Grouped Dirichlet distribution: A new tool for incomplete categorical data analysis

Kai Wang Ng\textsuperscript{a}, Man-Lai Tang\textsuperscript{b}, Ming Tan\textsuperscript{c}, Guo-Liang Tian\textsuperscript{c}, \textsuperscript{*}

\textsuperscript{a}Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, PR China
\textsuperscript{b}Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, PR China
\textsuperscript{c}Division of Biostatistics, University of Maryland Greenebaum Cancer Center, 22 South Greene Street, Baltimore, MD 21201, USA

Received 31 March 2006
Available online 26 January 2007

Abstract

Motivated by the likelihood functions of several incomplete categorical data, this article introduces a new family of distributions, \textit{grouped Dirichlet distributions} (GDD), which includes the classical \textit{Dirichlet distribution} (DD) as a special case. First, we develop distribution theory for the GDD in its own right. Second, we use this expanded family as a new tool for statistical analysis of incomplete categorical data. Starting with a GDD with two partitions, we derive its stochastic representation that provides a simple procedure for simulation. Other properties such as mixed moments, mode, marginal and conditional distributions are also derived. The general GDD with more than two partitions is considered in a parallel manner. Three data sets from a case-control study, a leprosy survey, and a neurological study are used to illustrate how the GDD can be used as a new tool for analyzing incomplete categorical data. Our approach based on GDD has at least two advantages over the commonly used approach based on the DD in both frequentist and conjugate Bayesian inference: (a) in some cases, both the maximum likelihood and Bayes estimates have closed-form expressions in the new approach, but not so when they are based on the commonly-used approach; and (b) even if a closed-form solution is not available, the EM and data augmentation algorithms in the new approach converge much faster than in the commonly-used approach.

\textcopyright{} 2007 Elsevier Inc. All rights reserved.

\textit{AMS 1991 subject classification:} 60E10; 62G35; 62H05; 62H10; 62J05

\textit{Keywords:} Beta–Liouville distribution; Data augmentation; Dirichlet distribution; EM algorithm; Grouped Dirichlet distribution; Incomplete categorical data

* Corresponding author. Fax: +1 410 3282578.
E-mail address: gtian2@ummaryland.edu (G.-L. Tian).

0047-259X/S - see front matter \textcopyright{} 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2007.01.010
1. Introduction

As a fundamental multivariate distribution, the Dirichlet distribution (DD) plays an important role in statistical modeling, distribution theory [7,3], and Bayesian statistical analysis of categorical data [1] and of compositional data [2]. It is well known that the covariance structure associated with the DD is completely non-positive. Hence, those compositional data that require positive covariance structures cannot be modeled by the DD. Aitchison [2] developed a class of logistic normal distributions partly in response to this shortcoming. It is noteworthy that the class of logistic normal distributions does not include the DD.

Extensions of the DD for various purposes have been studied in the literature. Amongst them, the Liouville and the generalized Liouville distributions are perhaps the most famous. For instance, Marshall and Olkin [18, Chapter 11] described the family of the Liouville distributions. Sivazlian [22,23] presented some results on marginal distributions and transformation properties for the class of Liouville distributions. Gupta and Richards [11–13] used the Weyl fractional integral and Deny’s theorem in measure theory on locally compact groups to derive some important results on the multivariate Liouville distributions. The integral related to the generalized Liouville distribution was first presented by Edwards [6, pp. 160–162] but received no attention from statisticians until Marshall and Olkin [18]. Fang et al. [7] provided an extensive study of the Liouville family. Rayens and Srinivasan [19] extended the Liouville family on the simplex further to the generalized Liouville family and studied its application to compositional data analysis.

The likelihood function of a complete categorical data set usually has the form of the density of a DD and thus finding the maximum likelihood estimate (MLE) is alike finding the mode. This likelihood form also lends Bayesian analysis with a conjugate prior relatively easily. In the case of incomplete categorical data, however, the likelihood has a very different form which is not easy to handle. Here the Liouville and generalized Liouville distributions as extensions of the DD cannot help neither. The most commonly adopted approach is the data augmentation (DA) algorithm [25]. This approach introduces latent variables which, together with the observed data, can make the augmented likelihood into a familiar form. In the case of incomplete categorical data, the augmented likelihood is in the form of a Dirichlet density and hence the EM algorithm [5] and the DA algorithm can be used, respectively, to obtain the MLE and the Bayes estimate. Although EM/DA algorithms can be implemented easily, they could converge slowly when the number of the introduced latent variables is large. In fact, the EM and DA algorithms based on the DD tend to introduce more latent variables than necessary.

This paper introduces a new family named grouped Dirichlet distributions (GDD), which includes the DD as a special case. It is motivated by the likelihood functions of incomplete categorical data which often occur in biomedical studies. Our purpose is two-fold. First, we develop distribution theory and explore important distribution properties. Second, we provide a new way to manage the statistical analysis of incomplete categorical data based on the GDD. We will show that the approach has two advantages when comparing with that based on the traditional DD for incomplete categorical data: (a) in some cases, both the MLE and Bayes estimates have closed-form expressions based on the new approach, but not so when based on the commonly used approach of augmented likelihood; and (b) in other cases when the closed-form solution is not available, the EM and DA algorithms based on the new approach converge much faster than that based on the augmented likelihood in the Dirichlet form.

This article is organized as follows. Section 2 presents three motivating examples. In Section 3, we define a special case of the GDD and derive its stochastic representation, mixed moments and mode. In Section 4, we derive marginal distributions in the form of the Liouville
and the beta–Liouville distributions. Section 5 considers conditional distributions related to the Liouville distributions as well. In Section 6 we study the GDD in full generality. Three data sets are analyzed in Section 7 to illustrate the application of the proposed distributions as a new tool in frequentist and conjugate Bayesian inference for incomplete categorical data. We conclude with a discussion in Section 8 and the lengthy proofs of some propositions will be presented in the Appendix.

2. Motivating examples

Example 1. Cervical cancer data: Williamson and Haber [27] reported a case-control study which examined the relationship between disease status of cervical cancer and the number of sex partners and other risk factors. Cases were 20–70 year-old women of Fulton or Dekalb county in Atlanta, Georgia. They were diagnosed and were ascertained to have invasive cervical cancer. Controls were randomly chosen from the same counties and the same age ranges. Table 1 gives the cross-classification of disease status (control or case, denoted by $X = 0$ or $X = 1$) and number of sex partners (“few” (0–3) or “many” (≥4), denoted by $Y = 0$ or $Y = 1$). Generally, a sizable proportion (13.5% in this example) of the responses would be missing because of “unknown” or “refused to answer” in a telephone interview. In this example, they assumed that it is missing at random (MAR, [21]); i.e., the absence of data is independent of both number of sex partners and disease status. The objective is to examine if association exists between the number of sex partners and disease status of cervical cancer.

Let $Y_{\text{obs}} = \{(n_1, \ldots, n_4); (n_{12}, n_{34})\}$ denote the observed counts, $\theta = (\theta_1, \ldots, \theta_4)^\top$ the cell probability vector and $\psi = \theta_1\theta_4/(\theta_2\theta_3)$ the odds ratio. The likelihood function is

$$L(\theta|Y_{\text{obs}}) \propto \left(\prod_{i=1}^{4} \theta_i^{n_i}\right) \cdot (\theta_1 + \theta_2)^{n_{12}}(\theta_3 + \theta_4)^{n_{34}},$$

$\theta \in \mathcal{T}_4,$

where $\mathcal{T}_n$ denotes the closed simplex $\{(\theta_1, \ldots, \theta_n)^\top : \theta_i \geq 0, \ i = 1, \ldots, n, \sum_{i=1}^{n} \theta_i = 1\}.$

Example 2. Leprosy survey data: Hocking and Oxspring [14] analyzed a data set (see, Table 2) on the use of drugs in the treatment of leprosy. A random sample of 196 patients was cross-classified by two categories of the degree of infiltration (little or much) and five categories of changes in clinical condition (marked, moderate, slight improvement, stationary or worse) after a fixed time over which treatments were administered. A supplementary sample of another 400 different patients was classified coarsely with respect to improvement in health. Again, MAR is assumed here. The goal is to estimate the cell probabilities.

<table>
<thead>
<tr>
<th>Disease status of cervical cancer</th>
<th>Number of sex partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = 0$ (control)</td>
<td>$Y = 0$ (few, 0–3)</td>
</tr>
<tr>
<td>$X = 0$ (control)</td>
<td>164 ($n_1$, $\theta_1$)</td>
</tr>
<tr>
<td>$X = 1$ (case)</td>
<td>103 ($n_3$, $\theta_3$)</td>
</tr>
</tbody>
</table>

Note: The observed counts and the corresponding cell probabilities are in parentheses.
Table 2
Leprosy survey data from Hocking and Oxspring [14]

<table>
<thead>
<tr>
<th>Main sample</th>
<th>Supplemental sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOI Clinical condition</td>
<td>DOI Clinical condition</td>
</tr>
<tr>
<td>Improvement Stationary Worse</td>
<td>Improvement Stationary Worse</td>
</tr>
<tr>
<td>Marked Moderate Slight</td>
<td>Marked Moderate Slight</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Little</th>
<th>(n1, ( \theta_1 ))</th>
<th>11</th>
<th>27</th>
<th>42</th>
<th>53</th>
<th>11</th>
<th>Little</th>
<th>(n123, ( \sum_{i=1}^{3} \theta_i ))</th>
<th>144</th>
<th>120</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n2, ( \theta_2 ))</td>
<td>(n3, ( \theta_3 ))</td>
<td>(n4, ( \theta_4 ))</td>
<td>(n5, ( \theta_5 ))</td>
<td>(n6, ( \theta_6 ))</td>
<td>(n7, ( \theta_7 ))</td>
<td>(n8, ( \theta_8 ))</td>
<td>(n9, ( \theta_9 ))</td>
<td>(n10, ( \theta_{10} ))</td>
<td>(n678, ( \sum_{i=6}^{8} \theta_i ))</td>
<td>92</td>
<td>24</td>
</tr>
<tr>
<td>(n11, ( \tilde{\theta}_1 ))</td>
<td>(n12, ( \tilde{\theta}_2 ))</td>
<td>(n13, ( \tilde{\theta}_3 ))</td>
<td>(n14, ( \tilde{\theta}_4 ))</td>
<td>(n15, ( \tilde{\theta}_5 ))</td>
<td>(n16, ( \tilde{\theta}_6 ))</td>
<td>(n17, ( \tilde{\theta}_7 ))</td>
<td>(n18, ( \tilde{\theta}_8 ))</td>
<td>(n19, ( \tilde{\theta}_{10} ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Much</th>
<th>7</th>
<th>15</th>
<th>16</th>
<th>13</th>
<th>1</th>
<th>Much</th>
<th>92</th>
<th>24</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n6, ( \theta_6 ))</td>
<td>(n7, ( \theta_7 ))</td>
<td>(n8, ( \theta_8 ))</td>
<td>(n9, ( \theta_9 ))</td>
<td>(n10, ( \theta_{10} ))</td>
<td>(n678, ( \sum_{i=6}^{8} \theta_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
<td>(n678, ( \sum_{i=6}^{8} \tilde{\theta}_i ))</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>42</td>
<td>58</td>
<td>66</td>
<td>12</td>
<td>Total</td>
<td>236</td>
<td>144</td>
<td>20</td>
</tr>
</tbody>
</table>

*Note: DOI = degree of infiltration. The observed counts and probabilities are in parentheses.*

Table 3
Neurological complication data from Choi and Stablein [4]

<table>
<thead>
<tr>
<th>( Y = 0 )</th>
<th>( Y = 1 )</th>
<th>Supplement on ( X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X = 0 )</td>
<td>( X = 1 )</td>
<td>( X = 1 )</td>
</tr>
<tr>
<td>6 ( (n_1, \theta_1) )</td>
<td>3 ( (n_2, \theta_2) )</td>
<td>2 ( (n_{12}, \theta_1 + \theta_2) )</td>
</tr>
<tr>
<td>8 ( (n_3, \theta_3) )</td>
<td>8 ( (n_4, \theta_4) )</td>
<td>4 ( (n_{13}, \theta_3 + \theta_4) )</td>
</tr>
</tbody>
</table>

Note: “\( X = 0 \) (1)” means that patient’s complication at the beginning of the treatment is absent (present), “\( Y = 0 \) (1)” means that patient’s complication at the end of the treatment is absent (present). The observed frequencies and probabilities are in parentheses.

Let \( Y_{\text{obs}} = \{(n_1, \ldots, n_{10}); (n_{123}, \tilde{n}_4, \tilde{n}_5, n_{678}, \tilde{n}_9, \tilde{n}_{10})\} \) denote the observed counts and \( \theta = (\theta_1, \ldots, \theta_{10})^T \) the cell probability vector. The likelihood function is

\[
L(\theta|Y_{\text{obs}}) \propto \left( \prod_{i=1}^{10} \theta_i^{n_i + \tilde{n}_i} \right) \cdot \left( \sum_{i=1}^{3} \theta_i \right)^{n_{123}} \left( \sum_{i=4}^{5} \theta_i \right)^{0} \left( \sum_{i=6}^{8} \theta_i \right)^{n_{678}} \left( \sum_{i=9}^{10} \theta_i \right)^{0},
\]

where \( \theta \in \mathcal{T}_{10} \) and \( \tilde{n}_i = 0 \) for \( i = 1, 2, 3, 6, 7, 8 \).

**Example 3.** Neurological complication data: Choi and Stablein [4] reported a neurological study in which 33 young meningitis patients at the St. Louis Children’s Hospital were given neurological tests at the beginning and the end of a standard treatment on neurological complication. The response of the tests is absent (denoted by 0) or present (denoted by 1) of any neurological complication. The data are reported in Table 3 and consist of two parts: the complete observations which form a \( 2 \times 2 \) table from correlated series, and the incomplete observations which form the row and column margins from two independent binomial populations. The primary objective of this study is to detect if there is a substantial proportion difference in proportion before and after the treatment.
Let $Y_{\text{obs}} = \{(n_1, \ldots, n_4); (n_{12}, n_{34}); (n_{13}, n_{24})\}$ be the observed frequencies and $\theta = (\theta_1, \ldots, \theta_4)^T$ the cell probability vector. We are interested in the parameter $\theta_{23} = \Pr(Y = 1) - \Pr(X = 1) = \theta_2 - \theta_3$, the difference between the proportions of patients with neurological complication before and after the treatment. Under the MAR assumption, the likelihood function of the observed data is

$$L(\theta|Y_{\text{obs}}) \propto \left\{ \prod_{i=1}^{4} \theta_{i}^{n_{i}} \cdot (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}} \right\} \cdot (\theta_1 + \theta_3)^{n_{13}} (\theta_2 + \theta_4)^{n_{24}},$$

where $\theta \in \mathcal{T}_4$. \hfill (2.3)

### 3. Definitions and properties

Motivated by the likelihood functions (2.1) and (2.3), we first define and study grouped Dirichlet distribution (GDD) with two partitions. An $n$-variate $x \in \mathcal{T}_n$ is said to follow a GDD with two partitions, if the density of $x_{-n} \sim (x_1, \ldots, x_{n-1})^T$ is [26]

$$GD_{n,2,s}(x_{-n}|a, b) = c_2^{-1} \cdot \left( \prod_{i=1}^{n} x_i^{a_i-1} \right) \cdot \left( \sum_{i=1}^{s} x_i \right)^{b_1} \cdot \left( \sum_{i=s+1}^{n} x_i \right)^{b_2}, \quad x_{-n} \in \mathcal{V}_{n-1}, \ (3.1)$$

where $a = (a_1, \ldots, a_n)^T$ and $b = (b_1, b_2)^T$ are two non-negative parameter vectors, $s$ is a known positive integer less than $n$, $\mathcal{V}_{n-1} \equiv \{ x_{-n} : x_i \geq 0, i = 1, \ldots, n-1, \sum_{i=1}^{n-1} x_i \leq 1 \}$ denotes the open simplex, and the normalizing constant is given by

$$c_2 = B_s(a_1, \ldots, a_s) \cdot B_{n-s}(a_{s+1}, \ldots, a_n) \cdot B \left( \sum_{i=1}^{s} a_i + b_1, \sum_{i=s+1}^{n} a_i + b_2 \right).$$

$B_n(a) \equiv \prod_{i=1}^{n} \Gamma(a_i)/\Gamma(\sum_{i=1}^{n} a_i)$ denotes the multivariate beta function. We will write $x \sim GD_{n,2,s}(a, b)$ on $\mathcal{T}_n$ or $x_{-n} \sim GD_{n,2,s}(a, b)$ on $\mathcal{V}_{n-1}$ to distinguish between the two equivalent representations if necessary. In particular, when $s = n - 1$, the GDD (3.1) reduces to an important special case, i.e.,

$$GD_{n,2,n-1}(x_{-n}|a, b) = \frac{\left( \prod_{i=1}^{n-1} x_i^{a_i-1} \right) \cdot ||x_{-n}||^{b_1} (1 - ||x_{-n}||)^{a_n+b_2-1}}{B_{n-1}(a_1, \ldots, a_{n-1}) \cdot B(\sum_{i=1}^{n-1} a_i + b_1, a_n + b_2)}, \ (3.2)$$

where $x_{-n} \in \mathcal{V}_{n-1}$ and $||x_{-n}|| \equiv \sum_{i=1}^{n-1} x_i$. In addition, when $b_1 = b_2 = 0$, the GDD (3.1) reduces to the Dirichlet distribution $D_n(a)$.

To understand the nature of the GDD, we partition $x_{n \times 1}$ into $(x^{(1)^T}, x^{(2)^T})^T$ each with $s$ and $n-s$ elements, respectively. We partition $y_{n \times 1} = (y^{(1)^T}, y^{(2)^T})^T$ and $a_{n \times 1} = (a^{(1)^T}, a^{(2)^T})^T$ in the same fashion. Furthermore, throughout this paper, we define the parametric functions, $z_1, z_2$ and $z_{12}$, as follows:

$$z_1 \equiv ||a^{(1)}|| + b_1, \quad z_2 \equiv ||a^{(2)}|| + b_2, \quad z_{12} = z_1 + z_2. \ (3.3)$$

Hence, the normalizing constant $c_2$ can be rewritten as

$$c_2 = B_s(a^{(1)}) \cdot B_{n-s}(a^{(2)}) \cdot B(z_1, z_2). \ (3.4)$$
The following proposition provides a stochastic representation (SR) of the GDD and hence a straightforward procedure for generating independently and identically distributed (i.i.d.) samples, which plays a crucial role in Bayesian analysis for incomplete categorical data.

**Proposition 1.** An $n$-vector $\mathbf{x} = (\mathbf{x}^{(1)} , \mathbf{x}^{(2)})^T \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on $\mathcal{T}_n$ if and only if

$$\mathbf{x} \sim \left( \frac{\mathbf{x}^{(1)}}{\mathbf{x}^{(2)}} \right) \sim \left( \frac{R \cdot y^{(1)}}{(1-R) \cdot y^{(2)}} \right),$$

where (i) $y^{(1)} \sim D_s(\mathbf{a}^{(1)})$ on $\mathcal{T}_s$, $y^{(2)} \sim D_{n-s}(\mathbf{a}^{(2)})$ on $\mathcal{T}_{n-s}$; (ii) $R \sim \text{Beta}(\alpha_1, \alpha_2)$; and (iii) $y^{(1)}$, $y^{(2)}$ and $R$ are mutually independent.

The notation “$\sim$” means having the same distribution. The proof of Proposition 1 is given in the Appendix. The following results are immediate consequences of Proposition 1.

**Proposition 2.** An $n$-vector $\mathbf{x} \sim GD_{n,2,n-1}(\mathbf{a}, \mathbf{b})$ on $\mathcal{T}_n$ if and only if

$$\mathbf{x} \sim \left( \frac{\mathbf{x}_{-n}}{1 - \|\mathbf{x}_{-n}\|} \right) \sim \left( \frac{R \cdot y_{-n}}{1 - R} \right),$$

where (i) $y_{-n} \sim D_{n-1}(\mathbf{a}_{-n})$ on $\mathcal{T}_{n-1}$; (ii) $R \sim \text{Beta}(\|\mathbf{a}_{-n}\| + b_1, a_n + b_2)$; and (iii) $y_{-n}$ and $R$ are mutually independent.

