Publ. Math. Debrecen
74/3-4 (2009), 321–339

Moments method approach to characterizations of Dirichlet tables through neutralities

By KONSTANCJA BOBECKA (Warszawa) and JACEK WESOŁOWSKI (Warszawa)

Abstract. The concept of neutrality of random probabilities with respect to a partition covers several properties of independence for unit simplex valued random vectors. In particular the Dirichlet random vector is neutral with respect to any partition of its indices. The main result of this paper simplifies and extends the characterization of the Dirichlet random table due to Geiger and Heckerman (1997). This characterization was based on independence conditions which can be viewed as neutrality with respect to row and column partitions of a two-way random table. Its proof was based on solving a functional equations for densities with the heavy use of advanced regularization techniques due to Járai (1986). Our approach is based on identification of moments through solution of a functional equation for functions of discrete arguments. Moreover the characterization is extended to multi-way tables.

1. Introduction

Let $E = \{1, 2, \dots, n, n+1\}$. A partition of set E is a set $\pi = \{P_1, \dots, P_K\}$ of nonempty pairwise disjoint subsets $P_1, \dots, P_K$ of E whose union is E. The members of π are called the blocks of π.

Let $\Theta = (\theta_1, \dots, \theta_n, \theta_{n+1})$ be a random probability on E, i.e. θ_j, $j = 1, \dots, n+1$, are nonnegative and $\sum_{j=1}^{n+1} \theta_j = 1$.

With any partition π one can associate an independence condition, which will be called neutrality.

Mathematics Subject Classification: 62H05, 62E10, 60G57, 39B40.
Key words and phrases: neutrality with respect to partition, Dirichlet distribution, method of moments, characterization, functional equations, independence, multi-way tables.
This research was partially supported by the Ministry of Science and Higher Education of Poland (Project no. 1P03A022427).

Definition 1. Let $\pi = \{P_1, \ldots, P_K\}$ be a partition of $E = \{1, 2, \ldots, n, n+1\}$. A vector $\Theta = \big(\theta_1, \ldots, \theta_{n+1}\big)$ of random probabilities is neutral with respect to π if the vectors

$$\bar{S} = \left(\sum_{i\in P_1} \theta_i, \ldots, \sum_{i\in P_K} \theta_i\right),$$

$$\bar{U}_1 = \left(\frac{\theta_j}{\sum_{i\in P_1} \theta_i};\ j \in P_1\right), \ldots,\ \bar{U}_K = \left(\frac{\theta_j}{\sum_{i\in P_K} \theta_i};\ j \in P_K\right) \tag{1.1}$$

are independent.

Several other concepts of neutrality exist in literature. Originally it was introduced by Connor and Mosimann (1969) as a neutrality of a subvector in a unit-simplex-valued random vector. They defined also complete neutrality which was adapted to stochastic processes as neutrality-to-the-right (ntr) and neutrality-to-the-left by Doksum (1974). It is worth to mention that ntr process are widely investigated in recent years, see for instance Walker and Muliere (1999), Epifani, Lijoi and Prünster (2003) or Doksum and James (2004).

In Section 2 we will show that neutrality of sub-vector and complete neutrality can be expressed in terms of neutralities with respect to partitions, introduced in Definition 1.

Neutralities appeared to be useful in studying properties of the Dirichlet distribution and processes. In particular the concepts defined by Connor and Mosimann (1969) were used for characterizations of the Dirichlet distribution, see for instance Darroch and Ratcliff (1971), Fabius (1973), James and Mosimann (1980), Bobecka and Wesołowski (2007) and others. For a recent review see Gupta and Richards (2001).

Recall that a random vector $(\theta_1, \ldots, \theta_n)$ has a Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_{n+1})$, $(\theta_1, \ldots, \theta_n) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_{n+1})$, if it has the density of the form

$$f(x_1, \ldots, x_n) = \frac{\Gamma\big(\sum_{i=1}^{n+1} \alpha_i\big)}{\prod_{i=1}^{n+1} \Gamma\big(\alpha_i\big)} \big(1 - \sum_{i=1}^{n} x_i\big)^{\alpha_{n+1}-1} \prod_{i=1}^{n} x_i^{\alpha_i - 1} I_{T_n}(x_1, \ldots, x_n),$$

where $\alpha_i > 0$, $i = 1, \ldots, n+1$ and $T_n = \{(x_1, \ldots, x_n) \in (0, \infty)^n : \sum_{i=1}^n x_i < 1\}$.

In Section 2 we also point out the fact that, not unexpectedly, the Dirichlet vector, understood as a vector of random probabilities, is neutral with respect to any partition of the set of its indices.

In Section 3 we give a direct extension of a characterization of the Dirichlet distribution obtained by Geiger and Heckerman (1997) (referred to by GH in the sequel). They characterized a $k \times n$ Dirichlet random table $[\theta_{ij}]$ (where $\sum_{i,j} \theta_{ij} = 1$) by some independence conditions, which in view of Definition 1, can be seen as neutrality of $[\theta_{ij}]$ with respect to the row and column partitions of the set of indices of the table $[\theta_{ij}]$. GH proved their main result under the additional assumptions that the distribution, they wanted to characterize, has a strictly positive density on $(kn-1)$-dimensional unit simplex. Their approach to the problem was through solution of a rather complicated functional equation, see (4) in GH. They solved the equation using, first, advanced regularization techniques developed by JÁRAI (1986), and then, differentiations which led to complicated partial differential equations, see Sections A2–A5 in GH. Moreover, the argument they provided is split into three separate cases: $k = n = 2$, $k, n \geq 3$, and $k = 2$, $n \geq 3$ or $k \geq 3$, $n = 2$. Even more advanced analytical techniques allowed JÁRAI (1998) to omit the assumption of strict positivity of the density. In this context it is proper to refer to a recent monograph by JÁRAI (2005) devoted to regularization methods for functional equations with Ch. 23 (*Characterization of the Dirichlet distribution*, p. 275–284) devoted exclusively to the approach we have just mentioned.

The proof we offer in this paper is free of the density assumption and is based on the method of moments. Consequently, it does not need any sophisticated regularization methods, as those developed in JÁRAI (2005) and any smoothness properties, as differentiability used in GH. It is also universal, i.e. covers all the above cases by one approach. It is still based on solving functional equations, see (3.2) in Section 3 and (4.1) in Section 4, but the unknown moments functions involved in those equations are defined on discrete domains.

It is worth to mention that some of the equations appearing in our proof are quite similar to those which are used in characterizations of the Dirichlet law by Johnson's "sufficientness" postulate – see ZABEL (1982), LO (1991) or WALKER and MULIERE (1999).

In Section 4 we extend the result of Section 3 to multi-way tables of probabilities. The proof is to a large extent based on the argument developed for the 2-way tables, that is again the method of moments is employed, though it is not only the notation problem due to multi-dimensionality, we have to cope with. Though the result of Section 4, is a direct extension of Theorem 2, we think that for clarity of presentation it is reasonable, to have the intermediate step, we offer in Section 3. The same approach was adapted by GH, where, in Section 3 they offer another multivariate version of their Theorem 2. Their extension is based on notions of global and local independence of multinomial parameters for Bayesian networks. This property can be interpreted as neutralities with respect to partitions applied in a hierarchical way to the original vector of random probabilities, i.e. in a given step we have independent vectors of random probabilities, each of them being further partitioned producing a collection of independent vectors of random probabilities according to the rule of Definition 1. The proof, by reduction of the multivariate case to two-dimensional situation, is given in Geiger, Heckerman, Chickering (1995). That paper is a good reference for understanding the role of the Dirichlet distribution as a distribution of the parameters associated with each node in a Bayesian network.

