Journal of Multivariate Analysis 129 (2014) 1–15

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Dirichlet distribution through neutralities with respect to two partitions

Agata Sakowicz *, Jacek Wesołowski

Politechnika Warszawska, Wydział Matematyki i Nauk Informacyjnych, ul. Koszykowa 75, 00-662 Warsaw, Poland

ARTICLE INFO

Article history:
Received 15 November 2013
Available online 18 April 2014

AMS 2010 subject classifications:
62H05
62E10
39B22

Keywords:
Neutrality with respect to partition
Dirichlet distribution
Method of moments
Characterization

ABSTRACT

Different concepts of neutrality have been studied in the literature in context of independence properties of vectors of random probabilities, in particular, for Dirichlet random vectors. Some neutrality conditions led to characterizations of the Dirichlet distribution. In this paper we provide a new characterization in terms of neutrality with respect to two partitions, which generalizes previous results. In particular, no restrictions on the size of the vector of random probabilities are imposed. In the proof we enhance the moments method approach proposed in Bobecka and Wesołowski (2009) [2] by combining it with some graph theoretic techniques.

1. Introduction

Let $\underline{X} = (X_1, \dots, X_n)$ be a random vector of probabilities, i.e. $X_i \geq 0$, $i = 1, \dots, n$, and $\sum_{i=1}^n X_i = 1$. A concept of neutrality was first introduced for such vectors by Connor and Mosimann in [5]. They indicated that given a vector of random probabilities $\underline{X} = (X_1, \dots, X_n)$ it is desirable in some situations to eliminate one of the proportions, say X_1, and to analyse its effects on proportions of the form $X_2/(1-X_1), \dots, X_n/(1-X_1)$. This led to the following definition (see [5]): X_1 is neutral in $\underline{X}$ whenever X_1 and the vector $(X_2/(1-X_1), \dots, X_n/(1-X_1))$ are independent. These authors defined also neutrality of a subvector in a random vector of proportions and complete neutrality of a vector. Similar notions of neutrality to the right and neutrality to the left were defined in Doksum [7]. There were also other related notions of neutrality studied in the literature. All these notions embed in the notion of neutrality with respect to partition of an index set (introduced in [2]) which we recall below.

We say that $\pi = \{P_1, \dots, P_K\}$ is a partition of a set E when $P_1, \dots, P_K$ are nonempty pairwise disjoint subsets of E, whose union is E. The elements of π are called blocks.

Definition 1.1. Let $\pi = \{P_1, \dots, P_K\}$ be a partition of $E = \{1, \dots, n\}$. We say that a vector of random probabilities $\underline{X} = (X_1, \dots, X_n)$ is neutral with respect to π (from here abbreviated nwrt π) if the following random vectors are mutually independent:

$$U = \left(\sum_{i \in P_1} X_i, \dots, \sum_{i \in P_K} X_i\right),$$

$$W_{P_1} = \left(\frac{X_j}{\sum_{i \in P_1} X_i}, j \in P_1\right), \dots, W_{P_K} = \left(\frac{X_j}{\sum_{i \in P_K} X_i}, j \in P_K\right).$$

* Corresponding author.
E-mail addresses: A.Sakowicz@mini.pw.edu.pl (A. Sakowicz), wesolo@mini.pw.edu.pl (J. Wesołowski).

http://dx.doi.org/10.1016/j.jmva.2014.04.004

The notion of neutrality appeared to be a useful tool in studying independence properties of the Dirichlet distribution. In particular, the Dirichlet distribution, which can be imposed on a vector of random probabilities $\underline{X}$, is neutral with respect to all possible partitions of the corresponding index set. For a recent accounts on Dirichlet distributions, including relations to neutrality concepts see e.g. Ng, Tian and Tang [13] (in particular, Ch. 2.6) or Chang, Gupta and Richards [3].

Recall that a random vector $\underline{X} = (X_1, \dots, X_{n-1})$ has the Dirichlet distribution $\mathrm{Dir}(\alpha_1, \dots, \alpha_n)$ if its density is of the form

$$f(x_1, \dots, x_{n-1}) = \frac{\Gamma\left(\sum_{i=1}^{n} \alpha_i\right)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n-1} x_i^{\alpha_i - 1} \left(1 - \sum_{i=1}^{n-1} x_i\right)^{\alpha_n - 1} \mathbb{1}_{T_{n-1}}(x_1, \dots, x_{n-1}),$$

where $\alpha_i > 0, i = 1, \dots, n$, and $T_{n-1} = \{(x_1, \dots, x_{n-1}) : x_i > 0, \sum_{i=1}^{n-1} x_i < 1\}$. In the sequel we will say that a vector of random probabilities $\underline{X} = (X_1, \dots, X_n)$ has the Dirichlet distribution if a subvector $(X_1, \dots, X_{n-1})$ has the density given by the above formula.

Characterizations of the Dirichlet distribution by different independence assumptions related to neutralities were discussed by several authors. All these results can be formulated in terms of neutrality with respect to partitions, although this notion had not been explicitly referred to. Darroch and Ratcliff proved in [6] a characterization of the Dirichlet distribution, using neutralities with respect to partitions $\pi_i = \{\{1\}, \dots, \{i-1\}, \{i+1\}, \dots, \{n-1\}, \{i, n\}\}, i = 1, \dots, n-1$. A result by Fabius, [8], concerned partitions $\pi_i = \{\{i\}, \{1, \dots, i-1, i+1, \dots, n\}\}, i = 1, \dots, n-1$. James and Mosimann presented in [10] a characterization by neutrality with respect to partitions $\pi_i = \{\{1\}, \dots, \{i\}, \{i+1, \dots, n\}\}, i = 1, \dots, n-2$, and $\pi_{n-1} = \{\{n-1\}, \{1, \dots, n-2, n\}\}$. Their result was further generalized in [1] by Bobecka and Wesołowski, where partitions $\pi_i = \{\{1\}, \dots, \{i\}, \{i+1, \dots, n\}\}, i = 1, \dots, n-2$, and $\pi_{n-1} = \{\{i_0\}, \{i_0+1\}, \dots, \{n-1\}, \{1, \dots, i_0-1, n\}\}$ for an arbitrary fixed i_0 were considered. Note that for a vector of size n all of these characterizations require exactly $n-1$ partitions. Another result, requiring only 2 partitions for any n not being a prime number, was presented in [9] by Geiger and Heckerman. These authors proved, assuming the existence of a density, a characterization of an $L \times M$ Dirichlet random matrix, with one partition determined by its rows and another one by its columns. The proof was based on solving a functional equation for densities (further developed by Járai in [11], see also Chapter 23 of [12]). See also Heckerman, Geiger and Chickering [4] for a thorough description of the Bayesian networks context of this characterization. Bobecka and Wesołowski, [2], refined this result proving an analogous characterization by means of moments method and, consequently, without additional density assumption. They also generalized it to multi-way tables. The result of Geiger and Heckerman (and its extension) has been also recently proved within the Bayesian framework by Ramamoorthi and Sangalli in [15].

The aim of this paper is to present a new characterization of the Dirichlet distribution, which generalizes all the previous results when neutrality with respect to only two partitions is assumed. Actually we determine a set of all pairs of partitions such that neutrality with respect to both elements of the pair characterizes the Dirichlet distribution for the vector of random probabilities. We use the moments method as in [2], hence no density assumption is needed. In the proof we also rely heavily on graph theoretic techniques.

The paper is organized as follows. Section 2 contains some definitions and facts from graph theory which are used in the proof of the main result. In Section 3 we state and prove the characterization of the Dirichlet law, which is our main result. In the proof we use also an auxiliary result on a functional equation which is formulated in Section 3 and proved in the Appendix. In Section 4 we illustrate the characterization with several examples.

2. Facts from graph theory

In this section we present some definitions and facts from graph theory that will be used in the proof of our main result.

Definition 2.1. Let $\mathcal{G} = (V, E)$, where V is the set of vertices and E is the set of edges, be a connected graph. The vertex $v \in V$ is called a cut vertex if removing v from $\mathcal{G}$ disconnects the graph. Otherwise, we say that v is a non-cut vertex. Alternatively, v is non-cut if for any $u, w \in V \setminus \{v\}$ there exists a path between u and w that does not contain v.

Below we state a fact on non-cut vertices, which belongs to the graph theory folklore. Since we were not able to find an exact reference, a short proof is also given.

Lemma 2.2. *In every connected graph $\mathcal{G} = (V, E)$, $|V| \geq 2$, there exist at least two non-cut vertices.*

Proof. Let $\mathcal{T}$ denote a spanning tree of the graph $\mathcal{G}$. Since $|V| \geq 2$, there exist at least two leaves u, v in $\mathcal{T}$. As the removal of leaves does not disconnect the tree, u, v are non-cut in $\mathcal{T}$, and hence they are non-cut in $\mathcal{G}$. □

For the purpose of this paper we introduce a notion of significance of a vertex.

Definition 2.3. Let C be a maximal clique in a graph $\mathcal{G}$ and $v \in C$. Denote by $N(v)$ the set of neighbours of v. We say that v is significant in C if $N(v) \cup \{v\} = C$.

Below we give some properties of significant vertices which are important in the proof of the characterization in Section 3.

Lemma 2.4. *Let C be a maximal clique in a connected graph $\mathcal{G}$ and $v \in C$ be significant in C. Then v is non-cut in $\mathcal{G}$.*

Proof. If C consists of two elements, the significance of v in C means that v is a leaf and hence non-cut. Assume that C has more than two elements. Then $N(v)$ contains at least two elements, and all of them are connected. Thus if between some vertices u, w of $\mathcal{G}$ there exists a path containing v, we can modify its part contained in C by replacing v with some of its neighbours. Hence v is non-cut in $\mathcal{G}$. □

Definition 2.5. Let $\mathcal{G} = (V, E)$. We say that a subgraph $\widetilde{\mathcal{G}} = (W, \widetilde{E})$ of $\mathcal{G}$ is induced by the set of vertices W (denote $\widetilde{\mathcal{G}} = G[W]$) if $W \subset V$ and $\widetilde{E}$ consists of all the edges from E whose endpoints are both in W.