**Proposition 3.** Let $\mathbf{x} = (\mathbf{x}^{(1)} , \mathbf{x}^{(2)})^T \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on $\mathcal{T}_n$, then (i) $\mathbf{x}^{(1)}/\|\mathbf{x}^{(1)}\| \sim D_1(\mathbf{a}^{(1)})$ on $\mathcal{T}_1$ and $\mathbf{x}^{(2)}/\|\mathbf{x}^{(2)}\| \sim D_{n-s}(\mathbf{a}^{(2)})$ on $\mathcal{T}_{n-s}$; (ii) $\|\mathbf{x}^{(1)}\| \sim \text{Beta}(\alpha_1, \alpha_2)$; and (iii) $\mathbf{x}^{(1)}/\|\mathbf{x}^{(1)}\|$, $\mathbf{x}^{(2)}/\|\mathbf{x}^{(2)}\|$ and $\|\mathbf{x}^{(1)}\|$ are mutually independent.

We now consider the mixed moments of $\mathbf{x}$ by means of the SR (3.5). The independence among $y^{(1)}$, $y^{(2)}$ and $R$ implies

$$E \left( \prod_{i=1}^{n} x_i^{r_i} \right) = E \left( \prod_{i=1}^{s} y_i^{r_i} \right) \cdot E \left( \prod_{i=s+1}^{n} y_i^{r_i} \right) \cdot E \left[ R \sum_{i=1}^{s} r_i (1-R)^{\sum_{i=s+1}^{n} r_i} \right].$$

Using Theorem 1.3 of Fang et al. [7, p. 18], we obtain the following results.

**Proposition 4.** Let $\mathbf{x} \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on $\mathcal{T}_n$. For any $r_1, \ldots, r_n \geq 0$, the mixed moments of $\mathbf{x}$ are given by

$$E \left( \prod_{i=1}^{n} x_i^{r_i} \right) = \frac{B_s(\mathbf{a}^{(1)} + \mathbf{r}^{(1)})}{B_s(\mathbf{a}^{(1)})} \cdot \frac{B_{n-s}(\mathbf{a}^{(2)} + \mathbf{r}^{(2)})}{B_{n-s}(\mathbf{a}^{(2)})} \cdot \frac{B(\alpha_1 + \|\mathbf{r}^{(1)}\|, \alpha_2 + \|\mathbf{r}^{(2)}\|)}{B(\alpha_1, \alpha_2)},$$

where $\mathbf{r} = (\mathbf{r}^{(1)} , \mathbf{r}^{(2)})^T$, $\mathbf{r}^{(1)} : s \times 1$, $\mathbf{r}^{(2)} : (n-s) \times 1$. In particular, we have

$$E(x_i) = \frac{a_i}{\alpha_1} \left( \frac{\alpha_1}{\|\mathbf{a}^{(1)}\|} \cdot I(1 \leq i \leq s) + \frac{\alpha_2}{\|\mathbf{a}^{(2)}\|} \cdot I(s+1 \leq i \leq n) \right),$$

where $a_i$ are the elements of the row vector $\mathbf{a}$. 

Proposition 5. The mode of the grouped Dirichlet density (3.1) is given by

$$\hat{x}_i = \frac{a_i - 1}{\|a\| + \|b\| - n} \left( \frac{\|a(1)\|}{\|a(2)\|} : I_{(i \leq s)} + \frac{\|a(1)\|}{\|a(2)\|} : I_{(s+1 \leq i \leq n)} \right).$$

(3.7)

where $1 \leq i \leq n$.

4. Marginal distributions

In this section, we first derive the marginal distributions of subvectors for the GDD with two partitions. It turns out that these marginal distributions are the Liouville distributions of the second kind, including the beta–Liouville distributions. Here we follow the notation in Fang et al. [7, Chapter 6].

A vector $z \in \mathbb{R}_+^m$ follows a Liouville distribution, denoted by $z \sim L_m(\gamma_1, \ldots, \gamma_m; f)$, if $z$ has the SR: $z \overset{d}{=} R \cdot y$, where $y \sim D_m(\gamma_1, \ldots, \gamma_m)$ on $T_m$ and $R$ is an independent random variable, called the generating variate, with density $f$, called the generating density. We see that a Liouville distribution $L_m(\gamma_1, \ldots, \gamma_m; f)$ has density

$$g(z) = \left[ \prod_{i=1}^m \frac{\gamma_i - 1}{\gamma_i} \right] \cdot f(\|z\|) \cdot \frac{\|z\|^{\gamma - 1}}{B_m(\gamma)}, \quad \gamma = (\gamma_1, \ldots, \gamma_m)^\top.$$  

(4.1)

If $f$ has an unbounded domain, the Liouville distribution is of the first kind. If $f$ is defined in the interval $[0, d]$, the Liouville distribution is of the second kind, being restricted in the simplex $\mathcal{V}_m(d) \equiv \{z : z \in \mathbb{R}_+^m \text{ and } \|z\| \leq d\}$. In particular, when $R \sim \text{Beta}(\alpha, \beta)$, we say that $z$ follows a beta–Liouville distribution and write $z \sim BL_m(\gamma_1, \ldots, \gamma_m; \alpha, \beta)$. 

Next we find the mode of the GDD directly by calculus of several variables as follows.

Proposition 5. The mode of the grouped Dirichlet density (3.1) is given by

$$\hat{x}_i = \frac{a_i - 1}{\|a\| + \|b\| - n} \left( \frac{\|a(1)\|}{\|a(2)\|} : I_{(i \leq s)} + \frac{\|a(1)\|}{\|a(2)\|} : I_{(s+1 \leq i \leq n)} \right).$$

(3.7)

where $1 \leq i \leq n$. 

4. Marginal distributions

In this section, we first derive the marginal distributions of subvectors for the GDD with two partitions. It turns out that these marginal distributions are the Liouville distributions of the second kind, including the beta–Liouville distributions. Here we follow the notation in Fang et al. [7, Chapter 6].

A vector $z \in \mathbb{R}_+^m$ follows a Liouville distribution, denoted by $z \sim L_m(\gamma_1, \ldots, \gamma_m; f)$, if $z$ has the SR: $z \overset{d}{=} R \cdot y$, where $y \sim D_m(\gamma_1, \ldots, \gamma_m)$ on $T_m$ and $R$ is an independent random variable, called the generating variate, with density $f$, called the generating density. We see that a Liouville distribution $L_m(\gamma_1, \ldots, \gamma_m; f)$ has density

$$g(z) = \left[ \prod_{i=1}^m \frac{\gamma_i - 1}{\gamma_i} \right] \cdot f(\|z\|) \cdot \frac{\|z\|^{\gamma - 1}}{B_m(\gamma)}, \quad \gamma = (\gamma_1, \ldots, \gamma_m)^\top.$$  

(4.1)

If $f$ has an unbounded domain, the Liouville distribution is of the first kind. If $f$ is defined in the interval $[0, d]$, the Liouville distribution is of the second kind, being restricted in the simplex $\mathcal{V}_m(d) \equiv \{z : z \in \mathbb{R}_+^m \text{ and } \|z\| \leq d\}$. In particular, when $R \sim \text{Beta}(\alpha, \beta)$, we say that $z$ follows a beta–Liouville distribution and write $z \sim BL_m(\gamma_1, \ldots, \gamma_m; \alpha, \beta)$.
First note that the special GDD with \( s = n - 1 \), namely \( GD_{n,2,n-1}(a, b) \), having the density (3.2) and the SR (3.6), is equivalent to a beta–Liouville distribution, namely \( BL_{n-1}(a_n; \|a_n\| + b_1, a_n + b_2) \). Furthermore, comparing the SR (3.5) with that of the Liouville distribution, we can see that \( x^{(1)} \) and \( x^{(2)} \) have beta–Liouville distributions, respectively, \( BL_s(a^{(1)}; \alpha_1, \alpha_2) \) and \( BL_{n-s}(a^{(2)}; \alpha_2, \alpha_1) \), where \( \alpha_1 \) and \( \alpha_2 \) are defined in (3.3). We therefore have the following results.

**Proposition 6.** Let \( x = (x^{(1)\top}, x^{(2)\top})\top \sim GD_{n,2,s}(a, b) \) on \( T_n \), and \( \alpha_1 \) and \( \alpha_2 \) be defined in (3.3), then (i) \( x^{(1)} \overset{d}{\sim} R \cdot y^{(1)} \sim BL_s(a^{(1)}; \alpha_1, \alpha_2) \) \( \overset{d}{=} GD_{s+1,2,s}(a^{(1)\top}, b^{(1)\top}) \), \( b \) inside \( V_s \) with density

\[
g(x^{(1)}) = \frac{\prod_{i=1}^{s} x_i^{\alpha_i-1} \cdot \|x^{(1)}\|^{b_1} (1 - \|x^{(1)}\|)^{b_2-1}}{B_s(a^{(1)}) \cdot B(\|a^{(1)}\| + b_1, \|a^{(2)}\| + b_2)}, \quad x^{(1)} \in V_s, \tag{4.2}
\]

and the SR is given by

\[
\left( x^{(1)} \begin{array}{c} \end{array} \right) \overset{d}{=} \left( \begin{array}{c} R \cdot y^{(1)} \begin{array}{c} \end{array} \end{array} \right); \tag{4.3}
\]

(ii) \( x^{(2)} \overset{d}{=} (1 - R) \cdot y^{(2)} \sim BL_{n-s}(a^{(2)}; \alpha_2, \alpha_1) \overset{d}{=} GD_{n-s+1,2,n-s}(a^{(2)\top}, b^{(2)\top}) \) inside \( V_{n-s} \) with density

\[
g(x^{(2)}) = \frac{\prod_{i=s+1}^{n} x_i^{\alpha_i-1} \cdot \|x^{(2)}\|^{b_2} (1 - \|x^{(2)}\|)^{b_1-1}}{B_{n-s}(a^{(2)}) \cdot B(\|a^{(2)}\| + b_2, \|a^{(1)}\| + b_1)}, \quad x^{(2)} \in V_{n-s},
\]

and the SR is given by

\[
\left( x^{(2)} \begin{array}{c} \end{array} \right) \overset{d}{=} \left( (1 - R) \cdot y^{(2)} \begin{array}{c} \end{array} \right). \tag{4.4}
\]

Note that the pair of subvectors in Proposition 6 are very special as they correspond naturally to the two components of the partitions of \( GD_{n,2,s}(a, b) \). We consider the marginal distributions of those subvectors \( x^{(1)} \) and \( x^{(2)} \) as well.