Finally, we would like to mention yet another approach to the GH characterization (and its extension) published recently in Ramamoorthi and Sangalli (2007). Their proof of the characterization of the Dirichlet law is done completely within the Bayesian framework and uses rather hermetic Bayesian language.

2. Neutralities, partitions and Dirichlet

Independence properties of the Dirichlet distribution can be summarized by referring to neutrality property. Namely, it appears that the Dirichlet random vector is neutral with respect to any partition of the set of its indices. Moreover the sub-vectors associated with such a partition have again Dirichlet distributions. We explain this below.

Theorem 1. *Let $(\theta_1, \ldots, \theta_{n+1})$ be a Dirichlet vector of random probabilities, i.e. $(\theta_1, \ldots, \theta_n) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_{n+1})$ and $\sum_{j=1}^{n+1} \theta_j = 1$. For any partition π of $\{1, 2, \ldots, n, n+1\}$ the random vector $(\theta_1, \ldots, \theta_{n+1})$ is neutral with respect to π.*

Remark 1. It follows that the random vectors $\bar{S}, \bar{U}_1, \ldots, \bar{U}_K$, as defined in (1.1), have Dirichlet distributions:

$$\bar{S} \sim \mathrm{Dir}\left(\sum_{i \in P_j} \alpha_i;\ j = 1, \ldots, K\right),$$

$$\bar{U}_j \sim \mathrm{Dir}\left(\alpha_{i_1 + \cdots + i_{j-1} + l};\ l = 1, \ldots, i_j\right), \quad j = 1, \ldots, K.$$

The proofs of both Theorem 1 and Remark 1 follow either by direct computations involving densities or by the standard gamma representation of the Dirichlet distribution and are skipped.

Now we will show how the notion of neutrality defined by Connor and Mosimann (1969) and independence conditions in GH can be expressed in terms of Definition 1. In Connor and Mosimann (1969), the notion of neutrality of a subvector $(\theta_1,\dots,\theta_k)$ in $(\theta_1,\dots,\theta_n)$, where $(\theta_1,\dots,\theta_n)$ is a T_n-valued random vector, was defined as independence of

$$(\theta_1,\dots,\theta_k) \quad \text{and} \quad \left(\frac{\theta_{k+1}}{1-\sum_{i=1}^k \theta_i},\dots,\frac{\theta_n}{1-\sum_{i=1}^k \theta_i}\right), \quad \text{where } k\in\{1,\dots,n-1\}.$$

Note that in view of Definition 1 it is neutrality of a vector $\Theta=(\theta_1,\dots,\theta_{n+1})$ of random probabilities with respect to the partition $\pi=\{\{1\},\dots,\{k\},\{k+1,\dots,n+1\}\}$. Similarly, the notion of complete neutrality, which means independence of the components of the random vector

$$\left(\theta_1,\frac{\theta_2}{1-\theta_1},\frac{\theta_3}{1-\sum_{i=1}^2\theta_i},\dots,\frac{\theta_n}{1-\sum_{i=1}^{n-1}\theta_i}\right),$$

is equivalent to neutrality of a vector $\Theta=(\theta_1,\dots,\theta_{n+1})$ of random probabilities with respect to partitions $\pi_i=\{\{1\},\dots,\{i\},\{i+1,\dots,n+1\}\}$, $i=1,\dots,n-1$.

The independence conditions used in GH to characterize a $k\times n$ Dirichlet random table $[\theta_{ij}]$ (where $\sum_{i,j}\theta_{ij}=1$) were the following:

$$\begin{aligned}&\left(\sum_{j=1}^n\theta_{1j},\sum_{j=1}^n\theta_{2j},\dots,\sum_{j=1}^n\theta_{kj}\right),\left(\frac{\theta_M}{\sum_{j=1}^n\theta_{1j}};\ M\in\{(1,1),\dots,(1,n)\}\right),\dots\\&\qquad\dots,\left(\frac{\theta_M}{\sum_{j=1}^n\theta_{kj}};\ M\in\{(k,1),\dots,(k,n)\}\right)\end{aligned}\tag{2.1}$$

are mutually independent and

$$\begin{aligned}&\left(\sum_{i=1}^k\theta_{i1},\sum_{i=1}^k\theta_{i2},\dots,\sum_{i=1}^k\theta_{in}\right),\left(\frac{\theta_M}{\sum_{i=1}^k\theta_{i1}};\ M\in\{(1,1),\dots,(k,1)\}\right),\dots\\&\qquad\dots,\left(\frac{\theta_M}{\sum_{i=1}^k\theta_{in}};\ M\in\{(1,n),\dots,(k,n)\}\right)\end{aligned}\tag{2.2}$$

are mutually independent. Note that (2.1) is equivalent to neutrality of $[\theta_{ij}]$ with respect to the row partition $\{\{(1,1),\dots,(1,n)\},\dots,\{(k,1),\dots,(k,n)\}\}$ of the set of indices of the table $[\theta_{ij}]$. Similarly (2.2) is equivalent to neutrality of $[\theta_{ij}]$ with respect to the column partition $\{\{(1,1),\dots,(k,1)\},\dots,\{(1,n),\dots,(k,n)\}\}$.

Also other existing conditions characterizing the Dirichlet distribution can be formulated in terms of neutralities with respect to partitions. Darroch and

Ratcliff (1971) used neutralities with respect to partitions $\pi_i = \{\{1\},\ldots,\{i-1\},\{i+1\},\ldots,\{n\},\{i,n+1\}\}$, $i = 1,\ldots,n$. Fabius (1973) used moments method to refine their proof and gave a new result using partitions $\pi_i = \{\{i\},\{1,\ldots,i-1,i+1,\ldots,n,n+1\}\}$, $i = 1,\ldots,n$. James and Mossiman (1980) proved a characterization by neutralities with respect to $\pi_i = \{\{1\},\ldots,\{i\},\{i+1,\ldots,n+1\}\}$, $i = 1,\ldots,n-1$ and $\pi_n = \{\{n\},\{1,\ldots,n-1,n+1\}\}$. Their result has been recently generalized in Bobecka and Wesołowski (2007), where neutralities with respect to $\pi_i = \{\{1\},\ldots,\{i\},\{i+1,\ldots,n+1\}\}$, $i = 1,\ldots,n-1$, were combined with a regression version of neutrality with respect to $\pi_n = \{\{i_0\},\{i_0+1\},\ldots,\{n\},\{1,\ldots,i_0-1,n+1\}\}$, where $i_0 \in \{2,\ldots,n\}$ is fixed.

Remark 2. A Liouville distribution, see for instance Fang, Kotz and Ng (1989), is defined as any distribution of a random vector $(X_1,\ldots,X_n)$ if only

(1) $\underline{X} = (X_1+\cdots+X_n)^{-1}(X_1,\ldots,X_{n-1})$ is Dirichlet;

(2) $\underline{X}$ and $X_1+\cdots+X_n$ are independent.

Therefore such distributions have independence properties induced by the Dirichlet law. For instance, complete neutrality of the Dirichlet distribution implies that if $(X_1,\ldots,X_n)$ is a Louiville random vector, then the random variables $X_i/(X_i+\cdots+X_n)$, $i = 1,\ldots,n$, are independent. Moreover, the Liouville distribution can be characterized by independencies specific for the Dirichlet law written in terms of suitable properties of the vector $\underline{X}$ combined with independence of $\underline{X}$ and $X_1+\cdots+X_n$. On the other hand a Liouville random vector is completely neutral iff it is Dirichlet – see Proposition 6.4 in Gupta and Richards (1987). That paper and subsequent papers of these authors (see the references in Gupta and Richards (2001)), give a comprehensive study of Liouville distributions.

