Structural Properties of the Generalized Dirichlet Distributions

Wan-Ying Chang, Rameshwar D. Gupta, and Donald St. P. Richards

Abstract. We study the general structure of the generalized Dirichlet distributions, deriving general formulas for the marginal and conditional probability density functions of those distributions. We develop the multivariate reverse rule properties of these distributions, apply those properties to derive probability inequalities, and derive stochastic representations and orderings for the distributions. Further, we study approaches for estimating the parameters of these distributions and recommend that parameter estimation be carried out by the maximum likelihood method.

1. Introduction

There are many statistical problems which involve an n-dimensional random vector, $(X_1, \ldots, X_n)$, taking values in a unit simplex. Examples of such problems arise in compositional data analysis [**1**], where $X_1, \ldots, X_n$ often represent proportions of a chemical or geological substance which has been decomposed into its constituent parts. Among probability distributions arising in compositional data analysis, the generalized Dirichlet distributions [**8**] play a prominent role. These distributions are useful for modeling proportions of substances, and they also arise in other contexts, including: random divisions of an interval [**25**], spacings [**28**], extreme value distributions [**29**], Bayesian inference for multinomial distributions [**12**], Bayesian life-testing problems [**23, 27**], probability and variance inequalities [**6, 7**], mixture models for high-dimensional pattern recognition [**4**], and machine learning for image processing [**5**].

In a study of independence concepts for random vectors defined on a unit simplex, Connor and Mosimann [**8**] defined the property of *neutrality*, which arises naturally in the following context. In an analysis of the proportions $X_1, \ldots, X_n$ of

1991 *Mathematics Subject Classification.* Primary 62H05, 62H10; Secondary 62E10, 62E15.

Key words and phrases. Complete neutrality, compositional data, Dirichlet distribution, generalized hypergeometric series, maximum likelihood estimation, neutrality, stochastic representation, multivariate reverse rule, multivariate total positivity, stochastic ordering.

R. D. Gupta is supported in part by the National Scientific and Engineering Research Council, grant A-4850.

D. St. P. Richards is supported in part by the National Science Foundation, grant DMS-0705210.

This paper is in final form and no version of it will be submitted for publication elsewhere.

a substance it may be desirable for some reason to eliminate a proportion, say X_1, and then to analyze the components $X_2, \ldots, X_n$ as proportions of the remaining material; i.e., we wish to analyze the proportions $X_2/(1 - X_1), \ldots, X_n/(1 - X_1)$. If the joint distribution of these remaining proportions is independent of the distribution of X_1, then X_1 is said to be *neutral*. Connor and Mosimann [**8**] further extended the concept of neutrality to more than one variable and defined a vector of proportions $(X_1, \ldots, X_n)$ to be *completely neutral* if the ratios

$$X_1, \frac{X_2}{1 - X_1}, \frac{X_3}{1 - X_1 - X_2}, \ldots, \frac{X_n}{1 - X_1 - \cdots - X_{n-1}} \tag{1.1}$$

are mutually independent. If, further, it is assumed that the marginal distribution of each of these ratios is a beta distribution then the random vector $(X_1, \ldots, X_n)$ is said to have a *generalized Dirichlet distribution.*

For the case in which $(X_1, \ldots, X_n)$ follows a generalized Dirichlet distribution, the property that the ratios in (1.1) are mutually independent, beta-distributed random variables leads easily to the evaluation of the moments of $(X_1, \ldots, X_n)$ and, in particular, to the covariance matrix of $(X_1, \ldots, X_n)$ [**8**], [**22**, p. 519 ff.]. Other than the evaluation of these moments, to date, very little appears to be known about the distributional properties of the generalized Dirichlet distributions. Indeed, by comparison, far more is known about the Dirichlet and Liouville distributions (cf. [**13, 15, 16**]) and the references given in those articles), and this raises the hope that the structure of the generalized Dirichlet distributions can be developed beyond current limits of knowledge.

In this paper we investigate the structure of the generalized Dirichlet distributions. In Section 2, we apply the theory of generalized hypergeometric functions [**2**] to obtain general representations for all marginal and conditional distributions of subsets of $(X_1, \ldots, X_n)$. In particular, we will show that for all $i = 1, \ldots, n$, the marginal and conditional distributions of $(X_1, \ldots, X_i)$ are also of generalized Dirichlet type.

In Section 3, we study the multivariate reverse rule properties of the generalized Dirichlet distributions. From the reverse rule properties, we deduce probability and expectation inequalities in the trivariate case, thereby generalizing results available for the multinomial, hypergeometric, Dirichlet and Liouville distributions [**14**]. Further, our approach to deriving these MRR results for the generalized Dirichlet distributions also are applicable to other generalizations of the Dirichlet distributions such as the hyper-Dirichlet distributions defined by Hankin [**17**].

By utilizing the property of complete neutrality, we derive in Section 4 stochastic representations for $(X_1, \ldots, X_n)$. Then, we obtain stochastic inequalities between the generalized and the classical Dirichlet distributions. These results lead to probability and expectation inequalities for all $n \geq 2$.

Finally, in Section 5 we study the problem of estimating the parameters of the distribution of $(X_1, \ldots, X_n)$. We compare estimators obtained by the method-of-moments and the maximum likelihood method. Although the method-of-moments estimators are simpler to compute, we will see that the maximum likelihood estimators are superior in that they have smaller variance than the method-of-moments estimators. Moreover, the maximum likelihood estimators are to be preferred because they innately are functions of the minimal sufficient statistics and best asymptotically normal.

2. Marginal and conditional distributions

For parameters $a_1, \ldots, a_n, b_1, \ldots, b_n > 0$, a random vector $(X_1, \ldots, X_n)$, taking values in the open unit simplex

$$\mathcal{S}_n = \{(x_1, \ldots, x_n) : x_1 > 0, \ldots, x_n > 0, \sum_{i=1}^{n} x_i < 1\},$$

is said to have a generalized Dirichlet distribution if its probability density function is of the form

$$c \prod_{i=1}^{n} \Big[x_i^{a_i - 1} \big(1 - \sum_{k=1}^{i} x_k\big)^{b_i - 1}\Big], \tag{2.1}$$

$(x_1, \ldots, x_n) \in \mathcal{S}_n$. To evaluate the normalizing constant c, we integrate sequentially over $x_n, x_{n-1}, \ldots, x_2, x_1$ (cf. [**8**]), deducing that

$$c = \prod_{i=1}^{n} \frac{\Gamma\big(1 + \sum_{k=i}^{n} (a_k + b_k - 1)\big)}{\Gamma(a_i)\Gamma\big(b_i + \sum_{k=i+1}^{n} (a_k + b_k - 1)\big)}.$$

Throughout, we write $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$ to denote that the vector $(X_1, \ldots, X_n)$ follows a generalized Dirichlet distribution with the density function (2.1).

The following result on the marginal distributions of arbitrary subsets of $X_1, \ldots,$ X_n is obtained by integration of the density function (2.1). In (2.2), we shall abide by the standard convention in the case $j_1 = 1$ that an empty product is identically equal to 1.