Lemma 2.6. *Let $\mathcal{G} = (V, E)$ be a connected graph and $\widetilde{C} = \{C_1, \ldots, C_k\}$ be any family of maximal cliques in $\mathcal{G}$ such that for every $i = 1, \ldots, k$ there exists at least one vertex $v_i \in C_i$ significant in C_i, and C_i consists of at least 3 elements. Let $\widetilde{\mathcal{G}} = G[V \setminus V']$, where $V' = \{v_i,\ i = 1, \ldots, k\}$, denote a subgraph of $\mathcal{G}$ induced by the set of vertices $V \setminus V'$. If a vertex is non-cut in $\widetilde{\mathcal{G}}$ then it is also non-cut in $\mathcal{G}$.*

Proof. Note first that $\widetilde{\mathcal{G}}$ is connected. Indeed, by Lemma 2.4 removal of any vertex $v_i \in V'$ does not disconnect $\mathcal{G}$, and the remaining vertices $V' \setminus \{v_i\}$ in $G[V \setminus \{v_i\}]$ are still significant in the corresponding cliques. Hence removing all of the vertices V', we do not disconnect $\mathcal{G}$. Suppose that there exists w which is non-cut in $\widetilde{\mathcal{G}}$, but it is cut in $\mathcal{G}$. We will show that it leads to a contradiction. Removing w from $\mathcal{G}$, we divide it into connectivity components with vertices in sets $V_1, \ldots, V_k$. Since w is non-cut in $\widetilde{\mathcal{G}}$, it follows that $V \setminus V' \subset V_i \cup \{w\}$ for some i, $1 \le i \le k$. Consequently, no vertex $u \in \bigcup_{j \ne i} V_j \subset V'$ can be adjacent in $\mathcal{G}$ to any – except for w – vertex from $V \setminus V'$. This means that either u is a leaf (adjacent to w) or u is adjacent to a vertex $y \in V'$. But u cannot be a leaf since the vertices in V' come from cliques of at least 3 elements. Therefore $u, y \in V'$ are adjacent. However, this is impossible since vertices from V' are significant in different cliques. □

3. The characterization

In this section we state and prove a characterization of the Dirichlet distribution which is the main result of this paper. First we present an auxiliary lemma on a solution of a system of functional equations for a function of multivariate variables satisfying a structure condition (3.1). The result of this lemma will be used in the proof of the characterization theorem.

Note that if $(\Theta_1, \ldots, \Theta_n)$ is a random vector such that $\sum_{i=1}^n \Theta_i = 1$, then F defined as $F(\underline{x}) = \mathbb{E}\,(\Theta_1^{x_1} \ldots \Theta_n^{x_n})$ (provided the moments are finite) satisfies the structural condition (3.1).

Lemma 3.1. *Let F be a positive function defined on the n-fold cartesian product of non-negative integers, such that $F(\underline{0}) = 1$ and*

$$F(\underline{x}) = \sum_{l=1}^{n} F(\underline{x} + \underline{\varepsilon}_l), \tag{3.1}$$

where $\underline{\varepsilon}_l$ is a unit vector of length n with 1 at the lth coordinate. Assume that there exist a set $A \subset \{1, \ldots, n\}$, $k = \#A > 1$, functions α_j, $j \in A$, and constants b_j, $j \in \{1, \ldots, n\} \setminus A$, such that for every $i \in A$

$$\frac{F(\underline{x}^A + \underline{\varepsilon}_i)}{F(\underline{x}^A + \underline{\varepsilon}_l)} = \begin{cases} \dfrac{\alpha_i(x_i)}{\alpha_l(x_l)}, & l \in A \quad \text{(a)} \\[2ex] \dfrac{\alpha_i(x_i)}{b_l}, & l \notin A \quad \text{(b)} \end{cases} \tag{3.2}$$

where $\underline{x}^A = (x_i^A)_{i=1}^n$:

$$x_i^A = \begin{cases} x_i, & i \in A \\ 0, & i \notin A \end{cases}.$$

Then there exist $a \in \mathbb{R}$ and $b_i \in \mathbb{R}$, $i \in A$, such that

$$\alpha_i(x) = ax + b_i, \quad i \in A. \tag{3.3}$$

If $a \ne 0$, then

$$F(\underline{x}^A) = \frac{\Gamma\left(|\underline{d}|\right)}{\Gamma\left(|\underline{d}| + |\underline{x}^A|\right)} \prod_{i \in A} \frac{\Gamma(x_i + d_i)}{\Gamma(d_i)},$$

where $|\underline{d}| = \sum_{i=1}^n d_i$ and $d_i = b_i/a$, $i = 1, \ldots, n$.

If $a = 0$, *then*

$$F(\underline{x}^A) = \prod_{i\in A} d_i^{x_i},$$

where $d_i = b_i/|\underline{b}|$.

The above result in the case $A = \{1, \dots, n\}$ can be read out from the proof of Theorem 2 of [2]. Actually, the proof of this result that we give in the Appendix, borrows a lot from the one given in [2].

We denote by $\vee$ and $\wedge$ the standard operations of taking maximum and minimum in the lattice of partitions of a given set.

Now we can state the characterization of the Dirichlet distribution through neutralities.

Theorem 3.2. *Let* $\underline{X} = (X_1, \dots, X_n)$ *be a vector of random non-degenerate probabilities and* π_1, π_2 *be two different partitions of the index set* $E = \{1, \dots, n\}$. *Let* $\widehat{\pi}_i$, $i = 1, 2$, *be a partition created from* π_i *by merging into one block all blocks of size one. Assume that*

$$\pi_1 \vee \pi_2 = \pi^* = \{\{1, \dots, n\}\} \tag{3.4}$$

and

$$\widehat{\pi}_1 \wedge \widehat{\pi}_2 = \pi_* = \{\{1\}, \dots, \{n\}\}. \tag{3.5}$$

If X *is neutral with respect to partitions* π_1 *and* π_2, *then it has the Dirichlet distribution.*

Note that considering $\widehat{\pi}_i$ instead of π_i in (3.5) is necessary to obtain the characterization. The crucial issue is that any block of one partition has to contain at most one block of size one of the other partition. This fact will play an important role in the proof. Also, in Section 4 below we will give an example showing that if (3.5) holds just for π_1 and π_2, there exist other distributions than Dirichlet satisfying all other assumptions of Theorem 3.2.

Note also that the characterization is invariant under permutation of elements of the set $E = \{1, \dots, n\}$, since (3.4) and (3.5) are preserved under any permutations. □

Proof. For any set $D \subset \{1, \dots, n\}$ we denote by $|D|$ the cardinality of D, and by $\underline{r}^D$—the vector of length $|D|$ indexed by elements from this set. For any vector $\underline{v} = (v_1, \dots, v_k)$ we denote $|\underline{v}| = v_1 + \cdots + v_k$. Further we define vectors related to the partition $\pi_1 = \{B_1, \dots, B_L\}$: $\underline{\varepsilon}_i$ denotes a unit vector of length L with 1 at the ith coordinate, $\underline{\varepsilon}_p^i$, $i = 1 \dots, L$, $p \in B_i \in \pi_1$, denotes a vector of length $|B_i|$ created from $\underline{r}^{B_i}$ by replacing all entries with 0 except for the pth entry, which is replaced with 1. Similarly, we define the unit vectors connected with the partition $\pi_2 = \{C_1, \dots, C_M\}$: $\underline{\delta}_i$ of length M and $\underline{\delta}_p^i$, $i = 1, \dots, M$ of length $|C_i|$, $C_i \in \pi_2$. Denote by $\underline{r}^{A,L} = (r_i^{A,L})_{i=1}^L$ for $A \subset \{1, \dots, L\}$ a vector of length L of the form

$$r_i^{A,L} = \begin{cases} r_i, & i \in A \\ 0, & i \notin A \end{cases}.$$

Let $\underline{Y} = (Y_1, \dots, Y_n)$ be a vector having the Dirichlet distribution $\mathrm{Dir}(z_1, \dots, z_n)$. Then for any $\rho_1, \dots, \rho_n \in \mathbb{N}_0$ we have

$$\mathbb{E}Y_1^{\rho_1} \dots Y_n^{\rho_n} = \frac{\Gamma(|\underline{z}|)}{\Gamma(|\underline{z}| + |\underline{\rho}|)} \prod_{i=1}^n \frac{\Gamma(z_i + \rho_i)}{\Gamma(z_i)}, \tag{3.6}$$

where $\underline{z} = (z_1, \dots, z_n)$ and $\underline{\rho} = (\rho_1, \dots, \rho_n)$. Since the Dirichlet distribution is characterized by its moments, it suffices to show that (3.6) holds for $\underline{X}$.

By neutrality of $\underline{X}$ with respect to π_1 and π_2 we have for any $r_1, \dots, r_n \in \mathbb{N}_0$:

$$\begin{aligned} \mathbb{E}X_1^{r_1} \dots X_n^{r_n} &= \mathbb{E}\left(\prod_{i\in B_1}\left(\frac{X_i}{R_1}\right)^{r_i}\right) \dots \mathbb{E}\left(\prod_{i\in B_L}\left(\frac{X_i}{R_L}\right)^{r_i}\right) \mathbb{E}(R_1^{s_1} \dots R_L^{s_L}) \\ &= \mathbb{E}\left(\prod_{i\in C_1}\left(\frac{X_i}{T_1}\right)^{r_i}\right) \dots \mathbb{E}\left(\prod_{i\in C_M}\left(\frac{X_i}{T_M}\right)^{r_i}\right) \mathbb{E}(T_1^{t_1} \dots T_M^{t_M}), \end{aligned} \tag{3.7}$$

where

$$R_i = \sum_{j\in B_i} X_j, \qquad s_i = \sum_{j\in B_i} r_j, \quad i = 1, \dots, L,$$

$$T_i = \sum_{j\in C_i} X_j, \qquad t_i = \sum_{j\in C_i} r_j, \quad i = 1, \dots, M.$$

For any vectors $\underline{m}_i = (m^i_1, \dots, m^i_{|B_i|})$, $\underline{l} = (l_1, \dots, l_L)$, $\underline{\widetilde{m}}_i = (\widetilde{m}^i_1, \dots, \widetilde{m}^i_{|C_i|})$, $\underline{\widetilde{l}} = (\widetilde{l}_1, \dots, \widetilde{l}_M)$ we define functions F, f_i, $i = 1, \dots, L, G, g_j, j = 1, \dots, M$, as follows:

$$f_i(\underline{m}_i) = \mathbb{E}\left(\prod_{j\in B_i}\left(\frac{X_j}{R_i}\right)^{m^i_j}\right), \qquad F(\underline{l}) = \mathbb{E}(R_1^{l_1}\dots R_L^{l_L}),$$

$$g_j(\underline{\widetilde{m}}_j) = \mathbb{E}\left(\prod_{i\in C_j}\left(\frac{X_i}{T_j}\right)^{\widetilde{m}^j_i}\right), \qquad G(\underline{\widetilde{l}}) = \mathbb{E}(T_1^{\widetilde{l}_1}\dots T_M^{\widetilde{l}_M}).$$

Note that the property (3.1) holds for all the functions defined above.