**Proposition 7.** Let \( x \sim GD_{n,2,s}(a, b) \) on \( T_n \), and for \( 1 \leq m < s, \ s + 1 \leq k < n \), define

\[
x^{(1)}_* = (x_1, \ldots, x_m)\top, \quad \delta_1 = \sum_{i=1}^{m} x_i, \quad y^{(1)}_* = (y_1, \ldots, y_m)\top,
\]

\[
a^{(1)}_* = (a_1, \ldots, a_m)\top, \quad \tau_1 = \sum_{i=1}^{m} a_i, \quad \tau_2 = \sum_{i=m+1}^{s} a_i,
\]

\[
x^{(2)}_* = (x_{s+1}, \ldots, x_k)\top, \quad \delta_2 = \sum_{i=s+1}^{k} x_i, \quad y^{(2)}_* = (y_{s+1}, \ldots, y_k)\top,
\]

\[
a^{(2)}_* = (a_{s+1}, \ldots, a_k)\top, \quad \tau_3 = \sum_{i=s+1}^{k} a_i, \quad \tau_4 = \sum_{i=k+1}^{n} a_i.
\]

The following are true: (i) \( x^{(1)}_* \sim L_m(a^{(1)}_*; f^{(1)}_*) \) with density

\[
g(x^{(1)}_*) = \left[ \frac{\prod_{i=1}^{m} x_i^{\alpha_i-1}}{B_m(a^{(1)}_*)} \right] \cdot \frac{1}{\delta_1^{\tau_1-1}} \cdot f^{(1)}_*(\delta_1), \quad x^{(1)}_* \in V_m, \tag{4.4}
\]
where the generating density \( f_\ast((1) \) is given by

\[
f_\ast((1)(\delta_1) = [B(\tau_1, \tau_2) B(x_1, x_2)]^{-1} \delta_1^{\tau_1-1} \int_0^{1-\delta_1} (\delta_1 + y)^b_1 \times (1 - \delta_1 - y)^{\tau_2-1} y^{\tau_2-1} dy.
\]

Furthermore, \( x_\ast((1) \) has the SR

\[
x_\ast((1) = \left(R \cdot \|y_\ast((1)\|, \left(y_\ast((1)/\|y_\ast((1)\|\right),
\]

where \( R \sim Beta(z_1, z_2), \|y_\ast((1)\| \sim Beta(\tau_1, \tau_2), \frac{y_\ast((1)}{\|y_\ast((1)\|} \sim D_m(a_\ast((1)), R_\ast((1) = R \cdot \|y_\ast((1)\| \sim f_\ast((1),
\]

and \( R, \|y_\ast((1)\| \) and \( y_\ast((1)/\|y_\ast((1)\| \) are mutually independent.

(ii) \( x_\ast((2) \sim L_{k-s}(a_\ast((2), f_\ast((2)) \) with density

\[
g(x_\ast((2)) = \frac{\prod_{i=s+1}^{k} x_i^{a_i-1}}{B_k-s(a_\ast((2)) \cdot \frac{1}{\delta_2^{t_3-1}}} \cdot f_\ast((2)(\delta_2), x_\ast((2) \in V_{k-s},
\]

where the generating density \( f_\ast((2) \) is given by

\[
f_\ast((2)(\delta_2) = [B(\tau_3, \tau_4) B(x_2, x_1)]^{-1} \delta_2^{t_3-1} \int_0^{1-\delta_2} (\delta_2 + y)^b_2 (1 - \delta_2 - y)^{\tau_1-1} y^{\tau_1-1} dy.
\]

Furthermore, \( x_\ast((2) \) has the SR

\[
x_\ast((2) = \left[(1 - R) \cdot \|y_\ast((2)\|, \left(y_\ast((2)/\|y_\ast((2)\|\right),
\]

where \( R \sim Beta(z_1, z_2), \|y_\ast((2)\| \sim Beta(\tau_3, \tau_4), \frac{y_\ast((2)}{\|y_\ast((2)\|} \sim D_k-s(a_\ast((2)), R_\ast((2) = (1 - R) \cdot \|y_\ast((2)\| \sim f_\ast((2), \) and \( R, \|y_\ast((2)\| \) and \( y_\ast((2)/\|y_\ast((2)\| \) are mutually independent.

The proof is given in the Appendix. In particular, setting \( m = 1 \) in Proposition 7 yields the following marginal density of \( x_i: \)

\[
g(x_1) = \frac{x_1^{a_1-1}}{B(a_1, \tau_2^s)} B(x_1, x_2) \int_0^{1-x_1} (x_1 + y)^b_1 (1 - x_1 - y)^{\tau_2-1} y^{\tau_2-1} dy,
\]

where \( \tau_2^s = a_2 + \cdots + a_s. \) When \( b_1 \) is a positive integer, applying the binomial expansion on \((x_1 + y)^b_1\) yields:

\[
g(x_1) = \sum_{\ell=0}^{b_1} \binom{b_1}{\ell} \frac{B(a_1 + b_1 - \ell, \tau_2^s + \ell)}{B(a_1, \tau_2^s)} \cdot Beta(x_1|a_1 + b_1 - \ell, \tau_2^s + \ell + x_2).
\]

Note that the coefficients in (4.8) constitute a beta–binomial distribution for variable \( \ell \) (i.e., a beta mixture of binomial distribution). Hence, (4.8) is a beta–binomial mixture of beta distributions when \( b_1 \) is a positive integer. The marginal density of \( x_i \) \((2 < i < n)\) can be obtained in a similar manner.
5. Conditional distributions

In this section, we consider conditional distributions of the GDD. We adopt the following notations: \( x = (x_1, \ldots, x_n) \top \in T_n, x_{-n} = (x_1, \ldots, x_{n-1}) \top \in V_{n-1}, \)

\[
\begin{align*}
x^{(1)} &= (x_1, \ldots, x_s) \top, \quad \Delta_1 = \|x^{(1)}\|, \quad x^{(1)}_{-s} = (x_1, \ldots, x_{s-1}) \top, \quad u_1 = \|x^{(1)}_{-s}\|, \\
x^{(2)} &= (x_{s+1}, \ldots, x_n) \top, \quad \Delta_2 = \|x^{(2)}\|, \quad x^{(2)}_{-n} = (x_{s+1}, \ldots, x_{n-1}) \top, \quad u_2 = \|x^{(2)}_{-n}\|, \\
a^{(1)} &= (a_1, \ldots, a_s) \top, \quad a^{(1)}_{-s} = (a_1, \ldots, a_{s-1}) \top, \quad v_1 = \|a^{(1)}\|, \\
a^{(2)} &= (a_{s+1}, \ldots, a_n) \top, \quad a^{(2)}_{-n} = (a_{s+1}, \ldots, a_{n-1}) \top, \quad v_2 = \|a^{(2)}\|.
\end{align*}
\]

Let \( x \) follow the GDD \((3.1)\). Our objective is to derive the conditional distributions of \( x^{(1)}|x^{(2)} \) and \( x^{(2)}|x^{(1)} \). Direct derivation of the conditional density of \( x^{(1)}|x^{(2)} \) via the joint and marginal densities will encounter the difficulty that \( x \) does not have a density as \( \|x\| = 1 \). Instead, we first consider the conditional distribution of \( x^{(1)}|x^{(2)} \), which is again a Liouville distribution whose generating density depends on \( x^{(2)} \) only through the \( \ell_1 \)-norm \( \|x^{(2)}\| \).

**Proposition 8.** Let \( x \sim GD_{n,2,s}(a, b) \) on \( T_n \), then (i) \( x^{(1)}_{-s}|x^{(2)} \sim L_m(a^{(1)}_{-s}; f^{(1)}_{-s}(1)) \) with density

\[
g(x^{(1)}_{-s}|x^{(2)}) = \left[ \prod_{i=1}^{s-1} x_i^{a_i-1} \right] \frac{f^{(1)}_{-s}(u_1|x^{(2)})}{B_{s-1}(a^{(1)}_{-s})} \frac{u_1^{v_1-1}}{u_1^{v_1-1} - 1} ,
\]

where the generating density is given by

\[
f^{(1)}_{-s}(u_1|x^{(2)}) = \frac{1}{(1 - \Delta_2)} \frac{u_1^{v_1-1}}{1 - \frac{u_1}{1 - \Delta_2}} ,
\]

which is a beta distribution with scale parameter \( 1 - \Delta_2 \).

(ii) \( x^{(2)}_{-n}|x^{(1)} \sim L_m(a^{(2)}_{-n}; f^{(2)}_{-n}(1)) \) with density

\[
g(x^{(2)}_{-n}|x^{(1)}) = \left[ \prod_{i=s+1}^{n} x_i^{a_i-1} \right] \frac{f^{(2)}_{-n}(u_2|x^{(1)})}{B_{n-1-s}(a^{(2)}_{-n})} \frac{u_2^{v_2-1}}{u_2^{v_2-1} - 1} ,
\]

where the generating density is given by

\[
f^{(2)}_{-n}(u_2|x^{(1)}) = \frac{1}{(1 - \Delta_1)} \frac{u_2^{v_2-1}}{1 - \frac{u_2}{1 - \Delta_1}} ,
\]

which is a beta distribution with scale parameter \( 1 - \Delta_1 \).

**Proof.** We only need to prove case (ii). Let \( g(x_{-n}) = g(x^{(1)}_{-n}, x^{(2)}_{-n}) \) denote the grouped Dirichlet density \((3.1)\) and \( g(x^{(1)}) \) the marginal density \((4.1)\). Hence, the conditional density

\[
g(x^{(2)}_{-n}|x^{(1)}) = \frac{g(x^{(1)}, x^{(2)}_{-n})}{g(x^{(1)})} = \prod_{i=s+1}^{n} x_i^{a_i-1} \left[ B_{n-1-s}(a^{(2)}_{-n}) \cdot (1 - \Delta_1)\|a^{(2)}\|^{-1} \right] .
\]

Simplifying \((5.3)\) yields \((5.1)\) and \((5.2)\). \( \Box \)
Proposition 9. Let \( x \sim GD_{n,2,s}(a, b) \) on \( \mathcal{T}_n \), then (i) \( \frac{x^{(1)}}{1-\Lambda_2} \bigg| x^{(2)} \sim D_{s-1}(a^{(1)}_s; a_s) \) on \( \mathcal{V}_{s-1} \); (ii) \( \frac{x^{(2)}}{1-\Lambda_1} \bigg| x^{(1)} \sim D_{n-1-s}(a^{(2)}_n; a_n) \) on \( \mathcal{V}_{n-1-s} \).