Proposition 2.1. *Let $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_{n-k}$ be complementary subsets of $\{1, \ldots, n\}$. If $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$ then the marginal density function of $(X_{i_1}, \ldots, X_{i_k})$ is of the form*

$$c' \prod_{r=1}^{k} x_{i_r}^{a_{i_r} - 1} \cdot \prod_{r=1}^{j_1 - 1} \Big(1 - \sum_{s=1}^{r} x_s\Big)^{b_{i_r} - 1} \cdot \varphi(x_{i_1}, \ldots, x_{i_k}), \tag{2.2}$$

where $(x_{i_1}, \ldots, x_{i_k}) \in \mathcal{S}_k$, and c' is the normalizing constant,

$$\varphi(x_{i_1}, \ldots, x_{i_k}) = \int \cdots \int \prod_{i=1}^{j_{n-k}} x_i^{a_i - 1} \cdot \prod_{r=j_1}^{n} \Big(1 - \sum_{s=1}^{r} x_s\Big)^{b_r - 1} \mathrm{d}x_{j_1} \cdots \mathrm{d}x_{j_{n-k}},$$

and the region of integration is the simplex

$$\{(x_{j_1}, \ldots, x_{j_{n-k}}) : x_{j_1}, \ldots, x_{j_{n-k}} > 0; x_{j_1} + \cdots + x_{j_{n-k}} < 1 - x_{i_1} - \cdots - x_{i_k}\}.$$

We remark that the function φ is a generalized hypergeometric function of the type studied by Aomoto [**2**] and Gelfand, et al. [**9, 10**]. The results provided by those authors contain reduction formulas, recurrence relations, and series expansions of hypergeometric type, and systems of differential equations for φ, and hence also for the marginal densities of subsets of $X_1, \ldots, X_n$. For the case in which the parameters b_r, $r \ge j_1$, all are positive integers, closed-form expressions can be obtained for φ by expanding each term, $\big(1 - \sum_{s=1}^{r} x_s\big)^{b_r - 1}$ using the binomial theorem, and integrating term-by-term using the classical Dirichlet integral.

In general, the evaluation of the normalizing constant c', or the function φ, can be done only by numerical methods. In this regard, we recommend the **hyperdirichlet** R package, developed by Hankin [**17**] for the purposes of computations for a generalization of the Dirichlet distribution.

For the case in which $i_1 = 1, \ldots, i_k = k$, the following result on the marginal distribution of $(X_1, \ldots, X_k)$ is due to Connor and Mosimann [**8**].

Corollary 2.2. (Connor and Mosimann [**8**]) If $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$ then $(X_1, \ldots, X_k) \sim GD(a_1, \ldots, a_k; b_1, \ldots, b_{k-1}, b_k + \sum_{i=k+1}^n (a_i + b_i - 1))$ for each $k = 1, \ldots, n$.

This result follows from Proposition 2.1 by successive integration of the variables $x_n, x_{n-1}, \ldots, x_{k+1}$. As a consequence, we obtain the following result on the conditional distributions.

Corollary 2.3. *Suppose* $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$, *and* $1 \le r \le n-1$. *For* $i = r+1, \ldots, n$, *define*

$$U_i = \frac{X_i}{1 - X_1 - \cdots - X_r}.$$

Then the conditional distribution of $(U_{r+1}, \ldots, U_n)$, *given* $(X_1, \ldots, X_r) = (x_1, \ldots, x_r)$, *is* $GD(a_{r+1}, \ldots, a_n; b_{r+1}, \ldots, b_n)$.

Proof. From Corollary 2.2, we already know the marginal distribution of $(X_1, \ldots, X_r)$. Therefore the conditional density function of $(X_{r+1}, \ldots, X_n)$, given $(X_1, \ldots, X_r) = (x_1, \ldots, x_r)$, is proportional to

$$\frac{\prod_{i=1}^n \left[x_i^{a_i-1}(1 - \sum_{k=1}^i x_k)^{b_i-1}\right]}{x_r^{a_r-1}(1 - \sum_{k=1}^r x_k)^{b_r + \sum_{i=r+1}^n (a_i+b_i-1)} \prod_{i=1}^{r-1} \left[x_i^{a_i-1}(1 - \sum_{k=1}^i x_k)^{b_i-1}\right]}$$
$$= \frac{\prod_{i=r+1}^n \left[x_i^{a_i-1}(1 - \sum_{k=1}^i x_k)^{b_i-1}\right]}{(1 - \sum_{k=1}^r x_k)^{1+\sum_{i=r+1}^n (a_i+b_i-1)}}$$
$$= \frac{\prod_{i=r+1}^n \left[x_i^{a_i-1}(1 - \sum_{k=1}^r x_k - \sum_{k=r+1}^i x_k)^{b_i-1}\right]}{(1 - \sum_{k=1}^r x_k)^{1+\sum_{i=r+1}^n (a_i+b_i-1)}}.$$

Now consider the random vector $(U_{r+1}, \ldots, U_n)$, which equals $(X_{r+1}, \ldots, X_n)$ multiplied by the constant $(1 - x_1 - \cdots - x_r)^{-1}$. By a transformation we deduce that the conditional density function of $(U_{r+1}, \ldots, U_n)$, given $(X_1, \ldots, X_r) = (x_1, \ldots, x_r)$, is proportional to

$$\prod_{i=r+1}^n \left[{u_i}^{a_i-1}\big(1 - \sum_{k=r+1}^i u_k\big)^{b_i-1}\right].$$

This proves that the conditional distribution of $(U_{r+1}, \ldots, U_n)$ is a generalized Dirichlet distribution. □

We remark also that Corollary 2.3 can be obtained as a consequence of the complete neutrality property of the generalized Dirichlet distributions.

Corollary 2.4. *Suppose* $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$. *Then* $X_1 \sim B(a_1, b_1 + \sum_{i=2}^n (a_i + b_i - 1))$. *Further, for any* $j = 1, \ldots, n-1$, *the conditional* *distribution of* $X_j/(1-X_1-\cdots-X_{j-1})$*, given* $(X_1,\ldots,X_{j-1})=(x_1,\ldots,x_{j-1})$*, is* $B(a_j, b_j+\sum_{i=j+1}^n(a_i+b_i-1))$*.*

3. Multivariate reverse rule properties

Various classical distributions on the simplex, including the multinomial, hypergeometric, Dirichlet, and Liouville distributions are well-known [**13, 20**] to satisfy certain multivariate reverse rule properties. From those reverse rule properties follow negative correlation inequalities and other probability inequalities. In this section we investigate the multivariate reverse rule properties of the generalized Dirichlet distributions with an eye toward related correlation and probability inequalities.

Following Karlin [**18**], we call a nonnegative function $f:\mathbb{R}^2\to\mathbb{R}$ totally positive of order 2 (TP_2) if

$$f(x_1,y_1)f(x_2,y_2)\ge f(x_1,y_2)f(x_2,y_1)$$

whenever $x_1\ge x_2$ and $y_1\ge y_2$. If the reverse inequality is valid for all $x_1\ge x_2$ and $y_1\ge y_2$ then we say that f is reverse rule of order 2 (RR_2).

A nonnegative function $\phi:\mathbb{R}\to\mathbb{R}$ is a Pólya frequency function of order 2 (PF_2) if the function $f(x,y)=\phi(x-y)$ is TP_2.

Some basic examples of functions satisfying these properties are the following. These examples will be utilized repeatedly in the sequel.

Example 3.1. ([**18**, p. 15]) (i) The function $f(x,y)=e^{xy}$, $x,y\in\mathbb{R}$, is TP_2.

