

## BAYESIAN COMPUTATION FOR CONTINGENCY TABLES WITH INCOMPLETE CELL-COUNTS

Guo-Liang Tian<sup>\*,†</sup>, Kai Wang Ng<sup>‡</sup> and Zhi Geng<sup>\*</sup>

*\*Peking University, †The University of Hong Kong  
and ‡St. Jude Children's Research Hospital*

*Abstract:* This article studies Bayesian analysis of contingency tables (or multinomial data) where the cell counts are not fully observed due to reasons such as nonresponse and misclassification, and derives the posterior distributions of the unknown cell probabilities in terms of various types of generalized Dirichlet distributions. For some special situations such as grouped and nested Dirichlet distributions, the posterior means of the unknown cell probabilities can be obtained in closed form by using inverse Bayes formulae and/or stochastic representation. When closed-form expressions do not exist, we suggest using importance sampling with a feasible proposal density to approximately compute the posterior quantities, and propose a procedure for choosing an effective proposal density. Applications are illustrated by sample surveys with nonresponse, crime survey data, death penalty attitude data, and misclassified multinomial data.

*Key words and phrases:* Bayesian inference, grouped and nested Dirichlet distributions, incomplete data, inverse Bayes formulae, stochastic representation.

### 1. Introduction

Statistical procedures for the treatment of missing value problems have received considerable attention in the past several decades. The advent of the EM algorithm (Dempster, Laird and Rubin (1977)) has virtually revolutionized the practice of frequentist statistics. In a Bayesian framework, the posterior density of the observed data may be difficult to calculate directly. By introducing latent variables or unobserved data, the data augmentation algorithm (Tanner and Wong (1987)) and the *Markov chain Monte Carlo* (MCMC) or the Gibbs sampler (Gelfand and Smith (1990)) can be used to deal with such problems. No closed-form expressions are obtained for these procedures because they are iterative.

The *inverse Bayes formulae* (IBF) method of Ng (1995, 1997) can be used to work out closed-form solutions to the incomplete-data problems for some situations. Tan and Tian (2001) obtained some extensive results on the applications of the IBF method to a wide variety of statistical problems, including bivariate normal/truncated normal/exponential distributions, a genetic linkage model,

misclassified multinomial data, a reliability growth model with missing data, and hierarchical models. With these applications it is argued that the IBF is a useful tool for incomplete data in Bayesian settings. One aim of the paper is to further show that Bayesian computation can be routinely performed by the IBF method when the posterior is a grouped Dirichlet distribution.

This article focuses on Bayesian analysis of contingency tables with incomplete cell-counts and derives the posteriors of the unknown cell probabilities in terms of various types of generalized Dirichlet distributions. For some special situations such as grouped and nested Dirichlet distributions, the posterior means of the unknown cell probabilities can be obtained in closed form by using the IBF or *stochastic representation* (SR). When closed-form expressions do not exist, we suggest using importance sampling with a feasible proposal density to approximately compute the posterior quantities, and propose a procedure for choosing an effective proposal density.

Beginning with the formulation of statistical problems, Section 2 provides a closed-form solution by IBF, derives the SRs for the grouped and nested Dirichlet distributions, and suggests importance sampling approximation for the generalized Dirichlet distributions. Section 3 presents applications of the proposed methods to sample surveys with nonresponse, crime survey data, death penalty attitude data, and misclassified multinomial data. In Section 4, we give an illustrative example. Section 5 proposes a procedure for choosing an effective proposal density in importance sampling. Finally, a discussion is given and some mathematical proofs are put into the Appendix.

## 2. Formulation of Problems and Development of Methodology

### 2.1. Formulation of problems

Consider Bayesian analysis for the contingency tables with incomplete cell-counts. Let  $Y_{\text{obs}} = \{y_1, \dots, y_n; y_1^*, \dots, y_m^*\}$  denote the observed cell counts and  $\theta$  be the cell probability vector of interest, where  $\theta \in T_n = \{(\theta_1, \dots, \theta_n)^\top : \theta_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \theta_i = 1\}$ . When there exist some missing cell-counts, the likelihood function  $L(\theta|Y_{\text{obs}})$  contains two parts:  $\prod_{i=1}^n \theta_i^{y_i}$  — the product of the powers of cell probabilities and  $\prod_{j=1}^m (\sum_{i=1}^n \gamma_{ij} \theta_i)^{y_j^*}$  — the product of powers of linear combinations of cell probabilities over sets of categories not distinguished. The Dirichlet distribution  $D(\alpha_1, \dots, \alpha_n)$  is a natural prior distribution with resulting posterior a *generalized Dirichlet distribution*. Its density is given by  $f(\theta|Y_{\text{obs}}) = f(\theta|a, b, \Gamma) = c^{-1} \cdot gD(\theta|a, b, \Gamma)$  with kernel

$$gD(\theta|a, b, \Gamma) = \left( \prod_{i=1}^n \theta_i^{a_i-1} \right) \cdot \prod_{j=1}^m \left( \sum_{i=1}^n \gamma_{ij} \theta_i \right)^{b_j-1}, \quad \theta \in T_n, \quad (2.1)$$

where  $a = (a_1, \dots, a_n)^\top$  with  $a_i = y_i + \alpha_i$  and  $b = (b_1, \dots, b_m)^\top$  with  $b_j = y_j^* + 1$  are two known vectors, and  $\Gamma = (\gamma_{ij})$  is an  $n \times m$  known scale matrix.

Our aim is to compute posterior moments. Denote the simplex by  $V_n = \{(x_1, \dots, x_n)^\top : x_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n x_i \leq 1\}$ . It is easy to see that  $\theta \in T_n$  is equivalent to  $\theta_{-n} := (\theta_1, \dots, \theta_{n-1})^\top \in V_{n-1}$ . The normalizing constant and the posterior moments are given by

$$c = c(a, b, \Gamma) = \int_{V_{n-1}} gD(\theta|a, b, \Gamma) d\theta_{-n}, \quad (2.2)$$

$$E\left(\prod_{i=1}^n \theta_i^{r_i}\right) = \frac{c(a+r, b, \Gamma)}{c(a, b, \Gamma)}, \quad \text{where } r = (r_1, \dots, r_n)^\top. \quad (2.3)$$

Dickey, Jiang and Kadane (1987) noted that (2.2) has a close relationship with Carlson (1977)'s multiple hypergeometric function and (2.3) can be expressed as ratios of such Carlson's functions. One method proposed by Kadane (1985) is multinomial expansion of the integrand, and the other is Laplace's integral method (Tierney and Kadane (1986)) which is approximate. However, both of them are inconvenient for users.

In what follows, we give the closed-form expressions of (2.3), or equivalently (2.2), for two special cases of grouped and nested Dirichlet distributions by IBF and SR. For the generalized Dirichlet distribution, we suggest using importance sampling with a feasible proposal density to approximately compute the posterior moments.

## 2.2. Inverse Bayes formulae

We briefly introduce the IBF in the context of the general observed/missing data. Tanner and Wong (1987) introduced the concept of data augmentation for calculating the observed posterior density  $f(\theta|Y_{\text{obs}})$  when the normalizing constant is difficult to compute. The idea is to introduce a latent variable  $z$ , which is not observable or missing, such that the complete-data posterior  $f(\theta|Y_{\text{obs}}, z)$  and the conditional predictive density  $f(z|Y_{\text{obs}}, \theta)$  are available. Then  $f(\theta|Y_{\text{obs}})$  can be obtained as an iterative solution of an integral equation. Ng (1995, 1997) noticed a simple analytic solution to that integral equation. Specifically, given  $f(\theta|Y_{\text{obs}}, z)$  and  $f(z|Y_{\text{obs}}, \theta)$ , we have

$$f(\theta|Y_{\text{obs}}) = \left\{ \int \frac{f(z|Y_{\text{obs}}, \theta)}{f(\theta|Y_{\text{obs}}, z)} dz \right\}^{-1} = \frac{f(\theta|Y_{\text{obs}}, z_0)}{f(z_0|Y_{\text{obs}}, \theta)} \left\{ \int \frac{f(\theta|Y_{\text{obs}}, z_0)}{f(z_0|Y_{\text{obs}}, \theta)} d\theta \right\}^{-1}. \quad (2.4)$$

The first equation of (2.4) is called a pointwise IBF and the last one a functionwise IBF. Note that the functionwise IBF holds for some arbitrary  $z = z_0$ . Section 3.1

will give the closed-form expression of the posterior mean for grouped Dirichlet distribution by using (2.4).

### 2.3. Grouped Dirichlet distribution

A generalized Dirichlet distribution (2.1) is called a *grouped Dirichlet distribution* if its density is given by

$$f(\theta|a, b) = c_1^{-1} \cdot \left( \prod_{i=1}^n \theta_i^{a_i-1} \right) \cdot \left( \sum_{j=1}^s \theta_j \right)^{b_1-1} \left( \sum_{j=s+1}^n \theta_j \right)^{b_2-1}, \quad \theta \in T_n, \quad (2.5)$$

where  $a = (a_1, \dots, a_n)^\top$  and  $b = (b_1, b_2)^\top$ . We write  $\theta \sim \text{GD}_{n,2}(a, b)$ . Motivated by the SR of a Dirichlet distribution (Fang, Kotz and Ng (1990), p.146), we obtain an SR of  $\theta \sim \text{GD}_{n,2}(a, b)$  as follows (see Appendix):

$$\begin{aligned} \theta_i &\stackrel{d}{=} \phi_i \phi_s, & i = 1, \dots, s-1, & \quad \theta_s &\stackrel{d}{=} (1 - \sum_{j=1}^{s-1} \phi_j) \phi_s, \\ \theta_i &\stackrel{d}{=} \phi_i (1 - \phi_s), & i = s+1, \dots, n-1, & \quad \theta_n &\stackrel{d}{=} (1 - \sum_{j=s+1}^{n-1} \phi_j) (1 - \phi_s), \end{aligned} \quad (2.6)$$

where  $(\phi_1, \dots, \phi_{s-1})^\top \sim \text{D}(a_1, \dots, a_{s-1}; a_s)$ ,  $\phi_s \sim \text{Beta}(\sum_{j=1}^s a_j + b_1 - 1, \sum_{j=s+1}^n a_j + b_2 - 1)$ ,  $(\phi_{s+1}, \dots, \phi_{n-1})^\top \sim \text{D}(a_{s+1}, \dots, a_{n-1}; a_n)$ , and they are independent. Further,  $(\xi_1, \dots, \xi_{n-1})^\top \sim \text{D}(d_1, \dots, d_{n-1}; d_n)$  implies that  $(\xi_1, \dots, \xi_n)^\top \sim \text{D}(d_1, \dots, d_n)$ , where  $\xi_n = 1 - \sum_{j=1}^{n-1} \xi_j$ . Using the moments of Dirichlet and Beta distributions, one can calculate the high-order moments of a grouped Dirichlet, for instance,

$$\begin{aligned} E(\theta_i) &= E(\phi_i) \cdot E(\phi_s) = \frac{a_i}{\sum_{j=1}^s a_j} \cdot \frac{(\sum_{j=1}^s a_j) + b_1 - 1}{(\sum_{j=1}^n a_j) + b_1 + b_2 - 2}, \quad i = 1, \dots, s-1, \\ E(\theta_i) &= E(\phi_i) \cdot E(1 - \phi_s) = \frac{a_i}{\sum_{j=s+1}^n a_j} \cdot \frac{(\sum_{j=s+1}^n a_j) + b_2 - 1}{(\sum_{j=1}^n a_j) + b_1 + b_2 - 2}, \quad i = s+1, \dots, n-1. \end{aligned} \quad (2.7)$$

It is easy to generalize these results to the more general case of a grouped Dirichlet distribution with  $t$  partitions, denoted by  $\theta \sim \text{GD}_{n,t}(a, b)$ . Its density is

$$c_2^{-1} \cdot \left( \prod_{i=1}^n \theta_i^{a_i-1} \right) \cdot \prod_{j=1}^t (\theta_{s_{j-1}+1} + \dots + \theta_{s_j})^{b_j-1}, \quad \theta \in T_n, \quad (2.8)$$

where  $0 = s_0 < 1 \leq s_1 < \dots < s_t = n$ . Similarly, an SR of  $\theta \sim \text{GD}_{n,t}(a, b)$  with parameter vectors  $a = (a_1, \dots, a_n)^\top$  and  $b = (b_1, \dots, b_t)^\top$  is given by (see Appendix)

$$\begin{aligned} \theta_i &\stackrel{d}{=} \phi_i \phi_{s_1}, \quad i = 1, \dots, s_1 - 1, & \theta_{s_1} &\stackrel{d}{=} (1 - \sum_{j=1}^{s_1-1} \phi_j) \phi_{s_1}, \\ \theta_i &\stackrel{d}{=} \phi_i \phi_{s_2}, \quad i = s_1 + 1, \dots, s_2 - 1, & \theta_{s_2} &\stackrel{d}{=} (1 - \sum_{j=s_1+1}^{s_2-1} \phi_j) \phi_{s_2}, \\ &\vdots & & \vdots \\ \theta_i &\stackrel{d}{=} \phi_i \phi_{s_t}, \quad i = s_{t-1} + 1, \dots, s_t - 1, & \theta_{s_t} &\stackrel{d}{=} (1 - \sum_{j=s_{t-1}+1}^{s_t-1} \phi_j) \phi_{s_t}, \end{aligned} \quad (2.9)$$

where  $(\phi_1, \dots, \phi_{s_1-1})^\top \sim D(a_1, \dots, a_{s_1-1}; a_{s_1})$ ,  $(\phi_{s_1+1}, \dots, \phi_{s_2-1})^\top \sim D(a_{s_1+1}, \dots, a_{s_2-1}; a_{s_2})$ ,  $\dots$ ,  $(\phi_{s_{t-1}+1}, \dots, \phi_{s_t-1})^\top \sim D(a_{s_{t-1}+1}, \dots, a_{s_t-1}; a_{s_t})$ ,  $(\phi_{s_1}, \phi_{s_2}, \dots, \phi_{s_t})^\top \sim D(\sum_{k=1}^{s_1} a_k + b_1 - 1, \sum_{k=s_1+1}^{s_2} a_k + b_2 - 1, \dots, \sum_{k=s_{t-1}+1}^{s_t} a_k + b_t - 1)$ , and they are independent. Then the moments of  $\theta$  can be obtained via (2.9).

#### 2.4. Nested Dirichlet distribution

A generalized Dirichlet distribution (2.1) is called a *nested Dirichlet distribution* if its density is given by

$$c_3^{-1} \cdot \left( \prod_{i=1}^n \theta_i^{a_i-1} \right) \cdot \prod_{j=1}^{n-1} \left( \sum_{k=1}^j \theta_k \right)^{b_j-1}, \quad \theta \in T_n, \quad (2.10)$$

where  $a = (a_1, \dots, a_n)^\top$  and  $b = (b_1, \dots, b_{n-1})^\top$ . We write  $\theta \sim \text{ND}_{n,n-1}(a, b)$ . As shown in the Appendix, we have the following SR:

$$\theta_i \stackrel{d}{=} (1 - \phi_{i-1}) \prod_{j=i}^{n-1} \phi_j, \quad \phi_0 \equiv 0, \quad i = 1, \dots, n, \quad (2.11)$$

where  $\phi_j \sim \text{Beta}(\sum_{k=1}^j (a_k + b_k - 1), a_{j+1})$ ,  $j = 1, \dots, n-1$ , and  $\phi_1, \dots, \phi_{n-1}$  are mutually independent. Furthermore, from (2.11), we have  $\theta_1 + \dots + \theta_i \stackrel{d}{=} \prod_{j=i}^{n-1} \phi_j$ ,  $i = 1, \dots, n-1$ . Then, for example, we obtain

$$E(\theta_i) = a_i \cdot E(\theta_1 + \dots + \theta_i) / \left\{ \sum_{k=1}^{i-1} (a_k + b_k - 1) + a_i \right\}, \quad i = 1, \dots, n,$$

$$E(\theta_1 + \dots + \theta_i) = \prod_{j=i}^{n-1} \left\{ \sum_{k=1}^j (a_k + b_k - 1) / \left[ \sum_{k=1}^j (a_k + b_k - 1) + a_{j+1} \right] \right\}, \quad i = 1, \dots, n-1.$$

#### 2.5. Generalized Dirichlet distribution

Now we calculate the posterior moments (2.3). Our suggestion is first to find a proposal density  $h(\cdot)$  defined on  $V_{n-1}$ , and then to estimate  $c(a, b, \Gamma) = \int_{V_{n-1}} \{gD(x|a, b, \Gamma)/h(x)\} h(x) dx$  by

$$\hat{c}(a, b, \Gamma) = \frac{1}{M} \sum_{k=1}^M \frac{gD(x^{(k)}|a, b, \Gamma)}{h(x^{(k)})}, \quad (2.12)$$

where  $x^{(1)}, \dots, x^{(M)}$  is an i.i.d. sample of size  $M$  from  $h(\cdot)$ . Feasible choices of  $h(\cdot)$  include a Dirichlet distribution  $D(a_1, \dots, a_{n-1}; a_n)$ , a grouped Dirichlet distribution with suitable parameter vectors  $a$  and  $b$ , and a nested Dirichlet distribution with corresponding parameter vectors  $a$  and  $b$ . They can be simulated from Beta distributions via (2.6), (2.9) and (2.11).

In Section 5, we propose an effective proposal density in importance sampling for better efficiency. Quasi-Monte Carlo methods can be used in calculating (2.12), readers are referred to Fang, Wang and Bentler (1994).

### 3. Applications

#### 3.1. Sample surveys with nonresponse

Let  $n$  denote the total number of questionnaires sent out. Suppose  $n_1$  individuals respond but  $n_2 = n - n_1$  do not. Of these  $n_1$  respondents, there are  $y_1$  individuals whose answers are classified into category  $A_1$  and the remaining  $y_2$  are in  $A_2$ . Denoting the respondents by  $R$  and the nonrespondents by  $NR$ , the observed counts and the corresponding cell probabilities may be summarized in Table 1 with  $\pi_1$  as the parameter of interest. Park and Brown (1994) used the frequentist approach and Albert and Gupta (1985) and Chiu and Sedransk (1986) used the Bayesian approach to study this nonresponse problem. Employing the IBF and the SR of the grouped Dirichlet distribution, we can obtain the exact expression for a Bayesian estimate of  $\pi_1$  in dichotomous and polytomous cases.

Table 1.  $2 \times 2$  observed counts and corresponding cell probabilities.

Categories	R	NR	Total	Categories	R	NR	Total
$A_1$	$y_1$	$z$		$A_1$	$\pi_{11}$	$\pi_{12}$	$\pi_1$
$A_2$	$y_2$	$n_2 - z$		$A_2$	$\pi_{21}$	$\pi_{22}$	$\pi_2$
Total	$n_1$	$n_2$	$n$	Total	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	1

First we consider the dichotomous case. The observed data is denoted by  $Y_{\text{obs}} = (y_1, y_2; n_2)^\top$ , where  $n_2 = n - (y_1 + y_2)$ . A natural latent variable  $z$  is introduced by writing  $n_2 = z + (n_2 - z)$  and the corresponding cell probability  $\pi_{\cdot 2} = \pi_{12} + \pi_{22}$ . The likelihood function for the complete-data  $(Y_{\text{obs}}, z)$  is  $L(Y_{\text{obs}}, z|\pi) \propto \pi_{11}^{y_1} \pi_{21}^{y_2} \pi_{12}^z \pi_{22}^{n_2 - z}$ . If we take  $D(\pi|\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22})$  as the prior of  $\pi$ , then the complete-data posterior distribution is  $f(\pi|Y_{\text{obs}}, z) = D(\pi|y_1 + \alpha_{11}, y_2 + \alpha_{21}, z + \alpha_{12}, n_2 - z + \alpha_{22})$ . Noting that the conditional predictive density of  $z$  given  $Y_{\text{obs}}$  and  $\pi$  is Binomial( $n_2, \pi_{12}/\pi_{\cdot 2}$ ), i.e.,  $f(z|Y_{\text{obs}}, \pi) = \binom{n_2}{z} \left(\frac{\pi_{12}}{\pi_{\cdot 2}}\right)^z \left(\frac{\pi_{22}}{\pi_{\cdot 2}}\right)^{n_2 - z}$ ,  $z = 0, 1, \dots, n_2$ , we have, using the pointwise IBF (2.4),

$$\begin{aligned}
 f(\pi|Y_{\text{obs}}) &= \left\{ \sum_{z=0}^{n_2} \frac{f(z|Y_{\text{obs}}, \pi)}{f(\pi|Y_{\text{obs}}, z)} \right\}^{-1} \\
 &= c^{-1}(\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}) \pi_{11}^{y_1 + \alpha_{11} - 1} \pi_{21}^{y_2 + \alpha_{21} - 1} \pi_{12}^{\alpha_{12} - 1} \pi_{22}^{\alpha_{22} - 1} \pi_{\cdot 2}^{n_2},
 \end{aligned}$$

where  $c(\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}) = \sum_{z=0}^{n_2} \binom{n_2}{z} B(y_1 + \alpha_{11}, y_2 + \alpha_{21}, z + \alpha_{12}, n_2 - z + \alpha_{22})$ .

Gunel (1984, p.742) showed that  $B(a_1, a_2, a_3, a_4) = B(a_1, a_2)B(a_3, a_4)B(a_1 + a_2, a_3 + a_4)$  for  $a_i > 0 (i = 1, \dots, 4)$ , and  $\sum_{s=0}^n \binom{n}{s} B(s + a, n - s + b) = B(a, b)$ .

By these identities, we obtain  $c(\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}) = B(y_1 + \alpha_{11}, y_2 + \alpha_{21})B(n_1 + \alpha_{.1}, n_2 + \alpha_{.2})B(\alpha_{12}, \alpha_{22})$ . The Bayesian estimate of  $\pi_{11}$  is  $c(\alpha_{11} + 1, \alpha_{21}, \alpha_{12}, \alpha_{22})$  over  $c(\alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22})$ , i.e.,  $\hat{\pi}_{11} = (y_1 + \alpha_{11}) / (n + \alpha_{.})$ , where  $\alpha_{.} = \alpha_{11} + \alpha_{21} + \alpha_{12} + \alpha_{22}$ . Similarly,  $\hat{\pi}_{12} = (n_2 + \alpha_{.2})\alpha_{12} / \{(n + \alpha_{.})\alpha_{.2}\}$ . Therefore, the Bayes estimator of  $\pi_{1.} = \pi_{11} + \pi_{12}$  is  $(y_1 + \alpha_{.1} + n_2\alpha_{12}/\alpha_{.2}) / (n + \alpha_{.})$ .

The generalization of the above IBF analysis to the polytomous case is straightforward. Here we apply the SR of a grouped Dirichlet distribution as an alternative approach. The corresponding observed frequencies and cell probabilities are displayed in Table 2 with  $\pi_{1.}, \dots, \pi_{k.}$  as the parameters of interest. The likelihood function for the observed data is proportional to  $(\prod_{i=1}^k \pi_{i1}^{y_i}) \cdot \pi_{.2}^{n_2}$ , and the prior of  $\pi$  can be taken as  $D(\alpha)$ . After introducing the reparametrization  $\theta = (\theta_1, \dots, \theta_k, \theta_{k+1}, \dots, \theta_{2k})^T$  with  $\theta_i = \pi_{i1}$  and  $\theta_{k+i} = \pi_{i2}$  for  $i = 1, \dots, k$ , we know that the observed posterior of  $\theta$  is proportional to  $(\prod_{i=1}^{2k} \theta_i^{y_i + \alpha_i - 1}) \cdot (\sum_{j=1}^k \theta_j)^0 (\sum_{j=k+1}^{2k} \theta_j)^{n_2}$ , where  $y_i = 0$  for  $i = k + 1, \dots, 2k$ . This means that  $\theta \sim \text{GD}_{2k,2}(y + \alpha, b)$ , where  $y = (y_1, \dots, y_{2k})^T$ ,  $\alpha = (\alpha_1, \dots, \alpha_{2k})^T$  and  $b = (1, n_2 + 1)^T$ . By (2.7), it is easy to see that the Bayes estimator of  $\pi_{1.}$  is  $\hat{\pi}_{1.} = E(\theta_1 + \theta_{k+1}) = (y_1 + \alpha_1 + \alpha_{k+1} + \frac{n_2\alpha_{k+1}}{\alpha_{k+1} + \dots + \alpha_{2k}}) / (n + \alpha_{.})$  with  $\alpha_{.} = \sum_{i=1}^{2k} \alpha_i$ , which coincides with Eq. (2.16) in Basu and Pereira ((1982), p.351).

Table 2.  $k \times 2$  observed counts and corresponding cell probabilities.

Categories	R	NR	Total	Categories	R	NR	Total
$A_1$	$y_1$			$A_1$	$\pi_{11}$	$\pi_{12}$	$\pi_{1.}$
$\vdots$	$\vdots$			$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_k$	$y_k$			$A_k$	$\pi_{k1}$	$\pi_{k2}$	$\pi_{k.}$
Total	$n_1$	$n_2$	$n$	Total	$\pi_{.1}$	$\pi_{.2}$	1

### 3.2. Crime survey data

Consider the data set in Table 3 obtained via the National Crime Survey conducted by the U.S. Bureau of the Census (Kadane (1985)). Households are interviewed to see if they had been victimized by crime in the preceding six-month period. The occupants of the same housing unit were reinterviewed again six months later to determine if they had been victimized in the intervening months, whether these were the same people or not. Discarding 115 households, which is equivalent to the assumption of *missing at random* (MAR) or ignorable missing mechanism (Little and Rubin (1987)), Schafer ((1997), p.45, p.271) analyzed

this data set by the EM algorithm from a frequentist perspective. In a Bayesian framework, this data set was originally analyzed by Kadane (1985). Now we denote the probability that a household is crime-free (victimized) in both periods by  $\theta_1$  ( $\theta_4$ ), that it is crime-free (victimized) in period 1 and victimized (crime-free) in period 2 by  $\theta_2$  ( $\theta_3$ ). Naturally,  $\theta = \sum_{j=1}^4 \theta_j = 1$ , and  $\theta_j > 0$  for  $j = 1, \dots, 4$ . One of the goals is to obtain the Bayes estimator of  $\theta_j$ .

Table 3. Victimization results from the national crime survey in Kadane (1985).

1st Visit\2nd Visit	Crime-free	Victims	Nonresponse
Crime-free	392 ( $n_1, \theta_1$ )	55 ( $n_2, \theta_2$ )	33 ( $n_{12}$ )
Victims	76 ( $n_3, \theta_3$ )	38 ( $n_4, \theta_4$ )	9 ( $n_{34}$ )
Nonresponse	31 ( $n_{13}$ )	7 ( $n_{24}$ )	115 ( $n_{1234}$ )

NOTE: Notations for the observed frequencies of households and probabilities are in parentheses.

### 3.2.1. Nonignorable missing mechanism

Under the assumption of a nonignorable missing mechanism, we have a total of 15 free-parameter  $\pi = (\pi_{ij})$ , a  $4 \times 4$  matrix, see Table 4. These  $\{\pi_{ij}\}$  are not identifiable unless there is a prior distribution for  $\pi$ . At present,  $\pi_{ij}$  can be decomposed as

$$\pi_{ij} = \theta_j \lambda_{ij}, \quad i, j = 1, \dots, 4, \quad (3.1)$$

where  $\lambda_{.j} = \lambda_{1j} + \dots + \lambda_{4j} = 1$  and  $\theta_j = \pi_{1j} + \dots + \pi_{4j}$  for  $j = 1, \dots, 4$ . Naturally,  $\lambda_{ij}$  denotes the corresponding conditional probability, reflecting the prior information of nonignorability. For instance,  $\lambda_{11}$  ( $\lambda_{41}$ ) is the conditional probability that a household responds (does not respond) in both interviews given that this household is crime-free in both periods. Therefore, in Table 4, responding set  $R_{1\bar{2}}$  represents that a household responds in the 1st interview but does not in the 2nd, and the other responding sets have analogous interpretations. Obviously,  $A_1$  ( $A_4$ ) represents the category that a household is crime-free (victimized) in both periods. In this way, we can write  $\lambda_{11} = \Pr(R_{12}|A_1)$ ,  $\lambda_{21} = \Pr(R_{1\bar{2}}|A_1)$ ,  $\lambda_{31} = \Pr(R_{\bar{1}2}|A_1)$  and  $\lambda_{41} = \Pr(R_{\bar{1}\bar{2}}|A_1)$ . The likelihood function is proportional to

$$\prod_{j=1}^4 \pi_{1j}^{n_j} \cdot (\pi_{21} + \pi_{22})^{n_{12}} (\pi_{23} + \pi_{24})^{n_{34}} (\pi_{31} + \pi_{33})^{n_{13}} (\pi_{32} + \pi_{34})^{n_{24}} \left( \sum_{j=1}^4 \pi_{4j} \right)^{n_{1234}}, \quad (3.2)$$

and the prior density can be taken as  $f(\pi) \propto \prod_{i=1}^4 \prod_{j=1}^4 \pi_{ij}^{\alpha_{ij}-1}$ . The posterior density is proportional to  $\prod_{j=1}^4 \pi_{1j}^{n_j + \alpha_{1j} - 1} \prod_{i=2}^4 \prod_{j=1}^4 \pi_{ij}^{\alpha_{ij} - 1} \cdot (\pi_{21} + \pi_{22})^{n_{12}} (\pi_{23} +$

$\pi_{24})^{n_{34}}(\pi_{31} + \pi_{33})^{n_{13}}(\pi_{32} + \pi_{34})^{n_{24}}(\sum_{j=1}^4 \pi_{4j})^{n_{1234}}$ , which can be rewritten as, by a straightforward reparametrization,

$$\prod_{i=1}^4 \xi_i^{n_i + \alpha_{i1} - 1} \cdot \prod_{i=5}^8 \xi_i^{\alpha_{2,i-4} - 1} \cdot \xi_9^{\alpha_{31} - 1} \xi_{10}^{\alpha_{33} - 1} \xi_{11}^{\alpha_{32} - 1} \xi_{12}^{\alpha_{34} - 1} \cdot \prod_{i=13}^{16} \xi_i^{\alpha_{4,i-12} - 1} \cdot \left(\sum_{j=1}^4 \xi_j\right)^0 \left(\sum_{j=5}^6 \xi_j\right)^{n_{12}} \left(\sum_{j=7}^8 \xi_j\right)^{n_{34}} \left(\sum_{j=9}^{10} \xi_j\right)^{n_{13}} \left(\sum_{j=11}^{12} \xi_j\right)^{n_{24}} \left(\sum_{j=13}^{16} \xi_j\right)^{n_{1234}}. \tag{3.3}$$

Compared with (2.8), we know that (3.3) is a grouped Dirichlet distribution with  $t = 6$  partitions. Then (2.9) can be employed to derive the expectation of  $\xi_i$ ,  $i = 1, \dots, 16$ . For instance,

$$\begin{aligned} E(\xi_1) &= \frac{n_1 + \alpha_{11}}{n + \alpha_{..}}, & E(\xi_5) &= \frac{\alpha_{21}(\alpha_{21} + \alpha_{22} + n_{12})}{(\alpha_{21} + \alpha_{22})(n + \alpha_{..})}, \\ E(\xi_9) &= \frac{\alpha_{31}(\alpha_{31} + \alpha_{33} + n_{13})}{(\alpha_{31} + \alpha_{33})(n + \alpha_{..})}, & E(\xi_{13}) &= \frac{\alpha_{41}(\alpha_{41} + n_{1234})}{\alpha_{4.}(n + \alpha_{..})}, \end{aligned} \tag{3.4}$$

where  $n = \sum_{i=1}^4 n_i + n_{12} + n_{34} + n_{13} + n_{24} + n_{1234}$ ,  $\alpha_{4.} = \sum_{j=1}^4 \alpha_{4j}$ , and  $\alpha_{..} = \sum_{i=1}^4 \sum_{j=1}^4 \alpha_{ij}$ . Therefore, the Bayes estimator for  $\theta_1$  is given by

$$\hat{\theta}_1 = E(\theta_1) = E(\xi_1) + E(\xi_5) + E(\xi_9) + E(\xi_{13}). \tag{3.5}$$

By analogy, the respective posterior means of  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  can also be obtained.

Table 4. Parameter structure of nonignorable missing mechanism.

Categories	$R_{12}$	$R_{1\bar{2}}$	$R_{\bar{1}2}$	$R_{\bar{1}\bar{2}}$	$R_{12}$	$R_{1\bar{2}}$	$R_{\bar{1}2}$	$R_{\bar{1}\bar{2}}$	$R_{1\bar{2}}$	$R_{\bar{1}\bar{2}}$	Prob.
$A_1$	$\pi_{11}$	0	0	0	$\pi_{21}$	0	$\pi_{31}$	0	$\pi_{41}$		$\theta_1$
$A_2$	0	$\pi_{12}$	0	0	$\pi_{22}$	0	0	$\pi_{32}$	$\pi_{42}$		$\theta_2$
$A_3$	0	0	$\pi_{13}$	0	0	$\pi_{23}$	$\pi_{33}$	0	$\pi_{43}$		$\theta_3$
$A_4$	0	0	0	$\pi_{14}$	0	$\pi_{24}$	0	$\pi_{34}$	$\pi_{44}$		$\theta_4$
Counts	$n_1$	$n_2$	$n_3$	$n_4$	$n_{12}$	$n_{34}$	$n_{13}$	$n_{24}$	$n_{1234}$		$n \setminus 1$

SOURCE: Kadane (1985). NOTE:  $R_{12}$ ,  $R_{1\bar{2}}$ ,  $R_{\bar{1}2}$ , and  $R_{\bar{1}\bar{2}}$  denote the responding sets.

How do we determine the values of all  $\alpha_{ij}$  in the prior density? In practice, what we know about is the joint prior of the original parameters  $\{\theta_j\}$  and  $\{\lambda_{ij}\}$ , rather than  $\pi$  specified by  $f(\theta, \lambda_1, \dots, \lambda_4)$ , where  $\theta = (\theta_1, \dots, \theta_4)^\top$  and  $\lambda_j = (\lambda_{1j}, \dots, \lambda_{4j})^\top$  for  $j = 1, \dots, 4$ , as defined in (3.1). We would like to clarify the relation between  $f(\theta, \lambda_1, \dots, \lambda_4)$  and  $f(\pi)$ . Consider the more general case for (3.1) with  $i = 1, \dots, k$  and  $j = 1, \dots, m$ . The Jacobian of the transformation (3.1) is  $\prod_{j=1}^m \theta_j^{k-1}$ . Paulino and Pereira (1992) showed that  $\pi \sim D(\{\alpha_{ij}\})$  is

equivalent to saying that

$$\begin{cases} \theta = (\theta_1, \dots, \theta_m)^\top \sim D(\alpha_{\cdot 1}, \alpha_{\cdot 2}, \dots, \alpha_{\cdot m}), \\ \lambda_j = (\lambda_{1j}, \dots, \lambda_{kj})^\top \sim D(\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{kj}), \\ \theta, \lambda_1, \dots, \lambda_m \quad \text{are mutually independent,} \end{cases} \quad (3.6)$$

where  $\theta_{\cdot} = 1$  and  $\lambda_{\cdot j} = 1$  for  $j = 1, \dots, m$ . In this way, all  $\alpha_{ij}$  in prior  $f(\pi)$  can be determined.

### 3.2.2. An ignorable missing mechanism

An ignorable missing mechanism implies that the elements in each column of the array  $(\lambda_{ij})$  are equal. Removing these  $\{\lambda_{ij}\}$  from the likelihood function, (3.2) is reduced to  $\prod_{j=1}^4 \theta_j^{n_j} \cdot (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}} (\theta_1 + \theta_3)^{n_{13}} (\theta_2 + \theta_4)^{n_{24}}$ . Now,  $D(\alpha_1, \dots, \alpha_4)$  is a natural prior for  $\theta = (\theta_1, \dots, \theta_4)^\top$ . The resulting posterior is a generalized Dirichlet distribution with kernel

$$gD(\theta|a, b, \Gamma) = \prod_{j=1}^4 \theta_j^{n_j + \alpha_j - 1} \cdot (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}} (\theta_1 + \theta_3)^{n_{13}} (\theta_2 + \theta_4)^{n_{24}}, \quad (3.7)$$

where  $a = (n_1 + \alpha_1, \dots, n_4 + \alpha_4)^\top$ ,  $b = (n_{12} + 1, n_{34} + 1, n_{13} + 1, n_{24} + 1)^\top$  and  $\Gamma = (\gamma_{ij})$  with first row  $(1, 0, 1, 0)$ , and so on. Consequently, the posterior moments for  $\theta_j$ ,  $j = 1, \dots, 4$ , can be obtained by (2.2), (2.3) and (2.12) with proposal density

$$h(\theta) = c_h^{-1} \cdot \prod_{j=1}^4 \theta_j^{n_j + \alpha_j - 1} \cdot (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}}. \quad (3.8)$$

The proposal density  $h(\theta)$  is a grouped Dirichlet distribution with normalizing constant  $c_h = B(n_1 + \alpha_1, n_2 + \alpha_2) \cdot B(n_1 + n_2 + n_{12} + \alpha_1 + \alpha_2, n_3 + n_4 + n_{34} + \alpha_3 + \alpha_4) \cdot B(n_3 + \alpha_3, n_4 + \alpha_4)$ , see (A.2).

The other parameter of interest is the odds ratio (Kadane (1985)), denoted by  $\psi = \theta_1 \theta_4 / (\theta_2 \theta_3)$ , one of the ways to measure association in a contingency table. Noting that  $\psi$  greater than 1 implies victimization is chronic, the mean and variance of  $\psi$  are of interest. In fact, both  $E(\psi)$  and  $E(\psi^2)$  are given by (2.3) with, respectively,  $r = (1, -1, -1, 1)$  and  $r = (2, -2, -2, 2)$ . Therefore a grouped Dirichlet proposal density  $h(\cdot)$  facilitates the computation.

### 3.3. Death penalty attitude data

Consider Kadane's data from two sample surveys of juror's attitudes on a death penalty (Kadane (1983)), in which respondents are classified into four categories:  $A_1$  — would not decide guilt versus innocence in a fair and impartial

manner;  $A_2$  — fair and impartial on guilt versus innocence and, when sentencing, would always vote for the death penalty regardless of circumstance;  $A_3$  — fair and impartial on guilt and, when sentencing, would never vote for the death penalty;  $A_4$  — fair and impartial on guilt and, when sentencing, would sometimes and sometimes not vote for the death penalty. The frequency data  $y_1 = 68$ ,  $y_3 = 97$  and  $y_{24} = 674$  were obtained by a survey of the Field Research Corporation;  $y_2 = 15$  and  $y_{134} = 1484$  by the Harris Survey Company.

Under the assumption of a nonignorable missing mechanism, the combination data of the two-count sets are exhibited in Table 5 which bears some analogy to Table 4. Especially, for the MAR case, the combined likelihood is  $(\prod_{j=1}^4 \theta_j^{n_j})(\theta_2 + \theta_4)^{n_{24}}(\theta_1 + \theta_3 + \theta_4)^{n_{134}}$ . The Dirichlet prior  $D(\alpha_1, \dots, \alpha_4)$  is adequate for  $\theta = (\theta_1, \dots, \theta_4)^\top$ , which leads to a posterior of a generalized Dirichlet distribution with kernel  $gD(\theta|a, b, \Gamma) = (\prod_{j=1}^4 \theta_j^{n_j + \alpha_j - 1})\theta_1^0(\theta_1 + \theta_3)^0(\theta_1 + \theta_3 + \theta_4)^{n_{134}} \cdot (\theta_2 + \theta_4)^{n_{24}}$ . Accordingly, the posterior moments of  $\theta_j$ ,  $j = 1, \dots, 4$ , can be obtained by (2.2), (2.3) and (2.12) with proposal density  $h(\theta) = c_h^{-1} \cdot (\prod_{j=1}^4 \theta_j^{n_j + \alpha_j - 1})\theta_1^0(\theta_1 + \theta_3)^0(\theta_1 + \theta_3 + \theta_4)^{n_{134}}$ , a nested Dirichlet with normalizing constant  $c_h = B(a_1, a_2)B(a_1 + a_2, a_3)B(a_1 + a_2 + a_3 + n_{134}, a_4)$ , where  $a_j = n_j + \alpha_j$  for  $j = 1, \dots, 4$ , see (A.4).

Table 5. Combination data for death penalty attitudes.

Categories	$R_1$	$R_2$	$R_3$	$R_4$	$R_{24}$	$R_{134}$	Prob.
$A_1$	$\pi_{11}$	0	0	0	0	$\pi_{31}$	$\theta_1$
$A_2$	0	$\pi_{12}$	0	0	$\pi_{22}$	0	$\theta_2$
$A_3$	0	0	$\pi_{13}$	0	0	$\pi_{33}$	$\theta_3$
$A_4$	0	0	0	$\pi_{14}$	$\pi_{24}$	$\pi_{34}$	$\theta_4$
Counts	68 ( $n_1$ )	15 ( $n_2$ )	97 ( $n_3$ )	0 ( $n_4$ )	674 ( $n_{24}$ )	1484 ( $n_{134}$ )	2338 ( $n$ )\1

SOURCE: Kadane (1983). NOTE:  $R_1$ – $R_4$ ,  $R_{24}$ , and  $R_{134}$  denote the responding sets.

### 3.4. Misclassified multinomial data

In this section, we demonstrate the potential of our approach for the Bayesian analysis of cell probabilities in categorical data with misclassifications. Geng and Asano (1989) considered a contingency table with binary error-free variables  $A$  and  $B$ , and denoted the corresponding error-prone variables as  $a$  and  $b$ , respectively. The observed counts of the main sample categorized imprecisely and a subsample categorized both imprecisely and precisely, and the corresponding cell probabilities, are shown in Table 6. The objective is to find the posterior means of cell probabilities of a contingency table categorized by error-free variables, i.e.,  $\Pr(A = 1, B = 1) = \theta_1 + \theta_5 + \theta_9 + \theta_{13}$ ,  $\Pr(A = 2, B = 1) = \theta_2 + \theta_6 + \theta_{10} + \theta_{14}$ ,  $\Pr(A = 1, B = 2) = \theta_3 + \theta_7 + \theta_{11} + \theta_{15}$ , and  $\Pr(A = 2, B = 2) = \theta_4 + \theta_8 + \theta_{12} + \theta_{16}$ .

Table 6. Counts and probabilities for main sample And subsample.

Main Sample				Subsample							
		$a = 1$		$a = 2$				$a = 1$		$a = 2$	
		$A=1$	$A=2$	$A=1$	$A=2$			$A=1$	$A=2$	$A=1$	$A=2$
$b=1$	$B=1$	$m_1$		$m_2$		$b=1$	$B=1$	$n_1(\theta_1)$	$n_2(\theta_2)$	$n_5(\theta_5)$	$n_6(\theta_6)$
	$B=2$						$n_3(\theta_3)$	$n_4(\theta_4)$	$n_7(\theta_7)$	$n_8(\theta_8)$	
$b=2$	$B=1$	$m_3$		$m_4$		$b=2$	$B=1$	$n_9(\theta_9)$	$n_{10}(\theta_{10})$	$n_{13}(\theta_{13})$	$n_{14}(\theta_{14})$
	$B=2$						$n_{11}(\theta_{11})$	$n_{12}(\theta_{12})$	$n_{15}(\theta_{15})$	$n_{16}(\theta_{16})$	

Under the assumption of MAR, we take a Dirichlet prior  $D(\alpha_1, \dots, \alpha_{16})$ , then the posterior of  $\theta = (\theta_1, \dots, \theta_{16})^\top$  is proportional to  $\prod_{i=1}^{16} \theta_i^{a_i-1} (\sum_{j=1}^4 \theta_j)^{m_1} \times (\sum_{j=5}^8 \theta_j)^{m_2} (\sum_{j=9}^{12} \theta_j)^{m_3} (\sum_{j=13}^{16} \theta_j)^{m_4}$  with  $a_i = n_i + \alpha_i$  for  $i = 1, \dots, 16$ , a grouped Dirichlet distribution. Using (2.9), we have

$$\begin{aligned}
 E(\theta_i) &= \frac{a_i}{a. + m.} \left( 1 + \frac{m_1}{\sum_{j=1}^4 a_j} \right), \quad i = 1, \dots, 4, \\
 E(\theta_i) &= \frac{a_i}{a. + m.} \left( 1 + \frac{m_2}{\sum_{j=5}^8 a_j} \right), \quad i = 5, \dots, 8, \\
 E(\theta_i) &= \frac{a_i}{a. + m.} \left( 1 + \frac{m_3}{\sum_{j=9}^{12} a_j} \right), \quad i = 9, \dots, 12, \\
 E(\theta_i) &= \frac{a_i}{a. + m.} \left( 1 + \frac{m_4}{\sum_{j=13}^{16} a_j} \right), \quad i = 13, \dots, 16,
 \end{aligned}$$

where  $a. = n. + \alpha. = \sum_{i=1}^{16} (n_i + \alpha_i)$  and  $m. = \sum_{i=1}^4 m_i$ .

Geng and Asano also considered the following case. Let  $A$  and  $B$  be two error-free binary variables. Suppose there is an error-prone variable  $b$  for  $B$ . Assume that observations in a main sample are categorized by  $A$  and  $b$ . To obtain information on misclassifications of the error-prone variable  $b$ , we observe, from the same population, a random supplemental sample which is categorized by  $B$  and  $b$ . The observations can be represented as in Table 7. The goal is to find the posterior means of cell probabilities  $\Pr(A = 1, B = 1) = \theta_1 + \theta_5$ ,  $\Pr(A = 2, B = 1) = \theta_3 + \theta_7$ ,  $\Pr(A = 1, B = 2) = \theta_2 + \theta_6$ , and  $\Pr(A = 2, B = 2) = \theta_4 + \theta_8$ .

Under the assumptions of MAR, we take  $D(\alpha_1, \dots, \alpha_8)$  as the prior. Then the posterior for  $\theta = (\theta_1, \dots, \theta_8)^\top$  is proportional to  $\prod_{j=1}^8 \theta_j^{\alpha_j-1} (\theta_1 + \theta_2)^{m_{12}} (\theta_3 + \theta_4)^{m_{34}} (\theta_5 + \theta_6)^{m_{56}} (\theta_7 + \theta_8)^{m_{78}} \cdot (\theta_1 + \theta_3)^{n_{13}} (\theta_2 + \theta_4)^{n_{24}} (\theta_5 + \theta_7)^{n_{57}} (\theta_6 + \theta_8)^{n_{68}}$ , a generalized Dirichlet distribution. The means  $E(\prod_{j=1}^8 \theta_j^{r_j})$  can be calculated by importance sampling (see, (2.2), (2.3), and (2.12)) with proposal density  $h(\theta) = c_h^{-1} \cdot \prod_{j=1}^8 \theta_j^{r_j + \alpha_j - 1} (\theta_1 + \theta_2)^{m_{12}} (\theta_3 + \theta_4)^{m_{34}} (\theta_5 + \theta_6)^{m_{56}} (\theta_7 + \theta_8)^{m_{78}}$ , a grouped Dirichlet with normalizing constant  $c_h = B(a_1, a_2)B(a_3, a_4)B(a_5, a_6)B(a_7, a_8) \times$

$B(a_1 + a_2 + m_{12}, a_3 + a_4 + m_{34}, a_5 + a_6 + m_{56}, a_7 + a_8 + m_{78})$ , where  $a_j = r_j + \alpha_j$  for  $j = 1, \dots, 8$ , see (A.3).

Table 7. Observations for main and supplemental samples.

Main Sample				Supplemental Sample				
		A = 1	A = 2			A = 1	A = 2	
b = 1	B = 1	m <sub>12</sub>	m <sub>34</sub>	b = 1	B = 1	(θ <sub>1</sub> )	n <sub>13</sub>	(θ <sub>3</sub> )
	B = 2				(θ <sub>2</sub> )	n <sub>24</sub>	(θ <sub>4</sub> )	
b = 2	B = 1	m <sub>56</sub>	m <sub>78</sub>	b = 2	B = 1	(θ <sub>5</sub> )	n <sub>57</sub>	(θ <sub>7</sub> )
	B = 2				(θ <sub>6</sub> )	n <sub>68</sub>	(θ <sub>8</sub> )	

#### 4. An Illustrative Example

In this section the crime survey data, Table 3, is used to demonstrate the proposed methods. The goal is to obtain Bayes estimates of  $\theta_i$ ,  $i = 1, \dots, 4$ . We first consider the situation of a nonignorable missing mechanism. Equations (3.4) and (3.5) give the Bayes estimator of  $\theta_1$ . Similarly, we have  $\hat{\theta}_2 = E(\theta_2) = E(\xi_2) + E(\xi_6) + E(\xi_{11}) + E(\xi_{14})$ ,  $\hat{\theta}_3 = E(\theta_3) = E(\xi_3) + E(\xi_7) + E(\xi_{10}) + E(\xi_{15})$  and  $\hat{\theta}_4 = E(\theta_4) = E(\xi_4) + E(\xi_8) + E(\xi_{12}) + E(\xi_{16})$ , where  $E(\xi_i) = (n_i + \alpha_{1i}) / (n + \alpha_{..})$ ,  $i = 1, 2, 3, 4$ ,  $E(\xi_5) = \alpha_{21}d_{12}$ ,  $E(\xi_6) = \alpha_{22}d_{12}$ ,  $E(\xi_7) = \alpha_{23}d_{34}$ ,  $E(\xi_8) = \alpha_{24}d_{34}$ ,  $E(\xi_9) = \alpha_{31}d_{13}$ ,  $E(\xi_{10}) = \alpha_{33}d_{13}$ ,  $E(\xi_{11}) = \alpha_{32}d_{24}$ ,  $E(\xi_{12}) = \alpha_{34}d_{24}$ ,  $E(\xi_{13}) = \alpha_{41}d_{1234}$ ,  $E(\xi_{14}) = \alpha_{42}d_{1234}$ ,  $E(\xi_{15}) = \alpha_{43}d_{1234}$ ,  $E(\xi_{16}) = \alpha_{44}d_{1234}$ , and

$$d_{12} = \frac{\alpha_{21} + \alpha_{22} + n_{12}}{(\alpha_{21} + \alpha_{22})(n + \alpha_{..})}, \quad d_{34} = \frac{\alpha_{23} + \alpha_{24} + n_{34}}{(\alpha_{23} + \alpha_{24})(n + \alpha_{..})},$$

$$d_{13} = \frac{\alpha_{31} + \alpha_{33} + n_{13}}{(\alpha_{31} + \alpha_{33})(n + \alpha_{..})}, \quad d_{24} = \frac{\alpha_{32} + \alpha_{34} + n_{24}}{(\alpha_{32} + \alpha_{34})(n + \alpha_{..})}, \quad d_{1234} = \frac{\alpha_{4.} + n_{1234}}{\alpha_{4.}(n + \alpha_{..})}.$$

The corresponding variance and *standard deviation* (SD) of  $\theta_i$  can be obtained by calculating the variance of  $\xi_i$  and the covariance of  $\xi_i$  and  $\xi_j$  with (2.9).

We consider two prior distributions. The first is a uniform prior with  $\alpha_{ij} = 1$  for all  $i, j = 1, \dots, 4$ . From (3.6), this is equivalent to saying  $\theta = (\theta_1, \dots, \theta_4)^T \sim D(4, 4, 4, 4)$ ,  $\lambda_j \sim D(1, 1, 1, 1)$ , for  $j = 1, \dots, 4$ , and  $\theta, \lambda_1, \dots, \lambda_4$  are mutually independent. The second prior represents the opinion of experts taken as  $\theta \sim D(10, 5, 5, 10)$ ,  $\lambda_1 \sim D(1, 3, 2, 4)$ ,  $\lambda_2 \sim D(1, 0.5, 2, 1.5)$ ,  $\lambda_3 \sim D(1.5, 2, 0.5, 1)$ ,  $\lambda_4 \sim D(4, 2, 3, 1)$ , and they are independent. Table 8 summarizes results that indicate that the posterior means are slightly sensitive to the choice of the prior.

Table 8. Posterior mean and SD under nonignorable missing mechanism.

Priors	$E(\theta_1)$	$E(\theta_2)$	$E(\theta_3)$	$E(\theta_4)$
Uniform	0.5916	0.1396	0.1668	0.1020
	(0.0125)	(0.0137)	(0.0180)	(0.0102)
Experts	0.6570	0.1152	0.1362	0.0916
	(0.0134)	(0.0142)	(0.0156)	(0.0187)

Now we consider the situation of an ignorable missing mechanism. We want to calculate the posterior mean and SD of  $\theta_i$  and the odds ratio  $\psi = \theta_1\theta_4/(\theta_2\theta_3)$ . The prior for  $\theta = (\theta_1, \dots, \theta_4)^\top$  is specified by  $D(\alpha)$ , where  $\alpha = (\alpha_1, \dots, \alpha_4)^\top$ . Six prior distributions are discussed by Kadane (1985). They are (i) a uniform prior with  $\alpha = (1, 1, 1, 1)^\top$ ; (ii) a Haldane prior with  $\alpha = (0, 0, 0, 0)^\top$ ; (iii) a Jeffreys prior with  $\alpha = (0.5, 0.5, 0.5, 0.5)^\top$ ; (iv) Kadane's information prior with  $\alpha = (7.5, 1, 1, 0.5)^\top$ ; (v) the prior corresponding the uniform prior in Table 8 with  $\alpha = (4, 4, 4, 4)^\top$ ; (vi) the experts prior with  $\alpha = (10, 5, 5, 10)^\top$ . Table 9 displays outcomes which show that the posterior means are robust to the choice of the prior.

Table 9. Posterior mean and SD under ignorable missing mechanism.

Priors		$E(\theta_1)$	$E(\theta_2)$	$E(\theta_3)$	$E(\theta_4)$	$\psi$
Uniform	(1, 1, 1, 1)	0.6914 (0.0125)	0.0953 (0.0182)	0.1354 (0.0104)	0.0779 (0.0130)	4.4153 (1.2403)
Jeffreys	(0, 0, 0, 0)	0.6927 (0.0131)	0.0943 (0.0183)	0.1350 (0.0102)	0.0780 (0.0139)	4.3811 (1.2655)
Haldane	(0.5, 0.5, 0.5, 0.5)	0.6920 (0.0127)	0.0953 (0.0183)	0.1352 (0.0103)	0.0775 (0.0134)	4.4006 (1.2515)
Information	(7.5, 1, 1, 0.5)	0.6930 (0.0131)	0.0957 (0.0186)	0.1347 (0.0099)	0.0766 (0.0141)	4.3540 (1.2779)
Uniform*	(4, 4, 4, 4)	0.6880 (0.0149)	0.0956 (0.0180)	0.1370 (0.0129)	0.0794 (0.0157)	4.4186 (1.2297)
Experts	(10, 5, 5, 10)	0.6899 (0.0126)	0.0891 (0.0118)	0.1355 (0.0099)	0.0855 (0.0148)	4.9954 (0.7231)

## 5. Choice of Effective Proposal Density

We return to (2.12) and consider the approximation of the normalizing constant. In importance sampling, the usual difficulty is finding a suitable proposal density  $h(\cdot)$  which mimics the target function  $gD(\cdot|a, b, \Gamma)$ . A multivariate split normal/Student proposal density suggested by Geweke (1989) seems infeasible for the present situation since  $\theta$  belongs to the hyperplane  $T_n$ . In Section 2.5, we suggest three feasible choices for  $h(\cdot)$ . Two questions emerge: (i) what is a natural class of proposal densities? (ii) which member of the class is the most effective? In what follows, we partially answer these questions.

Clearly, the functionwise IBF (2.4) provides a natural class of proposal densities: the complete-data posterior densities  $\{f(\theta|Y_{\text{obs}}, z_0) : z_0 \in \mathcal{S}_{(z|Y_{\text{obs}})}\}$ , where  $\mathcal{S}_{(z|Y_{\text{obs}})}$  denotes the conditional support of  $z$ . However, the efficiency for approximating the normalizing constant  $c = \int f(\theta|Y_{\text{obs}}, z_0)/f(z_0|Y_{\text{obs}}, \theta) d\theta$  by importance sampling depends on how well the proposal density  $f(\theta|Y_{\text{obs}}, z_0)$  mimics the target function  $f(\theta|Y_{\text{obs}}, z_0)/f(z_0|Y_{\text{obs}}, \theta) = c \cdot f(\theta|Y_{\text{obs}})$ . Let  $\hat{\theta}_{\text{obs}}$  denote the mode of

the observed posterior  $f(\theta|Y_{\text{obs}})$ . The EM algorithm shows that  $f(\theta|Y_{\text{obs}}, z_0)$  and  $f(\theta|Y_{\text{obs}})$  share the same mode  $\hat{\theta}_{\text{obs}}$ , where

$$z_0 = E(z|Y_{\text{obs}}, \hat{\theta}_{\text{obs}}). \quad (5.1)$$

Thus, there is substantial amount of overlapping area under the proposal density and the target function. Then  $f(\theta|Y_{\text{obs}}, z_0)$ , with  $z_0$  given by (5.1), is heuristically an effective proposal density.

Now we use the crime survey data under the assumption of an ignorable missing mechanism to illustrate our idea. Return to Section 3.2.2 and denote the observed data by  $Y_{\text{obs}} = \{n_1, n_2, n_3, n_4, n_{12}, n_{34}, n_{13}, n_{24}\}$ . Note that the observed posterior density  $f(\theta|Y_{\text{obs}})$  is proportional to  $gD(\theta|a, b, \Gamma)$  given in (3.7). We introduce a latent vector  $z = (z_{13}, z_{24})^\top$  such that the complete-data posterior is

$$f(\theta|Y_{\text{obs}}, z) \propto \theta_1^{n_1+z_{13}+\alpha_1-1} \theta_2^{n_2+z_{24}+\alpha_2-1} \theta_3^{n_3+n_{13}-z_{13}+\alpha_3-1} \theta_4^{n_4+n_{24}-z_{24}+\alpha_4-1} \cdot (\theta_1 + \theta_2)^{n_{12}} (\theta_3 + \theta_4)^{n_{34}}, \quad (5.2)$$

and the conditional predictive density is given by

$$f(z|Y_{\text{obs}}, \theta) = \text{Binomial} \left( z_{13} \mid n_{13}, \frac{\theta_1}{\theta_1 + \theta_3} \right) \cdot \text{Binomial} \left( z_{24} \mid n_{24}, \frac{\theta_2}{\theta_2 + \theta_4} \right). \quad (5.3)$$

Based on (5.2) and (5.3), the EM algorithm can be used to find the posterior mode  $\hat{\theta}_{\text{obs}}$  and  $z_0 = E(z|Y_{\text{obs}}, \hat{\theta}_{\text{obs}})$ . Then an effective proposal density is  $f(\theta|Y_{\text{obs}}, z_0)$ . Comparing  $f(\theta|Y_{\text{obs}}, z_0)$  with (3.8), we know that both of them belong to the same class of proposal densities and they are very closed. Therefore the proposal density (3.8) is feasible but not the best and  $f(\theta|Y_{\text{obs}}, z_0)$  is the best at the expense of running an EM algorithm.

## 6. Discussion

In this paper, we study the Bayesian computations of the posterior moments of the unknown cell probabilities for the contingency table with incomplete cell-counts. For some special cases where the posterior is a grouped or a nested Dirichlet distribution, the posterior means of the unknown cell probabilities can be obtained in closed form by using inverse Bayes formulae and stochastic representation.

When closed-form expressions do not exist, we suggest using importance sampling to approximately compute the posterior quantities. Three feasible proposal densities are suggested and propose a procedure for choosing an effective proposal density. We have noted that  $\text{Var}(\theta|Y_{\text{obs}}, z_0) \leq \text{Var}(\theta|Y_{\text{obs}})$  contradicts with the common request in importance sampling that the tails of proposal density do not decay more quickly than the tails of the target function (Geweke

(1989)). Our procedure is not perfect, but it provides a universal way to find an effective proposal density for the situation where  $\text{Var}(\theta|Y_{\text{obs}}, z_0)$  is not much less than  $\text{Var}(\theta|Y_{\text{obs}})$ . Since no methods currently exist for assessing the efficiency of a proposal density and the accuracy of an importance sampling estimate (Gelman, Carlin, Stern and Rubin (1995), p.307), it is a problem worthy of further study.

### Acknowledgements

The research was partially supported by Chinese NSF grants CNSF-19820224 and CNSF-19831010, by a University of Hong Kong CRGC grant, by a U.S. National Cancer Center grant CA21765 and by the American, Lebanese, Syrian Associated Charities (ALSAC). Part of the research was carried out while the first author was visiting the Department of Statistics and Actuarial Science of the University of Hong Kong. We are grateful to an associate editor and two referees for many helpful comments and suggestions on an earlier version of this article.

### Appendix

#### A.1. Derivation of (2.6)

Let  $\theta \sim \text{GD}_{n,2}(a, b)$  with density given by (2.5). The transformation  $\phi_i = \theta_i / \sum_{j=1}^s \theta_j$ ,  $i = 1, \dots, s-1$ ,  $\phi_s = \sum_{j=1}^s \theta_j$ ,  $\phi_i = \theta_i / \sum_{j=s+1}^n \theta_j$ ,  $i = s+1, \dots, n-1$ , has an inverse transformation given by (2.6). Noting that the Jacobian  $|J| = \phi_s^{s-1}(1 - \phi_s)^{n-s-1}$ , the joint density  $f(\phi_1, \dots, \phi_{n-1})$  is

$$c_1^{-1} \cdot \prod_{i=1}^{s-1} \phi_i^{a_i-1} \left(1 - \sum_{j=1}^{s-1} \phi_j\right)^{a_s-1} \cdot \phi_s^{a_1^*+b_1-2} (1 - \phi_s)^{a_2^*+b_2-2} \cdot \prod_{i=s+1}^{n-1} \phi_i^{a_i-1} \left(1 - \sum_{j=s+1}^{n-1} \phi_j\right)^{a_n-1}, \quad (\text{A.1})$$

where  $a_1^* = \sum_{j=1}^s a_j$  and  $a_2^* = \sum_{j=s+1}^n a_j$ . Therefore  $(\phi_1, \dots, \phi_{s-1})^\top$ ,  $\phi_s$  and  $(\phi_{s+1}, \dots, \phi_{n-1})^\top$  are independent Dirichlet distributions, and (2.6) follows. From (A.1), we obtain

$$c_1 = B(a_1, \dots, a_s) \cdot B\left(\sum_{j=1}^s a_j + b_1 - 1, \sum_{j=s+1}^n a_j + b_2 - 1\right) \cdot B(a_{s+1}, \dots, a_n). \quad (\text{A.2})$$

#### A.2. Derivation of (2.9)

Let  $\theta \sim \text{GD}_{n,t}(a, b)$  with density given by (2.8). Making the transformation

$$\begin{cases} \phi_i = \theta_i / (\theta_1 + \dots + \theta_{s_1}), & i = 1, \dots, s_1 - 1, & \phi_{s_1} = \theta_1 + \dots + \theta_{s_1}, \\ \phi_i = \theta_i / (\theta_{s_1+1} + \dots + \theta_{s_2}), & i = s_1 + 1, \dots, s_2 - 1, & \phi_{s_2} = \theta_{s_1+1} + \dots + \theta_{s_2}, \\ \vdots & & \\ \phi_i = \theta_i / (\theta_{s_{t-1}+1} + \dots + \theta_{s_t}), & i = s_{t-1} + 1, \dots, s_t - 1, & \phi_{s_t} = \theta_{s_{t-1}+1} + \dots + \theta_{s_t}, \end{cases}$$

the inverse transformation is given by (2.9) and the Jacobian is  $|J| = \prod_{j=1}^{t-1} \phi_{s_j}^{s_j - s_{j-1} - 1} (1 - \sum_{k=1}^{t-1} \phi_{s_k})^{s_t - s_{t-1} - 1}$ . Partition  $\phi = (\phi_1, \dots, \phi_n)^\top$  into  $(\phi_1^*, \phi_{s_1}, \phi_2^*, \phi_{s_2}, \dots, \phi_t^*, \phi_{s_t})^\top$ , where  $\phi_j^* = (\phi_{s_{j-1}+1}, \dots, \phi_{s_j-1})$ ,  $j = 1, \dots, t$ . We know  $f(\phi_1, \dots, \phi_{n-1}) = f(\theta_{-n}) * |J|$ , which leads to

$$\phi_j^{*\top} \sim D(a_{s_{j-1}+1}, \dots, a_{s_j-1}; a_{s_j}), \quad j = 1, \dots, t,$$

$$(\phi_{s_1}, \phi_{s_2}, \dots, \phi_{s_t})^\top \sim D\left(\sum_{k=1}^{s_1} a_k + b_1 - 1, \sum_{k=s_1+1}^{s_2} a_k + b_2 - 1, \dots, \sum_{k=s_{t-1}+1}^{s_t} a_k + b_t - 1\right),$$

and they are independent. Thus (2.9) follows. Similar to (A.2), we obtain the normalizing constant

$$c_2 = \prod_{j=1}^t B(a_{s_{j-1}+1}, \dots, a_{s_j}) \cdot B\left(\sum_{k=1}^{s_1} a_k + b_1 - 1, \dots, \sum_{k=s_{t-1}+1}^{s_t} a_k + b_t - 1\right).$$

### A.3. Derivation of (2.11)

Let  $\theta \sim \text{ND}_{n,n-1}(a, b)$  with density given by (2.10). Making the transformation  $\phi_i = \sum_{j=1}^i \theta_j / \sum_{j=1}^{i+1} \theta_j$ ,  $i = 1, \dots, n-2$ , and  $\phi_{n-1} = \sum_{j=1}^{n-1} \theta_j$ , the inverse transformation is given by (2.11) and the Jacobian is  $|J| = \prod_{j=1}^{n-1} \phi_j^{j-1}$ . Hence, the joint density  $f(\phi_1, \dots, \phi_{n-1}) = c_3^{-1} \cdot \prod_{j=1}^{n-1} \phi_j^{d_j-1} (1-\phi_j)^{a_{j+1}-1}$ , which indicates that  $\phi_j \sim \text{Beta}(d_j, a_{j+1})$  for  $j = 1, \dots, n-1$ , and  $\phi_1, \dots, \phi_{n-1}$  are independent, where  $d_j = \sum_{k=1}^j (a_k + b_k - 1)$ . Thus (2.11) follows. Similarly,

$$c_3 = \prod_{j=1}^{n-1} B\left(\sum_{k=1}^j (a_k + b_k - 1), a_{j+1}\right).$$

## References

- Albert, J. H. and Gupta, A. K. (1985). Bayesian methods for binomial data with applications to a nonresponse problem. *J. Amer. Statist. Assoc.* **80**, 167-174.
- Basu, D. and Pereira, C. A. de B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *J. Statist. Plann. Inference* **6**, 345-362.
- Carlson, B. C. (1977). *Special Functions of Applied Mathematics*. Academic Press, New York.
- Chiu, H. Y. and Sedransk, J. (1986). A Bayesian procedure for imputing missing values in sample surveys. *J. Amer. Statist. Assoc.* **81**, 667-676.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.
- Dickey, J. M., Jiang, J. M. and Kadane, J. B. (1987). Bayesian methods for censored categorical data. *J. Amer. Statist. Assoc.* **82**, 773-781.
- Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

- Fang, K. T., Wang, Y. and Bentler, P. M. (1994). Some applications of number-theoretic methods in statistics. *Statist. Sci.* **9**, 416-428.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.
- Geng, Z. and Asano, C. (1989). Bayesian estimation methods for categorical data with misclassification. *Comm. Statist. Theory Methods* **18**, 2935-2954.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**, 1317-1339.
- Gunel, E. (1984). A Bayesian analysis of the multinomial model for a dichotomous response with nonrespondents. *Comm. Statist. Theory Methods* **13**, 737-751.
- Kadane, J. B. (1983). Juries hearing death penalty cases: statistical analysis of a legal procedure. *J. Amer. Statist. Assoc.* **78**, 544-552.
- Kadane, J. B. (1985). Is victimization chronic? a Bayesian analysis of multinomial missing data. *J. Econometrics* **29**, 47-67.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.
- Ng, K. W. (1995). Explicit formulas for unconditional pdf. Research Report, No. 82 (revised). Department of Statistics, University of Hong Kong.
- Ng, K. W. (1997). