



Title	Further properties and new applications of the nested Dirichlet distribution
Author(s)	Tian, GL; Tang, ML; Yuen, KC; Ng, KW
Citation	Computational Statistics And Data Analysis, 2010, v. 54 n. 2, p. 394-405
Issued Date	2010
URL	http://hdl.handle.net/10722/65606
Rights	Computational Statistics & Data Analysis. Copyright © Elsevier BV.

Further Properties and New Applications of the Nested Dirichlet Distribution

Guo-Liang Tian^{†,*}, Man-Lai Tang[‡], Kam Chuen Yuen[†], Kai Wang Ng[†]

[†]Department of Statistics and Actuarial Science,
The University of Hong Kong,
Pokfulam Road, Hong Kong, P. R. China

*Corresponding author's email: gltian@hku.hk

[‡]Department of Mathematics,
Hong Kong Baptist University,
Kowloon Tong, Hong Kong, P. R. China

Abstract: Recently, Ng *et al.* (2009) studied a new family of distributions, namely the nested Dirichlet distributions. This family includes the traditional Dirichlet distribution as a special member and can be adopted to analyze incomplete categorical data. However, other important aspects of the family, such as marginal and conditional distributions and related properties are not yet available in the literature. Moreover, diverse applications of the family to the real world need to be further explored. In this paper, we first obtain the marginal and conditional distributions and other related properties of the nested Dirichlet Distribution. We then present new applications of the family in fitting competing-risks model, analyzing incomplete categorical data and evaluating cancer diagnosis tests. Three real data involving failure times of radio transmitter receivers, attitude toward the death penalty and ultrasound ratings for breast cancer metastasis are provided.

Key words: Dirichlet distribution; EM algorithm; Missing at random; Nested Dirichlet distribution; Stochastic representation.

1. Introduction

The Dirichlet distribution plays an important role in multivariate statistics, distribution theory and Bayesian inference for categorical data and compositional data (cf. Fang *et al.*, 1990; Gupta and Richards, 2001; Balakrishnan and Nevzorov, 2003). Generalizations of Dirichlet distribution for various purposes have been studied in the literature. Amongst them, Liouville distribution is perhaps the most famous. For instance, Marshall and Olkin (1979, Chapter 11) described the family of Liouville distributions. Sivazlian (1981a, 1981b) presented some results on marginal distributions and transformation properties for the class of Liouville distributions. Gupta and Richards (1987, 1990, 1991, 1992) used the Weyl fractional integral and Deny's theorem in measure theory on locally compact groups to derive some important results for the multivariate Liouville distributions. The integral related to the generalized Liouville distribution was presented as early as in Edwards (1922, p.160–162) but did not appear in the literature until Marshall and Olkin (1979). Fang, Kotz and Ng (1990, Chapter 6) provided an extensive study of the Liouville family, including discussions of its applications to compositional data modeling, nonparametric prediction of lifetimes, survival functions and multiple Type I error of ANOVA. Gupta and Richards (2001) provided a comprehensive review of the theory and applications of the Dirichlet and Liouville distributions. Thomas and George (2004) presented a review of the development of the Dirichlet distribution, some applications and its generalizations in various directions.

The nested Dirichlet distribution, defined on the closed n -dimensional simplex, includes the Dirichlet distribution as a special case. To our knowledge, the distribution was first briefly introduced by Tian, Ng and Geng (2003), in which only one stochastic representation was presented. Other basic properties such as mixed (or raw) moments and mode are obtained by Ng *et al.* (2009). Based on the distribution, they also examined large-sample likelihood inferences and small-sample Bayesian inferences for incomplete

categorical data. However, certain important components of distribution theory, including the marginal and conditional distributions for the family are not yet available in the literature. This state of development of the nested Dirichlet distribution is not helped to further applications to the real world.

This article is organized as follows. Section 2 reviews some properties of the nested Dirichlet distribution, including stochastic representation and mode. In Sections 3 and 4, we respectively derive the marginal and conditional distributions of the nested Dirichlet distribution and other related properties. In Section 5, applications of the distribution in the competing–risks model, sample survey and cancer diagnosis tests are introduced. Three real data involving failure times of radio transmitter receivers, attitude toward the death penalty and ultrasound ratings for breast cancer metastasis are presented. We conclude with a discussion in Section 6.

2. The nested Dirichlet distribution

Let \mathbb{T}_n denote the closed simplex $\{(x_1, \dots, x_n)^\top : x_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n x_i = 1\}$. A random vector $\mathbf{x} \in \mathbb{T}_n$ is said to follow a nested Dirichlet distribution, if the density of $\mathbf{x}_{-n} \stackrel{\wedge}{=} (x_1, \dots, x_{n-1})^\top$ is given by (Tian *et al.*, 2003)

$$\text{ND}_{n,n-1}(\mathbf{x}_{-n}|\mathbf{a}, \mathbf{b}) = c^{-1} \left(\prod_{i=1}^n x_i^{a_i-1} \right) \prod_{j=1}^{n-1} \left(\sum_{k=1}^j x_k \right)^{b_j}, \quad \mathbf{x}_{-n} \in \mathbb{V}_{n-1}, \quad (2.1)$$

where $\mathbf{a} = (a_1, \dots, a_n)^\top$ is a positive parameter vector, $\mathbf{b} = (b_1, \dots, b_{n-1})^\top$ non-negative parameter vector, $\mathbb{V}_{n-1} \stackrel{\wedge}{=} \{\mathbf{x}_{-n} : x_i \geq 0, i = 1, \dots, n-1, \sum_{i=1}^{n-1} x_i \leq 1\}$ denotes the open simplex, and c is the normalizing constant defined by

$$c = \prod_{j=1}^{n-1} B(d_j, a_{j+1}) \quad \text{with} \quad d_j \stackrel{\wedge}{=} \sum_{k=1}^j (a_k + b_k), \quad (2.2)$$

$B(\alpha, \beta) \stackrel{\wedge}{=} \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the beta function. We will write $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n or $\mathbf{x}_{-n} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{V}_{n-1} , where the first subscript n in the notation $\text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$

represents the effective dimension of the parameter vector \mathbf{a} and the second subscript $n-1$ denotes that of the parameter vector \mathbf{b} .

In particular, when $\mathbf{b} = \mathbf{0}_{n-1}$ the distribution in (2.1) reduces to the traditional Dirichlet distribution $D_n(\mathbf{a})$. On the other hand, when $\mathbf{b} = (0, \dots, 0, b_{n-1})^\top$ with $b_{n-1} > 0$, it reduces to a beta Liouville distribution (e.g., see the formula (6.13) of Fang *et al.*, 1990). Figure 1 shows some plots of Dirichlet and nested Dirichlet densities for $n = 3$ with various combinations of \mathbf{a} and \mathbf{b} . We observe that the latter provides more varieties than the traditional Dirichlet distribution in terms of skewness.

[Insert Figure 1 here]

The following result suggests that the nested Dirichlet distribution can be stochastically represented by, and thus generated via, a sequence of mutually independent beta random variables (Ng *et al.*, 2009).