(ii) For $a\ge 0$, let $f(x,y)=(x-y)^a$, $x\ge y$, and $f(x,y)=0$, otherwise. Then the function f is TP_2. Equivalently, the function $\phi(x)=x_+^a$ is PF_2, where $x_+^a=x^a$ or 0 according as $x>0$ or $x\le 0$, respectively.

(iii) For $a\ge 0$, let $f(x,y)=(k-x-y)^a$, $x+y<k$, and $f(x,y)=0$, otherwise. Then the function f is RR_2.

Starting with these examples, we can construct new TP_2 or RR_2 functions using the following result that is known as the Basic Composition Formula.

Theorem 3.2. (Karlin [**18**, p. 17]) *Let X, Y, and Z be subsets of $\mathbb{R}$; σ be a sigma-finite measure defined on Y; and K and L be nonnegative Borel-measurable functions on $X\times Y$ and $Y\times Z$, respectively. For $\xi\in X$ and $\eta\in Z$, define*

$$M(\xi,\eta)=\int_Y K(\xi,\zeta)L(\zeta,\eta)\,\mathrm{d}\sigma(\zeta),$$

where the integral is assumed to converge absolutely. Then:
(i) If K and L are both TP_2 *or both* RR_2 *then M is* TP_2 *on $X\times Z$.*
(ii) If K is TP_2 *and L is* RR_2 *then M is* RR_2*.*

Definition 3.3. (Karlin and Rinott [**19, 20**]) For $\mathbf{x}=(x_1\ldots,x_n)$ and $\mathbf{y}=(y_1,\ldots,y_n)$ in $\mathbb{R}^n$, let

$$\mathbf{x}\vee\mathbf{y}=\big(\max(x_1,y_1),\ldots,\max(x_n,y_n)\big),\quad \mathbf{x}\wedge\mathbf{y}=\big(\min(x_1,y_1),\ldots,\min(x_n,y_n)\big).$$

A nonnegative function $f:\mathbb{R}^n\to\mathbb{R}$ is called multivariate totally positive of order 2 (MTP_2) if

$$f(\mathbf{x}\vee\mathbf{y})f(\mathbf{x}\wedge\mathbf{y})\ge f(\mathbf{x})f(\mathbf{y}), \tag{3.1}$$

for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$.

If the reverse inequality in (3.1) is valid for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then we say that f is multivariate reverse rule of order 2 (MRR_2).

A random vector $(X_1, \ldots, X_n) \in$ is said to be MTP_2 (resp. MRR_2) if its density function is MTP_2 (resp. MRR_2).

We now establish the multivariate reverse rule properties of the generalized Dirichlet distributions.

THEOREM 3.4. *Let* $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$, *where* $b_j \geq 1$ *for all* $j = 2, \ldots, n$. *Then* $(X_1, \ldots, X_n)$ *is* MRR_2.

PROOF. Because the density function of $(X_1, \ldots, X_n)$ is strictly positive on $\mathcal{S}_n$ then, by Karlin and Rinott [**20**], it is sufficient to show that $(X_1, \ldots, X_n)$ is pairwise RR_2; i.e., we need only show that the density function (2.1) is RR_2 in each pair of variables chosen from $x_1, \ldots, x_n$, with all other variables held fixed.

To that end, choose a pair (X_i, X_j) where, without loss of generality, $i < j$. By ignoring in the density function all terms which are free of the pair (x_i, x_j), it follows that $(X_1, \ldots, X_n)$ is MRR_2 if and only if the function

$$g(x_i, x_j) = \prod_{m=j}^{n} (1 - \sum_{k=1}^{m} x_k)^{b_m - 1} \tag{3.2}$$

is RR_2 in (x_i, x_j), where $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ are held fixed. Noting that the function g in (3.2) depends on (x_i, x_j) only through $x_i + x_j$, it then follows by [**18**, p. 158] that it is sufficient to prove that g is log-concave in $x_i + x_j$.

To show this, we observe that

$$\begin{aligned} \log g(x_i, x_j) &= \sum_{m=j}^{n} (b_m - 1) \log \Big(1 - \sum_{k=1}^{m} x_k\Big) \\ &= \sum_{m=j}^{n} (b_m - 1) \log \big(1 - w_m - (x_i + x_j)\big), \end{aligned}$$

where $w_m = \sum_{k=1, k \neq i,j}^{m} x_k > 0$. Because $b_m \geq 1$ for all $m = j, \ldots, n$ and the function $\log(1 - w_m - (x_i + x_j))$ is concave in $x_i + x_j$, it follows that $g(x_i, x_j)$ is log-concave in $x_i + x_j$. Therefore $g(x_i, x_j)$ is RR_2 in (x_i, x_j). The pair (i, j) having been chosen arbitrarily, the proof is complete. $\square$

DEFINITION 3.5. (Karlin and Rinott [**20**]) Let the random vector $(X_1, \ldots, X_n)$ be MRR_2 with density function f. Then $(X_1, \ldots, X_n)$ is said to be strongly MRR_2 ($\mathrm{S\text{-}MRR}_2$) if for any set of PF_2 functions $\phi_1, \ldots, \phi_n$ and any sets $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_{n-k}\}$ of complementary subsets of indices in $\{1, \ldots, n\}$, the function

$$g(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \int \cdots \int f(x_1, \ldots, x_n) \prod_{r=1}^{n-k} \phi_r(x_{j_r}) \, \mathrm{d}x_{j_r} \tag{3.3}$$

is MRR_2 in the variables $(x_{i_1}, \ldots, x_{i_k})$.

We now derive the $\mathrm{S\text{-}MRR}_2$ property of the generalized Dirichlet distributions in the trivariate case.

PROPOSITION 3.6. *Suppose that* $(X_1, X_2, X_3) \sim GD(a_1, a_2, a_3; b_1, b_2, b_3)$ *where* $a_i, b_i \geq 1$ *for all* $i = 1, 2, 3$. *Then* (X_1, X_2, X_3) *is* $\mathrm{S\text{-}MRR}_2$.

Proof. We proceed in a case-by-case manner, proving that the function g in (3.4) is RR_2 in each of the pairs (i) (x_1, x_2), (ii) (x_1, x_3), and (iii) (x_2, x_3). Throughout, we let ϕ be a PF_2 function.

In Case (i), (3.3) reduces to

$$\begin{aligned} g(x_1,x_2) &= x_1^{a_1-1}x_2^{a_2-1}(1-x_1)^{b_1-1}(1-x_1-x_2)^{b_2-1} \\ &\quad \times \int_0^{1-x_1-x_2} \phi(x_3)x_3^{a_3-1}(1-x_1-x_2-x_3)^{b_3-1}\,\mathrm{d}x_3. \end{aligned}$$

Substituting $x_3 = w - x_2$, and recalling the notation x_+^a which equals x^a or 0 according as $x > 0$ or $x \le 0$, we obtain

$$\begin{aligned} g(x_1,x_2) &= x_1^{a_1-1}x_2^{a_2-1}(1-x_1)^{b_1-1}(1-x_1-x_2)^{b_2-1} \\ &\quad \times \int_0^{1} \phi(w-x_2)(w-x_2)_+^{a_3-1}(1-x_1-w)_+^{b_3-1}\,\mathrm{d}w. \end{aligned}$$

By the definition of a PF_2 function, $\phi(w-x_2)$ is TP_2 in (w, x_2); and by Example 3.2 (ii), $(w-x_2)_+^{a_3-1}$ is TP_2 in (w, x_2). Because the product of two TP_2 functions is also TP_2, $\phi(w-x_2)(w-x_2)_+^{a_3-1}$ is TP_2. Next, by Example 3.2 (iii), the function $(1 - x_1 - w)^{b_3-1}$ is RR_2 in (x_1, w) because $b_3 \ge 1$. Therefore, by the Basic Composition Formula (Theorem 3.2), the function

$$\int_0^{1} \phi(w-x_2)(w-x_2)_+^{a_3-1}(1-x_1-w)_+^{b_3-1}\,\mathrm{d}w$$

is RR_2 in (x_1, x_2). Because $b_2 \ge 1$, the function $(1-x_1-x_2)_+^{b_2-1}$ is RR_2 in (x_1, x_2). Finally, because the product of two positive RR_2 functions is RR_2, we deduce that g is RR_2.