After plugging these functions into (3.7) we obtain the equation

$$F(|\underline{r}^{B_i}|, i = 1, \dots, L)\prod_{i=1}^{L} f_i(\underline{r}^{B_i}) = G(|\underline{r}^{C_i}|, i = 1, \dots, M)\prod_{i=1}^{M} g_i(\underline{r}^{C_i}). \tag{3.8}$$

We will show that the moments of $\underline{X}$ are of the form (3.6). The proof proceeds in the following steps:

I. Determining the form of function F (and G):
  (1) at point $\underline{r}^{A,L}$ for some $A \subset \{1, \dots, L\}$:
    (a) construction of the set A,
    (b) showing that the assumptions of Lemma 3.1 hold for A,
  (2) at any point $\underline{r}$.
II. Determining the form of functions g_l, $l = 1, \dots, M$ (and f_l, $l = 1, \dots, L$).
III. Identification of the parameters of functions F, G, g_l, $l = 1, \dots, M$.

I. *Determining the form of function F*

(1)(a) First we identify a k-element set $A \subset \{1, \dots, L\}$ that satisfies the assumptions of Lemma 3.1 in order to determine the form of $F(\underline{r}^{A,L})$.

For any blocks $B_p, B_q \in \pi_1$ we denote

$$\mathcal{C}(B_p, B_q) = \{C \in \pi_2 : B_p \cap C \neq \emptyset \wedge B_q \cap C \neq \emptyset\}.$$

Consider two different blocks $B_p, B_q \in \pi_1$. By (3.4) there exists a set of blocks $\mathcal{D}^{pq} = \{B_p = B_{i_0}, B_{i_1}, \dots, B_{i_m}, B_{i_{m+1}} = B_q\}$ such that

$$\mathcal{C}(B_{i_{j-1}}, B_{i_j}) \neq \emptyset, \quad j = 1, \dots, m+1. \tag{3.9}$$

The set $\mathcal{D}^{pq}$ will be further referred to as a path between B_p and B_q.

It will be convenient to use the following graph representation for partitions π_1, π_2. We assume that the blocks $B_i \in \pi_1$, $i = 1, \dots, L$, correspond to vertices of the graph $\mathcal{G} = (V, E)$, and the existence of an edge between two vertices B_i, B_j is defined by the condition $\mathcal{C}(B_i, B_j) \neq \emptyset$. Note that for vertices of the graph $\mathcal{G}$ we will alternatively use symbols $v \in \{1, \dots, L\}$ or $B \in \pi_1$. Then (3.4) is equivalent to $\mathcal{G}$ being connected. Additionally, we associate with each vertex its type being a number of elements in the corresponding block. Note that for every vertex B of type one the set $N(B) \cup \{B\}$ forms a maximal clique (whose all edges may be assumed to come from one block $C \in \pi_2$). Every vertex of type one is then always significant in a clique (but not necessarily the only one in the clique with this property). In addition, by (3.5) in every clique there exists at most one vertex of type one.

Let us first notice that every set A satisfying the assumptions of Lemma 3.1 must consist of at least two elements. We will identify $A \subset \{1, \dots, L\}$ by choosing a family $\mathcal{B}_A$ of blocks of π_1, i.e. by choosing a subset of vertices of graph $\mathcal{G}$. This will be done in two steps. First we will choose a single special vertex B_1. Then we will choose remaining vertices through an algorithm with a starting point in B_1.

Let B_1 be a non-cut vertex of $\mathcal{G}$ such that either it is a leaf or it is not of type one. We will show that such a vertex always exists. To this end, we will use Lemma 2.6. Let $\widetilde{C} = \{C_1, \dots, C_h\}$ denote a subfamily of all maximal cliques such that each of them contains a vertex of type one not being a leaf. Note that each such a clique has at least 3 elements. Let v_i be the vertex of type one from the ith clique, $i = 1, \dots, h$. Define $V' = \{v_i,\ i = 1, \dots, h\}$. Consider a subgraph $\widetilde{\mathcal{G}}$ of $\mathcal{G}$ induced by a vertex set $V \setminus V'$. Since v_i is of type one, it is significant in C_i, $i = 1, \dots, h$. Hence the assumptions of Lemma 2.6 are satisfied for $\widetilde{C}$, V' and $\widetilde{\mathcal{G}}$. Therefore, to find a non-cut vertex of $\mathcal{G}$ which either is a leaf or is not of type one, it suffices to find a vertex that is non-cut in $\widetilde{\mathcal{G}}$ and has this property. Since every $v_i \in V'$ is significant in C_i, $i = 1, \dots, h$, by Lemma 2.4 graph $G[V \setminus \{v_i\}]$ is connected, and vertices $V' \setminus \{v_i\}$ are still significant in the corresponding cliques. Then the same reasoning for all of the remaining vertices $V' \setminus \{v_i\}$ leads to $\widetilde{\mathcal{G}}$ being connected. Hence by Lemma 2.2 there exist in $\widetilde{\mathcal{G}}$ at least two non-cut vertices. Let us choose any of them. By definition of V', if this vertex is of type one, it has to be a leaf in $\mathcal{G}$, which completes the proof.

To identify the remaining elements of $\mathcal{B}_A$, we will first specify appropriate unique paths $\mathcal{D}^{1q}$ for all $q \neq 1$. To this end, we assign weights to the edges of $\mathcal{G}$ and use a greedy algorithm to find a minimal spanning tree $\mathcal{T}_{\mathcal{G}}$ in $\mathcal{G}$ (in general $\mathcal{T}_{\mathcal{G}}$ is not unique for $\mathcal{G}$). The algorithm (for details see [14]) starts from an arbitrary vertex, and until all vertices from V are in the tree, in every step one edge of minimal weight, connecting one of the already chosen vertices with one of the remaining ones, is added (if there are multiple edges with the same weight, any of them may be chosen). Let B_1 be a starting vertex of the algorithm (the root in the resulting spanning tree). Weights are assigned as follows.

1. In every maximal clique that contains a vertex B of type one edges incident to B are given weight w_0;
2. If the vertex B_1 does not belong to any clique considered in (1), we assign weight $w_1 > w_0$ to one fixed edge incident to B_1, and to the rest of them—weight $w_2 > w_1$;
3. All remaining edges in $\mathcal{G}$ are given weight w_1.

This method of assigning weights implies in particular that:

(i) the resulting tree contains only one edge incident to the root B_1 (note that the fact that B_1 was chosen to be non-cut is important here),
(ii) if a maximal clique K of $\mathcal{G}$ contains a vertex B of type one, the only edges in K that remain in the tree are the ones incident to B. This means that two vertices from K that are not of type one are not adjacent in the tree.

The resulting spanning tree $\mathcal{T}_{\mathcal{G}}$ with a root in B_1 determines uniquely the set of paths $\{\mathcal{D}^{1q},\ q = 2, \dots, L\}$. Since the root is fixed, we can consider the tree directed. Then we define the set $\mathcal{B}_A$ as the union of $\{B_1\}$ and the set of leaves of $\mathcal{T}_{\mathcal{G}}$. Note that $\mathcal{B}_A$ is then symmetric with respect to its elements, i.e. if we chose another vertex $B_i \in \mathcal{B}_A$ to be the root, B_1 would become a leaf.

(b) Let $\mathcal{B}_A$ be as defined above. Without loss of generality we will assume that $\mathcal{B}_A = \{B_1, \dots, B_k\}$, $k \ge 2$. We will show that the corresponding set $A = \{1, \dots, k\}$ satisfies the assumptions of Lemma 3.1. To this end, we will consider all pairs of vertices adjacent in $\mathcal{T}_{\mathcal{G}}$. Consider B_1 and the path $\mathcal{D}^{1q} = \{B_1, B_{i_1}, \dots, B_{i_m}, B_q\}$ to a certain leaf B_q in $\mathcal{T}_{\mathcal{G}}$. Let $B_u, B_v \in \mathcal{D}^{1q}$ be two adjacent vertices. Then there exists a block $C_l \in \mathcal{C}(B_u, B_v)$. It follows from (3.5) that C_l can contain at most one block $B \in \pi_1$ of size one. In addition, if none of B_u, B_v is of size one, C_l contains no 1-element blocks. Indeed, if there existed 1-element block $B \subset C_l$, $B \notin \{B_u, B_v\}$, then vertices B_u, B_v and B would belong to a maximal clique K (with edges defined by the block C_l) in $\mathcal{G}$. But then (since B is of type one) – by (ii) – vertices B_u, B_v would not be adjacent in $\mathcal{T}_{\mathcal{G}}$, which leads to a contradiction. Hence the only possible situations are:

(1) $|B_v| = 1$ and $|B_u| > 1$;
(2) $|B_u| = 1$ and $|B_v| > 1$;
(3) The block C_l does not contain any blocks of size one from π_1.

Consider first (1). Then since $B_v \subset C_l$, and no other blocks are contained in C_l,

$$\forall B_s \in \pi_1, B_s \ne B_v, \quad \exists \widetilde{s} \in B_s \setminus C_l. \tag{3.10}$$

Again by (3.5), the last condition is also satisfied under (3) (in the latter case additionally there exists an element $\widetilde{v} \in B_v \setminus C_l$). We will discuss these two cases together, using the condition (3.10) only.