PROPOSITION 1. A random vector $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n if and only if

$$\begin{cases} x_i \stackrel{d}{=} (1 - y_{i-1}) \prod_{j=i}^{n-1} y_j, & y_0 \equiv 0, \quad i = 1, \dots, n-1, \\ x_n \stackrel{d}{=} 1 - y_{n-1}, \end{cases} \quad (2.3)$$

where

$$y_j \sim \text{Beta}(d_j, a_{j+1}), \quad j = 1, \dots, n-1, \quad d_j = \sum_{k=1}^j (a_k + b_k). \quad (2.4)$$

and y_1, \dots, y_{n-1} are mutually independent. ¶

Remark 1. (i) Proposition 1 provides a stochastic representation of the nested Dirichlet distribution and hence give a simple procedure for generating independently and identically distributed samples from nested Dirichlet distribution, which in turn plays a crucial role in Bayesian analysis for incomplete categorical data. (ii) When all $\mathbf{b} = \mathbf{0}_{n-1}$ (hence, d_j in (2.4) is equal to $\sum_{k=1}^j a_k$ according to the definition (2.2) of d_j), the stochastic representation (2.3) reduces to the stochastic representation of Dirichlet distribution, which

was first utilized by Gupta and Richards (1987) within the context of the Dirichlet and Liouville distributions. ¶

Ng *et al.* (2009) also obtain a closed-form expression for the mode of a nested Dirichlet density, implying that explicit *maximum likelihood estimates* (MLEs) of cell probabilities are available in the frequentist analysis of incomplete categorical data.

PROPOSITION 2. Let $a_i \geq 1$ for all $i \in \{1, \dots, n\}$ and $b_j \geq 0$ for all $j \in \{1, \dots, n-1\}$ but there exists at least one j such that $b_j > 0$. Then the mode of the nested Dirichlet density (2.1) is given by

$$\begin{cases} \hat{x}_n &= \frac{a_n - 1}{d_{n-1} + a_n - n}, \\ \hat{x}_i &= \frac{(a_i - 1)(1 - \hat{x}_{i+1} - \hat{x}_{i+2} - \dots - \hat{x}_n)}{d_{i-1} + a_i - i}, \quad i = 2, \dots, n-1, \\ \hat{x}_1 &= 1 - \hat{x}_2 - \dots - \hat{x}_n. \end{cases} \quad (2.5)$$

where $\{d_j\}$ are defined in (2.2). ¶

3. Marginal distributions

Let $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n and $1 \leq m < n$. We partition $\mathbf{x}_{n \times 1}$ into two components $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top$, each with m and $n - m$ elements, respectively. Similarly, we partition $\mathbf{z}_{n \times 1} = (\mathbf{z}^{(1)\top}, \mathbf{z}^{(2)\top})^\top$ and $\mathbf{a}_{n \times 1} = (\mathbf{a}^{(1)\top}, \mathbf{a}^{(2)\top})^\top$ in the same fashion as \mathbf{x} . Furthermore, we let $\mathbf{b}_{(n-1) \times 1} = (\mathbf{b}^{(1)\top}, b_m, \mathbf{b}^{(2)\top})^\top$, where $\mathbf{b}^{(1)} : (m-1) \times 1$ and $\mathbf{b}^{(2)} : (n-m-1) \times 1$. Finally, we define $\|\mathbf{x}\| = \mathbf{1}^\top \mathbf{x} = \sum_{i=1}^n x_i$. We have the following results.

PROPOSITION 3. Let $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n . We have the following stochastic representation:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} R \mathbf{z}^{(1)} \\ (1-R) \mathbf{z}^{(2)} \end{pmatrix}, \quad (3.1)$$

where

- (i) $\mathbf{z}^{(1)} \sim \text{ND}_{m,m-1}(\mathbf{a}^{(1)}, \mathbf{b}^{(1)})$ on \mathbb{T}_m ;
- (ii) $\mathbf{z}^{(1)}$ is independent of $(R, \mathbf{z}^{(2)})$;
- (iii) $R \stackrel{d}{=} \prod_{j=m}^{n-1} y_j$ with $\{y_j\}$ being defined in (2.4); and
- (iv) $\mathbf{z}^{(2)}$ follows a mixture of nested Dirichlet distributions with density

$$f(\mathbf{z}^{(2)}) = \sum_{k_{m+1}=0}^{b_{m+1}} \cdots \sum_{k_{n-1}=0}^{b_{n-1}} \omega(\mathbf{k}^{(2)}) \times \text{ND}_{n-m,n-m-1}(\mathbf{z}^{(2)} | \mathbf{a}^{(2)}, \mathbf{b}^{(2)} - \mathbf{k}^{(2)}), \quad (3.2)$$

$\mathbf{z}^{(2)} \in \mathbb{T}_{n-m}$, where $\mathbf{k}^{(2)} \stackrel{\wedge}{=} (k_{m+1}, \dots, k_{n-1})^\top$, the weights are given by

$$\begin{aligned} \omega(\mathbf{k}^{(2)}) &= \left[\prod_{j=m+1}^{n-1} \binom{b_j}{k_j} \right] \frac{B(d_m + \|\mathbf{k}^{(2)}\|, \|\mathbf{a}^{(2)}\| + \|\mathbf{b}^{(2)}\| - \|\mathbf{k}^{(2)}\|)}{B(d_m, a_{m+1})} \\ &\times \prod_{j=m+1}^{n-1} \frac{B(\sum_{\ell=m+1}^j (a_\ell + b_\ell - k_\ell), a_{j+1})}{B(d_j, a_{j+1})}, \end{aligned} \quad (3.3)$$

and $\{d_j\}$ are defined in (2.2). ¶

PROOF. If $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n , then the density of \mathbf{x}_{-n} is given by (2.1). Noting that

$$\mathbf{x}^{(1)} = \|\mathbf{x}^{(1)}\| \frac{\mathbf{x}^{(1)}}{\|\mathbf{x}^{(1)}\|}, \quad \mathbf{x}^{(2)} = \|\mathbf{x}^{(2)}\| \frac{\mathbf{x}^{(2)}}{\|\mathbf{x}^{(2)}\|}, \quad (3.4)$$

and $\|\mathbf{x}\| = 1$, we make the transformation

$$\mathbf{z}^{(1)} = \mathbf{x}^{(1)}/R, \quad R = \|\mathbf{x}^{(1)}\|, \quad \mathbf{z}^{(2)} = \mathbf{x}^{(2)}/(1 - R). \quad (3.5)$$

The Jacobian determinant is then given by

$$J(\mathbf{x}_{-n} \rightarrow z_1, \dots, z_{m-1}, R, z_{m+1}, \dots, z_{n-1}) = R^{m-1}(1 - R)^{n-m-1}.$$

Hence, the joint density of $(z_1, \dots, z_{m-1}, R, z_{m+1}, \dots, z_{n-1})$ can be written as

$$\frac{(\prod_{i=1}^m z_i^{a_i-1}) \prod_{j=1}^{m-1} (\sum_{k=1}^j z_k)^{b_j}}{\prod_{j=1}^{m-1} B(d_j, a_{j+1})} f(R, \mathbf{z}^{(2)}), \quad (3.6)$$

where $\mathbf{z}^{(1)} \in \mathbb{T}_m$, $R \in [0, 1]$, $\mathbf{z}^{(2)} \in \mathbb{T}_{n-m}$, and

$$f(R, \mathbf{z}^{(2)}) = \frac{R^{d_m-1} (1-R)^{\|\mathbf{a}^{(2)}\|-1}}{\prod_{j=m}^{n-1} B(d_j, a_{j+1})} \times \left(\prod_{i=m+1}^n z_i^{a_i-1} \right) \prod_{j=m+1}^{n-1} \left[R + (1-R) \sum_{k=m+1}^j z_k \right]^{b_j}. \quad (3.7)$$

Therefore, (3.3) follows from (3.4) and (3.5).