Proof. For case (ii), we consider the transformation \( w^{(2)}_{-n} = (w_{s+1}, \ldots, w_{n-1})^\top = \frac{x^{(2)}}{1-\Lambda_1} \). Then, the Jacobian is \((1 - \Delta_1)^{n-1-s}\) and \( w^{(2)}_{-n} \in \mathcal{V}_{n-1-s} \). From (5.3), we have
\[
g(w^{(2)}_{-n}|x^{(1)}) = g(x^{(2)}|x^{(1)}) (1 - \Delta_1)^{n-1-s} = \prod_{i=s+1}^{n-1} w_i^{a_i-1} \frac{(1 - \sum_{i=s+1}^{n-1} w_i)^{a_n-1}}{B_{n-s}(a^{(2)}_n)},
\]
which completes the proof of case (ii). Case (i) can be proved similarly. \( \square \)

If we set \( w_n = 1 - \sum_{i=s+1}^{n-1} w_i \) in (5.4), then we have \( w_n = x_n/(1 - \Delta_1) \) and \( w^{(2)} = (w_{s+1}, \ldots, w_{n-1}, w_n)^\top = x^{(2)}/(1 - \Delta_1) \in \mathcal{T}_{n-s} \). From (5.4), we immediately have \( w^{(2)}|x^{(1)} \sim D_{n-s}(a^{(2)}_n) \) and the following result.

Proposition 10. Let \( x \sim GD_{n,2,s}(a, b) \) on \( \mathcal{T}_n \), then (i) \( \frac{x^{(1)}}{1-\Lambda_2} \bigg| x^{(2)} \sim D_s(a^{(1)}_1) \) on \( \mathcal{T}_s \); (ii) \( \frac{x^{(2)}}{1-\Lambda_1} \bigg| x^{(1)} \sim D_{n-s}(a^{(2)}_n) \) on \( \mathcal{T}_{n-s} \).

The above proposition tells that the conditional distribution of \( x^{(1)}|x^{(2)} \) is a DD with scale parameter \( 1 - \Delta_2 \), which is a constant when given \( x^{(2)} \). That is, \( x^{(1)}|x^{(2)} \overset{d}{=} (1 - \Delta_2) \cdot \xi^{(1)} \) with \( \xi^{(1)} \sim D_s(a^{(1)}_1) \) on \( \mathcal{T}_s \). Similarly, we have \( x^{(2)}|x^{(1)} \overset{d}{=} (1 - \Delta_1) \cdot \xi^{(2)} \) with \( \xi^{(2)} \sim D_{n-s}(a^{(2)}_n) \) on \( \mathcal{T}_{n-s} \).

6. Extension to multiple partitions

Motivated by (2.2), we extend the GDD with two partitions to the general GDD with multiple partitions and then develop the corresponding distribution properties.

6.1. Definition and properties

An \( n \)-variate \( x \in \mathcal{T}_n \) is said to follow a GDD with \( m \) partitions, if the density of \( x_{-n} \) is given by [26]
\[
GD_{n,m,s}(x_{-n}|a, b) = c_m^{-1} \cdot \left( \prod_{i=1}^{n} x_i^{a_i-1} \right) \cdot \prod_{j=1}^{m} \left( \sum_{k=s_j-1+1}^{s_j} x_k \right)^{b_j}, \quad x_{-n} \in \mathcal{V}_{n-1},
\]
where \( a = (a_1, \ldots, a_n)^\top, b = (b_1, \ldots, b_m)^\top \) are two non-negative parameter vectors, \( 0 \leq s_0 < 1 \leq s_1 < \cdots < s_m \equiv n, s = (s_1, \ldots, s_m)^\top \), and the normalizing constant is
\[
c_m = \left\{ \prod_{j=1}^{m} B_{s_j-s_{j-1}}(a_{s_{j-1}+1}, \ldots, a_{s_j}) \right\} \cdot B_m \left( \sum_{k=1}^{s_1} a_k + b_1, \ldots, \sum_{k=s_{m-1}+1}^{s_m} a_k + b_m \right).
\]
We will write \( x \sim GD_{n,m,s}(a, b) \) on \( \mathcal{T}_n \) or \( x_{-n} \sim GD_{n,m,s}(a, b) \) on \( \mathcal{V}_{n-1} \) accordingly. In particular, when \( m = 2 \), (6.1) reduces to the GDD (3.1).
Let \( t_j = s_j - s_{j-1}, \ j = 1, \ldots, m \). We partition \( x \) into \( m \) parts where \( x = (x^{(1)} \top, \ldots, x^{(m)} \top) \top \) with \( r_1, \ldots, r_m \) components, respectively, and correspondingly partition \( y = (y^{(1)} \top, \ldots, y^{(m)} \top) \top \) and \( a = (a^{(1)} \top, \ldots, a^{(m)} \top) \top \) in the same manner. Let \( \beta = (\beta_1, \ldots, \beta_m) \top \) with

\[
\beta_j = \|a^{(j)}\| + b_j = \sum_{k=s_{j-1}+1}^{s_j} a_k + b_j, \quad j = 1, \ldots, m. \tag{6.2}
\]

Hence, the normalizing constant \( c_m \) can be rewritten as

\[
c_m = \prod_{j=1}^{m} B_{t_j}(a^{(j)}) \cdot B_m(\beta). \tag{6.3}
\]

Similar to Propositions 1 and 3, we have the following propositions.

**Proposition 11.** An \( n \)-vector \( x = (x^{(1)} \top, \ldots, x^{(m)} \top) \top \sim GD_{n,m,s}(a, b) \) on \( \mathcal{T}_n \) iff

\[
(x^{(1)} \top, \ldots, x^{(m)} \top) \top \overset{d}{=} (R_1 \cdot y^{(1)} \top, \ldots, R_m \cdot y^{(m)} \top) \top, \tag{6.4}
\]

where (i) \( y^{(j)} \sim D_{t_j}(a^{(j)}) \) on \( \mathcal{T}_{t_j}, \ j = 1, \ldots, m; \) (ii) \( R = (R_1, \ldots, R_m) \top \sim D_m(\beta) \) on \( \mathcal{T}_m; \) and (iii) the \( m \) subvectors \( y^{(1)}, \ldots, y^{(m)} \) and \( R \) are mutually independent.

**Proposition 12.** Let \( x = (x^{(1)} \top, \ldots, x^{(m)} \top) \top \sim GD_{n,m,s}(a, b) \) on \( \mathcal{T}_n \), then \( R_j \overset{d}{=} \|x^{(j)}\|, y^{(j)} \overset{d}{=} x^{(j)}/\|x^{(j)}\|, j = 1, \ldots, m. \)

Let \( r_1, \ldots, r_n \geq 0 \) and \( r = (r_1, \ldots, r_n) \top = (r^{(1)} \top, \ldots, r^{(m)} \top) \top \) have the same partition as \( x = (x^{(1)} \top, \ldots, x^{(m)} \top) \top \sim GD_{n,m,s}(a, b) \). The SR (6.4) can be used to calculate the mixed moments of \( x \). The independence among \( y^{(1)}, \ldots, y^{(m)} \) and \( R \) implies

\[
E\left( \prod_{i=1}^{n} x_i^{r_i} \right) = \prod_{j=1}^{m} E\left( \prod_{i=s_{j-1}+1}^{s_j} y_i^{r_i} \right) \cdot E\left( \prod_{j=1}^{m} R_j^{\|r^{(j)}\|} \right) = \prod_{j=1}^{m} B_{t_j}(a^{(j)} + r^{(j)}) \cdot B_m(\beta_1 + \|r^{(1)}\|, \ldots, \beta_m + \|r^{(m)}\|) / B_m(\beta). \tag{6.5}
\]

Analogous to Proposition 5 we can find the mode, as follows.

**Proposition 13.** The mode of the GDD with \( m \) partitions, as defined by (6.1), is

\[
\hat{x}_i = \frac{a_i - 1}{\|a\| + \|b\| - n} \left( 1 + \frac{b_1}{\|a^{(1)}\| - t_1} \right), \quad 1 \leq i \leq s_1,
\]

\[
\hat{x}_i = \frac{a_i - 1}{\|a\| + \|b\| - n} \left( 1 + \frac{b_2}{\|a^{(2)}\| - t_2} \right), \quad s_1 + 1 \leq i \leq s_2.
\]

\[
\vdots
\]

\[
\hat{x}_i = \frac{a_i - 1}{\|a\| + \|b\| - n} \left( 1 + \frac{b_m}{\|a^{(m)}\| - t_m} \right), \quad s_m - 1 \leq i \leq s_m. \tag{6.6}
\]
6.2. Marginal distributions

Let \( \mathbf{x} \) follow the grouped Dirichlet (6.1). Proposition 14 below tells that the marginal distributions of the \( t_j \)-subvector \( \mathbf{x}^{(j)} \) and the \( s_r \)-subvector \( (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(r)\top})^\top \), \( 1 \leq r < m \), still belong to the family of the GDD.

**Proposition 14.** Let \( \mathbf{x} = (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(m)\top})^\top \sim GD_{n,m,s}(\mathbf{a}, \mathbf{b}) \) on \( \mathcal{T}_n \), then (i) for any fixed \( j \) (\( 1 \leq j \leq m \)), \( \mathbf{x}^{(j)} \) follows GDD (6.7) on \( \mathcal{V}_{ij} \) with two partitions and the SR is

\[
\mathbf{x}^{(j)} \sim GD_{t_j + 1.2,t_j} \left( \left( \frac{\mathbf{a}^{(j)}}{\parallel \mathbf{a} \parallel - \parallel \mathbf{a}^{(j)} \parallel} \right), \left( \frac{\mathbf{b} - b_j}{\parallel \mathbf{b} \parallel - b_j} \right) \right),
\]

and (ii) for any fixed \( r \) (\( 1 \leq r < m \)), \( (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(r)\top})^\top \) follows GDD (6.8) on \( \mathcal{V}_{sr} \) with \( r + 1 \) partitions and the SR is given by

\[
\begin{pmatrix}
\mathbf{x}^{(1)} \\
\vdots \\
\mathbf{x}^{(r)} \\
1 - \sum_{j=1}^{r} \parallel \mathbf{x}^{(j)} \parallel \\
\end{pmatrix} \overset{d}{=} 
\begin{pmatrix}
R_1 \cdot \mathbf{y}^{(1)} \\
\vdots \\
R_r \cdot \mathbf{y}^{(r)} \\
1 - \sum_{j=1}^{r} R_j \\
\end{pmatrix}
\sim GD_{s_r + 1,r+1,s_r} \left( \left( \sum_{i=s_r+1}^{n} a_i \right), \left( \sum_{j=r+1}^{m} b_j \right) \right),
\]

where \( s_r = (s_1, \ldots, s_r, s_r + 1)^\top \).