Let us go back to Eq. (3.8), and consider B_u, B_v as defined above (without loss of generality we can assume that B_u is the predecessor of B_v). By (3.5) there exist elements a, b such that $\{a\} = B_u \cap C_l$ and $\{b\} = B_v \cap C_l$. (Actually it follows from the fact that $\pi_1 \wedge \pi_2 = \pi_*$, which is a weaker condition than (3.5).) First we add 1 to r_a in (3.8). Then we divide the resulting equation by (3.8) with 1 added to r_b. After cancellations we get

$$\frac{f_u(\underline{r}^{B_u} + \underline{\varepsilon}_a^u) f_v(\underline{r}^{B_v})}{f_u(\underline{r}^{B_u}) f_v(\underline{r}^{B_v} + \underline{\varepsilon}_b^v)} \frac{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_u}| + 1, \dots, |\underline{r}^{B_L}|\right)}{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_v}| + 1, \dots, |\underline{r}^{B_L}|\right)} = \frac{g_l(\underline{r}^{C_l} + \underline{\delta}_a^l)}{g_l(\underline{r}^{C_l} + \underline{\delta}_b^l)}. \tag{3.11}$$

Next, using (3.10), we choose from each block $B_s \in \pi_1$, $B_s \ne B_v$, an element $\widetilde{s}$ that does not belong to the block C_l, and from B_v we choose the element b. We define $\underline{r}$ as a vector of length L with $r_{\widetilde{s}}$ at the sth coordinate, $s = 1, \dots, L$, $s \ne v$, and r_b at the vth one. In Eq. (3.11) we replace with zeros all r_κ that are not the elements of $\underline{r}$. We get after some transformations

$$\frac{F(\underline{r} + \underline{\varepsilon}_u)}{F(\underline{r} + \underline{\varepsilon}_v)} = \frac{f_u(r_{\widetilde{u}} \underline{\varepsilon}_{\widetilde{u}}^u) f_v((r_b + 1)\underline{\varepsilon}_b^v)}{f_u(r_{\widetilde{u}} \underline{\varepsilon}_{\widetilde{u}}^u + \underline{\varepsilon}_a^u) f_v(r_b \underline{\varepsilon}_b^v)} \frac{g_l(\underline{\delta}_a^l + r_b \underline{\delta}_b^l)}{g_l((r_b + 1)\underline{\delta}_b^l)} = \frac{\alpha_u^2(r_{\widetilde{u}})}{\alpha_{v,u}^1(r_b)}. \tag{3.12}$$

The function

$$\alpha_{v,u}^1(r_b) = \frac{f_v((r_b + 1)\underline{\varepsilon}_b^v)}{f_v(r_b \underline{\varepsilon}_b^v)} \frac{g_l(\underline{\delta}_a^l + r_b \underline{\delta}_b^l)}{g_l((r_b + 1)\underline{\delta}_b^l)}$$

depends on the block B_u through a in $g_l(\underline{\delta}_a^l + r_b \underline{\delta}_b^l)$, however, the vertex B_u – as the predecessor of B_v in the tree $\mathcal{T}_{\mathcal{G}}$ – is uniquely determined for B_v. Hence we can write $\alpha_{v,u}^1(r_b) = \alpha_v^1(r_b)$.

Now we consider the case (2). Then the condition (3.10) with v replaced by u holds. By similar to the above procedure we get

$$\frac{F(\underline{r} + \underline{\varepsilon}_u)}{F(\underline{r} + \underline{\varepsilon}_v)} = \frac{f_u(r_a \underline{\varepsilon}_a^u) f_v(r_{\widetilde{v}} \underline{\varepsilon}_{\widetilde{v}}^v + \underline{\varepsilon}_b^v)}{f_u((r_a + 1)\underline{\varepsilon}_a^u) f_v(r_{\widetilde{v}} \underline{\varepsilon}_{\widetilde{v}}^v)} \frac{g_l((r_a + 1)\underline{\delta}_a^l)}{g_l(r_a \underline{\delta}_a^l + \underline{\delta}_b^l)} = \frac{\alpha_{u,v}^2(r_a)}{\alpha_v^1(r_{\widetilde{v}})}. \tag{3.13}$$

Note that this time the function $\alpha^2_{u,v}$ may depend on B_v since in general the successor B_v of B_u is not uniquely determined for all paths $\mathcal{D}^{1t}$, $t \in A$. Hence if there exists another successor B_z of B_u on a certain path $\mathcal{D}^{1t}$, we will be also interested in determining the quotient of the form (3.13) for the pair B_u, B_z, and the corresponding function $\alpha^2_{u,z}$ appearing in the numerator of the right-hand side may differ from $\alpha^2_{u,v}$.

If in the latter case $B_u = B_1$, Eq. (3.13) can be rewritten as

$$\frac{F(\underline{r}+\underline{\varepsilon}_1)}{F(\underline{r}+\underline{\varepsilon}_v)} = \frac{\alpha^2_{1,v}(r_a)}{\alpha^1_v(r_{\widetilde{v}})}.$$

Since as noted in (i) there exists for the root B_1 only one successor, the function $\widetilde{\alpha}_{1,v}$ is uniquely determined for B_1, so it does not depend on v. We denote it by $\widetilde{\alpha}_1$.

Summarizing all the three cases, we see that for any adjacent B_u, B_v (where B_u is the predecessor of B_v), $u = i_j$, $v = i_{j+1}$, $j = 1, \ldots, m$ (where $i_{m+1} = q$) we have

$$(1)\ \frac{F(\underline{r}+\underline{\varepsilon}_u)}{F(\underline{r}+\underline{\varepsilon}_v)} = \frac{\alpha^2_u(r_u)}{\alpha^1_v(r_v)} \ \vee \qquad (2)\ \frac{F(\underline{r}+\underline{\varepsilon}_u)}{F(\underline{r}+\underline{\varepsilon}_v)} = \frac{\alpha^2_{u,v}(r_u)}{\alpha^1_v(r_v)}$$

and

$$\frac{F(\underline{r}+\underline{\varepsilon}_1)}{F(\underline{r}+\underline{\varepsilon}_{i_1})} = \frac{\alpha_1(r_1)}{\alpha^1_{i_1}(r_{i_1})}$$

for some functions $\alpha_1, \alpha^1_{i_1}, \alpha^2_u\ (\alpha^2_{u,v}), \alpha^1_v, u = i_1, \ldots, i_m, v = i_2, \ldots, i_{m+1} = q$. Since (1) is a special case of (2), we can assume without loss of generality that for the elements of the path $\mathcal{D}^{1q}$ we have:

$$\frac{F(\underline{r}+\underline{\varepsilon}_1)}{F(\underline{r}+\underline{\varepsilon}_{i_1})} = \frac{\alpha_1(r_1)}{\alpha^1_{i_1}(r_{i_1})},$$
$$\frac{F(\underline{r}+\underline{\varepsilon}_{i_j})}{F(\underline{r}+\underline{\varepsilon}_{i_{j+1}})} = \frac{\alpha^2_{i_j,i_{j+1}}(r_{i_j})}{\alpha^1_{i_{j+1}}(r_{i_{j+1}})}, \quad j = 1, \ldots, m,\ i_{m+1} = q.$$

Note that α_1 does not depend on the path, because the root B_1 has the same successor in every path. Multiplying consecutively the above equations, we get

$$\frac{F(\underline{r}+\underline{\varepsilon}_1)}{F(\underline{r}+\underline{\varepsilon}_{i_1})} = \frac{\alpha_1(r_1)}{\alpha^1_{i_1}(r_{i_1})}, \tag{3.14}$$
$$\frac{F(\underline{r}+\underline{\varepsilon}_1)}{F(\underline{r}+\underline{\varepsilon}_{i_{j+1}})} = \frac{\alpha_1(r_1)}{\alpha^1_{i_1}(r_{i_1})} \prod_{t=1}^{j} \frac{\alpha^2_{i_t,i_{t+1}}(r_{i_t})}{\alpha^1_{i_{t+1}}(r_{i_{t+1}})}, \quad j = 1, \ldots, m,\ i_{m+1} = q.$$

Denote $\underline{r}^k = (r_1, \ldots, r_k, 0, \ldots, 0)$. Note that the only elements from the path $\mathcal{D}^{1q}$ that belong to the set $\mathcal{B}_A$ are the blocks B_1 and B_q. Hence after replacing with zeros all r_λ, $\lambda \notin A = \{1, \ldots, k\}$, in the above equations we get

$$\frac{F(\underline{r}^k+\underline{\varepsilon}_1)}{F(\underline{r}^k+\underline{\varepsilon}_{i_j})} = \frac{\alpha_1(r_1)}{b_{i_j}}, \quad j = 1, \ldots, m,$$
$$\frac{F(\underline{r}^k+\underline{\varepsilon}_1)}{F(\underline{r}^k+\underline{\varepsilon}_q)} = C\frac{\alpha_1(r_1)}{\alpha^1_q(r_q)},$$

where

$$b_{i_j} = \frac{F(\underline{\varepsilon}_{i_j})\alpha_1(0)}{F(\underline{\varepsilon}_1)}, \quad j = 1, \ldots, m, \quad \text{and} \quad C = \frac{F(\underline{\varepsilon}_1)\alpha^1_q(0)}{F(\underline{\varepsilon}_q)\alpha_1(0)}.$$

In order to obtain a condition of the form (3.2)(a), we then substitute $\alpha_q(r_q) := \alpha^1_q(r_q)/C$.