By noting that (3.6) can be factorized into two independent parts, we have (i) and (ii). Combining (3.5) with (2.3), we obtain

$$R = x_1 + \cdots + x_m \stackrel{d}{=} \prod_{j=m}^{n-1} y_j,$$

which implies (iii). To derive the marginal density of $\mathbf{z}^{(2)}$, we assume all b_j ($j = m+1, \dots, n-1$) are positive integers. By integrating out the density of R from (3.7) and using the Taylor expansion, we obtain (3.2) and (3.3) easily. \parallel

Remark 2. From (3.3), we can see that

$$\mathbf{x}^{(1)} \stackrel{d}{=} R \mathbf{z}^{(1)}$$

is a mixture of nested Dirichlet distributions. In addition,

$$\mathbf{x}^{(2)} \stackrel{d}{=} (1-R) \mathbf{z}^{(2)}$$

follows a double-mixture distribution of nested Dirichlet distributions in the sense that $\mathbf{z}^{(2)}$ itself is a mixture of nested Dirichlet distributions. This is not surprising because of the asymmetry of x_i in (2.1). \blacktriangleright

From (3.7), setting $b_{m+1} = \cdots = b_{n-1} = 0$ immediately yields the following result.

PROPOSITION 4. If $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n with $\mathbf{b}^{(2)} = \mathbf{0}$, the stochastic representation in (3.3) still holds, and

- (i) $\mathbf{z}^{(1)} \sim \text{ND}_{m,m-1}(\mathbf{a}^{(1)}, \mathbf{b}^{(1)})$ on \mathbb{T}_m , $\mathbf{z}^{(2)} \sim D_{n-m}(\mathbf{a}^{(2)})$ on \mathbb{T}_{n-m} ;
- (ii) $R \sim \text{Beta}(d_m, \|\mathbf{a}^{(2)}\|)$; and
- (iii) $\mathbf{z}^{(1)}$, $\mathbf{z}^{(2)}$ and R are mutually independent. ¶

Remark 3. According to Fang, Kotz and Ng (1990, p.147), Proposition 4 implies that $\mathbf{x}^{(2)}$ follows a beta-Liouville distribution and we write

$$\mathbf{x}^{(2)} \sim \text{BL}_{n-m}(\mathbf{a}^{(2)}; \|\mathbf{a}^{(2)}\|, d_m). \quad \text{¶}$$

By letting $m = n - 1$ in Proposition 3, we have the following result.

PROPOSITION 5. Given that $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n , we have the following stochastic representation

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{-n} \\ 1 - \|\mathbf{x}_{-n}\| \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} R \mathbf{z}_{-n} \\ 1 - R \end{pmatrix}$$

where

- (i) $\mathbf{z}_{-n} \sim \text{ND}_{n-1,n-2}(\mathbf{a}_{-n}, \mathbf{b}_{-(n-1)})$ on \mathbb{T}_{n-1} ;
- (ii) $R \sim \text{Beta}(d_{n-1}, a_n)$; and
- (iii) \mathbf{z}_{-n} and R are mutually independent. ¶

4. Conditional distributions

Now, we consider the conditional distributions of $\mathbf{x}^{(1)}|\mathbf{x}^{(2)}$ and $\mathbf{x}^{(2)}|\mathbf{x}^{(1)}$. For the sake of convenience, we denote $\|\mathbf{x}^{(1)}\|$ and $\|\mathbf{x}^{(2)}\|$ by Δ_1 and Δ_2 , respectively.

PROPOSITION 6. Given that $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n , we have

- (i) $\frac{\mathbf{x}^{(1)}}{1-\Delta_2}|\mathbf{x}^{(2)} \sim \text{ND}_{m,m-1}(\mathbf{a}^{(1)}, \mathbf{b}^{(1)})$ on \mathbb{T}_m ; and
- (ii) $\frac{\mathbf{x}^{(2)}}{1-\Delta_1}|\mathbf{x}^{(1)} \sim g(\cdot|\mathbf{x}^{(1)})$ on \mathbb{T}_{n-m} , where

$$g(\mathbf{z}^{(2)}|\mathbf{x}^{(1)}) \propto \left(\prod_{i=m+1}^n z_i^{a_i-1} \right) \prod_{j=m+1}^{n-1} \left[\Delta_1 + (1 - \Delta_1) \sum_{k=m+1}^j z_k \right]^{b_j}. \quad \text{¶} \quad (4.1)$$

PROOF. From (3.3), we have

$$\mathbf{x}^{(1)} \stackrel{d}{=} R \mathbf{z}^{(1)} \stackrel{d}{=} \|\mathbf{x}^{(1)}\| \mathbf{z}^{(1)} = (1 - \Delta_2) \mathbf{z}^{(1)}.$$

Result (ii) in Proposition 3 implies that

(a) $\mathbf{z}^{(1)}$ and Δ_2 are mutually independent; and

(b) $\mathbf{z}^{(1)}$ is independent of $\mathbf{x}^{(2)}$.

Therefore,

$$\frac{\mathbf{x}^{(1)}}{1 - \Delta_2} \stackrel{d}{=} \mathbf{z}^{(1)}$$

and

$$\frac{\mathbf{x}^{(1)}}{1 - \Delta_2} \Big|_{\mathbf{x}^{(2)} \stackrel{d}{=} \mathbf{z}^{(1)}} \Big|_{\mathbf{x}^{(2)} \stackrel{d}{=} \mathbf{z}^{(1)}} \sim \text{ND}_{m,m-1}(\mathbf{a}^{(1)}, \mathbf{b}^{(1)}) \text{ on } \mathbb{T}_m.$$

Similarly, we have

$$\frac{\mathbf{x}^{(2)}}{1 - \Delta_1} \Big|_{\mathbf{x}^{(1)} \stackrel{d}{=} \mathbf{z}^{(2)}} \Big|_{\mathbf{x}^{(1)} \stackrel{d}{=} \mathbf{z}^{(2)}} \Big|_{\Delta_1 \stackrel{d}{=} \mathbf{z}^{(2)}} \Big|_R.$$

From (3.7), we immediately obtain (4.1). ||

By combining Proposition 4 with Proposition 6, we have the following result.

COROLLARY 7. Given that $\mathbf{x} \sim \text{ND}_{n,n-1}(\mathbf{a}, \mathbf{b})$ on \mathbb{T}_n and $\mathbf{b}^{(2)} = \mathbf{0}$, we have

$$\frac{\mathbf{x}^{(1)}}{1 - \Delta_2} \Big|_{\mathbf{x}^{(2)}} \sim \text{ND}_{m,m-1}(\mathbf{a}^{(1)}, \mathbf{b}^{(1)}) \text{ on } \mathbb{T}_m$$

and

$$\frac{\mathbf{x}^{(2)}}{1 - \Delta_1} \Big|_{\mathbf{x}^{(1)}} \sim D_{n-m}(\mathbf{a}^{(2)}) \text{ on } \mathbb{T}_{n-m}. \quad \blacktriangleleft$$

Remark 4. Result (i) in Proposition 6 suggests that the conditional density $g(\mathbf{x}^{(1)}|\mathbf{x}^{(2)})$ depends on $\mathbf{x}^{(2)}$ only through the ℓ_1 -norm $\|\mathbf{x}^{(2)}\|$ and is a nested Dirichlet distribution with scale parameter $1 - \Delta_2$, which is a constant when $\mathbf{x}^{(2)}$ is fixed. We do not have similar conclusion for $g(\mathbf{x}^{(1)}|\mathbf{x}^{(2)})$ because of the asymmetry between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$. However, for the special case (i.e., $\mathbf{b}^{(2)} = \mathbf{0}$), Corollary 7 indicates that the conditional distribution

$g(\mathbf{x}^{(2)}|\mathbf{x}^{(1)})$ depends on $\mathbf{x}^{(1)}$ only through the ℓ_1 -norm $\|\mathbf{x}^{(1)}\|$ and is a Dirichlet distribution with scale parameter $1 - \Delta_1$, which is a constant when $\mathbf{x}^{(1)}$ is given. \blacksquare

5. Applications

5.1 Competing-risks model: Failure data for radio transmitter receivers

5.1.1 Competing-risks model

In reliability, clinical and epidemiologic studies, we often encounter competing-risks problem (Prentice *et al.*, 1978). The term ‘competing-risks’ has come to encompass the study of any failure process (e.g., survival study) in which there is more than one distinct type of failure or cause of death. For the purpose of illustration, we assume that there are only two possible causes of failure, indexed by i ($i = 1, 2$). Suppose that the notional times to failure of a unit under those two risks are denoted by the random variables X and Y , respectively. The variables X and Y cannot be observed. Available data on each unit typically include the time of failure $T = \min(X, Y) \geq 0$ which may be right censored and the corresponding cause of failure $C \in \{1, 2\}$ which will be unknown if T is censored.

To connect the likelihood function for the competing-risks model with nested Dirichlet distribution, we assume that the time to failure, T , is discrete with m possible values, say t_1, \dots, t_m . For continuous failure times, one can classify the failure times into a finite number of intervals with $\{t_j\}_{j=1}^m$, for instance, being the endpoints of those intervals. The cause-specific hazard rate for the i -th cause, $g_i(t_j)$, is defined as the rate of failure at time t_j from the i -th cause given that the unit has been working to time t_j , i.e.,

$$g_i(t_j) = \Pr(C = i, T = t_j | T \geq t_j), \quad i = 1, 2, \quad j = 1, \dots, m. \quad (5.1)$$

Define $p_{ij} = \Pr(C = i, T = t_j)$. Hence, $p_{ij} \geq 0$, $\sum_{i=1}^2 \sum_{j=1}^m p_{ij} = 1$, and (5.1) becomes

$$g_i(t_j) = \frac{p_{ij}}{\sum_{i=1}^2 \sum_{j'=j}^m p_{ij'}}. \quad (5.2)$$

One of the objective is to estimate the cause-specific hazard rate.

Let $Y_{\text{obs}} = \{n_{ij} : i = 1, 2, j = 1, \dots, m\} \cup \{r_1, \dots, r_m\}$ denote the observed data, where n_{ij} is the number of failure from cause i in the time period indexed by j and r_j is the number of right-censored items during that time period. The likelihood function for the unknown cell probabilities $\{p_{ij}\}$ is given by (Dykstra, Kochar and Robertson, 1998)

$$L(\{p_{ij}\} | Y_{\text{obs}}) \propto \prod_{i=1}^2 \prod_{j=1}^m p_{ij}^{n_{ij}} \times \prod_{j=1}^m \left(\sum_{j'=j+1}^m [p_{1j'} + p_{2j'}] \right)^{r_j}.$$

If we reparameterize $\{p_{ij}\}$ by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{2m})^\top$ with

$$\theta_{2j-1} = p_{1,m-j+1} \quad \text{and} \quad \theta_{2j} = p_{2,m-j+1}, \quad j = 1, \dots, m, \quad (5.3)$$

then the above likelihood function can be expressed as

$$L(\boldsymbol{\theta} | Y_{\text{obs}}) \propto \prod_{i=1}^{2m} \theta_i^{a_i-1} \times \prod_{j=1}^{2m-1} \left(\sum_{k=1}^j \theta_k \right)^{b_j}, \quad \boldsymbol{\theta} \in \mathbb{T}_{2m}, \quad (5.4)$$

where $\mathbf{a} = (a_1, \dots, a_{2m})^\top$ and $\mathbf{b} = (b_1, \dots, b_{2m-1})^\top$ with

$$\begin{aligned} a_{2j-1} &= n_{1,m-j+1} + 1, & a_{2j} &= n_{2,m-j+1} + 1, & j &= 1, \dots, m, \\ b_{2j-1} &= 0, & b_{2j} &= r_{m-j}, & j &= 1, \dots, m-1. \end{aligned}$$

Therefore, the likelihood function (5.4) is a nested Dirichlet distribution up to a normalizing constant, i.e.,

$$\boldsymbol{\theta} \sim \text{ND}_{2m,2m-1}(\mathbf{a}, \mathbf{b}) \quad \text{on} \quad \mathbb{T}_{2m} \quad (5.5)$$

if $\boldsymbol{\theta}$ is regarded as a random variable. The MLEs of $\boldsymbol{\theta}$ can be obtained analytically¹ by using Proposition 2. After some straightforward algebras, we find that the MLEs of the cause-specific hazard rates at time t_j are given by

$$\hat{g}_i(t_j) = \frac{n_{ij}}{\sum_{j'=j}^m (n_{1j'} + n_{2j'} + r_{j'})}, \quad i = 1, 2, \quad j = 1, \dots, m, \quad (5.6)$$

which coincide with the results obtained by Davis and Lawrence (1989).

¹Although Dykstra *et al.* (1998) obtained the analytical expressions for the MLEs of the parameters of interest based on the approach of Dykstra *et al.* (1991), they seem to be unaware that the likelihood function (5.4) is indeed a density function with closed-form mode.

5.1.2 Analyzing failure data of radio transmitter receivers

Next, consider a study of the failure times of 369 radio transmitter receivers (Medenhall and Hader, 1958; Cox, 1959). The failures were classified into two types, those confirmed (Type I) and unconfirmed (Type II) on arrival at the maintenance center. Forty-four of the 369 receivers were censored, i.e., they did not fail during the test period of 630 hours. Following Cox (1959), we exclude the 44 items in our analysis as the information contained in them would be not helpful in the estimation of the cause-specific hazard rates for the two types of failure. These data are listed in Table 1.