3. Row and column neutralities for two-way tables

In this section we formulate and prove a refinement of the GH characterization of the Dirichlet distribution in the sense that we do not assume existence of densities. The proof is much more elementary than the original one. It is based on identification of the Dirichlet distribution by multivariate moments of any order. But first, neutralities are translated into a functional equation for moments functions. The results of this section can be regarded as an intermediate step towards the investigation of multi-way tables, which is done in Section 4.

Theorem 2. *Let $[\theta_{ij}]$ be a $k \times n$ matrix with entries being positive random variables such that $\sum_{i,j} \theta_{ij} = 1$. Let*

$$\pi_1 = \{\{(1,1),\dots,(1,n)\},\dots,\{(k,1),\dots,(k,n)\}\} \qquad \textit{and}$$

$$\pi_2 = \{\{(1,1),\dots,(k,1)\},\dots,\{(1,n),\dots,(k,n)\}\}$$

be partitions of the set $\{(i,j);\ i = 1,\dots,k,\ j = 1,\dots,n\}$ of indices of $[\theta_{ij}]$. If $[\theta_{ij}]$ is neutral with respect to the partitions π_1 and π_2 then it is a Dirichlet random probabilities table.

Proof. As already observed in Section 2, the neutrality conditions with respect to partitions π_1 and π_2 are equivalent to (2.1) and (2.2). For any natural (r_{ij}), $i = 1,\dots,k$, $j = 1,\dots,n$, from these conditions we obtain two different factorizations of the joint moment

$$\begin{aligned} E\left[\prod_{i=1}^{k}\prod_{j=1}^{n}\theta_{ij}^{r_{ij}}\right] &= E\left[\prod_{j=1}^{n}\left(\frac{\theta_{1j}}{R_1}\right)^{r_{1j}}\right]\dots E\left[\prod_{j=1}^{n}\left(\frac{\theta_{kj}}{R_k}\right)^{r_{kj}}\right]E[R_1^{s_1}\dots R_k^{s_k}] \\ &= E\left[\prod_{i=1}^{k}\left(\frac{\theta_{i1}}{C_1}\right)^{r_{i1}}\right]\dots E\left[\prod_{i=1}^{k}\left(\frac{\theta_{in}}{C_n}\right)^{r_{in}}\right]E[C_1^{t_1}\dots C_n^{t_n}], \qquad (3.1) \end{aligned}$$

where

$$R_i = \sum_{j=1}^{n}\theta_{ij}, \quad s_i = \sum_{j=1}^{n} r_{ij}, \quad i = 1,\dots,k, \qquad \text{and}$$

$$C_j = \sum_{i=1}^{k}\theta_{ij}, \quad t_j = \sum_{i=1}^{k} r_{ij}, \quad j = 1,\dots,n.$$

For arbitrary vectors $\bar{m} = (m_1,\dots,m_n)$ and $\bar{l} = (l_1,\dots,l_k)$ with non-negative integer components we define

$$f_i(\bar{m}) = E\left[\prod_{j=1}^{n}\left(\frac{\theta_{ij}}{R_i}\right)^{m_j}\right], \quad i = 1,2,\dots,k, \qquad F(\bar{l}) = E\left[\prod_{i=1}^{k} R_i^{l_i}\right],$$

$$g_j(\bar{l}) = E\left[\prod_{i=1}^{k}\left(\frac{\theta_{ij}}{C_j}\right)^{l_i}\right], \quad j = 1,2,\dots,n, \qquad G(\bar{m}) = E\left[\prod_{j=1}^{n} C_j^{m_j}\right].$$

Inserting the above definitions in (3.1) we get our basic functional equation

$$\begin{aligned} &\left[\prod_{i=1}^{k} f_i(r_{i1}, r_{i2},\dots,r_{in})\right] F\big(s_1, s_2,\dots,s_k\big) \\ &\qquad = \left[\prod_{j=1}^{n} g_j(r_{1j}, r_{2j},\dots,r_{kj})\right] G(t_1,t_2,\dots,t_n), \qquad (3.2) \end{aligned}$$

for any $k \times n$ matrix $r = [r_{ij}]$ of non-negative integers $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$.

We rewrite (3.2) as

$$\left[\prod_{i=1}^{k} f_i(\bar{r}_{i\cdot})\right] F(\bar{s}) = \left[\prod_{j=1}^{n} g_j(\bar{r}_{\cdot j})\right] G(\bar{t}), \tag{3.3}$$

where $\bar{r}_{i\cdot} = (r_{i1}, \dots, r_{in})$, $\bar{r}_{\cdot j} = (r_{1j}, \dots, r_{kj})$ and $\bar{s} = (s_1, \dots, s_k)$, $\bar{t} = (t_1, \dots, t_n)$. Dividing two equations (3.3), taken in $r + \xi_{pj}$ and $r + \xi_{qj}$ for some fixed $p \neq q$ and $j \in \{1, \dots, n\}$, where ξ_{ij} is the matrix with all entries equal zero except the ij-th entry which equals 1, we get

$$\frac{f_p(\bar{r}_{p\cdot} + \bar{\delta}_j)}{f_p(\bar{r}_{p\cdot})} \frac{f_q(\bar{r}_{q\cdot})}{f_q(\bar{r}_{q\cdot} + \bar{\delta}_j)} \frac{F(\bar{s} + \bar{\varepsilon}_p)}{F(\bar{s} + \bar{\varepsilon}_q)} = \frac{g_j(\bar{r}_{\cdot j} + \bar{\varepsilon}_p)}{g_j(\bar{r}_{\cdot j} + \bar{\varepsilon}_q)},$$

where $(\bar{\delta}_j)_{j=1,\dots,n}$, $(\bar{\varepsilon}_i)_{i=1,\dots,k}$ are the canonical basis of $\mathbf{R}^n$ and $\mathbf{R}^k$, respectively. In this equation we insert zeros for all r_{ij}'s except the lth column, where $l \neq j$ is taken arbitrarily. We denote it by $\bar{l}$. Thus we get

$$\frac{F(\bar{l} + \bar{\varepsilon}_p)}{F(\bar{l} + \bar{\varepsilon}_q)} = \frac{\alpha_p(l_p)}{\alpha_q(l_q)}, \tag{3.4}$$

where

$$\alpha_i(x) = \frac{f_i(x\ \bar{\delta}_l)}{f_i(x\bar{\delta}_l + \bar{\delta}_j)} g_j(\bar{\varepsilon}_i), \quad i = 1, \dots, k.$$