In Case (ii), (3.3) reduces to

$$\begin{aligned} g(x_1,x_3) &= x_1^{a_1-1}x_3^{a_3-1}(1-x_1)^{b_1-1} \\ &\quad \times \int_0^{1-x_1-x_3} \phi(x_2)x_2^{a_2-1}(1-x_1-x_2)_+^{b_2-1}(1-x_1-x_2-x_3)^{b_3-1}\,\mathrm{d}x_2. \end{aligned}$$

Substituting $x_2 = w - x_1$, we obtain

$$\begin{aligned} g(x_1,x_3) &= x_1^{a_1-1}x_3^{a_3-1}(1-x_1)^{b_1-1} \\ &\quad \times \int_0^{1} \phi(w-x_1)(w-x_1)_+^{a_2-1}(1-w)_+^{b_2-1}(1-w-x_3)_+^{b_3-1}\,\mathrm{d}w. \end{aligned}$$

Arguing as before we deduce that $\phi(w-x_1)(w-x_1)_+^{b_2-1}$ is TP_2 in (w, x_1), and $(1-w-x_3)_+^{b_3-1}$ is RR_2 in (w, x_3). Again by the Basic Composition Formula, it follows that g is RR_2 in (x_1, x_3).

In Case (iii), (3.3) becomes

$$\begin{aligned} g(x_2,x_3) &= x_2^{a_2-1}x_3^{a_3-1}\int_0^1 \phi(x_1)x_1^{a_1-1}(1-x_1)_+^{b_1-1} \\ &\qquad\qquad \times (1-x_1-x_2)_+^{b_2-1}(1-x_1-x_2-x_3)_+^{b_3-1}\,\mathrm{d}x_1 \\ &= x_2^{a_2-1}x_3^{a_3-1}\int_0^1 \phi(w-x_2)(w-x_2)^{a_1-1}(1-w+x_2)_+^{b_1-1} \\ &\qquad\qquad \times (1-w)_+^{b_2-1}(1-w-x_3)_+^{b_3-1}\,\mathrm{d}w. \end{aligned}$$

By a similar argument, we find that $\phi(w-x_2)(w-x_2)^{a_1-1}(1-w+x_2)^{b_1-1}(1-w)^{b_2-1}$ is TP_2 in (w, x_2), and $(1-w-x_3)^{b_3-1}$ is RR_2 in (w, x_3). By the Basic Composition Formula, we deduce that g is RR_2 in (x_2, x_3). $\square$

PROPOSITION 3.7. *Suppose that* $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$, *where* $n \geq 4$; $a_i \geq 1$, $i = 1, \ldots, n$; $b_j = 1$, $j = 2, \ldots, n-2$; *and* $b_j \geq 1$, $j = 1, n-1, n$. *Then* $(X_1, \ldots, X_n)$ *is* S-MRR_2.

PROOF. The proof will proceed by induction on n. First, for any PF_2 function ϕ, we show that

$$g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \int_0^1 f(x_1, \ldots, x_n)\phi(x_i)\, \mathrm{d}x_i \tag{3.4}$$

is MRR_2 for $i = 1, \ldots, n$, where $f(x_1, \ldots, x_n)$ is the probability density function of $(X_1, \ldots, X_n)$. Because g is positive on the simplex $\mathcal{S}_{n-1}$, it suffices to show that g is pairwise RR_2. Consider a pair (x_p, x_q), with $p < q$, chosen arbitrarily from $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$. By arguments similar to those in the proof of Proposition 3.6, we shall show that g is RR_2 in (x_p, x_q) while holding x_k fixed, $k \neq i, p, q$, as follows.

For Case (i), in which $p < q < i$, we have

$$\begin{aligned} g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \quad &= \quad \prod_{j \neq i} x_j^{a_j-1} \cdot \int \phi(x_i) x_i^{a_i-1}(1-x_1)^{b_1-1} \\ &\quad \times \Big(1 - \sum_{k=1}^{n-1} x_k\Big)^{b_{n-1}-1} \Big(1 - \sum_{k=1}^{n} x_k\Big)^{b_n-1} \mathrm{d}x_i. \end{aligned}$$

If $i = n$ then, by making the substitution $x_n = w - x_p$, we obtain

$$\begin{aligned} g(x_1, \ldots, x_{n-1}) \quad &= \quad \prod_{j \neq n} x_j^{a_j-1} \cdot (1-x_1)^{b_1-1} \Big(1 - \sum_{k=1}^{n-1} x_k\Big)^{b_{n-1}-1} \\ &\quad \times \int \phi(w-x_p)(w-x_p)^{a_n-1} \Big(1 - \sum_{k \neq n,p} x_k - w\Big)^{b_n-1} \mathrm{d}w. \end{aligned}$$

Since $b_{n-1}, b_n, a_n \geq 1$ then, by applying the Basic Composition Formula, we deduce that $g(x_1, \ldots, x_n)$ is RR_2 in (x_p, x_q). On the other hand, if $i \neq n$ then

$$\begin{aligned} &g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\ &\qquad = \prod_{j \neq i} x_j^{a_j-1} \cdot (1-x_1)^{b_1-1} \int \phi(w-x_p)(w-x_p)^{a_i-1} \\ &\qquad\qquad \times \Big(1 - \sum_{k \neq i,p}^{n-1} x_k - w\Big)^{b_{n-1}-1} \Big(1 - \sum_{k \neq i,p}^{n} x_k - w\Big)^{b_n-1} \mathrm{d}w. \end{aligned}$$

Because $a_i \geq 1$, $i = 1, \ldots, n$, $b_{n-1}, b_n \geq 1$ then it follows by an application of the Basic Composition Formula that g is RR_2 in (x_p, x_q)

Next, consider Case (ii), in which $p < i < q$. If $q = n$ then

$$\begin{aligned} g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) &= \prod_{j \neq i} x_j^{a_j - 1} \cdot (1 - x_1)^{b_1 - 1} \int \phi(w - x_p)(w - x_p)^{a_i - 1} \\ &\qquad \times \Big(1 - \sum_{k \neq i,p}^{n-1} x_k - w\Big)^{b_{n-1} - 1} \Big(1 - \sum_{k \neq i,p}^{n} x_k - w\Big)^{b_n - 1} \, \mathrm{d}w. \end{aligned}$$

Because $a_i \geq 1$ for all i and $b_n \geq 1$ then, by applying the Basic Composition Formula, we deduce that (3.5) is RR_2 in (x_p, x_q). If $q \neq n$ then g is of the form arising as in Case (i) with $i \neq n$.