Table 1: *Observed cell counts of the failure times for radio transmitter receivers*

Index j	Time period $[t_{j-1}, t_j)$	Type I failures (n_{1j})	Type II failures (n_{2j})	Total	$\dagger r_j$
1	0-50	26	15	41	328
2	50-100	29	15	44	284
3	100-150	28	22	50	234
4	150-200	35	13	48	186
5	200-250	17	11	28	158
6	250-300	21	8	29	129
7	300-350	11	7	18	111
8	350-400	11	5	16	95
9	400-450	12	3	15	80
10	450-500	7	4	11	69
11	500-550	6	1	7	62
12	550-600	9	2	11	51
13	600-629	6	1	7	44
	Not fail at 630h	-	-	44	-
Total	-	218	107	369	-

Source: Medenhall and Hader (1958).

\dagger : r_j is the number of right-censored items during the time period indexed by j .

Based on (5.6), we calculate the estimates $\hat{g}_i(t_j)$ for $i = 1, 2$ and $j = 1, \dots, 13$ (see the 2-nd and 5-th column of Table 2). Figure 2 shows the comparison of the two hazard

rates.

[Insert Figure 2 here]

To describe the variability of the estimates $\hat{g}_i(t_j)$, we need to compute their standard errors. As a complicated function of $\{p_{ij}\}$, $g_i(t_j)$ defined in (5.2) is related to the parameter vector $\boldsymbol{\theta}$ through the relationship (5.3). Hence, the delta-method is quite difficult to be applied. However, the Bayesian approach is rather straightforward to be applied in the current situation. In fact, if we utilize the uniform prior distribution of $\boldsymbol{\theta}$, then the posterior distribution of $\boldsymbol{\theta}$ is still given by (5.5). Using the stochastic representation in Proposition 1, we generate 20,000 posterior samples of $\boldsymbol{\theta}$ from (5.5) and calculate 20,000 values of $g_i(t_j)$ via (5.2) and (5.3). The corresponding Bayesian means and standard deviations are given in Table 2.

Table 2: *MLEs and Bayesian estimates of $g_i(t_j)$ for the failure data of radio transmitter receivers*

Index	$g_1(t_j)$			$g_2(t_j)$		
	Classical	Bayesian		Classical	Bayesian	
j	MLE	Mean	SD	MLE	Mean	SD
1	0.01205	0.01268	0.00246	0.00695	0.00743	0.00184
2	0.01622	0.01690	0.00307	0.00839	0.00905	0.00223
3	0.01919	0.02017	0.00363	0.01507	0.01608	0.00335
4	0.02978	0.03129	0.00520	0.01106	0.01212	0.00324
5	0.01806	0.01965	0.00458	0.01168	0.01306	0.00383
6	0.02781	0.03020	0.00636	0.01059	0.01238	0.00402
7	0.01842	0.02113	0.00596	0.01172	0.01405	0.00486
8	0.02350	0.02759	0.00783	0.01068	0.01384	0.00554
9	0.03361	0.03998	0.01083	0.00840	0.01233	0.00618
10	0.02671	0.03538	0.01218	0.01526	0.02207	0.00982
11	0.03296	0.04853	0.01765	0.00549	0.01392	0.00962
12	0.07964	0.08667	0.03946	0.01769	0.04118	0.02255
13	0.11764	0.12205	0.05921	0.01960	0.01794	0.01921

SD = Standard Deviation.

5.2 Sample surveys: Two data sets for death penalty attitude

Kadane (1983) analyzed a data set from two sample surveys of jurors' attitudes toward the death penalty, where respondents are classified into one and only one of the following four groups:

- C_1 : Would not decide guilt versus innocence in a fair and impartial manner;
- C_2 : Fair and impartial on guilt and, when sentencing, would sometimes and sometimes not vote for the death penalty;
- C_3 : Fair and impartial on guilt and, when sentencing, would never vote for the death penalty;
- C_4 : Fair and impartial on guilt versus innocence and when sentencing, would always vote for the death penalty regardless of circumstance.

Let n_i be the frequency of group i , $i = 1, 2, 3, 4$. In some cases, some jurors found it difficult to classify their attitudes into one and only one of the aforementioned four categories. Instead, these jurors would consider themselves as a member of an union of some of the categories. For instance, some jurors may consider themselves to be fair and impartial, and would consider the death penalty and would at least sometimes vote for it, if the defendant is found guilty (i.e., $C_2 \cup C_4$). Let n_{24} and n_{123} be the frequencies of groups $C_2 \cup C_4$ and $C_1 \cup C_2 \cup C_3$, respectively. In the present death penalty study, the frequency data were given by $n_1 = 68$, $n_3 = 97$ and $n_{24} = 674$ according to the survey from the Field Research Corporation while $n_4 = 15$ and $n_{123} = 1484$ according to the survey by the Harris Survey Company. The goal is to estimate the cell probabilities.

Let $Y_{\text{obs}} = \{Y_{\text{obs},1}, Y_{\text{obs},2}\}$ denote the combined data and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_4)^\top$ the cell probabilities, where $Y_{\text{obs},1} = \{(n_1, n_3); n_{24}\}$ and $Y_{\text{obs},2} = \{n_4; n_{123}\}$ are the counts obtained

by the Field Research Corporation and the Harris Survey Company, respectively. The likelihood corresponding to the Field survey is

$$L(\boldsymbol{\theta}|Y_{\text{obs},1}) \propto \theta_1^{n_1} \theta_3^{n_3} (\theta_2 + \theta_4)^{n_{24}}.$$

Similarly, the likelihood based on the Harris survey is

$$L(\boldsymbol{\theta}|Y_{\text{obs},2}) \propto \theta_4^{n_4} (\theta_1 + \theta_2 + \theta_3)^{n_{123}}.$$

The combined likelihood for the observed data Y_{obs} is then given by (Dickey *et al.*, 1987)

$$L(\boldsymbol{\theta}|Y_{\text{obs}}) \propto \left\{ (\prod_{j=1}^4 \theta_j^{n_j}) \theta_1^0 (\theta_1 + \theta_2)^0 (\theta_1 + \theta_2 + \theta_3)^{n_{123}} \right\} \times (\theta_2 + \theta_4)^{n_{24}}, \quad (5.7)$$

where $\boldsymbol{\theta} \in \mathbb{T}_4$ and $n_2 \hat{=} 0$. We observe that the first term in 5.7) follows the $\text{ND}_{4,3}(\mathbf{a}, \mathbf{b})$ with $\mathbf{a} = (n_1 + 1, n_2 + 1, n_3 + 1, n_4 + 1)^\top$ and $\mathbf{b} = (0, 0, n_{123})^\top$ up to a normalizing constant while the second term is simply a power of a linear combination of the components of $\boldsymbol{\theta}$.