Since q is arbitrary in A, the above procedure allows to determine the quotients

$$\frac{F(\underline{r}^k+\underline{\varepsilon}_1)}{F(\underline{r}^k+\underline{\varepsilon}_j)} \tag{3.15}$$

for all $j = 1, \ldots, L$. Thus for a fixed index $1 \in A$ there exist uniquely determined functions $\alpha_1, \ldots, \alpha_k$ and constants $b_{k+1}, \ldots, b_L$ such that

$$\frac{F(\underline{r}^k + \underline{\varepsilon}_1)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{\alpha_1(r_1)}{\alpha_j(r_j)}, \quad j = 1, \ldots, k, \qquad (3.16)$$

$$\frac{F(\underline{r}^k + \underline{\varepsilon}_1)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{\alpha_1(r_1)}{b_j}, \quad j = k+1, \ldots, L.$$

Hence the assumptions (3.2)(a)–(3.2)(b) of Lemma 3.1 hold for $i = 1$. After appropriate multiplications of the above equations we obtain that they hold for all $i = 1, \ldots, k$ with the same functions $\alpha_j, j = 1, \ldots, k$, and constants $b_j, j = k+1, \ldots, L$. Indeed, we have for any $n \in A$

$$\frac{F(\underline{r}^k + \underline{\varepsilon}_n)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{F(\underline{r}^k + \underline{\varepsilon}_n)}{F(\underline{r}^k + \underline{\varepsilon}_1)}\frac{F(\underline{r}^k + \underline{\varepsilon}_1)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{\alpha_n(r_n)}{\alpha_1(r_1)}\frac{\alpha_1(r_1)}{\alpha_j(r_j)} = \frac{\alpha_n(r_n)}{\alpha_j(r_j)}, \quad j \in A \setminus \{n\}$$

and

$$\frac{F(\underline{r}^k + \underline{\varepsilon}_n)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{\alpha_n(r_n)}{\alpha_1(r_1)}\frac{\alpha_1(r_1)}{b_j} = \frac{\alpha_n(r_n)}{b_j}, \quad j = k+1, \ldots, L.$$

From Lemma 3.1 we conclude that

$$F(\underline{r}^k) = \frac{\Gamma\left(|\underline{d}|\right)}{\Gamma\left(|\underline{d}| + \sum_{i=1}^{k} r_i\right)} \prod_{i=1}^{k} \frac{\Gamma(r_i + d_i)}{\Gamma(d_i)} \qquad (3.17)$$

with some $\underline{d} \in \mathbb{R}^L$, or

$$F(\underline{r}^k) = \prod_{i=1}^{k} d_i^{r_i} \qquad (3.18)$$

with some $\underline{d} \in \mathbb{R}^k$.

(2) In the next steps we will complement the arguments of F with other non-zero coordinates until we arrive at its form at any point $\underline{r}$. To this end, we define $\widetilde{\mathcal{T}_{\mathcal{G}}}$ as a tree made from $\mathcal{T}_{\mathcal{G}}$ by removing all of the leaves (i.e. all of the elements from $\mathcal{B}_A$, except for B_1). We will show that if B_i is a leaf in $\widetilde{\mathcal{T}_{\mathcal{G}}}$ then we can determine the form of F at the point $\underline{r}^k + r_i\underline{\varepsilon}_i$ for any r_i. Since $\widetilde{\mathcal{T}_{\mathcal{G}}}$ is a tree, by removing one leaf in each step we obtain eventually a tree consisting of one vertex B_1, and thereby determine the form of F at any point $\underline{r}$.

Let us first go back to Eq. (3.14) and consider r_{i_m}. Assume without loss of generality that after removing the leaves (including B_q) from the tree $\mathcal{T}_{\mathcal{G}}$, B_{i_m} became a leaf in $\widetilde{\mathcal{T}_{\mathcal{G}}}$. Note that the quotients

$$\frac{F(\underline{r} + \underline{\varepsilon}_1)}{F(\underline{r} + \underline{\varepsilon}_j)} \qquad (3.19)$$

for $j \in \{i_1, \ldots, i_{m-1}\}$ do not depend on r_{i_m}. Obviously, the quotients of this form for other $j \notin A = \{1, \ldots, k\}$ do not depend on r_{i_m} either. Assume without loss of generality that $i_m = k+1$.

Consider first the case of (3.17). We will show by induction that for any r_{k+1} and any $\underline{r}^k$ we have

$$F(\underline{r}^{k+1}) = F(\underline{r}^k)\frac{\Gamma(d_{k+1} + r_{k+1})}{\Gamma(d_{k+1})}\frac{\Gamma\left(|\underline{d}| + |\underline{r}^k|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + r_{k+1}\right)}$$
$$= \frac{\Gamma\left(|\underline{d}|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^{k+1}|\right)} \prod_{i=1}^{k+1} \frac{\Gamma(r_i + d_i)}{\Gamma(d_i)}. \qquad (3.20)$$

The equality holds for $r_{k+1} = 0$. Assume that it holds for $r_{k+1} = 0, \ldots, l-1$ and consider $r_{k+1} = l$.

Let us rewrite (3.1) for $F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1})$ as

$$F(\underline{r}^k + l\underline{\varepsilon}_{k+1}) = F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1}) - \sum_{i=1}^{k} F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1} + \underline{\varepsilon}_i) - \sum_{i=k+2}^{L} F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1} + \underline{\varepsilon}_i). \qquad (3.21)$$

We start with computing the last sum in the above expression. Since the quotients (3.19) for $j \notin A$ do not depend on $r_{i_m} = r_{k+1}$, we get by (3.16)

$$\frac{F(\underline{r}^k + \underline{\varepsilon}_1)}{F(\underline{r}^k + \underline{\varepsilon}_j)} = \frac{F(\underline{r}^k + \underline{\varepsilon}_1 + (l-1)\underline{\varepsilon}_{k+1})}{F(\underline{r}^k + \underline{\varepsilon}_j + (l-1)\underline{\varepsilon}_{k+1})} = \frac{\alpha_1(r_1)}{b_j}$$

for $k+1 < j \le L$. After plugging into the above equation the form (3.3) of the function α_q for $q = 1$ we obtain

$$F(\underline{r}^k + \underline{\varepsilon}_j + (l-1)\underline{\varepsilon}_{k+1}) = F(\underline{r}^k + \underline{\varepsilon}_1 + (l-1)\underline{\varepsilon}_{k+1}) \frac{d_j}{r_1 + d_1}. \tag{3.22}$$

Using the induction assumption and (3.17), we can rewrite the right-hand side as

$$\begin{aligned} F(\underline{r}^k + \underline{\varepsilon}_1 + (l-1)\underline{\varepsilon}_{k+1}) &= F(\underline{r}^k + \underline{\varepsilon}_1) \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1 + l - 1\right)} \\ &= F(\underline{r}^k) \frac{d_1 + r_1}{|\underline{d}| + |\underline{r}^k|} \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right)}. \end{aligned}$$

After plugging the above into (3.22) we get

$$F(\underline{r}^k + \underline{\varepsilon}_j + (l-1)\underline{\varepsilon}_{k+1}) = F(\underline{r}^k) \frac{d_j}{|\underline{d}| + |\underline{r}^k|} \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right)} \tag{3.23}$$

for all $j = k+2, \ldots, L$.

For the remaining elements of (3.21), by induction assumption we have

$$F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1}) = F(\underline{r}^k) \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l - 1\right)}$$

and

$$F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1} + \underline{\varepsilon}_i) = F(\underline{r}^k) \frac{d_i + r_i}{|\underline{d}| + |\underline{r}^k|} \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right)},$$

$i = 1, \ldots, k$, where to write the last expression we used also (3.17). Plugging the above into (3.21), we get

$$\begin{aligned} F(\underline{r}^k + l\underline{\varepsilon}_{k+1}) &= F(\underline{r}^k) \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \\ &\quad \times \left[\frac{\Gamma\left(|\underline{d}| + |\underline{r}^k|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l - 1\right)} - \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k| + 1\right)}{(|\underline{d}| + |\underline{r}^k|)(\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right))} \left(\sum_{i=1}^{k} (d_i + r_i) + \sum_{i=k+2}^{L} d_i \right) \right] \\ &= F(\underline{r}^k) \frac{\Gamma(d_{k+1} + l - 1)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right)} \left[(|\underline{d}| + |\underline{r}^k| + l - 1) - \left(\sum_{i=1}^{k} (d_i + r_i) + \sum_{i=k+2}^{L} d_i \right) \right] \\ &= F(\underline{r}^k) \frac{\Gamma(d_{k+1} + l)}{\Gamma(d_{k+1})} \frac{\Gamma\left(|\underline{d}| + |\underline{r}^k|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}^k| + l\right)}. \end{aligned}$$

Hence the proof of (3.20) is complete.

Similarly, we determine the form of F at the point $\underline{r}^{k+1} + r_i\underline{\varepsilon}_i$ for the next leaf B_i of the tree $\widetilde{\mathcal{T}}_{\mathcal{G}}$, and we proceed until we determine the form of F at any point $\underline{r}$. We obtain eventually

$$F(\underline{r}) = \frac{\Gamma\left(|\underline{d}|\right)}{\Gamma\left(|\underline{d}| + |\underline{r}|\right)} \prod_{i=1}^{L} \frac{\Gamma(d_i + r_i)}{\Gamma(d_i)}. \tag{3.24}$$

Consider now the case of (3.18). We will follow the inductive proof of (3.20) to show that

$$F(\underline{r}^{k+1}) = F(\underline{r}^k) d_{k+1}^{r_{k+1}},$$

where $d_{k+1} = \frac{b_{k+1}}{|\underline{b}|}$.

In order to determine the terms of (3.21), we first observe that (3.22) now becomes

$$F(\underline{r}^k + \underline{\varepsilon}_j + (l-1)\underline{\varepsilon}_{k+1}) = F(\underline{r}^k + \underline{\varepsilon}_1 + (l-1)\underline{\varepsilon}_{k+1}) \frac{b_j}{b_1} = F(\underline{r}^k) d_1 d_{k+1}^{l-1} \frac{b_j}{b_1},$$

for all $j = k+2, \dots, L$, where in the second equality we used (3.18) and the induction assumption. Similarly, by the induction assumption we obtain

$$F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1}) = F(\underline{r}^k) d_{k+1}^{l-1},$$
$$F(\underline{r}^k + (l-1)\underline{\varepsilon}_{k+1} + \underline{\varepsilon}_i) = F(\underline{r}^k) d_i d_{k+1}^{l-1}, \quad i = 1, \dots, k.$$

Plugging all the terms into (3.21), we get

$$F(\underline{r}^k + l\underline{\varepsilon}_{k+1}) = F(\underline{r}^k) d_{k+1}^{l-1} \left(1 - \sum_{i=1}^{k} d_i - \sum_{i=k+2}^{L} \frac{b_i}{\sum_{j=1}^{L} b_j} \right)$$
$$= F(\underline{r}^k) d_{k+1}^{l-1} \left(1 - \frac{\sum_{i=1}^{k} b_i + \sum_{i=k+1}^{L} b_i - b_{k+1}}{\sum_{j=1}^{L} b_j} \right) = F(\underline{r}^k) d_{k+1}^{l}.$$

We proceed for the remaining coordinates until we arrive at

$$F(\underline{r}) = \prod_{i=1}^{L} d_i^{r_i}, \tag{3.25}$$

where $d_i = \frac{b_i}{|\underline{b}|}$.