Observe that

$$f_i(\bar{m}) = \sum_{j=1}^{n} f_i\big(\bar{m} + \bar{\delta}_j\big), \quad i = 1, 2, \dots, k, \tag{3.5}$$

$$F(\bar{l}) = \sum_{i=1}^{k} F\big(\bar{l} + \bar{\varepsilon}_i\big), \tag{3.6}$$

$$g_j(\bar{l}) = \sum_{i=1}^{k} g_j\big(\bar{l} + \bar{\varepsilon}_i\big), \qquad j = 1, 2, \dots, n,$$

$$G\big(\bar{m}\big) = \sum_{j=1}^{n} G\big(\bar{m} + \bar{\delta}_j\big),$$

where $\bar{m} = \big(m_1, m_2, \dots, m_n\big)$ and $\bar{l} = \big(l_1, l_2, \dots, l_k\big)$ are arbitrary vectors with nonnegative integer components. From (3.6) we get for any $p \in \{1, \dots, k\}$

$$F(\bar{l}) = \frac{F(\bar{l} + \bar{\varepsilon}_p)}{\alpha_p(l_p)} \sum_{q=1}^{k} \alpha_q(l_q)$$

for any k-dimensional vector $\bar{l}$ with nonnegative integer components. The above equation can be written as

$$F(\bar{l}) = \frac{\alpha_p(l_p-1)}{\alpha_p(l_p-1) + \sum_{w=1,w\neq p}^{k} \alpha_w(l_w)} F(\bar{l} - \bar{\varepsilon}_p) \tag{3.7}$$

if only $l_p \geq 1$. We repeat this procedure until at the right hand side the pth component of the argument of F becomes zero, getting

$$F(\bar{l}) = \prod_{u=0}^{l_p-1} \frac{\alpha_p(u)}{\alpha_p(u) + \sum_{w=1,w\neq p}^{k} \alpha_w(l_w)} F(\bar{l} - l_p\bar{\varepsilon}_p).$$

After iterating wrt to the qth component we get

$$\begin{aligned} F(\bar{l}) = & \prod_{u=0}^{l_p-1} \frac{\alpha_p(u)}{\alpha_p(u) + \sum_{w=1,w\neq p}^{k} \alpha_w(l_w)} \\ & \times \prod_{v=0}^{l_q-1} \frac{\alpha_q(v)}{\alpha_p(0) + \alpha_q(v) + \sum_{w=1,w\neq p,q}^{k} \alpha_w(l_w)} F(\bar{l} - l_p\bar{\varepsilon}_p - l_q\bar{\varepsilon}_q). \end{aligned} \tag{3.8}$$

Changing the role of p and q we get

$$\begin{aligned} F(\bar{l}) = & \prod_{v=0}^{l_q-1} \frac{\alpha_q(v)}{\alpha_q(v) + \sum_{w=1,w\neq q}^{k} \alpha_w(l_w)} \\ & \times \prod_{u=0}^{l_p-1} \frac{\alpha_p(u)}{\alpha_q(0) + \alpha_p(u) + \sum_{w=1,w\neq p,q}^{k} \alpha_w(l_w)} F(\bar{l} - l_p\bar{\varepsilon}_p - l_q\bar{\varepsilon}_q). \end{aligned} \tag{3.9}$$

Comparing (3.8) and (3.9) we have

$$\begin{aligned} & \prod_{u=0}^{l_p-1} \left[\alpha_p(u) + \sum_{w=1,w\neq p}^{k} \alpha_w(l_w)\right] \prod_{v=0}^{l_q-1} \left[\alpha_p(0) + \alpha_q(v) + \sum_{w=1,w\neq p,q}^{k} \alpha_w(l_w)\right] \\ = & \prod_{v=0}^{l_q-1} \left[\alpha_q(v) + \sum_{w=1,w\neq q}^{k} \alpha_w(l_w)\right] \prod_{u=0}^{l_p-1} \left[\alpha_q(0) + \alpha_p(u) + \sum_{w=1,w\neq p,q}^{k} \alpha_w(l_w)\right]. \end{aligned} \tag{3.10}$$

Inserting $l_p = l_q = 1$ we get

$$\alpha_p(1) - \alpha_p(0) = \alpha_q(1) - \alpha_q(0). \tag{3.11}$$

We will prove by induction that for any $j \geq 1$ and any $p \in \{1, \dots, k\}$

$$\alpha_p(j) - \alpha_p(j-1) = \alpha_p(1) - \alpha_p(0). \tag{3.12}$$

The statement for $j = 1$ is a tautology. We assume that it holds for any $j = 1, \dots, l$ and prove it for $j = l+1$. We write (3.10) for $l_p = l+1$ and $l_q = 1$

$$\left[\prod_{u=0}^{l}(\alpha_p(u) + \alpha_q(1) + A)\right][\alpha_p(0) + \alpha_q(0) + A]$$
$$= [\alpha_p(l+1) + \alpha_q(0) + A]\prod_{u=0}^{l}[\alpha_p(u) + \alpha_q(0) + A],$$

where

$$A = \sum_{w=1, w\neq p,q}^{k} \alpha_w(l_w).$$

By the induction assumption we get

$$\left[\prod_{u=0}^{l-1}(\alpha_p(u) + \alpha_q(1) + A)\right][\alpha_p(0) + \alpha_q(0) + A] = \prod_{u=0}^{l}[\alpha_p(u) + \alpha_q(0) + A].$$

Hence

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

Thus the formula (3.12) is proved. Consequently, there exist constants a_p and b_p such that

$$\alpha_p(j) = a_p j + b_p \tag{3.13}$$

for any $p \in \{1, \dots, k\}$ and any $j \geq 0$. Moreover, by (3.11) it follows that

$$a_p = a_q = a. \tag{3.14}$$

Note that taking the original formula (3.7) for $p = 1$ and then iterating with respect to subsequent variables we obtain

$$F(\bar{l}) = \prod_{i=1}^{k}\prod_{u=0}^{l_i-1}\frac{\alpha_i(u)}{\sum_{w=1}^{i-1}\alpha_w(0) + \alpha_i(u) + \sum_{w=i+1}^{k}\alpha_w(l_w)}. \tag{3.15}$$

Inserting (3.13) into (3.15), and using also (3.14) we have

$$
\begin{aligned}
F(\bar{l}) &= \prod_{i=1}^{k} \prod_{u=0}^{l_i-1} \frac{au+b_i}{\sum_{w=1}^{i-1} b_w + au + b_i + \sum_{w=i+1}^{k}(al_w+b_w)} \\
&= \prod_{i=1}^{k} \prod_{u=0}^{l_i-1} \frac{u+\frac{b_i}{a}}{\sum_{w=1}^{k} \frac{b_w}{a} + u + \sum_{w=i+1}^{k} l_w}.
\end{aligned}
$$

Thus

$$
F(\bar{l}) = \frac{\Gamma\big(\sum_{w=1}^{k} d_w\big)}{\Gamma\big(\sum_{w=1}^{k}(l_w+d_w)\big)} \prod_{i=1}^{k} \frac{\Gamma(l_i+d_i)}{\Gamma(d_i)}, \tag{3.16}
$$

where $d_w = b_w/a \geq 0$, $w = 1,\dots,k$.

Similarly using (3.3) we can find G, i.e. there exist constants $c_v \geq 0$, $v = 1,\dots,n$, such that

$$
G(\bar{m}) = \frac{\Gamma\big(\sum_{v=1}^{n} c_v\big)}{\Gamma\big(\sum_{v=1}^{n}(m_v+c_v)\big)} \prod_{j=1}^{n} \frac{\Gamma(m_j+c_j)}{\Gamma(c_j)}. \tag{3.17}
$$

To find the remaining functions f_i's and g_j's we return to (3.3) again. We divide two versions of (3.3), one taken in $r+\xi_{ij_1}$ and the other taken in $r+\xi_{ij_2}$, for some fixed $j_1 \neq j_2$ and $i \in \{1,\dots,k\}$. Thus we get

$$
\frac{f_i(\bar{r}_{i\cdot}+\bar{\delta}_{j_1})}{f_i(\bar{r}_{i\cdot}+\bar{\delta}_{j_2})} = \frac{g_{j_1}(\bar{r}_{\cdot j_1}+\bar{\varepsilon}_i)}{g_{j_1}(\bar{r}_{\cdot j_1})} \frac{g_{j_2}(\bar{r}_{\cdot j_2})}{g_{j_2}(\bar{r}_{\cdot j_2}+\bar{\varepsilon}_i)} \frac{G(\bar{t}+\bar{\delta}_{j_1})}{G(\bar{t}+\bar{\delta}_{j_2})}.
$$