Finally, consider Case (iii) in which $i < p < q$. Suppose that $i = 1$ and $q \neq n$; then,

$$\begin{aligned} g(x_2, \ldots, x_n) &= \prod_{j \neq 1} x_j^{a_j - 1} \cdot \int \phi(w - x_p)(1 - w + x_p)^{b_1 - 1}(w - x_p)^{a_1 - 1} \\ &\qquad \times \Big(1 - \sum_{k \neq 1,p}^{n-1} x_k - w\Big)^{b_{n-1} - 1} \Big(1 - \sum_{k \neq 1,p}^{n} x_k - w\Big)^{b_n - 1} \, \mathrm{d}w. \end{aligned}$$

Applying the Basic Composition Formula as before, we deduce that g is RR_2 in (x_p, x_q). If $i = 1$ and $q = n$, then

$$\begin{aligned} g(x_2, \ldots, x_n) &= \prod_{j \neq 1} x_j^{a_j - 1} \cdot \int \phi(w - x_p)(1 - w + x_p)^{b_1 - 1}(w - x_p)^{a_1 - 1} \\ &\qquad \times \Big(1 - \sum_{k \neq 1,p,q}^{n-1} x_k - w\Big)^{b_{n-1} - 1} \Big(1 - \sum_{k \neq 1,p}^{n} x_k - w\Big)^{b_n - 1} \, \mathrm{d}w. \end{aligned}$$

As usual, it now follows by the Basic Composition Formula that g is RR_2 in (x_p, x_q). To close this case, if $i \neq 1$, then g reduces to a form seen in Case (ii).

Now assume by the inductive hypothesis that for any collection $\phi_1, \ldots, \phi_k$ of PF_2 functions, the function

$$g(x_{u_1}, \ldots, x_{u_{n-k+1}}) = \int \cdots \int f(x_1, \ldots, x_n) \prod_{i=1}^{k-1} \phi_i(x_{v_i}) \, \mathrm{d}x_{v_i}$$

is MRR_2, where $\{v_1, \ldots, v_{k-1}\} \cup \{u_1, \ldots, u_{n-k+1}\} = \{1, \ldots, n\}$ and, as usual, the integral is taken over the simplex

$$\{(x_{v_1}, \ldots, x_{v_{k-1}}) : x_{v_1} > 0, \ldots, x_{v_{k-1}} > 0, x_{v_1} + \cdots + x_{v_{k-1}} < 1 - x_{u_1} - \cdots - x_{u_{n-k+1}}\}.$$

We need to show that the function

$$g(x_{u_1}, \ldots, x_{u_{n-k}}) = \int \cdots \int f(x_1, \ldots, x_n) \prod_{i=1}^{k} \phi_i(x_{v_i}) \, \mathrm{d}x_{v_i} \tag{3.5}$$

is MRR_2; by positivity, it suffices to show that $g(x_{u_1}, \ldots, x_{u_{n-k}})$ is pairwise RR_2.

By (3.5), we may express g in the form

$$g(x_{u_1},\ldots,x_{u_{n-k}}) = \prod_{i=1}^{n-k} x_{u_i}^{a_{u_i}-1} \cdot \int x_{u_{n-k+1}}^{a_{u_{n-k+1}}-1} h(x_{u_1},\ldots,x_{u_{n-k+1}}) \phi_k(x_{u_{n-k+1}})\, \mathrm{d}x_{u_{n-k+1}},$$

for a function h which, by the inductive hypothesis, is MRR$_2$. Moreover, if $n \in \{v_1,\ldots,v_{k-1}\}$ then $h(x_{u_1},\ldots,x_{u_{n-k+1}}) = h^*(\sum_{i=1}^{n-k+1} x_{u_i})$ for some function $h^* : \mathbb{R} \to \mathbb{R}$; otherwise, h is a function of the pair $(\sum_{i=1}^{n-k} x_{u_i}, \sum_{i=1}^{n-k} x_{u_i} + x_n)$.

Consider a pair of variables (x_{u_i}, x_{u_j}) with $i, j \neq n-k+1$. We now show that the function g in (3.5) is pairwise RR$_2$ in (x_{u_i}, x_{u_j}) while holding fixed all x_{u_l}, $l \neq i, j, n-k+1$.

Suppose, first, that $n \in \{v_1,\ldots,v_{k-1}\}$, in which case $h(x_{u_1},\ldots,x_{u_{n-k+1}}) = h^*(\sum_{i=1}^{n-k+1} x_{u_i})$. Then it suffices to show that

$$\int h^*\Big(\sum_{i=1}^{n-k+1} x_{u_i}\Big)\phi_k(x_{u_{n-k+1}})\, \mathrm{d}x_{u_{n-k+1}} \tag{3.6}$$

is RR$_2$ in (x_{u_i}, x_{u_j}). Because h is MRR$_2$ and is a function of $x_{u_i}+x_{u_j}+x_{u_{n-k+1}}$, h is RR$_2$ in $(x_{u_i}+x_{u_{n-k+1}}, x_{u_j})$. By an application of the Basic Composition Formula, we find that (3.6) is RR$_2$ in (x_{u_i}, x_{u_j}).

Finally, consider the case in which $n \notin \{v_1,\ldots,v_{k-1}\}$. In this case, as noted earlier, h is a function of $(\sum_{i=1}^{n-k} x_{u_i}, \sum_{i=1}^{n-k} x_{u_i} + x_n)$. If $u_{n-k+1} = n$ then we can rearrange the order of $v_1,\ldots,v_{k-1},u_{n-k+1}$ so that it reduces to the case resolved in the previous paragraph. On the other hand, if $u_{n-k+1} \neq n$ then, because h is MRR$_2$ and is a function of $x_{u_i}+x_{u_{n-k+1}}$ and x_{u_j}, then h is RR$_2$ in $(x_{u_i}+x_{u_{n-k+1}}, x_{u_j})$. By application of the Basic Composition Formula, it follows that (3.6) is RR$_2$ in (x_{u_i}, x_{u_j}). Thus we proved that g is pairwise RR$_2$. Therefore, by induction, $(X_1,\ldots,X_n)$ is S-MRR$_2$. $\square$

From the general theory of probability inequalities for S-MRR$_2$ functions [**19**, **20**], we obtain the following results.

Corollary 3.8. *Suppose that* $(X_1,\ldots,X_n) \sim GD(a_1,\ldots,a_n; b_1,\ldots,b_n)$, *and that the hypotheses of Propositions 3.6 or 3.7 are valid.*
(i) If $\alpha_j \geq 0$, $j = 1,\ldots,n$ *then* $E\big(\prod_{j=1}^n X_j^{\alpha_j}\big) \leq \prod_{j=1}^n E(X_j^{\alpha_j})$.
(ii) For $t \geq 0$, $E\big(\prod_{j=1}^n e^{-tX_j}\big) \leq \prod_{j=1}^n E(e^{-tX_i})$.
(iii) For $c_1,\ldots,c_n > 0$,

$$P(X_1 \leq c_1,\ldots,X_n \leq c_n) \leq \prod_{j=1}^n P\{X_j \leq c_j\},$$

and

$$P(X_1 \geq c_1,\ldots,X_n \geq c_n) \leq \prod_{j=1}^n P\{X_j \geq c_j\}.$$

Similar results apply to higher-dimensional generalized Dirichlet distributions satisfying the assumptions in Proposition 3.7.

