha_2^+)$. □