We use the *expectation-maximization* (EM) algorithm to calculate the MLEs of $\boldsymbol{\theta}$. By introducing a latent variable z to split $(\theta_2 + \theta_4)^{n_{24}}$, the conditional predictive density is

$$f(z|Y_{\text{obs}}, \boldsymbol{\theta}) = \text{Binomial}\left(z \mid n_{24}, \theta_2/(\theta_2 + \theta_4)\right), \quad z = 0, 1, \dots, n_{24}. \quad (5.8)$$

The likelihood function for the complete-data is then given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}}, z) = \text{ND}_{4,3}(\boldsymbol{\theta} | (n_1 + 1, n_2 + 1 + z, n_3 + 1, n_4 + 1 + n_{24} - z)^\top, (0, 0, n_{123})^\top). \quad (5.9)$$

Note that the MLEs of $\boldsymbol{\theta}$ based on the complete-data $\{Y_{\text{obs}}, z\}$ can be obtained analytically via Proposition 2. Using $\boldsymbol{\theta}^{(0)} = (1/4, 1/4, 1/4, 1/4)^\top$ as the initial value, the EM algorithm based on (5.8) and (5.9) converges in 11 iterations. The resultant MLEs are given by $\hat{\theta}_1 = 0.08105$, $\hat{\theta}_2 = 0.79332$, $\hat{\theta}_3 = 0.11561$ and $\hat{\theta}_4 = 0.01002$ with the corresponding standard errors being 0.00942, 0.01396, 0.01104 and 0.00257, which are obtained by the direct computation of the observed information matrix evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

5.3 Disease diagnosis: Ultrasound rating data for breast cancer metastasis

5.3.1 The ROC curve and the AUC

Diagnostic tests play an important role in medical studies and contribute significantly to health care costs. A *receiver operating characteristic* (ROC) curve is a plot of a diagnostic test's sensitivity (i.e., the test's ability to detect the condition of interest) versus its false-positive rate (i.e., the test's inability to recognize normal anatomy and physiology as normal). The curve is usually adopted to illustrate how different criteria for interpreting a test produce different values for the test's false-positive rate and sensitivity (Zhou, Obuchowski and McClish, 2002, p.5). The ROC curve provides an overall accuracy for a diagnosis test (Pepe, 2003). In existing literature, several numerical indices are proposed to summarize ROC curves. The most commonly used summary measure is the *area under ROC curve* (AUC), which is defined as

$$\text{AUC} = \int_0^1 \text{ROC}(t) dt.$$

The AUC gives a measure of the overall accuracy of the diagnostic test.

Let Y_D and $Y_{\bar{D}}$ denote independent and randomly chosen test results from the diseased and non-diseased populations, respectively. If the diagnostic score is considered to be continuous (e.g., the coronary artery calcium score), it can be shown that (Result 4.6 in Pepe, 2003)

$$\text{AUC} = \Pr(Y_D \geq Y_{\bar{D}}). \tag{5.10}$$

The AUC given by (5.10) has an interesting interpretation. It is equal to the probability that test results from a randomly selected pair of diseased and non-diseased subjects are correctly ordered.

5.3.2 Estimating AUC for ordinal diagnostic tests

When the diagnostic result is discrete (or ordinal), it can be shown that (Result 4.10 in Pepe, 2003)

$$\text{AUC} = \Pr(Y_D > Y_{\bar{D}}) + \frac{1}{2} \Pr(Y_D = Y_{\bar{D}}). \quad (5.11)$$

Here, we assume that the possible values of Y_D and $Y_{\bar{D}}$ are the consecutive integers $1, 2, \dots, n$. Let

$$\begin{aligned} \theta_i &= \Pr(Y_D = i), \quad i = 1, \dots, n, \quad \text{and} \\ \phi_j &= \Pr(Y_{\bar{D}} = j), \quad j = 1, \dots, n, \end{aligned}$$

with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top \in \mathbb{T}_n$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^\top \in \mathbb{T}_n$. From (5.11) and the independence between Y_D and $Y_{\bar{D}}$, we obtain (Broemeling, 2007, p.82)

$$\begin{aligned} \text{AUC}(\boldsymbol{\theta}, \boldsymbol{\phi}) &= \sum_{i=2}^n \sum_{j=1}^{i-1} \theta_i \phi_j + \frac{1}{2} \sum_{i=1}^n \theta_i \phi_i \\ &= \sum_{i=1}^n \sum_{j=1}^i \theta_i \phi_j - \frac{1}{2} \sum_{i=1}^n \theta_i \phi_i \\ &= \boldsymbol{\theta}^\top (\Omega_n - 0.5 I_n) \boldsymbol{\phi}, \end{aligned} \quad (5.12)$$

where I_n denotes the $n \times n$ identity matrix and

$$\Omega_n \triangleq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

5.3.3 Bayesian analysis of the ultrasound rating data

Consider the rating data for the study of breast cancer metastasis described in Peng & Hall (1996) and Hellmich *et al.* (1998). The diagnosis of metastasis is made with ultrasound by a radiologist to read 9 images from subjects with metastasis and 14 without metastasis.

The rating scale is from 1 to 4, where 1 denotes definitely abnormal, 2 probably abnormal, 3 probably normal, and 4 definitely normal. In addition, the rating scale 23 denotes equivocal (i.e., probably abnormal or probably normal). The ultrasound ratings of the twenty-three patients are given in Table 3.

Table 3 *Ultrasound rating data for detection of breast cancer metastasis*

Rating scale		1	2	23	3	4	Total
Metastasis	Yes	$5(y_1, \theta_1)$	$2(y_2, \theta_2)$	$0(y_{23}, \theta_2 + \theta_3)$	$2(y_3, \theta_3)$	$0(y_4, \theta_4)$	9
	No	$0(z_1, \phi_1)$	$1(z_2, \phi_2)$	$2(z_{23}, \phi_2 + \phi_3)$	$5(z_3, \phi_3)$	$6(z_4, \phi_4)$	14

Source: Peng & Hall (1996). The observed counts and the corresponding cell probabilities are in parentheses.

Let y_i denote the frequency of category ‘ i ’ in the 9 diseased patients and z_j denote the frequency of category ‘ j ’ in the 14 non-diseased patients. Hence, $\sum_{i=1}^4 y_i + y_{23} = 9$ and $\sum_{j=1}^4 z_j + z_{23} = 14$. Under the assumption of missing at random (Rubin, 1976), the likelihood function for the observed data $Y_{\text{obs}} = \{y_1, \dots, y_4, y_{23}\} \cup \{z_1, \dots, z_4, z_{23}\}$ is

$$L(\boldsymbol{\theta}, \boldsymbol{\phi} | Y_{\text{obs}}) \propto \left\{ \left(\prod_{i=1}^4 \theta_i^{y_i} \right) (\theta_2 + \theta_3)^{y_{23}} \right\} \times \left\{ \left(\prod_{j=1}^4 \phi_j^{z_j} \right) (\phi_2 + \phi_3)^{z_{23}} \right\},$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_4)^\top \in \mathbb{T}_4$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_4)^\top \in \mathbb{T}_4$. When the joint prior is the product of two independent Dirichlet densities:

$$\prod_{i=1}^4 \theta_i^{a_i-1} \times \prod_{j=1}^4 \phi_j^{b_j-1},$$

the joint posterior density is given by

$$\left\{ \left(\prod_{i=1}^4 \theta_i^{y_i+a_i-1} \right) (\theta_2 + \theta_3)^{y_{23}} \right\} \times \left\{ \left(\prod_{j=1}^4 \phi_j^{z_j+b_j-1} \right) (\phi_2 + \phi_3)^{z_{23}} \right\}. \quad (5.13)$$