In the above equation we insert zeros for all the entries of r except for the i-th row. Using (3.17) we conclude that

$$
\frac{f_i(\bar{r}_{i\cdot}+\bar{\delta}_{j_1})}{f_i(\bar{r}_{i\cdot}+\bar{\delta}_{j_2})} = \frac{\beta_{j_1}(r_{ij_1})}{\beta_{j_2}(r_{ij_2})},
$$

where β_{j_1} and β_{j_2} are some functions. Now, in view of (3.5), following the argument used in derivation of the formula for F, we arrive at

$$
f_i(\bar{m}) = \frac{\Gamma\big(\sum_{v=1}^{n} \mu_{iv}\big)}{\Gamma\big(\sum_{v=1}^{n}(m_v+\mu_{iv})\big)} \prod_{j=1}^{n} \frac{\Gamma(m_j+\mu_{ij})}{\Gamma(\mu_{ij})}, \quad i = 1,\dots,k, \tag{3.18}
$$

with $\mu_{iv} \geq 0$, $i = 1,\dots,k$, $v = 1,\dots,n$.

Analogously,

$$
g_j(\bar{l}) = \frac{\Gamma\big(\sum_{w=1}^{k} \nu_{wj}\big)}{\Gamma\big(\sum_{w=1}^{k}(l_w+\nu_{wj})\big)} \prod_{i=1}^{k} \frac{\Gamma(l_i+\nu_{ij})}{\Gamma(\nu_{ij})}, \quad j = 1,\dots,n, \tag{3.19}
$$

with $\nu_{wj} \geq 0$, $w = 1, \dots, k$, $j = 1, \dots, n$.

To identify parameters we rewrite (3.3) for arbitrary $\bar{r}_{1\cdot}$ and for other entries of the matrix r equal zero, using (3.18), (3.16), (3.19) and (3.17):

$$\frac{\Gamma\big(\sum_{v=1}^{n}\mu_{1v}\big)}{\Gamma\big(\sum_{v=1}^{n}(r_{1v}+\mu_{1v})\big)}\left(\prod_{j=1}^{n}\frac{\Gamma(r_{1j}+\mu_{1j})}{\Gamma(\mu_{1j})}\right)\frac{\Gamma\big(\sum_{w=1}^{k}d_w\big)}{\Gamma\big(\sum_{j=1}^{n}r_{1j}+\sum_{w=1}^{k}d_w\big)}$$
$$\times\frac{\Gamma(\sum_{j=1}^{n}r_{1j}+d_1)}{\Gamma(d_1)}=\left(\prod_{j=1}^{n}\frac{\Gamma\big(\sum_{w=1}^{k}\nu_{wj}\big)}{\Gamma\big(r_{1j}+\sum_{w=1}^{k}\nu_{wj}\big)}\frac{\Gamma(r_{1j}+\nu_{1j})}{\Gamma(\nu_{1j})}\right)$$
$$\times\frac{\Gamma\big(\sum_{v=1}^{n}c_v\big)}{\Gamma\big(\sum_{v=1}^{n}(r_{1v}+c_v)\big)}\prod_{j=1}^{n}\frac{\Gamma(r_{1j}+c_j)}{\Gamma(c_j)}. \tag{3.20}$$

Now, we leave only two non-zero components of the vector $\bar{r}_{1\cdot}$. Without loss of generality we assume that r_{11} and r_{12} are non-zero. Then the above equation takes the form

$$\frac{\Gamma\big(\sum_{v=1}^{n}\mu_{1v}\big)}{\Gamma\big(r_{11}+r_{12}+\sum_{v=1}^{n}\mu_{1v}\big)}\left(\prod_{j=1}^{2}\frac{\Gamma(r_{1j}+\mu_{1j})}{\Gamma(\mu_{1j})}\right)\frac{\Gamma\big(\sum_{w=1}^{k}d_w\big)}{\Gamma\big(r_{11}+r_{12}+\sum_{w=1}^{k}d_w\big)}$$
$$\times\frac{\Gamma(r_{11}+r_{12}+d_1)}{\Gamma(d_1)}=\left(\prod_{j=1}^{2}\frac{\Gamma\big(\sum_{w=1}^{k}\nu_{wj}\big)}{\Gamma\big(r_{1j}+\sum_{w=1}^{k}\nu_{wj}\big)}\frac{\Gamma(r_{1j}+\nu_{1j})}{\Gamma(\nu_{1j})}\right)$$
$$\times\frac{\Gamma\big(\sum_{v=1}^{n}c_v\big)}{\Gamma\big(r_{11}+r_{12}+\sum_{v=1}^{n}c_v\big)}\prod_{j=1}^{2}\frac{\Gamma(r_{1j}+c_j)}{\Gamma(c_j)}. \tag{3.21}$$

From the above equation it follows that

$$h(r_{11}+r_{12})=\frac{\Gamma(r_{11}+r_{12}+d_1)\Gamma\big(r_{11}+r_{12}+\sum_{v=1}^{n}c_v\big)}{\Gamma\big(r_{11}+r_{12}+\sum_{v=1}^{n}\mu_{1v}\big)\Gamma\big(r_{11}+r_{12}+\sum_{w=1}^{k}d_w\big)}$$

is a product of a function of r_{11} and of a function of r_{12}. Consequently $h(i) = \alpha\beta^i$, since it satisfies a multiplicative version of the Pexider equation, see e.g. ACZÉL (1966), Ch. 3. On the other hand, by its definition, h is a rational function. Thus $h(i) = \alpha$, for any $i \geq 0$. By inspecting the form of h we conclude that either $d_1 = \sum_{v=1}^{n}\mu_{1v}$ or $d_1 = \sum_{w=1}^{k}d_w$. Note that the second identity, which means $d_w = 0$, $w = 2, \dots, k$, is impossible. To see this take $\bar{l} = (l_1, 0, \dots, 0)$ in (3.16). Then we get $F(l_1, 0, \dots, 0) = 1$ for any l_1. On the other hand, by the definition $F(l_1, 0, \dots, 0) = E\big[\big(\sum_{j=1}^{n}\theta_{1j}\big)^{l_1}\big]$, which is always < 1. Thus, we get

$$\sum_{v=1}^{n}c_v=\sum_{w=1}^{k}d_w \quad\text{and}\quad d_1=\sum_{v=1}^{n}\mu_{1v}.$$

Since the first row $\bar{r}_{1\cdot}$ was taken arbitrarily we conclude also that

$$d_i = \sum_{v=1}^{n} \mu_{iv}, \quad i = 1, \ldots, k.$$

By symmetry of (3.20) we get also

$$c_j = \sum_{w=1}^{n} \nu_{wj}, \quad j = 1, \ldots, n. \tag{3.22}$$

We return now to (3.21) getting

$$\prod_{j=1}^{2} \frac{\Gamma(r_{1j} + \mu_{1j})}{\Gamma(\mu_{1j})} = \left(\prod_{j=1}^{2} \frac{\Gamma\big(\sum_{w=1}^{k} \nu_{wj}\big)}{\Gamma\big(r_{1j} + \sum_{w=1}^{k} \nu_{wj}\big)} \frac{\Gamma(r_{1j} + \nu_{1j})}{\Gamma(\nu_{1j})} \right) \prod_{j=1}^{2} \frac{\Gamma(r_{1j} + c_j)}{\Gamma(c_j)}.$$

Using (3.22) we obtain

$$\prod_{j=1}^{2} \frac{\Gamma(r_{1j} + \mu_{1j})}{\Gamma(\mu_{1j})} = \prod_{j=1}^{2} \frac{\Gamma(r_{1j} + \nu_{1j})}{\Gamma(\nu_{1j})}.$$

Consequently $\mu_{1j} = \nu_{1j}$, $j = 1, 2$, and by symmetry we conclude that $\mu_{ij} = \nu_{ij}$ for any $i \in \{1, \ldots, k\}$ and any $j \in \{1, \ldots, n\}$.