Because of the symmetry between the assumptions for π_1 and π_2 we determine the form of G as one of the following

$$G(\underline{r}) = \frac{\Gamma(|\underline{c}|)}{\Gamma(|\underline{c}| + |\underline{r}|)} \prod_{i=1}^{M} \frac{\Gamma(c_i + r_i)}{\Gamma(c_i)}, \tag{3.26}$$
$$G(\underline{r}) = \prod_{i=1}^{M} c_i^{r_i},$$

for $\underline{r} = (r_1, \dots, r_M) \in \{0, 1, \dots\}^M$ and some $\underline{c} = (c_1, \dots, c_M) \in \mathbb{R}^M$.

II. *Determining the form of functions g_l, $l = 1, \dots, M$ and f_l, $l = 1, \dots, L$*

We will now determine the form of g_l for every C_l, $l = 1, \dots, M$. Let us first note that if C_l is 1-element then $g_l \equiv 1$. Assume then that C_l has at least two elements. Consider any $a, b \in C_l$. By (3.5) there exist blocks $B_u, B_v \in \pi_1$ such that $\{a\} = B_u \cap C_l$ and $\{b\} = B_v \cap C_l$. In Eq. (3.8) we add 1 first to r_a and then to r_b. Dividing the first equation by the second one, we arrive at (3.11). After applying the known forms of F and plugging $r_\lambda = 0$ for all $\lambda \notin C_l$ the quotient of functions F appearing in (3.11) can be reduced to one of two forms, depending on whether F admits (3.24) or (3.25). Suppose first that (3.24) holds. Then

$$\frac{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_u}| + 1, \dots, |\underline{r}^{B_L}|\right)}{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_v}| + 1, \dots, |\underline{r}^{B_L}|\right)} = \frac{\left(\prod_{k \neq u} \frac{\Gamma(d_k + |\underline{r}^{B_k}|)}{\Gamma(d_k)} \right) \frac{\Gamma(d_u + |\underline{r}^{B_u}| + 1)}{\Gamma(d_u)}}{\left(\prod_{k \neq v} \frac{\Gamma(d_k + |\underline{r}^{B_k}|)}{\Gamma(d_k)} \right) \frac{\Gamma(d_v + |\underline{r}^{B_v}| + 1)}{\Gamma(d_v)}}$$
$$= \frac{\Gamma(r_b + d_v)}{\Gamma(r_a + d_u)} \frac{\Gamma(r_a + d_u + 1)}{\Gamma(r_b + d_v + 1)} = \frac{r_a + d_u}{r_b + d_v}. \tag{3.27}$$

Hence with $r_\lambda = 0$ for all $\lambda \notin C_l$ we have

$$\frac{g_l(\underline{r}^{C_l} + \underline{\delta}_a^l)}{g_l(\underline{r}^{C_l} + \underline{\delta}_b^l)} = \frac{f_u\left((r_a+1)\underline{\varepsilon}_a^u\right) f_v(r_b \underline{\varepsilon}_b^v)}{f_u(r_a \underline{\varepsilon}_a^u) f_v\left((r_b+1)\underline{\varepsilon}_b^v\right)} \frac{r_a + d_u}{r_b + d_v} = \frac{\gamma_{a,u}(r_a)}{\gamma_{b,v}(r_b)}, \tag{3.28}$$

where

$$\gamma_{a,u}(r_a) = \frac{f_u\left((r_a+1)\underline{\varepsilon}_a^u\right)(r_a + d_u)}{f_u(r_a \underline{\varepsilon}_a^u)}$$

is a function potentially depending on a and the block B_u. However, it follows from (3.5) that for every element $s \in C_l$ there exists exactly one block $B_t \in \pi_1$ such that $B_t \cap C_l = \{s\}$, and consequently we will write $\gamma_{a,u} = \gamma_a$.

Thus for all pairs of elements $\{e, f\} \subset C_l$ we can rewrite (3.28) as

$$\frac{g_l(\underline{r}^{C_l} + \underline{\delta}_e^l)}{g_l(\underline{r}^{C_l} + \underline{\delta}_f^l)} = \frac{\gamma_e(r_e)}{\gamma_f(r_f)}$$

for some functions $\gamma_\lambda, \lambda \in C_l$. By Lemma 3.1 for any $l = 1, \dots, M$ there exists a vector $\underline{z}^l = (z_1^l, \dots, z_{|C_l|}^l)$ such that for any $\underline{x} = (x_1, \dots, x_{|C_l|})$

$$g_l(\underline{x}) = \frac{\Gamma(|\underline{z}^l|)}{\Gamma(|\underline{z}^l| + |\underline{x}|)} \prod_{i=1}^{|C_l|} \frac{\Gamma(z_i^l + x_i)}{\Gamma(z_i^l)},$$

or

$$g_l(\underline{x}) = \prod_{i=1}^{|C_l|} (z_i^l)^{x_i}.$$

In the case where F is given by (3.25), the quotient (3.27) becomes

$$\frac{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_u}| + 1, \dots, |\underline{r}^{B_L}|\right)}{F\left(|\underline{r}^{B_1}|, \dots, |\underline{r}^{B_v}| + 1, \dots, |\underline{r}^{B_L}|\right)} = \frac{d_u}{d_v},$$

and the conclusion holds.

Similarly, we determine the functions $f_l, l = 1, \dots, L$, to be given by

$$f_l(\underline{x}) = \frac{\Gamma(|\underline{t}^l|)}{\Gamma(|\underline{t}^l| + |\underline{x}|)} \prod_{i=1}^{|B_l|} \frac{\Gamma(t_i^l + x_i)}{\Gamma(t_i^l)},$$

or

$$f_l(\underline{x}) = \prod_{i=1}^{|B_l|} (t_i^l)^{x_i},$$

for any $\underline{x} = (x_1, \dots, x_{|B_l|})$ and some vector $\underline{t}^l = (t_1^l, \dots, t_{|B_l|}^l)$.

We showed that each of the functions $F, G, f_i, i = 1, \dots, L, g_j, j = 1, \dots, M$, can take one of two forms—either the gamma form, or the exponential form. Note that if all the functions are of the exponential form, the moments (3.7) of the vector $\underline{X}$ are given by

$$\mathbb{E}X_1^{r_1} \dots X_n^{r_n} = \prod_{i=1}^{n} D_i^{r_i},$$

for some constants $D_i, i = 1, \dots, n$, which corresponds to the case of all variables $X_i, i = 1, \dots, n$, being degenerate. This means that for the non-degenerate case at least one function must take the gamma form. We will show that this results in all other functions being of the gamma form. In order to do it, let us go back to Eq. (3.8), and assume that f_i is of the gamma form. Then

$$f_i(\underline{r}^{B_i}) = \frac{\Gamma(|\underline{t}^i|)}{\Gamma(|\underline{t}^i| + |\underline{r}^{B_i}|)} \prod_{j=1}^{|B_i|} \frac{\Gamma(t_j^i + r_j)}{\Gamma(t_j^i)},$$

appears on the left-hand side of (3.8). Since none of the blocks $C_j \in \pi_2, j = 1, \dots, M$, can be equal to the block $B_i \in \pi_1$, a term of the form $\Gamma(|\underline{t}^i| + |\underline{r}^{B_i}|)$ (depending on the sum $|\underline{r}^{B_i}|$) will never appear on the right-hand side of (3.8), regardless of the form of $G, g_l, l = 1, \dots, M$. Hence for (3.8) to be true, this term must cancel with a term coming from another function on the left-hand side, and it can only happen when F takes the gamma form (3.24). Then on the left-hand side of (3.8) we obtain terms of the form $\Gamma(d_j + |\underline{r}^{B_j}|)$ for all other $j = 1, \dots, L, j \neq i$. As they can never appear on the right-hand side, they must cancel, which in turn forces the gamma form for all other functions $f_j, j = 1, \dots, L, j \neq i$. Now the equality (3.8) implies the gamma form for G and $g_l, l = 1, \dots, M$.

III. *Identification of the parameters of functions $F, G, g_l, l = 1, \dots, M$*

To show that the moments of $\underline{X}$ are of the form (3.6), it remains to find certain relations between the components of vectors $\underline{c}, z^j, j = 1, \dots, M$. After plugging into (3.8) the forms of functions F, G and $g_l, l = 1, \dots, M$, obtained in the previous steps we get

$$\prod_{i=1}^{L}\left[f_i(\underline{r}^{B_i})\right]\frac{\Gamma(|\underline{d}|)}{\Gamma\left(|\underline{d}|+\sum_{i=1}^{L}|\underline{r}^{B_i}|\right)}\prod_{i=1}^{L}\frac{\Gamma(d_i+|\underline{r}^{B_i}|)}{\Gamma(d_i)}$$
$$=\prod_{i=1}^{M}\left[\prod_{j=1}^{|C_i|}\left(\frac{\Gamma\left(z_j^i+r_j^{C_i}\right)}{\Gamma(z_j^i)}\right)\frac{\Gamma\left(|\underline{z}^i|\right)}{\Gamma\left(|\underline{z}^i|+|\underline{r}^{C_i}|\right)}\right]\frac{\Gamma(|\underline{c}|)}{\Gamma\left(|\underline{c}|+\sum_{i=1}^{M}|\underline{r}^{C_i}|\right)}\prod_{i=1}^{M}\frac{\Gamma(c_i+|\underline{r}^{C_i}|)}{\Gamma(c_i)}. \quad (3.29)$$

Consider again $C_l\in\pi_2$. After replacing with zeros all r_λ for $\lambda\notin\{a,b\}$, where $\{a\}=B_u\cap C_l$, $\{b\}=B_v\cap C_l$, we obtain

$$f_u(r_a\underline{\varepsilon}_a^u)f_v(r_b\underline{\varepsilon}_b^v)\frac{\Gamma(|\underline{d}|)\Gamma(r_a+d_u)\Gamma(r_b+d_v)}{\Gamma(|\underline{d}|+r_a+r_b)\Gamma(d_u)\Gamma(d_v)}$$
$$=\frac{\Gamma(r_a+z_a^l)}{\Gamma(z_a^l)}\frac{\Gamma(r_b+z_b^l)}{\Gamma(z_b^l)}\frac{\Gamma(|z^l|)}{\Gamma(|z^l|+r_a+r_b)}\frac{\Gamma(|\underline{c}|)\Gamma(r_a+r_b+c_l)}{\Gamma(|\underline{c}|+r_a+r_b)\Gamma(c_l)}. \quad (3.30)$$