Note that the second term in (5.13) can be rewritten as

$$\phi_2^{z_2+b_2-1} \phi_3^{z_3+b_3-1} \phi_4^{z_4+b_4-1} \phi_1^{z_1+b_1-1} \times \phi_2^0 (\phi_2 + \phi_3)^{z_{23}} (\phi_2 + \phi_3 + \phi_4)^0.$$

In other words, $(\phi_2, \phi_3, \phi_4, \phi_1)^\top \sim \text{ND}_{4,3}(\mathbf{a}^*, \mathbf{b}^*)$ with $\mathbf{a}^* = (z_2 + b_2, z_3 + b_3, z_4 + b_4, z_1 + b_1)^\top$ and $\mathbf{b}^* = (0, z_{23}, 0)^\top$. We have similar result for the first term in (5.13). The stochastic representation in Proposition 1 can be employed to generate i.i.d. posterior samples of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ according to (5.13). Therefore, the posterior distribution of the AUC is readily determined since (5.12) is a function of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

For the ultrasound rating data in Table 3, we adopt uniform prior distributions (i.e., all $a_i = b_j = 1$) for the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. From (5.13), we can see that the posterior distribution of $\boldsymbol{\theta}$ is Dirichlet with parameters $(6, 3, 3, 1)^\top$ and is independent of $(\phi_2, \phi_3, \phi_4, \phi_1)^\top$, which is nested Dirichlet with parameters $(2, 6, 7, 1)^\top$ and $(0, 2, 0)^\top$. Using the stochastic representation in Proposition 1, we generate 20,000 samples from the two independent posterior distributions and calculate 20,000 values, via (5.12), for the AUC. The posterior density and histogram of the AUC are shown in Figure 3, and the corresponding results are reported in Table 4.

[Insert Figure 3 here]

Table 4 *Bayesian estimates of AUC for the ultrasound rating data*

Investigator	Median	Mean	std	95% Credible Interval
Peng & Hall (1996)		0.987		[0.927, 0.999]
Hellmich <i>et al.</i> (1998)		0.903	0.073	[0.720, 0.990]
Our method	0.700	0.693	0.098	[0.487, 0.864]

It should be noted that Peng & Hall (1006) and Hellmich *et al.* (1998) used a regression model for the ordinal responses, normality assumption for a latent variable, and MCMC methods for computing the posterior distribution of the AUC. In addition, their results were based on a *five*-point ordinal scale, i.e., the rating scales are 1, 2, 23, 3 and 4. Therefore, there is a great difference between their results and ours. However, this is not surprising because we expect that the Bayesian estimate would produce underestimated

area as it is based on the linear interpolation of *four* points on the graph (Broemeling, 2007, p.84).

6. Discussion

The goal of this article is to further develop important distributional theory for the nested Dirichlet distribution and to explore some new applications in practice. Recently, Ng *et al.* (2008) studied a new family of distributions called grouped Dirichlet distributions, which also includes the Dirichlet distribution as a special case. We believe that both the grouped Dirichlet distributions and the nested Dirichlet distributions would provide two new tools for incomplete categorical data analyses (Tang *et al.* 2007).

Acknowledgment

The research of ML Tang was fully supported by a grant (HKBU261007) from the Research Grant Council of the Hong Kong Special Administrative Region.

References

- Balakrishnan, N. and Nevzorov, V. B. (2003). *A Primer on Statistical Distributions*. New York: John Wiley and Sons.
- Broemeling, L. D. (2007). *Bayesian Biostatistics and Diagnostic Medicine*. Boca Raton: Chapman & Hall/CRC.
- Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with two types of failures. *Journal of the Royal Statistical Society, B* **21**, 411–421.
- Davis, T. P. and Lawrence, A. J. (1989). The likelihood for competing risks survival analysis. *Scandinavian Journal of Statistics* **16**, 23–30.
- Dickey, J. M., Jiang J. M. and Kadane, J. B (1987). Bayesian methods for censored categorical data. *Journal of the American Statistical Association* **82**, 773–781.

- Dykstra, R., Kochar, S. and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Ann. Statist.* **19**, 870–888.
- Dykstra, R., Kochar, S. and Robertson, T. (1998). Restricted tests for testing independence of time to failure and cause of failure in a competing-risks model. *The Canadian Journal of Statistics* **26**(1), 57–68.
- Edwards, J. (1922). *A Treatise on the Integral Calculus, Vol. II*. New York: Macmillan.
- Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall.
- Gupta, R. D. and Richards, D. St. P. (1987). Multivariate Liouville distributions. *Journal of Multivariate Analysis* **23**, 233–256.
- Gupta, R. D. and Richards, D. St. P. (1990). The Dirichlet distributions and polynomial regression. *Journal of Multivariate Analysis* **32**, 95–102.
- Gupta, R. D. and Richards, D. St. P. (1991). Multivariate Liouville distributions, II. *Probability and Mathematical Statistics* **12**, 291–309.
- Gupta, R. D. and Richards, D. St. P. (1992). Multivariate Liouville distributions, III. *Journal of Multivariate Analysis* **43**, 29–57.
- Gupta, R. D. and Richards, D. St. P. (2001). The history of the Dirichlet and Liouville distributions. *International Statistical Review* **69**(3), 433–446.
- Hellmich, M., Abrams, K. R., Jones, D. R. and Lambert, P. C. (1998). A Bayesian approach to a general regression model for ROC curves. *Medical Decision Making* **18**, 436–443.
- Kadane, J. B. (1983). Juries hearing death penalty cases: Statistical analysis of a legal procedure. *Journal of the American Statistical Association* **78**, 544–552.
- Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press.
- Medenhall, W. and Hader, R. J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika* **45**,

504–520.

- Ng, K. W., Tang, M. L., Tan, M. and Tian, G. L. (2008). Grouped Dirichlet distribution: A new tool for incomplete categorical data analysis. *Journal of Multivariate Analysis* **99**(3), 490–509.
- Ng, K. W., Tang, M. L., Tian, G. L. and Tan, M. (2009). The nested Dirichlet distribution and incomplete categorical data analysis. *Statistica Sinica* **19**(1), 251–271.
- Peng, F. and Hall, W. J. (1996). Bayesian analysis of ROC curves using Markov Chain Monte Carlo methods. *Medical Decision Making* **16**, 404–411.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press.
- Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Sivazlian, B. D. (1981a). On a multivariate extension of gamma and beta distributions. *SIAM Journal of Applied Mathematics* **41**, 205–209.
- Sivazlian B. D. (1981b). A class of multivariate distributions. *Australian Journal of Statistics* **23**, 251–255.
- Tang, M. L., Ng, K. W., Tian, G. L. and Tan, M. (2007). On improved EM algorithm and confidence interval construction for incomplete $r \times c$ tables. *Computational Statistics and Data Analysis* **51**(6), 2919–2933.
- Thomas, S. and George, S. (2004). A review of Dirichlet distribution and its generalizations. *Journal of the Indian Society for Probability & Statistics* **8**, 72–91.
- Tian, G. L., Ng, K. W. and Geng, Z. (2003). Bayesian computation for contingency tables with incomplete cell-counts. *Statistica Sinica* **13**, 189–206.
- Zhou, X. H., Obuchowski, N. A. and McClish, D. K. (2002). *Statistical Methods in Diagnostic Medicine*. New York: John Wiley and Sons.