Now we are in a position to obtain a general expression for any joint moment of any order of the elements of the matrix $[\theta_{ij}]$. Combining (3.1) and (3.3) we get

$$\begin{aligned} E\left[\prod_{i=1}^{k} \prod_{j=1}^{n} \theta_{ij}^{r_{ij}} \right] &= \left[\prod_{i=1}^{k} f_i(\bar{r}_{i\cdot}) \right] F(\bar{s}) \\ &= \left[\prod_{i=1}^{k} \prod_{j=1}^{n} \frac{\Gamma(r_{ij} + \mu_{ij})}{\Gamma(\mu_{ij})} \right] \frac{\Gamma\big(\sum_{i=1}^{k} \sum_{j=1}^{n} \mu_{ij}\big)}{\Gamma\big(\sum_{i=1}^{k} \sum_{j=1}^{n} (r_{ij} + \mu_{ij})\big)}, \end{aligned}$$

which is the joint moment of order $[r_{ij}]$ for the Dirichlet matrix with the parameter $\mu = [\mu_{ij}]$. Since the Dirichlet distribution is identified by its moments the proof is completed. □

Remark 3. Let $(\theta_1, \ldots, \theta_L)$ be a vector of random probabilities and

$$P(X_1 = k_1, \ldots, X_n = k_n | \theta_1, \ldots, \theta_L) = \theta_1^{n_1} \ldots \theta_L^{n_L}$$

for any $n = 1, 2, \ldots$ and any $k_1, \ldots, k_n \in \{1, \ldots, L\}$, where $n_j = \sum_{i=1}^{n} I_j(k_i)$, i.e. n_j is the number of k_i's equal to j, $j = 1, \ldots, L$. Then Johnson's "sufficientness" postulate (see Zabel (1982)) has the form

$$P\big(X_{n+1} = k | X_1 = k_1, \ldots, X_n = k_n\big) = f\big(k, n_k, n\big) \tag{3.23}$$

for any $k = 1, \dots, L$, any $n = 1, 2, \dots$ and any $k_1, \dots, k_n \in \{1, \dots, L\}$, where f is an unknown function. Zabel (1982) characterized the Dirichlet distribution (for $(\theta_1, \dots, \theta_L)$) by the condition (3.23). This result was extended to the Dirichlet process by Lo (1991) and for the ntr processes by Walker and Muliere (1999).

Note that (3.23) can be equivalently rewritten as

$$\frac{E(\theta_1^{n_1} \dots \theta_k^{n_k+1} \dots \theta_L^{n_L})}{E(\theta_1^{n_1} \dots \theta_k^{n_k} \dots \theta_L^{n_L})} = f(k, n_k, n)$$

for any $k = 1, \dots, L$, any $n_1, \dots, n_L \in \{0, \dots, n\}$ and any $n = 1, 2, \dots$, which implies

$$\frac{E(\theta_1^{n_1} \dots \theta_k^{n_k+1} \dots \theta_L^{n_L})}{E(\theta_1^{n_1} \dots \theta_j^{n_j+1} \dots \theta_L^{n_L})} = \frac{f(k, n_k, n)}{f(j, n_j, n)} \tag{3.24}$$

for any j, k different. The last condition is quite similar to

$$\frac{E(\theta_1^{n_1} \dots \theta_k^{n_k+1} \dots \theta_L^{n_L})}{E(\theta_1^{n_1} \dots \theta_j^{n_j+1} \dots \theta_L^{n_L})} = \frac{\alpha_k(n_k)}{\alpha_j(n_j)}. \tag{3.25}$$

Note that (3.25) is equivalent to (3.4). Thus from the argument in the proof above (the function F which stands for $E(\theta_1^{n_1} \dots \theta_k^{n_k} \dots \theta_L^{n_L})$ has been completely identified there) it follows that (3.25) characterizes the Dirichlet distribution. On the other hand we doubt whether (3.24), which is visibly weaker then (3.25), suffices to obtain the characterization.

4. Multi-way tables

In this section we present the main result which extends Theorem 2 of GH type to multi-way tables. The proof to a large extent uses the argument developed in Section 3 for two-way tables. It appears that for k-way tables neutralities with respect to natural k partitions characterizes the Dirichlet distribution. The technical part of the proof relies on a functional equation (4.1) which is a multivariate version of the equation (3.2). Another extension is given in GH.

Theorem 3. *Let $\mathcal{I} = \{\bar{\imath} = (i_1, i_2, \dots, i_k),\ i_j = 1, 2, \dots, I_j,\ j = 1, 2, \dots, k\}$ be the set of indices of a k-way table of random probabilities $\Theta = [\theta_{\bar{\imath}},\ \bar{\imath} \in \mathcal{I}]$. Let π_j, $j = 1, 2, \dots, k$, be partitions of the set $\mathcal{I}$ of the following form*

$$\pi_j = \{P_1^{(j)}, P_2^{(j)}, \dots P_{I_j}^{(j)}\},$$

where

$$P_l^{(j)} = \{\bar{\imath} \in \mathcal{I} : i_j = l\} \quad \textit{for } l = 1, 2, \ldots, I_j.$$

If Θ is neutral with respect to π_j, $j = 1, 2, \ldots, k$, then it is a Dirichlet k-way table.

This result is rather different than the extension given in GH (see also Heckerman, Geiger and Chickering (1995)). That one was based on independence conditions related to nested partitions, called global and local independence, defined for random probabilities $\Theta = (\theta_{\bar{\imath}})_{\bar{\imath} \in \mathcal{I}}$, in the following way: For a complete undirected graph G having k nodes we associate with the node i a random variable X_i, $i = 1, \ldots, k$, such that

$$P(X_1 = i_1, \ldots, X_k = i_k | \Theta) = \theta_{\bar{\imath}} \quad \text{for } \bar{\imath} = (i_1, i_2, \ldots, i_k) \in \mathcal{I}.$$

Define independence for Θ related to the ordered path $(j_1, \ldots, j_s)$ in G, $s = 1, \ldots, k$, as joint independence of the vectors

$$\big(P\big(X_{j_l} = i | X_{j_1} = i_1, \ldots, X_{j_{l-1}} = i_{l-1}, \Theta\big)\big)_{i=1,\ldots,I_{j_l}},$$
$$i_m = 1, \ldots, I_m,\ m = 1, \ldots, l-1,\ l = 1, \ldots, s.$$

Global and local independence for Θ amounts to independence for Θ related to any ordered path of length k in G. Theorem 3 of GH states that under smoothness assumptions, global and local independence for Θ related to two ordered paths $(1, 2, \ldots, k)$ and $(k, 1, \ldots, k-1)$ implies that Θ is Dirichlet.

In this language assumptions of our Theorem 3 above can be expressed as independencies for Θ related to all degenerate paths of length one in G, i.e. the paths (s), $s = 1, \ldots, k$.