The proof is given in the Appendix. The following proposition describes the relationship between the marginal distributions of the GDD and the beta–Liouville distribution.

**Proposition 15.** Given that \( \mathbf{x} = (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(m)\top})^\top \sim GD_{n,m,s}(\mathbf{a}, \mathbf{b}) \) on \( \mathcal{T}_n \), we have \( \mathbf{x}^{(j)} \overset{d}{=} R_j \cdot \mathbf{y}^{(j)} \sim BL_{t_j}(\mathbf{a}^{(j)}; \beta_j, \parallel \mathbf{b} \parallel - \beta_j) \), \( j = 1, \ldots, m \).

6.3. Conditional distributions

Now we consider conditional pdfs of \( \mathbf{x}^{[1]}|\mathbf{x}^{[2]} \) and \( \mathbf{x}^{[2]}|\mathbf{x}^{[1]} \). Proposition 16 below shows that the conditional pdfs of \( \frac{x^{[1]}(1-\Delta^{[1]})}{1-\Delta^{[1]}} \) and \( \frac{x^{[2]}(1-\Delta^{[2]})}{1-\Delta^{[2]}} \) still belong to the family of the GDD.

**Proposition 16.** Let \( \mathbf{x} = (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(m)\top})^\top \sim GD_{n,m,s}(\mathbf{a}, \mathbf{b}) \) on \( \mathcal{T}_n \). Define

\[
\begin{align*}
\mathbf{x}^{[1]} &= (\mathbf{x}^{(1)\top}, \ldots, \mathbf{x}^{(r)\top})^\top, \quad \mathbf{x}^{[2]} = (\mathbf{x}^{(r+1)\top}, \ldots, \mathbf{x}^{(m)\top})^\top, \quad 1 \leq r < m, \\
\Delta^{[1]} &= \parallel \mathbf{x}^{[1]} \parallel = \sum_{j=1}^{r} \parallel \mathbf{x}^{(j)} \parallel, \quad \Delta^{[2]} = \parallel \mathbf{x}^{[2]} \parallel = \sum_{j=r+1}^{m} \parallel \mathbf{x}^{(j)} \parallel, \\
\mathbf{a}^{[1]} &= (\mathbf{a}^{(1)\top}, \ldots, \mathbf{a}^{(r)\top})^\top, \quad \mathbf{a}^{[2]} = (\mathbf{a}^{(r+1)\top}, \ldots, \mathbf{a}^{(m)\top})^\top, \\
\mathbf{b}^{[1]} &= (b_1, b_2, \ldots, b_r)^\top, \quad \mathbf{b}^{[2]} = (b_{r+1}, \ldots, b_m)^\top, \\
\mathbf{s}^{[1]} &= (s_1, \ldots, s_r)^\top, \quad \mathbf{s}^{[2]} = (s_{r+1}, \ldots, s_m)^\top.
\end{align*}
\]
We have the following conditional distributions:

(i) \( \frac{x_1^{[1]}}{1 - \Delta^{[2]}} | x_2^{[2]} \sim GD_{sr,r,s^{[1]}}(a^{[1]}, b^{[1]}) \) on \( T_{sr} \) with SR

\[
\frac{x_1^{[1]}}{1 - \Delta^{[2]}} | x_2^{[2]} \overset{d}{=} \left( \frac{R_1 \cdot y^{(1)\top}}{1 - \Delta^{[2]}}, \ldots, \frac{R_{r-1} \cdot y^{(r-1)\top}}{1 - \Delta^{[2]}}, \frac{(1 - \Delta^{[2]} - \sum_{j=1}^{r-1} R_j) \cdot y^{(r)\top}}{1 - \Delta^{[2]}} \right)^\top | x_2^{[2]},
\]

(ii) \( \frac{x_2^{[2]}}{1 - \Delta^{[1]}} | x_1^{[1]} \sim GD_{n-sr,m-r,s^{[2]}}(a^{[2]}, b^{[2]}) \) on \( T_{n-sr} \) with SR

\[
\frac{x_2^{[2]}}{1 - \Delta^{[1]}} | x_1^{[1]} \overset{d}{=} \left( \frac{R_{r+1} \cdot y^{(r+1)\top}}{1 - \Delta^{[1]}}, \ldots, \frac{R_{m-1} \cdot y^{(m-1)\top}}{1 - \Delta^{[1]}}, \frac{(1 - \Delta^{[1]} - \sum_{j=r+1}^{m-1} R_j) \cdot y^{(m)\top}}{1 - \Delta^{[1]}} \right)^\top | x_1^{[1]},
\]

The proof is given in the Appendix. From Proposition 16, we can see that the conditional distribution of \( x_1^{[1]} | x_2^{[2]} \) is a GDD with scale 1 - \( \Delta^{[2]} \), which is a constant when given \( x_2^{[2]} \). Namely, \( x_1^{[1]} | x_2^{[2]} \overset{d}{=} (1 - \Delta^{[2]}) \cdot \zeta^{[1]} \) where \( \zeta^{[1]} \sim GD_{sr,r,s^{[1]}}(a^{[1]}, b^{[1]}) \) on \( T_{sr} \). Similarly, \( x_2^{[2]} | x_1^{[1]} \overset{d}{=} (1 - \Delta^{[1]}) \cdot \zeta^{[2]} \), where \( \zeta^{[2]} \) is distributed as \( GD_{n-sr,m-r,s^{[2]}}(a^{[2]}, b^{[2]}) \) on \( T_{n-sr} \).

7. Applications

In this section, we illustrate the applications of the proposed distribution by the three data sets presented in Section 2. With a moment’s reflection of viewing the parameter as a variate, we see that the likelihood function of the first example has the pdf form of a GDD with two partitions, immediately obtaining the MLE and conjugate Bayesian solutions in closed-form. Analogously, the second example admits closed-form MLE and conjugate Bayesian solutions based on a GDD with three partitions. For the third example, our EM algorithm based on the GDD works on an augmented likelihood with fewer latent variables than the conventional EM algorithm that is based on the DD. Furthermore, this augmented likelihood admits a closed-form expression of MLE, and is thus more efficient in getting the iterative solution for the MLE of the target likelihood. Besides, the use of the GDD makes the conjugate Bayesian analysis more easily done.

7.1. Cervical cancer data

By comparing (2.1) with (3.1), it is easy to see that the MLE of \( \theta \) is exactly the mode of the grouped Dirichlet distribution \( GD_{4,2,2}(a, b) \) with \( a = (n_1 + 1, \ldots, n_4 + 1)^\top \) and \( b = (n_{12}, n_{34})^\top \). From (3.7), the MLEs of \( \theta \) are given by

\[
\hat{\theta}_i = \frac{n_i}{\sum_{i=1}^4 n_i} \left( \frac{n_1 + n_2 + n_{12}}{n_1 + n_2} \cdot I(1 \leq i \leq 2) + \frac{n_3 + n_4 + n_{34}}{n_3 + n_4} \cdot I(3 \leq i \leq 4) \right).
\]

Hence, the MLE of the odds ratio \( \psi \) is \( \hat{\psi} = \hat{\theta}_1 \hat{\theta}_4 / (\hat{\theta}_2 \hat{\theta}_3) \). Let \( \theta_{-4} = (\theta_1, \theta_2, \theta_3)^\top \). The asymptotic variance–covariance matrix of the MLE \( \hat{\theta}_{-4} \) is then given by \( I^{-1}(\theta_{-4} | Y_{\text{obs}}) \), where \( I(\theta_{-4} | Y_{\text{obs}}) = -\hat{\psi}^2 \log L(\theta | Y_{\text{obs}}) / \hat{\psi} \theta_{-4} \hat{\theta}_{-4}^\top \), with \( L(\theta | Y_{\text{obs}}) \) being given in (2.1). Hence, the delta method (e.g., [25], p. 34) can be used to approximate the standard error of \( \hat{\psi} \) and the 95% asymptotic confidence...
interval (CI) of $\psi$ can be constructed as $[\hat{\psi} - 1.96 \cdot \text{se} (\hat{\psi}), \hat{\psi} + 1.96 \cdot \text{se} (\hat{\psi})]$, where $\hat{\psi} = (i\hat{\psi}/i\theta_{-4})^{-1} I^{-1}(\theta_{-4}|Y_{\text{obs}})(i\hat{\psi}/i\theta_{-4})|_{\hat{\theta}}^{1/2}$. For the cervical cancer data, we have $\hat{\theta}_1 = 0.2460, \hat{\theta}_2 = 0.2460, \hat{\theta}_3 = 0.1615, \hat{\theta}_4 = 0.3465$ and $\hat{\psi} = 2.1456$. The corresponding standard errors are given by (0.0163, 0.0163, 0.0143, 0.0181) and 0.3488. Therefore, the 95% asymptotic CI of $\psi$ is [1.4620, 2.8293].

For Bayesian analysis, the GDD is a natural conjugate prior distribution. Multiplying the likelihood function (2.1) by the prior distribution

$$\theta \sim GD_{4,2,2}((n_1^*, \ldots, n_4^*)^\top, (n_{12}^*, n_{34}^*)^\top),$$

yields the grouped Dirichlet posterior distribution

$$\theta|Y_{\text{obs}} \sim GD_{4,2,2}((n_1 + n_1^*, \ldots, n_4 + n_4^*)^\top, (n_{12} + n_{12}^*, n_{34} + n_{34}^*)^\top).$$

The exact first- and second-order posterior moments of $\theta$ can be obtained explicitly by Proposition 4. The posterior means are similar to the MLEs. The marginal posterior density for each component of $\theta$ can be obtained from (4.7). Although we cannot obtain closed-form expressions of the first- and second-order posterior moments for the odds ratio $\psi$ (or an arbitrary function of $\theta$, say $h(\theta)$), an i.i.d. posterior sample of $\psi$ (or $h(\theta)$) can be obtained provided that an i.i.d. sample of $\theta$ is available. Fortunately, the i.i.d. posterior samples of $\theta$ in (7.3) can be drawn by using the SR (3.5).