Define $m(r_a,r_b)$ as

$$m(r_a,r_b)=\frac{\Gamma(|\underline{c}|+r_a+r_b)}{\Gamma(|\underline{d}|+r_a+r_b)}\frac{\Gamma(|z^l|+r_a+r_b)}{\Gamma(r_a+r_b+c_l)}. \quad (3.31)$$

Then by (3.30) $m(r_a,r_b)=\widetilde{m}(r_a+r_b)=m_1(r_a)m_2(r_b)$ for some functions m_1 and m_2. Thus functions $\widetilde{m}$, m_1, m_2 satisfy the Pexider equation, and hence $\widetilde{m}=\rho\nu^x$, where ρ, ν are some constants. However, $\widetilde{m}$ as the quotient of Gamma functions is rational, which implies $\widetilde{m}=\rho$. It follows from (3.31) that either $c_l=|c|$ or $c_l=|z^l|$. By (3.26) the first case implies $G(r_l\delta_l)=1$, which contradicts the definition of G. Thus $c_l=|z^l|$. We follow this reasoning for all other blocks $C_\lambda\in\pi_2$, $\lambda=1,\dots,M$, and get $c_\lambda=|z^\lambda|$. Then the formula (3.7) on the moments of X can be expressed as

$$\mathbb{E}X_1^{r_1}\dots X_n^{r_n}$$
$$=\prod_{i=1}^{M}\left[\prod_{j=1}^{|C_i|}\left(\frac{\Gamma\left(z_j^i+r_j^{C_i}\right)}{\Gamma(z_j^i)}\right)\frac{\Gamma\left(|\underline{z}^i|\right)}{\Gamma\left(|\underline{z}^i|+|\underline{r}^{C_i}|\right)}\right]\frac{\Gamma(|\underline{c}|)}{\Gamma\left(|\underline{c}|+\sum_{i=1}^{M}|\underline{r}^{C_i}|\right)}\prod_{i=1}^{M}\frac{\Gamma(c_i+|\underline{r}^{C_i}|)}{\Gamma(c_i)}$$
$$=\prod_{i=1}^{M}\left[\prod_{j=1}^{|C_i|}\left(\frac{\Gamma\left(z_j^i+r_j^{C_i}\right)}{\Gamma(z_j^i)}\right)\right]\frac{\Gamma\left(\sum_{i=1}^{M}|\underline{z}^i|\right)}{\Gamma\left(\sum_{i=1}^{M}|\underline{z}^i|+\sum_{i=1}^{M}|\underline{r}^{C_i}|\right)}=\prod_{i=1}^{n}\frac{\Gamma(z_i+r_i)}{\Gamma(z_i)}\frac{\Gamma(|\underline{z}|)}{\Gamma(|\underline{z}|+|\underline{r}|)},$$

where $\underline{z}=(z_1,\dots,z_n)$ is a vector made of elements of vectors $z^1,\dots,z^M$ so that the jth element of $\underline{z}$ corresponds to r_j, $j=1,\dots,n$. Hence the distribution of $\underline{X}$ is Dirichlet $\mathrm{Dir}(z_1,\dots,z_n)$. □

4. Examples

It is clear that the assumptions of Theorem 3.2 are satisfied for the known 2-partition characterizations. As mentioned before, the characterizations involving *only two* partitions presented in [6,8,10,1] work only for vectors of dimension 3. On the other hand, matrix characterization [9,2], involving two partitions, can be applied to any n-element vector, provided that n is not prime. Actually, in these characterizations it was important that any block of one partition has non-empty intersection with any block of the other partition. Additionally, no 1-element blocks were allowed.

An example of a class of two partitions leading to the characterization, which is not covered by any previous results, is given below.

Example 4.1. Let $n>3$ be any odd number. Define the partitions $\pi_1=\{\{1\},\{2,3\},\{4,5\},\dots,\{n-1,n\}\}$ and $\pi_2=\{\{1,2\},\dots,\{n-2,n-1\},\{n\}\}$. Then π_1 and π_2 satisfy the assumptions (3.4)–(3.5).

Note that the assumptions (3.4)–(3.5) are minimal in the sense that none of them is separately sufficient to imply the Dirichlet distribution for a given vector. We will illustrate it with two following examples.

Example 4.2. Consider first a random vector $\underline{X}=(X_1,\dots,X_6)$ and two partitions $\pi_1=\{\{1,2\},\{3\},\{4,5\},\{6\}\}$, $\pi_2=\{\{1\},\{2,3\},\{4\},\{5,6\}\}$. Then $\widehat{\pi}_1\wedge\widehat{\pi}_2=\pi_*$, but $\pi_1\vee\pi_2=\{\{1,2,3\},\{4,5,6\}\}$. Let (Y_1,Y_2,Y_3) and (Y_4,Y_5,Y_6) be two independent Dirichlet vectors and define $X_i = Y_i/2$, $i = 1, \ldots, 6$. Then $\underline{X}$ is nwrt π_1 and π_2. The first neutrality means that $(Z_1, Z_2) = (X_1 + X_2, X_4 + X_5)$, $Z_3 = \frac{X_1}{X_1+X_2}$ and $Z_4 = \frac{X_4}{X_4+X_5}$ are mutually independent, which is obvious due to the independence of (X_1, X_2, X_3) and (X_4, X_5, X_6) and neutralities of Dirichlet vectors. The neutrality with respect to π_2 follows by symmetry.

Example 4.3. Consider now a vector $\underline{X} = (X_1, X_2, X_3, X_4)$ and partitions $\pi_1 = \{\{1, 2\}, \{3\}, \{4\}\}$, $\pi_2 = \{\{1\}, \{2, 3, 4\}\}$. Then $\pi_1 \vee \pi_2 = \pi^*$, but $\widehat{\pi}_1 \wedge \widehat{\pi}_2 = \{\{1\}, \{2\}, \{3, 4\}\}$. Let $\underline{X}$ be defined as follows: $(X_1, X_2) \sim \mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$, $X_3 = U(1-X_1-X_2)$, where U is a uniform random variable on $(0, 1)$ independent of (X_1, X_2), and $X_4 = 1-X_1-X_2-X_3$. Neutrality with respect to π_1 is equivalent to the independence of $(X_1 + X_2, U(1 - X_1 - X_2))$ – which is a function of $(X_1 + X_2, U)$ – and $\frac{X_1}{X_1+X_2}$. In order to prove this neutrality, it suffices to notice that $(X_1 + X_2, \frac{X_1}{X_1+X_2})$ is independent of U by definition, and $X_1 + X_2$, $\frac{X_1}{X_1+X_2}$ are independent by the property of the Dirichlet distribution. Hence $X_1 + X_2, U, \frac{X_1}{X_1+X_2}$ are mutually independent. Similarly, neutrality with respect to π_2 is equivalent to the independence of X_1 and $(\frac{X_2}{1-X_1}, \frac{U(1-X_1-X_2)}{1-X_1})$, which is a function of $(U, \frac{X_2}{1-X_1})$. The vector $(X_1, \frac{X_2}{1-X_1})$ is independent of U by definition, and X_1 is independent of $\frac{X_2}{1-X_1}$ by the property of the Dirichlet distribution. Hence $X_1, \frac{X_2}{1-X_1}, U$ are mutually independent.

Note that – as illustrated by the above example – assuming $\pi_1 \wedge \pi_2 = \pi_*$ instead of $\widehat{\pi}_1 \wedge \widehat{\pi}_2 = \pi_*$ while keeping the assumption (3.4) is not sufficient for the characterization.

To better illustrate a gain from Theorem 3.2 let us consider vectors of random probabilities consisting of 4 and 5 elements.

Remark 4.4. In the case of 4 elements the only pairs of partitions π_1, π_2 that are sufficient to characterize a Dirichlet distribution of a vector $\underline{X}$, whenever $\underline{X}$ is neutral with respect to π_1 and π_2, are (up to a permutation)

1. $\pi_1 = \{\{1, 2\}, \{3, 4\}\}, \pi_2 = \{\{1, 3\}, \{2, 4\}\}$,
2. $\pi_1 = \{\{1, 2\}, \{3, 4\}\}, \pi_2 = \{\{1, 3\}, \{2\}, \{4\}\}$.

The first characterization has been already known from [9], and the second one is due to Theorem 3.2.

Remark 4.5. In the case of 5 elements the only pairs of partitions π_1, π_2 leading to the characterization are (up to a permutation)

1. $\pi_1 = \{\{1, 2\}, \{3, 4, 5\}\}, \pi_2 = \{\{3\}, \{1, 4\}, \{2, 5\}\}$,
2. $\pi_1 = \{\{1\}, \{2, 3\}, \{4, 5\}\}, \pi_2 = \{\{2\}, \{1, 4\}, \{3, 5\}\}$,
3. $\pi_1 = \{\{1\}, \{2\}, \{3, 4, 5\}\}, \pi_2 = \{\{3\}, \{1, 4\}, \{2, 5\}\}$.

All the three cases are new and follow from Theorem 3.2.

For all other pairs of partitions of 4 or 5 elements it is easy to construct vectors that are not Dirichlet and are neutral with respect to these partitions.

Acknowledgments

The authors are indebted to P. Naroski for helpful comments on facts from the graph theory which are used in the proof of the main result.

Appendix

Proof of Lemma 3.1. We follow the approach from the proof of the main result in [2].