Proof of Theorem 3. Decomposing the joint moment of the order $\mathcal{R} = (r_{\bar{\imath}},\ \bar{\imath} \in \mathcal{I})$ similarly as in the proof of Theorem 2 we arrive at the following system of functional equations

$$F^{(j)}(s_1^{(j)}, s_2^{(j)}, \ldots, s_{I_j}^{(j)}) \prod_{i=1}^{I_j} f_i^{(j)}(r_{\bar{\imath}},\ \bar{\imath} \in P_i^{(j)}) = E\left[\prod_{\bar{\imath} \in \mathcal{I}} \theta_{\bar{\imath}}^{r_{\bar{\imath}}}\right], \qquad (4.1)$$
$$j = 1, 2, \ldots, k,$$

where $s_i^{(j)} = \sum_{\bar{\imath} \in P_i^{(j)}} r_{\bar{\imath}}$, $i = 1, \ldots, I_j$.

For any fixed $l \in \{1, 2, \ldots, k\}$ we consider the above system of equations for $\mathcal{R} + \mathcal{E}_{\bar{m}^{(l)}}$ and $\mathcal{R} + \mathcal{E}_{\bar{n}^{(l)}}$, where $\mathcal{E}_{\bar{m}^{(l)}}$ and $\mathcal{E}_{\bar{n}^{(l)}}$ are k-way tables with all zeros except for the elements with multi-indices $\bar{m}^{(l)} = (i_1^{(0)}, \dots, i_{l-1}^{(0)}, m, i_{l+1}^{(0)}, \dots, i_k^{(0)})$ and $\bar{n}^{(l)} = (i_1^{(0)}, \dots, i_{l-1}^{(0)}, n, i_{l+1}^{(0)}, \dots, i_k^{(0)})$ which are equal one, respectively. Dividing one of these equations by the other for any $j = 1, \dots, k$, we get

$$\begin{aligned} &\frac{F^{(l)}\big(s_1^{(l)}, \dots, s_m^{(l)} + 1, \dots, s_{I_l}^{(l)}\big)}{F^{(l)}\big(s_1^{(l)}, \dots, s_n^{(l)} + 1, \dots, s_{I_l}^{(l)}\big)} \frac{f_m^{(l)}\big(r_{\bar\imath} + \delta_{\bar\imath}(\bar m^{(l)}),\ \bar\imath \in P_m^{(l)}\big)}{f_m^{(l)}\big(r_{\bar\imath},\ \bar\imath \in P_m^{(l)}\big)} \\ &\times \frac{f_n^{(l)}\big(r_{\bar\imath},\ \bar\imath \in P_n^{(l)}\big)}{f_n^{(l)}\big(r_{\bar\imath} + \delta_{\bar\imath}(\bar n^{(l)}),\ \bar\imath \in P_n^{(l)}\big)} = \frac{f_{i_j^{(0)}}^{(j)}\big(r_{\bar\imath} + \delta_{\bar\imath}(\bar m^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\big)}{f_{i_j^{(0)}}^{(j)}\big(r_{\bar\imath} + \delta_{\bar\imath}(\bar n^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\big)} \end{aligned} \tag{4.2}$$

for any $j \in \{1, 2, \dots, k\} \setminus \{l\}$, where $\delta_{\bar\imath}(\bar\jmath) = 0$ if $\bar\imath \neq \bar\jmath$ and $\delta_{\bar\imath}(\bar\imath) = 1$.

The part of the proof, which leads to identification of functions $F^{(j)}$, $j = 1, \dots, k$, translates the argument used in the proof of Theorem 2 into the multivariate setting and for a function $F^{(l)}$ uses a single equation only. To this end, in the above system we insert $r_{\bar\imath} = 0$ for all $\bar\imath$-th except $\bar\imath_g = (i_1^{(1)}, \dots, i_{l-1}^{(1)}, g, i_{l+1}^{(1)}, \dots, i_k^{(1)})$, $g = 1, \dots, I_l$, where $i_j^{(1)} \neq i_j^{(0)}$, $j \in \{1, 2, \dots, k\} \setminus \{l\}$, getting

$$\begin{aligned} &\frac{F^{(l)}\big(r_{\bar\imath_1}, \dots, r_{\bar\imath_m} + 1, \dots, r_{\bar\imath_{I_l}}\big)}{F^{(l)}\big(r_{\bar\imath_1}, \dots, r_{\bar\imath_n} + 1, \dots, r_{\bar\imath_{I_l}}\big)} \frac{f_m^{(l)}\big(r_{\bar\imath_m}\delta_{\bar\imath}(\bar\imath_m) + \delta_{\bar\imath}(\bar m^{(l)}),\ \bar\imath \in P_m^{(l)}\big)}{f_m^{(l)}\big(r_{\bar\imath_m}\delta_{\bar\imath}(\bar\imath_m),\ \bar\imath \in P_m^{(l)}\big)} \\ &\times \frac{f_n^{(l)}\big(r_{\bar\imath_n}\delta_{\bar\imath}(\bar\imath_n),\ \bar\imath \in P_n^{(l)}\big)}{f_n^{(l)}\big(r_{\bar\imath_n}\delta_{\bar\imath}(\bar\imath_n) + \delta_{\bar\imath}(\bar n^{(l)}),\ \bar\imath \in P_n^{(l)}\big)} = \frac{f_{i_j^{(0)}}^{(j)}\big(\delta_{\bar\imath}(\bar m^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\big)}{f_{i_j^{(0)}}^{(j)}\big(\delta_{\bar\imath}(\bar n^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\big)} \end{aligned}$$

for some $j \neq l$. Thus

$$\frac{F^{(l)}\big(x_1, \dots x_m + 1, \dots, x_{I_l}\big)}{F^{(l)}\big(x_1, \dots x_n + 1, \dots, x_{I_l}\big)}$$

is a function of the quotient of a function of x_m and a function of x_n for any m and n different. Moreover,

$$F^{(l)}\big(\bar x\big) = \sum_{i=1}^{I_l} F^{(l)}\big(\bar x + \bar e_i^{(l)}\big)$$

for any $\bar x = (x_1, \dots, x_{I_l})$, where $\bar e_i^{(l)}$ is the I_l-dimensional vector with all entries equal to zero except i-th entry, which is equal one, $i = 1, \dots, I_l$. Hence as in the proof of Theorem 2 we get

$$F^{(l)}\big(\bar x\big) = \frac{\Gamma\big(\sum_{i=1}^{I_l} d_i^{(l)}\big)}{\Gamma\big(\sum_{i=1}^{I_l}\big(d_i^{(l)} + x_i\big)\big)} \prod_{i=1}^{I_l} \frac{\Gamma\big(d_i^{(l)} + x_i\big)}{\Gamma\big(d_i^{(l)}\big)}, \quad l = 1, \dots, k. \tag{4.3}$$

Now we will identify functions $f_i^{(j)}$'s. Without loss of generality we will consider the whole system (4.2) and identify the functions $f_{i_j^{(0)}}^{(j)}$, $j \in \{1,\dots,k\}\setminus\{l\}$ for an arbitrary but fixed l.