In form, it seems that the prior distribution specified by (7.2) depends on the pattern of incomplete data. Noting that both uniform and DDs are special members of the GDD, we usually use the uniform prior: $n_1^* = \cdots = n_4^* = 1$ and $n_{12}^* = n_{34}^* = 0$. We generate 30,000 i.i.d. posterior samples from (7.3), and the Bayes estimates of $\theta$ and the odds ratio $\psi$ are given by (0.2460, 0.2459, 0.1619, 0.3459) and 2.1708. Therefore, the estimated odds of cervical cancer for patients with many sex partners is about $2.1708$ times the estimated odds for patients with few sex partners. The corresponding Bayes standard errors are (0.0163, 0.0163, 0.0142, 0.0179) and 0.3557. The 95% Bayes CI of $\psi$ is then [1.5671, 2.9527]. Since both the asymptotic and Bayes lower bounds are larger than the value of 1, there is stronger association between the number of sex partners and disease status of cervical cancer.

### 7.2. Leprosy survey data

The first objective is to obtain the MLE of $\theta$ in (2.2). By comparing (2.2) with (6.1), we know that the MLE of $\theta$ is exactly the mode of the grouped Dirichlet distribution $GD_{10,4,8}((n_1 + \tilde{n}_1 + 1, \ldots, n_{10} + \tilde{n}_{10} + 1)^\top, (n_{123}, 0, n_{678}, 0)^\top)$ with $s = (s_1, \ldots, s_4)^\top = (3, 5, 8, 10)^\top$. From (6.6), we obtain the explicit MLEs of $\theta$. For the leprosy survey data, the MLEs of $\theta$ and the corresponding standard errors are given in Table 4.

For Bayesian analysis, a GDD with four partitions is a natural conjugate prior for (2.2). We take $\theta|Y_{\text{obs}} \sim GD_{4,4,8}((n_1^*, \ldots, n_{10}^*)^\top, (n_{123}, n_{45}, n_{678}, n_{9_{10}}^*)^\top)$ as the prior. The posterior distribution is $\theta|Y_{\text{obs}} \sim GD_{4,4,8}(a^*, b^*)$, where $a^* = (n_1 + \tilde{n}_1 + n_1^*, \ldots, n_{10} + \tilde{n}_{10} + n_{10}^*)^\top$ and $b^* = (n_{123} + n_{123}^*, n_{45}, n_{678} + n_{678}^*, n_{9_{10}}^*)^\top$. We adopt uniform prior: $n_1^* = \cdots = n_{10}^* = 1$ and $n_{123}^* = n_{45}^* = n_{678}^* = n_{9_{10}}^* = 0$ and generate 30,000 i.i.d. samples of $\theta$ by using the SR (6.4). The Bayesian means and standard errors of $\theta$ are also reported in Table 4.
Table 4
Frequentist and Bayesian estimates of parameters for leprosy survey data

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Frequentist approach</th>
<th>Bayesian approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Standard error</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.0516</td>
<td>0.0147</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.1268</td>
<td>0.0209</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.1973</td>
<td>0.0234</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>0.2902</td>
<td>0.0185</td>
</tr>
<tr>
<td>$\theta_5$</td>
<td>0.0453</td>
<td>0.0085</td>
</tr>
<tr>
<td>$\theta_6$</td>
<td>0.0401</td>
<td>0.0140</td>
</tr>
<tr>
<td>$\theta_7$</td>
<td>0.0861</td>
<td>0.0185</td>
</tr>
<tr>
<td>$\theta_8$</td>
<td>0.0918</td>
<td>0.0188</td>
</tr>
<tr>
<td>$\theta_9$</td>
<td>0.0620</td>
<td>0.0098</td>
</tr>
<tr>
<td>$\theta_{10}$</td>
<td>0.0083</td>
<td>0.0037</td>
</tr>
</tbody>
</table>

7.3. Neurological complication data

In this example, we are unable to find the closed-form solutions and will use the EM and DA algorithms in dealing with the likelihood of $\theta$ as given in (2.3). Writing $n_{13} = Z_1 + (n_{13} - Z_1)$ and $n_{24} = Z_2 + (n_{24} - Z_2)$, we introduce a two-dimensional latent vector $Z = (Z_1, Z_2)^T$, and obtain the following augmented-likelihood function:

$$L(\theta|Y_{\text{obs}}, Z) \propto \left( \prod_{i=1}^{4} \theta_1^{n_{1i}+Z_i} \cdot (\theta_1 + \theta_2)^{n_{12}}(\theta_3 + \theta_4)^{n_{34}} \right), \quad \theta \in \mathcal{T}_4,$$

where $Z_3 \equiv n_{13} - Z_1$ and $Z_4 \equiv n_{24} - Z_2$. We see that this likelihood has the same functional form of (3.1) and thus from (3.7), we have the following MLEs for $\theta$ based on (7.6):

$$\begin{align*}
\theta_i &= \frac{n_i + Z_i}{\sum_{i=1}^{4} (n_i + Z_i) + n_{12} + n_{34}} \left( 1 + \frac{n_{12}}{n_1 + Z_1 + n_2 + Z_2} \right), \quad i = 1, 2, \\
\theta_i &= \frac{n_i + Z_i}{\sum_{i=1}^{4} (n_i + Z_i) + n_{12} + n_{34}} \left( 1 + \frac{n_{34}}{n_3 + Z_3 + n_4 + Z_4} \right), \quad i = 3, 4.
\end{align*}$$

(7.5)

The conditional predictive distributions are given by

$$Z_1|Y_{\text{obs}}, \theta) \sim \text{Binomial} \left( n_{13}, \frac{\theta_1}{\theta_1 + \theta_3} \right),$$

$$Z_2|Y_{\text{obs}}, \theta) \sim \text{Binomial} \left( n_{24}, \frac{\theta_2}{\theta_2 + \theta_4} \right).$$

(7.6)

Therefore, the E-step of the EM is to compute the conditional expectations

$$E(Z_1|Y_{\text{obs}}, \theta) = n_{13} \theta_1 / (\theta_1 + \theta_3), \quad E(Z_2|Y_{\text{obs}}, \theta) = n_{24} \theta_2 / (\theta_2 + \theta_4),$$

(7.7)

and the M-step updates (7.5) by replacing $\{Z_i\}_{i=1}^{4}$ with their conditional expectations. The MLE of $\theta_{23}$ is then given by $\hat{\theta}_{23} = \hat{\theta}_2 - \hat{\theta}_3$. The 95% asymptotic CI for $\theta_{23}$ can be built as $[\hat{\theta}_{23} - 1.96 \cdot$
Again, we adopt the uniform prior assuming no prior information, i.e., \( \text{se}(\hat{\theta}_{23}) = 1.96 \cdot \text{se}(\hat{\theta}_{23}) \), where \( \text{se}(\hat{\theta}_{23}) = \sqrt{\text{Var}(\hat{\theta}_2) + \text{Var}(\hat{\theta}_3) - 2\text{Cov}(\hat{\theta}_2, \hat{\theta}_3)} \). However, such a CI depends on the large-sample theory. For the neurological complication data, using \( \theta^{(0)} = (0.25, 0.25, 0.25, 0.25) \) as the initial values, the EM algorithm based on (7.5) and (7.7) converged in six iterations. The resultant MLEs are \( \hat{\theta}_1 = 0.2495, \hat{\theta}_2 = 0.1094, \hat{\theta}_3 = 0.3422, \hat{\theta}_4 = 0.2989 \), and \( \hat{\theta}_{23} = -0.2328 \). The corresponding standard errors are given by (0.0819, 0.0585, 0.0925, 0.0865) and 0.1191. Therefore, the 95% asymptotic CI of \( \theta_{23} \) is \([-0.4663, 0.0007]\). Since the asymptotic CI includes the value of 0, we can conclude that the incidence rates of neurological complication before and after the standard treatment are essentially the same.

For Bayesian analysis, we choose the same prior (7.2). Hence, the complete-data posterior is a GDD and we have

\[
\theta | (Y_{\text{obs}}, Z) \sim GD_{4,2,2}\left((n_1 + n_1^* + Z_1, \ldots, n_4 + n_4^* + Z_4), (n_{12} + n_{12}^*, n_{34} + n_{34}^*)\right).
\]

Again, we adopt the uniform prior assuming no prior information, i.e., \( n_1^* = \cdots = n_4^* = 0 \). Based on (7.6) and (7.8), we implement the DA algorithm to obtain posterior samples for \( \theta \). The criterion of potential scale reduction proposed by Gelman and Rubin [8] is utilized to check the convergence of the Markov chain. Potential scale reduction values close to 1 are indicative of convergence of the Markov chain to the target distribution. We run 10 multiple chains with length 3000 by using the dispersed initial values. The 10 \( \times \) 1500 samples from the second half of each sequence will be used to calculate the Bayes estimates of \( \theta \) and \( \theta_{23} \) which are given by (0.2232, 0.1281, 0.3571, 0.2915) and -0.2290. The corresponding Bayes standard errors are (0.0691, 0.0576, 0.0829, 0.0786) and 0.1129. The 95% Bayes CI of \( \theta_{23} \) is \([-0.4442, 0.0047]\), which includes the value of 0. This lends support to the belief that the incidence rates of neurological complication are essentially the same before and after the standard treatment.

The conventional DA scheme, we need to introduce four latent variables \( Z_1, Z_2, W_1 \) and \( W_3 \) such that the augmented-likelihood function

\[
L(\theta | Y_{\text{obs}}, Z_1, Z_2, W_1, W_3) \propto \prod_{i=1}^{4} \theta_i^{n_i+Z_i+W_i}, \quad \theta \in \mathcal{T}_4
\]

has the pdf form of a DD, where \( W_2 \sim \tilde{n}_{12} - W_1 \) and \( W_4 \sim \tilde{n}_{34} - W_3 \). It is known that the fewer the latent variables, the faster the EM [15,16]. Thus, the resulting EM/DA algorithm converges slowly because of the introduction of two unnecessary latent variables \( W_1 \) and \( W_3 \). For the more general \( r \times c \) table with two supplemental margins, Tang et al. [24] theoretically proved that the convergence speed of the new EM based on the GDD with only \( r(c-1) \) latent variables is faster than that of the conventional EM based on the DD with a total of \( 2rc - r - c \) latent variables. Their simulation studies further supported this conclusion.

8. Discussion

We extended the classical DD to a new family of GDD and developed some useful distribution theory. Using the three motivating real examples, we demonstrated a new approach for incomplete categorical data analysis. As shown by the examples considered in Section 7, the new approach has two advantages over the commonly-used approach, which is based on the augmented likelihood in the Dirichlet form, in both frequentist and conjugate Bayesian inference: (a) in cases like Examples 1 and 2, the MLE and the Bayes estimate possess closed-form expressions when the
new approach is adopted; and (b) in other cases like Example 3 where no closed-form solution is available, the EM and DA algorithms based on the GDD need fewer latent variables to work and converge much faster.

In the context of incomplete categorical data analysis, the basic idea behind the GDD and the resulting conclusions are closely related to factorization of likelihood function [20], the partial imputation EM (PIEM) and the collapsed Gibbs sampler [17]. For example, for graphical models with general missing pattern, Geng et al. [10] theoretically proved that the convergence speed of the PIEM is faster than the traditional EM algorithm that requires imputing a great amount of missing data. Furthermore, under the Bayesian framework, Geng and Li [9] discussed the factorization of a posterior distribution and presented a partial imputation Gibbs sampler. However, in practice, these authors mainly used the monotone missing pattern, while in this paper, we discussed a grouped missing pattern.

Acknowledgments

The authors thank the Associate Editor and two referees for their helpful comments and suggestions. The research of M.L. Tang was fully supported by a grant (CUHK4371/04M) from the Research Grant Council of the Hong Kong Special Administrative Region. The research of M. Tan and G.L. Tian was supported in part by U.S. National Cancer Institute Grants CA106767 and CA119758.

Appendix

Proof of Proposition 1. If \( x \sim GD_{n,2,s}(a, b) \), then the pdf of \( x_{-n} \) is given by (3.1). Noting that
\[
\mathbf{x}^{(1)} = \| \mathbf{x}^{(1)} \| \cdot (\mathbf{x}^{(1)}/\| \mathbf{x}^{(1)} \|), \quad \mathbf{x}^{(2)} = \| \mathbf{x}^{(2)} \| \cdot (\mathbf{x}^{(2)}/\| \mathbf{x}^{(2)} \|),
\]
and \( \| \mathbf{x} \| = 1 \), we make the transformation
\[
y^{(1)} = \mathbf{x}^{(1)}/R, \quad y^{(2)} = \mathbf{x}^{(2)}/(1 - R), \quad R = \| \mathbf{x}^{(1)} \|.
\]
The Jacobian is
\[
\left| \frac{\partial \mathbf{x}_{-n}}{\partial \mathbf{y}} \right| = \frac{R^{s-1} (1 - R)^{n-s-1}}{\prod_{i=1}^{s} x_{i}^{a_{i} - 1} \cdot \prod_{i=s+1}^{n} x_{i}^{a_{i} - 1} \cdot R^{x_{1}-1} (1 - R)^{x_{2}-1}}.
\]
which has been factorized into independent Dirichlet and beta distributions for \( y^{(1)}, y^{(2)} \) and \( R \). Combining (A.1) and (A.2), we immediately obtain (3.5). Conversely, suppose that (3.5) holds. Therefore, the joint density of \( y_{1}, \ldots, y_{s-1}, y_{s+1}, \ldots, y_{n-1} \) and \( R \) is given by (A.3). It is easy to show that the density of \( x_{-n} \) is given by (3.1). \( \square \)

Proof of proposition 7. To derive (4.4), we first introduce a special case of the famous integral formula of Joseph Liouville (see, e.g., [7], p. 21): let \( h(\cdot) \) be a real function defined on the interval \([0, d]\). Then
\[
\int_{\sum_{i=1}^{n} x_{i}^{d} \prod_{i=1}^{n} x_{i}^{a_{i} - 1} d x_{i}} \mathbf{B}_{n}(a_{1}, \ldots, a_{n}) \cdot \int_{0}^{d} h(y) y^{\sum_{i=1}^{n} a_{i} - 1} dy.
\]
From (4.2), the marginal density of $x_s^{(1)}$ is given by

$$g(x_s^{(1)}) = \frac{\prod_{i=1}^{m} x_i^{a_i-1}}{B_s(a^{(1)})B(x_1, x_2)} \int h \left( \sum_{i=m+1}^{s} x_i \right) \left( \prod_{i=m+1}^{s} x_i^{a_i-1} \right) dx_i,$$

where $h(y) = (\delta_1 + y)^{\beta_1}(1 - \delta_1 - y)^{\beta_2-1}$, $y \in [0, 1 - \delta_1]$. Applying (A.4), we immediately obtain (4.4) with $f_s^{(1)}$ given by (4.5). Comparing (4.1) with (4.4), we know that (4.4) is a Liouville distribution with $f_s^{(1)}$ being the generating density.

Alternatively, from the SR point of view, we can arrive at the same conclusion. In fact, from (3.5) we have $x^{(1)} \overset{d}{=} R \cdot y^{(1)}$. Hence, $x_s^{(1)} \overset{d}{=} R_s \cdot y_s^{(1)} = (R \cdot \|y_s^{(1)}\|) \cdot (y_s^{(1)}/\|y_s^{(1)}\|)$, which implies (4.6). Since $y^{(1)} = (y_1, \ldots, y_m)^T \sim D_s(a^{(1)})$ on $T_s$, from the properties of the Dirichlet distribution. We know that $y_s^{(1)} = (y_1, \ldots, y_m)^T \sim D_m(a_1, \ldots, a_m; \sum_{i=m+1}^{s} a_i = D_m(a_s^{(1)}; \tau_2)$ on $V_m$. Making the transformations $\|y_s^{(1)}\| = \sum_{j=1}^{m} y_j$ and $w_i = y_i / \sum_{j=1}^{m} y_j$ for $i = 1, \ldots, m-1$, we obtain the Jacobian as $\|y_s^{(1)}\|^{-m}$. Note that the joint density of $\|y_s^{(1)}\|$ and $(w_1, \ldots, w_{m-1})$ can be factorized as a product of a beta density and a Dirichlet density. We therefore have $\|y_s^{(1)}\| \sim \text{Beta}(\tau_1, \tau_2)$ and $y_s^{(1)}/\|y_s^{(1)}\| \sim D_m(a_s^{(1)})$. In addition, it is easy to verify that $R_s^{(1)} \overset{d}{=} R \cdot \|y_s^{(1)}\|$ has the density $f_s^{(1)}$. We complete the proof for case (i). Similarly, we can prove case (ii). □

**Proof of Proposition 14.** Proposition 11 implies that $x^{(j)} \overset{d}{=} R_j \cdot y^{(j)}$, where $y^{(j)} \sim D_{t_j}(a^{(j)})$ and $R_j \sim \text{Beta}(\beta_j, \|\beta\| - \beta_j)$. From (6.2), we can rewrite $R_j \sim \text{Beta}(\|a^{(j)}\| + b_j, \|a\| - \|a^{(j)}\| + \|b\| - b_j)$. By using Proposition 2, we obtain (6.7). In addition, we note that the SR in (6.8) is an immediate result of (6.4), and

$$\left( R_1, \ldots, R_r, 1 - \sum_{j=1}^{r} R_j \right)^T \sim D_{r+1} \left( \beta_1, \ldots, \beta_r, \sum_{j=r+1}^{m} \beta_j \right)^T = D_{r+1} \left( \beta_1, \ldots, \beta_r, \sum_{i=s_r+1}^{n} a_i + \sum_{j=r+1}^{m} b_j \right)^T.$$  

Note that $1 - \sum_{j=1}^{r} R_j$ in (6.8) can be written as $(1 - \sum_{j=1}^{r} R_j) \cdot 1$, where $1 \sim D_1(\sum_{i=s_r+1}^{n} a_i)$ is the degenerate DD. By comparing (6.4) with (6.8), we obtain the last part of (6.8). □

**Proof of Proposition 16.** Here we only need to show the case (i). From (6.4), since $1 - \Delta^{[2]}$ is a constant when $x^{[2]}$ is given, we have

$$\frac{x^{[1]}}{1 - \Delta^{[2]}} \bigg| x^{[2]} \overset{d}{=} \left( \frac{R_1 \cdot y^{(1)}^T}{1 - \Delta^{[2]}}, \ldots, \frac{R_r \cdot y^{(r)}^T}{1 - \Delta^{[2]}} \right)^T x^{[2]},$$

(A.5)

where $y^{(j)} \sim D_{t_j}(a^{(j)})$ on $T_{t_j}$, $j = 1, \ldots, r, (R_1, \ldots, R_m)^T \sim D_m(\beta)$ on $T_m$. Since $1 - \Delta^{[2]} = 1 - \sum_{j=r+1}^{m} \|x^{(j)}\| \overset{d}{=} 1 - \sum_{j=r+1}^{m} R_j$, we have $R_r = 1 - \sum_{j=1}^{r-1} R_j - \sum_{j=r+1}^{m} R_j \overset{d}{=} 1 - \Delta^{[2]} - \sum_{j=1}^{r-1} R_j$. Hence, from (A.5) we obtain (6.9). When $x^{[2]}$ is given, from Theorem 1.6 of
Fang et al. [7, p. 21], we obtain

\[
\left(\frac{R_1}{1 - \Delta[2]}, \ldots, \frac{R_r}{1 - \Delta[2]}\right) \sim D_r(\beta_1, \ldots, \beta_r) \quad \text{on } T_r.
\]

Combining (A.5) with (A.6), we have

\[
\frac{x^{[1]}}{1 - \Delta[2]} \mid x^{[2]} \sim GD_{sr,r,s}^{[1]}(a^{[1]}, b^{[1]}) \quad \text{on } T_{sr}.
\]

\[\square\]

References