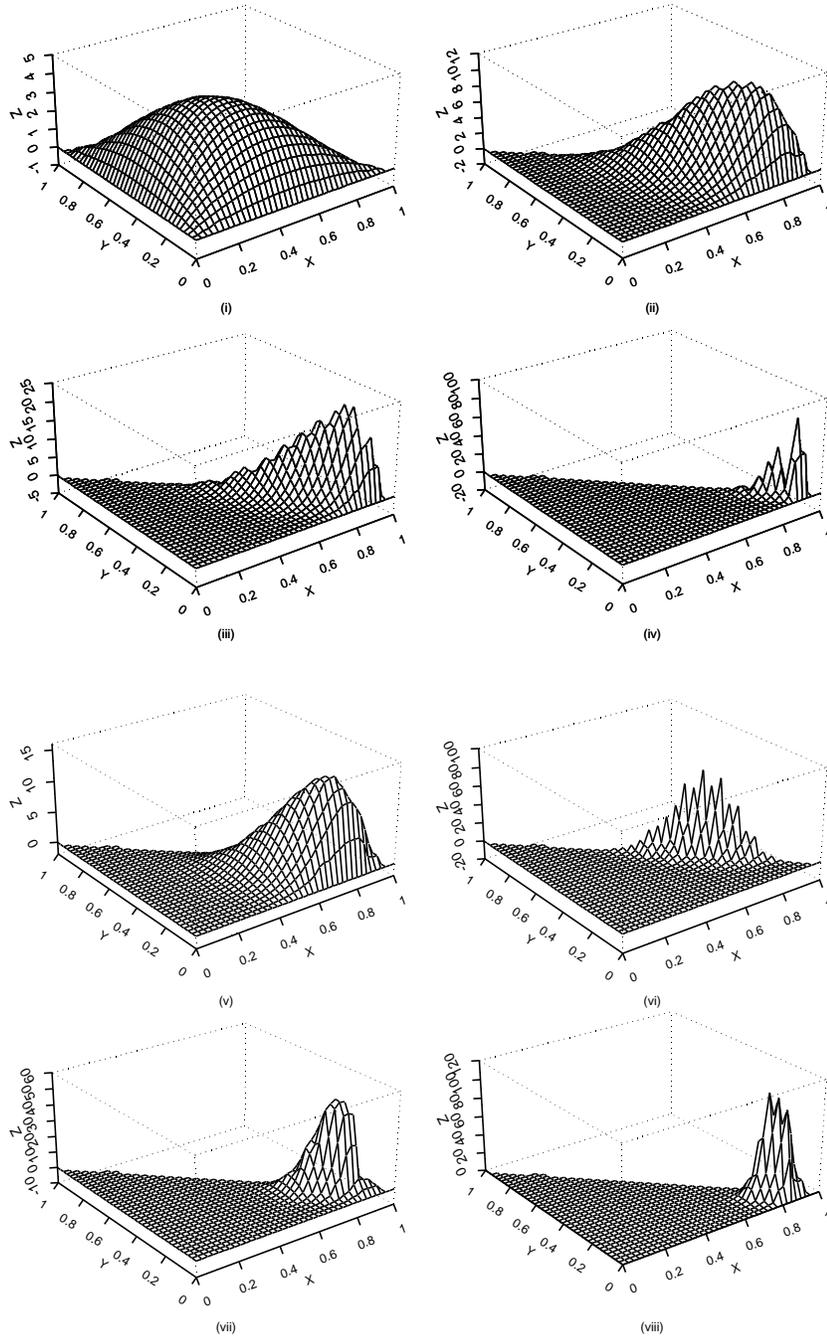


Figure 1. Perspective plots of Dirichlet and nested Dirichlet densities $ND_{3,2}(\mathbf{x}_{-3}|\mathbf{a}, \mathbf{b})$. (i) $\mathbf{a} = (2, 2, 2)^\top$ and $\mathbf{b} = (0, 0)^\top$; (ii) $\mathbf{a} = (2, 2, 2)^\top$ and $\mathbf{b} = (3, 3)^\top$; (iii) $\mathbf{a} = (2, 2, 2)^\top$ and $\mathbf{b} = (5, 10)^\top$; (iv) $\mathbf{a} = (2, 2, 2)^\top$ and $\mathbf{b} = (20, 30)^\top$; (v) $\mathbf{a} = (1, 2, 3)^\top$ and $\mathbf{b} = (6, 3)^\top$; (vi) $\mathbf{a} = (7, 9, 1)^\top$ and $\mathbf{b} = (10, 5)^\top$; (vii) $\mathbf{a} = (10, 6, 4)^\top$ and $\mathbf{b} = (20, 5)^\top$; (viii) $\mathbf{a} = (16, 7, 8)^\top$ and $\mathbf{b} = (50, 0)^\top$.

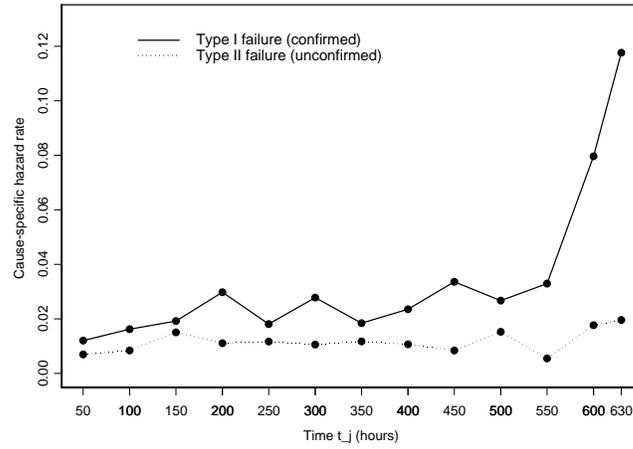


Figure 2. Comparison of the two cause-specific hazard rates for the radio transmitter receivers data.

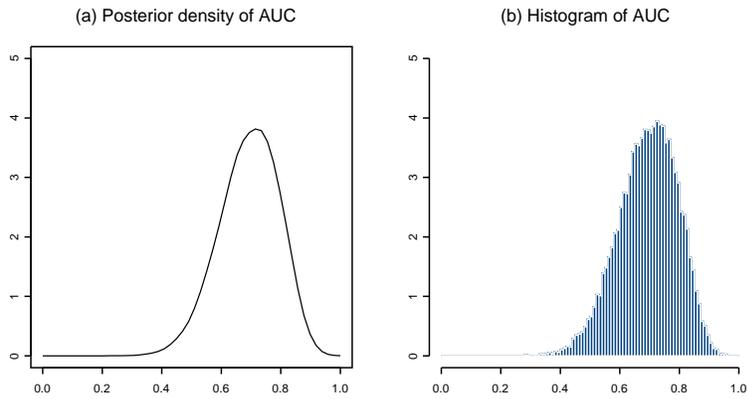


Figure 3. (a) Posterior density of the AUC given by (5.12) for the ultrasound rating data of breast cancer metastasis, where the density curve is estimated by a kernel density smoother based on 20,000 i.i.d. samples. (b) The histogram of the AUC.