Note that (4.3) implies that the left hand side of (4.2) is a function of $\big(r_{\bar\imath}, \bar\imath \in P_m^{(l)} \cup P_n^{(l)}\big)$, and the right hand sides are functions of $\big(r_{\bar\imath}, \bar\imath \in P_{i_j^{(0)}}^{(j)}\big)$ for $j \in \{1,\dots,k\} \setminus \{l\}$. Since

$$P_m^{(l)} \cap \bigcap_{j\in\{1,\dots,k\}\setminus\{l\}} P_{i_j^{(0)}}^{(j)} = \big\{\bar m^{(1)}\big\}, \quad P_n^{(1)} \cap \bigcap_{j\in\{1,\dots,k\}\setminus\{l\}} P_{i_j^{(0)}}^{(j)} = \big\{\bar n^{(1)}\big\},$$

we conclude that each side of the system (4.2), in particular,

$$\frac{f_{i_j^{(0)}}^{(j)}\Big(r_{\bar\imath} + \delta_{\bar\imath}(\bar m^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\Big)}{f_{i_j^{(0)}}^{(j)}\Big(r_{\bar\imath} + \delta_{\bar\imath}(\bar n^{(l)}),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\Big)}, \quad j \in \{1,\dots,k\} \setminus \{l\}, \tag{4.4}$$

is a function of $\big(r_{\bar m^{(l)}}, r_{\bar n^{(l)}}\big)$ only. Inserting now $r_{\bar\imath} = 0$ for all $\bar\imath$'s except $\bar\imath \in \big\{\bar m^{(l)}, \bar n^{(l)}\big\}$ in (4.2) we see that (4.4) is a quotient of a function of $r_{\bar m^{(l)}}$ and of a function of $r_{\bar n^{(l)}}$. Moreover,

$$f_{i_j^{(0)}}^{(j)}\Big(x_{\bar\imath},\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\Big) = \sum_{\bar g \in P_{i_j^{(0)}}^{(j)}} f_{i_j^{(0)}}^{(j)}\Big(x_{\bar\imath} + \delta_{\bar\imath}(\bar g),\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\Big)$$

for any $x_{\bar\imath}$, $\bar\imath \in P_{i_j^{(0)}}^{(j)}$. Hence again using the argument developed in the proof of Theorem 2 we get

$$f_{i_j^{(0)}}^{(j)}\Big(x_{\bar\imath},\ \bar\imath \in P_{i_j^{(0)}}^{(j)}\Big) = \frac{\Gamma\Big(\sum_{\bar\imath\in P_{i_j^{(0)}}^{(j)}} \mu_{\bar\imath}^{(l,j)}\Big)}{\Gamma\Big(\sum_{\bar\imath\in P_{i_j^{(0)}}^{(j)}} \big(\mu_{\bar\imath}^{(l,j)} + x_{\bar\imath}\big)\Big)} \prod_{\bar\imath\in P_{i_j^{(0)}}^{(j)}} \frac{\Gamma\big(\mu_{\bar\imath}^{(l,j)} + x_{\bar\imath}\big)}{\Gamma\big(\mu_{\bar\imath}^{(l,j)}\big)},$$

$$j \in \{1,\dots,k\} \setminus \{l\}.$$

Since the left hand side of the above expression does not depend on l, we can suppress the double superscript (l,j) of $\mu_{\bar\imath}^{(l,j)}$ into (j). Since l, $j \in \{1,\dots,k\} \setminus \{l\}$ and $i_j^{(0)}$ were taken arbitrarily we have

$$f_l^{(j)}\big(x_{\bar\imath},\ \bar\imath \in P_l^{(j)}\big) = \frac{\Gamma\big(\sum_{\bar\imath\in P_l^{(j)}} \mu_{\bar\imath}^{(j)}\big)}{\Gamma\big(\sum_{\bar\imath\in P_l^{(j)}} \big(\mu_{\bar\imath}^{(j)} + x_{\bar\imath}\big)\big)} \prod_{\bar\imath\in P_l^{(j)}} \frac{\Gamma\big(\mu_{\bar\imath}^{(j)} + x_{\bar\imath}\big)}{\Gamma\big(\mu_{\bar\imath}^{(j)}\big)},$$

$$l = 1,\dots,I_j,\ j = 1,\dots,k.$$

Considering now the fact that the right hand sides of (4.2) are equal we conclude that $\mu_{\bar{\imath}}^{(j)}$'s do not depend on j, that is $\mu_{\bar{\imath}}^{(j)} = \mu_{\bar{\imath}}$, for any $\bar{\imath} \in \mathcal{I}$.

Now, similarly as in the proof of Theorem 2 we identify the parameters of the functions $F^{(j)},\ j = 1, \dots, k$, as

$$d_l^{(j)} = \sum_{\bar{\imath} \in P_l^{(j)}} \mu_{\bar{\imath}}, \quad l = 1, \dots, I_j.$$

Thus $\Theta = [\theta_{\bar{\imath}},\ \bar{\imath} \in \mathcal{I}]$ is a Dirichlet k-way table with the parameter $\big(\mu_{\bar{\imath}},\ \bar{\imath} \in \mathcal{I}\big)$. $\square$

Acknowledgement The authors would like to thank Gérard Letac for inspiring discussions on partitions and neutrality. Remark 2 was inspired by referee's kind comment on similarities between Dirichlet and Liouville distributions.

References

[1] J. Aczél, Lectures on Functional Equations and Their Applications, *Academic Press*, *New York*, 1966.

[2] K. Bobecka and J. Wesołowski, The Dirichlet distribution and process through neutralities, *J. Theoret. Probab.* **20**, no. 2 (2007), 295–308.

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

[4] J. N. Darroch and D. Ratcliff, A characterization of the Dirichlet distribution, *J. Amer. Statist. Soc.* **66** (1971), 641–643.

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

[6] K. Doksum and L. F. James, On spatial neutral to the right processes and their posterior distributions, Mathematical Reliability: An Expository Perspective, Vol. 67, Internat. Ser. Oper. Res. Management Sci., *Kluwer*, *Boston*, 2004, 87–103.

[7] I. Epifani, A. Lijoi and I. Prünster, Exponential functionals and means of neutral–to-the-right priors, *Biometrika* **90** (2003), 791–808.

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

[9] K.-T. Fang, S. Kotz and K.-W. Ng, Symmetric Multivariate and Related Distributions, *Chapmann & Hall/CRC*, 1989.

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

[11] R. D. Gupta and D. S. P. Richards, Multivariate Liouiville distributions, *J. Multivar. Anal.* **23** (1987), 233–256.

[12] R. D. Gupta and D. S. P. Richards, The history of the Dirichlet and Liouville distributions, *Intern. Statist. Rev.* **69** (2001), 433–446.

[13] D. Heckerman, D. Geiger and D. M. Chickering, Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning* **20** (1995), 197–243.

[14] I. R. James and J. E. Mosimann, A new characterization of the Dirichlet distribution through neutrality, *Ann. Statist.* **8** (1980), 183–189.

[15] A. Járai, On regular solutions of functional equations, *Aequationes Math.* **30** (1986), 21–54.

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

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

[18] A.Y. Lo, A characterisation of the Dirichlet process, *Statist. Probab. Lett.* **12** (1991), 185–187.

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

[20] S. Walker and P. Muliere, A characterization of a neutral to the right prior via an extension of Johnson's sufficientness postulate, *Ann. Statist.* **27** (1999), 589–599.

[21] S. L. Zabell, W. E. Johnson's "sufficientness" postulate, *Ann. Statist.* **10** (1982), 1091–1099.

KONSTANCJA BOBECKA
WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH
POLITECHNIKA WARSZAWSKA
00-661 WARSZAWA
POLAND

E-mail: bobecka@mini.pw.edu.pl

JACEK WESOŁOWSKI
WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH
POLITECHNIKA WARSZAWSKA
00-661 WARSZAWA
POLAND

E-mail: wesolo@mini.pw.edu.pl

(Received August 4, 2008; revised December 22, 2008)