In closing this section, we remark that the methods developed here to derive the MRR properties of the generalized Dirichlet distributions can also be applied to derive MRR results for other generalizations of the Dirichlet distributions. We mention, in particular, the hyperdirichlet distributions of Hankin [**17**] and their special cases, the grouped Dirichlet distributions of Ng, *et al.* [**26**].

4. Stochastic representations and orderings

In this section we apply the property of complete neutrality to derive stochastic representations for the generalized Dirichlet distributions. In so doing, we are motivated by stochastic representations for the Dirichlet and Liouville distributions, as developed in [**14**]. We will also establish some stochastic orderings between the generalized Dirichlet distributions and the classical Dirichlet distributions, and provide some examples and applications of these orderings.

Given real-valued random variables U and V, we say that U is stochastically greater than V, denoted $U \overset{\mathcal{L}}{\geq} V$, if $P(U \leq t) \leq P(V \leq t)$ for all $t \in \mathbb{R}$.

LEMMA 4.1. *Let X and Y be random variables with strictly positive, continuous density functions f_X and f_Y, respectively. If the function $f_X(t)/f_Y(t)$ is nondecreasing then $X \overset{\mathcal{L}}{\geq} Y$.*

PROOF. Since f_X and f_Y are density functions, we have $\int_{\mathbb{R}}(f_X(t) - f_Y(t))\,\mathrm{d}t = 0$. Since f_X and f_Y are continuous, $f_X - f_Y$ changes sign at least once on $\mathbb{R}$, therefore there exists at least one $x_0 \in \mathbb{R}$ such that $f_X(x_0) = f_Y(x_0)$, equivalently, $f_X(x_0)/f_Y(x_0) = 1$. By the monotonicity of $f_X(t)/f_Y(t)$, it follows that $f_X(t) \leq f_Y(t)$ for all $t \leq x_0$. Integrating this latter inequality over the interval $(-\infty, t)$ where $t \leq x_0$, we deduce that $P(X \leq t) \leq P(Y \leq t)$ for $t \leq x_0$.

For $t \geq x_0$, it again follows from the monotonicity of $f_X(t)/f_Y(t)$ that, for $t \geq x_0$, $f_X(t) \geq f_Y(t)$. Integrating this inequality over an interval (t, ∞) where $t \geq x_0$, we obtain $P(X \geq t) \geq P(Y \geq t)$ for all $t \geq x_0$, i.e., $P(X \leq t) \leq P(Y \leq t)$ for all $t \geq x_0$. Thus we have proved the desired result. □

EXAMPLE 4.2. Suppose that X and Y are beta-distributed random variables, $X \sim B(a_1, a_2)$ and $Y \sim B(b_1, b_2)$. Then, $f_X(t)/f_Y(t)$ is proportional to $t^{a_1 - b_1}(1 - t)^{a_2 - b_2}$, a function which is nondecreasing on $(0, 1)$ if and only if $a_1 \geq b_1$ and $a_2 \leq b_2$. Therefore, $X \overset{\mathcal{L}}{\geq} Y$ if and only if $a_1 \geq b_1$ and $a_2 \leq b_2$.

LEMMA 4.3. *Suppose that $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$, and $Z_1, \ldots, Z_n$ are mutually independent beta-distributed variables with $Z_i \sim B(a_i, b_i + \sum_{k=i+1}^{n}(b_k + a_k - 1))$, $i = 1, \ldots, n$. Then*

$$(X_1, \ldots, X_n) \overset{\mathcal{L}}{=} (Z_1, Z_2(1 - Z_1), \ldots, Z_n \prod_{i=1}^{n-1}(1 - Z_i))$$

PROOF. The result follows from the definition of complete neutrality. □

To extend the notion of stochastic ordering from scalars to random vectors, $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, we apply the following approach (cf. [**24**, p. 485]).

DEFINITION 4.4. (Veinott [**30**]) A random vector $(X_1, \ldots, X_n)$ is stochastically greater than a random vector $(Y_1, \ldots, Y_n)$ if $E\phi(X_1, \ldots, X_n) \geq E\phi(Y_1, \ldots, Y_n)$ for any function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that ϕ is monotone increasing in each component, and for which the expectations exist.

In practice, this definition is applied by means of the following result.

PROPOSITION 4.5. (Veinott [**30**]) *Let* $(X_1,\dots,X_n)$ *and* $(Y_1,\dots,Y_n)$ *be random vectors such that for all* $t\in\mathbb{R}$,
(i) $P\{X_1\le t\}\le P\{Y_1\le t\}$, *and*
(ii) For all $x_1\le y_1,\dots,x_{j-1}\le y_{j-1}$ *and for all* $j=2,\dots,n-1$,

$$P\{X_j\le t|X_1=x_1,\dots,X_{j-1}=x_{j-1}\}\le P\{Y_j\le t|Y_1=y_1,\dots,Y_{j-1}=y_{j-1}\}.$$

Then, $(X_1,\dots,X_n)\overset{\mathcal{L}}{\ge}(Y_1,\dots,Y_n)$.

We now apply this result to the generalized Dirichlet distributions.

THEOREM 4.6. *Suppose that* $(X_1,\dots,X_n)\sim GD(a_1,\dots,a_n;b_1,\dots,b_n)$ *and* $(Y_1,\dots,Y_n)\sim GD(c_1,\dots,c_n;d_1,\dots,d_n)$, *where* $a_i\ge c_i$ *and* $b_i+\sum_{k=i+1}^n(a_k+b_k)\le d_i+\sum_{k=i+1}^n(c_k+d_k)$ *for all* $i=1,\dots,n$. *Then,* $(X_1,\dots,X_n)\overset{\mathcal{L}}{\ge}(Y_1,\dots,Y_n)$.

PROOF. By Corollary 2.4, the conditional distribution of $X_j/(1-X_1-\cdots-X_{j-1})$, given $(X_1,\dots,X_{j-1})=(x_1,\dots,x_{j-1})$, is a beta distribution, $B(a_j,b_j+\sum_{i=j+1}^n(a_i+b_i-1))$. Similarly, the conditional distribution of $Y_j/(1-Y_1-\cdots-Y_{j-1})$ given $(Y_1,\dots,Y_{j-1})=(y_1,\dots,y_{j-1})$ is $B(c_j,d_j+\sum_{i=j+1}^n(c_i+d_i-1))$, $j=1,\dots,n$. By assumption, $a_j\ge c_j$ and $d_j+\sum_{i=j+1}^n(c_i+d_i)\ge b_j+\sum_{i=j+1}^n(a_i+b_i)$, $j=1,\dots,n$; therefore, by Example 4.2 and Lemma 4.3, we have $X_1\overset{\mathcal{L}}{\ge}Y_1$ and $X_j|\{X_1=x_1,\dots,X_{j-1}=x_{j-1}\}\overset{\mathcal{L}}{\ge}Y_j|\{Y_1=y_1,\dots,Y_{j-1}=y_{j-1}\}$, $j=2,\dots,n-1$. Thus, by Proposition 4.5, $(X_1,\dots,X_n)\overset{\mathcal{L}}{\ge}(Y_1,\dots,Y_n)$. □