Fix $p \in A$. From Eqs. (3.2)(a)–(3.2)(b) (after substituting $i = p$) we determine $F(\underline{x}^A + \underline{\varepsilon}_l)$ for all $l = 1, \ldots, n$, and we get with $\underline{x} = \underline{x}^A$ in (3.1):

$$F(\underline{x}^A) = F(\underline{x}^A + \underline{\varepsilon}_p)\left(1 + \sum_{l\in A\setminus\{p\}} \frac{\alpha_l(x_l)}{\alpha_p(x_p)} + \sum_{l\notin A} \frac{b_l}{\alpha_p(x_p)}\right),$$

which for $x_p \geq 1$ can be rewritten as

$$F(\underline{x}^A) = F(\underline{x}^A - \underline{\varepsilon}_p)\frac{\alpha_p(x_p - 1)}{\alpha_p(x_p - 1) + \sum_{l\in A\setminus\{p\}} \alpha_l(x_l) + \sum_{l\notin A} b_l}. \tag{A.1}$$

Iterating the equality $x_p - 1$ times, we obtain

$$F(\underline{x}^A) = F(\underline{x}^A - x_p\underline{\varepsilon}_p) \prod_{i=0}^{x_p-1} \frac{\alpha_p(i)}{\alpha_p(i) + \sum_{l\in A\setminus\{p\}} \alpha_l(x_l) + \sum_{l\notin A} b_l}.$$

Similarly, we iterate x_q times the expression $F(\underline{x}^A - x_p\underline{\varepsilon}_p)$ with respect to $q \in A$, $q \neq p$, and we plug it into the above equation:

$$F(\underline{x}^A) = F(\underline{x}^A - x_p\underline{\varepsilon}_p - x_q\underline{\varepsilon}_q) \prod_{i=0}^{x_p-1} \frac{\alpha_p(i)}{\alpha_p(i) + \sum_{l\in A\setminus\{p\}} \alpha_l(x_l) + \sum_{l\notin A} b_l} \prod_{j=0}^{x_q-1} \frac{\alpha_q(j)}{\alpha_p(0) + \alpha_q(j) + \sum_{l\in A\setminus\{p,q\}} \alpha_l(x_l) + \sum_{l\notin A} b_l}.$$

After changing the order of iteration with respect to p and q we get:

$$F(\underline{x}^A) = F(\underline{x}^A - x_q\underline{\varepsilon}_p - x_q\underline{\varepsilon}_q) \prod_{j=0}^{x_q-1} \frac{\alpha_q(j)}{\alpha_q(j) + \sum_{l\in A\setminus\{q\}} \alpha_l(x_l) + \sum_{l\notin A} b_l} \prod_{i=0}^{x_p-1} \frac{\alpha_p(i)}{\alpha_q(0) + \alpha_p(i) + \sum_{l\in A\setminus\{p,q\}} \alpha_l(x_l) + \sum_{l\notin A} b_l}.$$

We compare both equations to obtain

$$\prod_{j=0}^{x_p-1} \left(\alpha_p(j) + \alpha_q(x_q) + C\right) \prod_{i=0}^{x_q-1} \left(\alpha_p(0) + \alpha_q(i) + C\right) = \prod_{j=0}^{x_q-1} \left(\alpha_q(j) + \alpha_p(x_p) + C\right) \prod_{i=0}^{x_p-1} \left(\alpha_q(0) + \alpha_p(i) + C\right), \tag{A.2}$$

where

$$C = \sum_{l\in A\setminus\{p,q\}} \alpha_l(x_l) + \sum_{l\notin A} b_l.$$

Since p and q were chosen arbitrarily, the above equality holds for all $p, q \in A$. Now we substitute $x_p = x_q = 1$ in (A.2) and arrive at

$$\alpha_q(0) + \alpha_p(1) = \alpha_p(0) + \alpha_q(1). \tag{A.3}$$

We will show by induction on j that

$$\alpha_q(j) - \alpha_q(j-1) = \alpha_q(1) - \alpha_q(0). \tag{A.4}$$

The equation holds for $j = 1$. Assume that it holds for all $j = 1, \dots, l$ and consider $j = l + 1$. We rewrite (A.2) for $x_q = l + 1$ and $x_p = 1$:

$$(\alpha_p(0) + \alpha_q(l+1) + C) \prod_{i=0}^{l} (\alpha_p(0) + \alpha_q(i) + C) = \prod_{j=0}^{l} (\alpha_q(j) + \alpha_p(1) + C)(\alpha_q(0) + \alpha_p(0) + C). \tag{A.5}$$

On the other hand

$$\begin{aligned}\prod_{i=0}^{l} (\alpha_p(0) + \alpha_q(i) + C) &= (\alpha_p(0) + \alpha_q(0) + C) \prod_{i=0}^{l-1} (\alpha_p(0) + \alpha_q(i+1) + C) \\ &= (\alpha_p(0) + \alpha_q(0) + C) \prod_{i=0}^{l-1} (\alpha_p(0) + \alpha_q(i) + \alpha_q(1) - \alpha_q(0) + C) \\ &= (\alpha_p(0) + \alpha_q(0) + C) \prod_{i=0}^{l-1} (\alpha_q(i) + \alpha_p(1) + C),\end{aligned} \tag{A.6}$$

where the two last equalities follow consecutively from the induction assumption and (A.3). After dividing (A.5) by (A.6) we get

$$\alpha_p(0) + \alpha_q(l+1) = \alpha_q(l) + \alpha_p(1),$$

which by (A.3) implies (A.4).

It follows from (A.4) that for every $q \in A$ there exist real numbers a_q and b_q such that for every j

$$\alpha_q(j) = a_q j + b_q. \tag{A.7}$$

Additionally, by (A.3) $a_q = a$ (it does not depend on q).

Let us go back to (A.1) and number elements of the set $A = \{s_1, \dots, s_k\}$. Substituting $p = s_1$ in (A.1) and iterating it consecutively with respect to all other variables $s_2, \dots, s_k$, we get

$$F(\underline{x}^A) = \prod_{j=1}^{k} \prod_{i=0}^{x_{s_j}-1} \frac{\alpha_{s_j}(i)}{\sum_{l=1}^{j-1} \alpha_{s_l}(0) + \alpha_{s_j}(i) + \sum_{l=j+1}^{k} \alpha_{s_l}(x_{s_l}) + \sum_{l \notin A} b_l}.$$

After plugging into the above equation the form of $\alpha_{s_j}, j = 1, \dots, k$, we obtain

$$F(\underline{x}^A) = \prod_{j=1}^{k} \prod_{i=0}^{x_{s_j}-1} \frac{ai + b_{s_j}}{\sum_{l=1}^{j-1} b_{s_l} + ai + b_{s_j} + \sum_{l=j+1}^{k} (ax_{s_l} + b_{s_l}) + \sum_{l \notin A} b_l}.$$

If $a = 0$, the equation reduces to

$$F(\underline{x}^A) = \prod_{j=1}^{k} \prod_{i=0}^{x_{s_j}-1} \frac{b_{s_j}}{\sum_{l=1}^{j-1} b_{s_l} + b_{s_j} + \sum_{l=j+1}^{k} b_{s_l} + \sum_{l \notin A} b_l} = \prod_{j=1}^{k} d_j^{x_{s_j}},$$

where

$$d_j = \frac{b_{s_j}}{\sum_{l=1}^{k} b_{s_l} + \sum_{l \notin A} b_l} = \frac{b_{s_j}}{|\underline{b}|}.$$

Otherwise, we can rewrite it as

$$F(\underline{x}^A) = \prod_{j=1}^{k} \prod_{i=0}^{x_{s_j}-1} \frac{i + \frac{b_{s_j}}{a}}{\sum_{l=1}^{k} \frac{b_{s_l}}{a} + i + \sum_{l=j+1}^{k} x_{s_l} + \sum_{l \notin A} \frac{b_l}{a}}.$$

Hence

$$F(\underline{x}^A) = \frac{\Gamma\left(\sum_{i=1}^{n} d_i\right)}{\Gamma\left(\sum_{i=1}^{n} d_i + \sum_{i \in A} x_i\right)} \prod_{i \in A} \frac{\Gamma(d_i + x_i)}{\Gamma(d_i)},$$

where $d_i = \frac{b_i}{a}$. □

References

[1] K. Bobecka, J. Wesołowski, The Dirichlet disribution and process through neutralities, J. Theoret. Probab. 20 (2007) 295–308.

[2] K. Bobecka, J. Wesołowski, Moments method approach to characterizations of Dirichlet tables through neutralities, Publ. Math. Debrecen 74/3–4 (2009) 321–339.

[3] W.Y. Chang, R.D. Gupta, D.St.P. Richards, Structural properties of the generalized Dirichlet distributions, in: M.A.G. Viana, H.P. Wynn (Eds.), Algebraic Methods in Statistics and Probability, II, Contemp. Math. 516 (2010), 109–124.

[4] D. Chickering, D. Geiger, D. Heckerman, Learning bayesian networks: the combination of knowledge and statistical data, Mach. Learn. 20 (1995) 197–243.

[5] J.R. Connor, J.E. Mosimann, Concepts of independence for proportions with a generalization od the Dirichlet distribution, J. Amer. Statist. Assoc. 64 (1969) 194–206.

[6] J. Darroch, D. Ratcliff, A characterization of the Dirichlet distribution, J. Amer. Statist. Assoc. 66 (1971) 641–643.

[7] K. Doksum, Tailfree and neutral random probabilities and their posterior distributions, Ann. Probab. 2 (1974) 183–201.

[8] J. Fabius, Two characterizations of the Dirichlet distribution, Ann. Statist. 1 (1973) 583–587.

[9] D. Geiger, D. Heckerman, A characterization of the Dirichlet distribution through global and local parameter independence, Ann. Statist. 25 (1997) 1344–1369.

[10] I.R. James, J.E. Mosimann, A new characterization of the Dirichlet distribution through neutrality, Ann. Statist. 1 (1980) 183–189.

[11] A. Járai, Regularity property of the functional equation of the Dirichlet distribution, Aequationes Math. 56 (1998) 37–46.

[12] A. Járai, Regularity Properties of Functional Equations in Several Variables, Springer, New York, 2005.

[13] K.W. Ng, G.-L. Tian, M.-L. Tang, Dirichlet and Related Distributions, Theory, Methods and Applications, Wiley, New York, 2011.

[14] R.C. Prim, Shortest connection networks and some generalizations, Bell Syst. Tech. J. 36 (1957) 1389–1401.

[15] R.V. Ramamoorthi, L.M. Sangalli, On a characterization of Dirichlet distribution, in: S.K. Upadhyay, U. Singh, D.K. Dey (Eds.), Bayesian Statistics and Its Applications, Anamaya Publishers, New Delhi, 2007, pp. 385–397.