COROLLARY 4.7. *Suppose that* $(X_1,\dots,X_n)\sim GD(a_1,\dots,a_n;b_1,\dots,b_n)$ *and* $(Y_1,\dots,Y_n)\sim GD(c_1,\dots,c_n;1,\dots,1,d)$ *i.e.,* $(Y_1,\dots,Y_n)$ *has a classical Dirichlet distribution, and assume that*
(i) $a_i\ge c_i$, $i=1,\dots,n$;
(ii) $d\ge\sum_{k=i+1}^n(a_k-c_k)+\sum_{k=i}^n b_k-n+i$, $i=1,\dots,n-1$; *and*
(iii) $d\ge b_n$.
Then, $(X_1,\dots,X_n)\overset{\mathcal{L}}{\ge}(Y_1,\dots,Y_n)$ *and, for* $k_1,\dots,k_n\ge 0$,

$$P(X_1\ge k_1,\dots,X_n\ge k_n)\le\prod_{i=1}^n P(Y_i\ge k_i).$$

PROOF. The stochastic inequality $(X_1,\dots,X_n)\overset{\mathcal{L}}{\ge}(Y_1,\dots,Y_n)$ follows by substituting $d_i=1$, $i=1,\dots,n-1$ and $d_n=d$ in Theorem 4.6. Because the Dirichlet distributions are S-MRR$_2$ (see [**20**]), we have $P(Y_1\ge k_1,\dots,Y_n\ge k_n)\le\prod_{i=1}^n P(Y_i\ge k_i)$, $k_1,\dots,k_n\ge 0$. From the definition of stochastic ordering, we also obtain $P(X_1\ge k_1,\dots,X_n\ge k_n)\le P(Y_1\ge k_1,\dots,Y_n\ge k_n)$, and therefore

$$P(X_1\ge k_1,\dots,X_n\ge k_n)\le P(Y_1\ge k_1,\dots,Y_n\ge k_n)\le\prod_{i=1}^n P(Y_i\ge k_i).$$

and the proof now is complete. □

The usefulness of the above result stems from the fact that although an exact analytical expression for the cumulative distribution function of a generalized Dirichlet distributed vector will be complicated, bounds for that function may be obtained in terms of the cumulative distribution function of a classical Dirichlet distribution.

5. Parameter estimation

For $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$, there are $2n$ parameters appearing in the probability density function. In this section we comment on the method-of-moments and maximum likelihood methods for estimating the parameters, and we also derive formulas for the method-of-moments estimators. Under either approach, we construct estimators of the parameters $a_1, \ldots, a_n, b_1, \ldots, b_n$ using the stochastic representations given in Lemma 4.3.

5.1. The method of moments. For $(X_1, \ldots, X_n) \sim GD(a_1, \ldots, a_n; b_1, \ldots, b_n)$ we apply Lemma 4.3 to express $(X_1, \ldots, X_n)$ in the form

$$(X_1, \ldots, X_n) \stackrel{\mathcal{L}}{=} (Z_1, Z_2(1 - Z_1), \ldots, Z_n \prod_{i=1}^{n-1} (1 - Z_i)),$$

where $Z_1 = X_1$ and $Z_i = X_i/(1-\sum_{k=1}^{i-1} X_k)$, $i = 2, \ldots, k$, are mutually independent beta variables. We write $Z_i \sim B(a_i, c_i)$, $i = 1, \ldots, n$, where c_i is given in Lemma 4.3 in terms of the a_i and b_i.

Connor and Mosimann [**8**] provide the following formulas for the moments of $X_1, \ldots, X_n$ in terms of the moments of $Z_1, \ldots, Z_n$: For $i = 1, \ldots, n$, define

$$\mu_{i1} = E(Z_i) = \frac{a_i}{a_i + b_i}, \qquad \mu_{i2} = E(Z_i^2) = \frac{a_i(a_i + 1)}{(a_i + b_i)(a_i + b_i + 1)}, \tag{5.1}$$

the first and the second moments of Z_i, respectively; then,

$$E(X_i) = E\Big(Z_i \prod_{j=1}^{i-1} (1 - Z_j)\Big) = \mu_{i1} \prod_{j=1}^{i-1} (1 - \mu_{j1}), \tag{5.2}$$

and

$$E(X_i^2) = E\Big(Z_i^2 \prod_{j=1}^{i-1} (1 - Z_j)^2\Big) = \mu_{i2} \prod_{j=1}^{i-1} (1 - 2\mu_{j1} + \mu_{j2}). \tag{5.3}$$

From (5.1)-(5.3), we obtain

$$a_i = \frac{\mu_{i1}(\mu_{i1} - \mu_{i2})}{\mu_{i2} - \mu_{i1}^2}, \qquad b_i = \frac{(1 - \mu_{i1})(\mu_{i1} - \mu_{i2})}{\mu_{i2} - \mu_{i1}^2}, \tag{5.4}$$

$i = 1, \ldots, n$.

Given a random sample $(X_{1j}, \ldots, X_{nj})$, $j = 1, \ldots, N$, from $X_1, \ldots, X_n$, the method-of-moment estimators of $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ can be calculated in two ways. By solving (5.4) and proceeding in the usual way, we obtain the method-of-moments estimators,

$$\widehat{a}_i = \frac{m_{i1}(m_{i1} - m_{i2})}{m_{i2} - m_{i1}^2}, \qquad \widehat{b}_i = \frac{(1 - m_{i1})(m_{i1} - m_{i2})}{m_{i2} - m_{i1}^2},$$

$i = 1, \ldots, n$, where

$$m_{i1} = \frac{1}{N} \sum_{j=1}^{N} \frac{Z_{ij}}{1 - \sum_{l=1}^{i-1} Z_{lj}}, \qquad m_{i2} = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{Z_{ij}}{1 - \sum_{l=1}^{i-1} Z_{lj}} \right)^2,$$

and $Z_{ij} = X_{ij}/(1 - \sum_{l=1}^{i-1} X_{il})$.

We can also use the moment estimators of $E(X_i)$ and $E(X_i^2)$ directly. By (5.2) and (5.3),

$$\mu_{i1} = \frac{E(X_i)}{1 - \sum_{j=1}^{i-1} E(X_j)},$$

and then μ_{i2} can be calculated recursively using the equations,

$$\mu_{i2} = \frac{E(X_i^2)}{\prod_{j=1}^{i-1}(1 - 2\mu_{j1} + \mu_{j2})},$$

$i = 1, \ldots, n$. Note that m_{i1} and m_{i2} are, respectively, the method-of-moments estimators of $E(Z_i)$ and $E(Z_i^2)$. It is also clear that $1 \geq m_{i1} \geq m_{i2}$ and by the Cauchy-Schwarz inequality, $m_{i2} \geq m_{i1}^2$; hence $0 \leq m_{i1}^2 \leq m_{i2} \leq m_{i1} \leq 1$, hence $\widehat{a}_i, \widehat{b}_i > 0$, almost surely.

5.2. The method of maximum likelihood. Instead of attempting to solve the $2n$ log-likelihood equations simultaneously we again apply the stochastic representation, Lemma 4.1, and reduce the problem to working with n likelihood equations in pairs. Thus, the problem is reduced to finding n pairs of maximum likelihood estimators for the parameters of n beta distributions.

From the previous discussion, we have $Z_i = X_i/(1 - \sum_{j=1}^{i-1} X_j) \sim B(a_i, c_i)$, $i = 1, \ldots, n$. For a random sample $(X_{1j}, \ldots, X_{nj})$, $j = 1, \ldots, N$, from $X_1, \ldots, X_n$, the corresponding likelihood function is

$$\prod_{l=1}^{N} \prod_{i=1}^{n} \left[\frac{\Gamma(a_i + c_i)}{\Gamma(a_i)\Gamma(c_i)} \left(\frac{x_{il}}{1 - \sum_{j=1}^{i-1} x_{jl}} \right)^{a_i - 1} \left(\frac{\sum_{j=1}^{i} x_{jl}}{1 - \sum_{j=1}^{i-1} x_{jl}} \right)^{c_i - 1} \right],$$

where we have expressed the X_{ij} in terms of the mutually independent Z_{ij}, and then reversed the process. Differentiating with respect to the parameters, the likelihood equations are obtained as

$$\begin{aligned} 0 &= \frac{\partial \log L}{\partial a_i} = \psi(a_i) - \psi(a_i + c_i) - \frac{1}{N} \sum_{l=1}^{N} \log x_{il} \\ 0 &= \frac{\partial \log L}{\partial c_i} = \psi(b_i) - \psi(a_i + c_i) - \frac{1}{N} \sum_{l=1}^{N} \log(1 - x_{il}), \end{aligned}$$

$i = 1, \ldots, n$, where $\psi(x) = d \log \Gamma(x)/dx$ is the digamma function.

There clearly is no closed-form solution to this system of equations. Gnanadesikan, Pinkham, and Hughes [**11**] developed a numerical approach to solving the system of equations; they apply Newton's method, with the method-of-moments estimates used as the initial values of the iterative scheme. Beckman and Tietjen [**3**] developed an improved approach in which no starting values are required and no convergence problems generally have been encountered.

Kottas and Lau [**21**], in comparing the method-of-moments and the maximum likelihood estimators of parameters of the beta distributions, concluded that the maximum likelihood estimators are superior to the method-of-moments estimators in terms of relative efficiency and consistency. Kottas and Lau also show that the method-of-moments estimators have larger variances than the maximum likelihood estimators and, moreover, for certain regions in which a_i and b_i are small, or their difference is large, the variances of the method-of-moments estimators actually exceed the variances of the maximum likelihood estimators by at least 25 percent.

Because all maximum likelihood estimators are functions of the minimal sufficient statistics and are asymptotically best normal estimators, we recommend that the maximum likelihood method be used for estimation of the parameters of the generalized Dirichlet distribution.

References

[1] J. Aitchison, *The Statistical Analysis of Compositional Data*, Chapman & Hall, London, 1986.

[2] K. Aomoto, *On the structure of integrals of power product of linear functions*, Sci. Papers College Gen. Ed. Univ. Tokyo, **27** (1977), 49–61.

[3] R. J. Beckman and G. L. Tietjen, *Maximum likelihood estimation for the beta distribution*, J. Statist. Comput. Simul., **7** (1978), 253–258.

[4] N. Bouguila and D. Ziou, *A hybrid SEM algorithm for high-dimensional unsupervised learning using a finite generalized Dirichlet mixture*, IEEE Trans. Image Process., **15** (2006), 2657–2668.

[5] N. Bouguila and D. Ziou, *High-dimensional unsupervised selection and estimation of a finite generalized Dirichlet mixture model based on minimum message length*, IEEE Trans. Pattern Anal. Mach. Intell., **29** (2007), 1716–1731.

[6] W.-Y. Chang, *Generalized Dirichlet Distributions and Variance Inequalities*, Doctoral Dissertation, University of Virginia, 1996.

[7] W.-Y. Chang and D. St. P. Richards, *Variance inequalities for functions of multivariate random variables*, Contemp. Math., **234** (1999), 43–67.

[8] R. J. 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.

[9] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Generalized Euler integrals and A-hypergeometric functions*, Adv. Math., **84** (1990), 255–271.

[10] I. M. Gelfand, M. I. Graev, and V. S. Retakh, *General hypergeometric systems of equations and series of hypergeometric type*, Russian Math. Surveys, **47** (1992), 1–88.

[11] R. Gnanadesikan, R. S. Pinkham, and L. P. Hughes, *Maximum likelihood estimation of the parameters of the beta distributions from the smallest order statistics*, Technometrics, **9** (1967), 607–620.

[12] I. J. Good, *The Estimation of Probabilities*, MIT Press, Cambridge, MA, 1965.

[13] R. D. Gupta and D. St. P. Richards, *Multivariate Liouville distributions*, J. Multivariate Anal., **23** (1987), 233–256.

[14] R. D. Gupta and D. St. P. Richards, *Multivariate Liouville distributions, III*, J. Multivariate Anal. **43** (1992), 29–57.

[15] R. D. Gupta and D. St. P. Richards, *Multivariate Liouville distributions, V*, In: *Advances in the Theory and Practice of Statistics* (N. L. Johnson and N. Balakrishnan, eds.), pp. 377–396, Wiley, New York, 1997.

[16] R. D. Gupta and D. St. P. Richards, *The covariance structure of the multivariate Liouville distributions*, Contemp. Math., **287** (2001), 125–138.

[17] R. K. S. Hankin, *A generalization of the Dirichlet distribution*, J. Statist. Software, 2009, to appear; URL: http://cran.r-project.org/web/packages/hyperdirichlet/index.html.

[18] S. Karlin, *Total Positivity*, Volume I, Stanford University Press, Stanford, CA, 1968.

[19] S. Karlin and Y. Rinott, *Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions*, J. Multivariate Anal., **10** (1980), 467–498.

[20] S. Karlin and Y. Rinott, *Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distributions*, J. Multivariate Anal., **10** (1980), 499–516.

[21] J. F. Kottas and H. S. Lau, *On estimating parameters for beta distributions*, Decision Sciences, **9** (1978), 526–531.

[22] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions: Models and Applications*, second edition. Wiley, New York, 2000.

[23] R. H. Lochner, *A generalized Dirichlet distribution in Bayesian life testing*, J. Roy. Statist. Soc., Ser. B, **37** (1975), 103–113.

[24] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[25] J. G. Mauldon, *Random division of an interval*, Proc. Cambridge Phil. Soc., **47** (1951), 331–336.

[26] K.-W. Ng, M.-L. Tang, M. Tan, and G.-L. Tian, *Grouped Dirichlet distribution: A new tool for incomplete categorical data analysis*, J. Multivariate Anal., **99** (2008), 490–509.

[27] M. R. Novick and J. E. Grizzle, *A Bayesian approach to the analysis of data from clinical trials*, J. Amer. Statist. Assoc., **60** (1965), 81–96.

[28] R. Pyke, *Spacings*, J. Roy. Stat. Soc., Ser. B, **27** (1965), 395–449.

[29] R. L. Smith, *Multivariate threshold methods*, In: *Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simiu, eds.), pp. 225–248, Kluwer, Dordrecht, 1994.

[30] A. F. Veinott, *Optimal policy in a dynamic single product, non-stationary inventory model with several demand classes*, Operations Res., **13** (1965), 761–778.

U.S. Bureau of the Census, Washington, DC 20233, U.S.A.

Department of Applied Statistics and Computer Science, University of New Brunswick, Saint John, New Brunswick, E2L 4L5, Canada
E-mail address: `gupta@unbsj.ca`

Department of Statistics, Penn State University, University Park, PA 16802
E-mail address: `richards@stat.psu.edu`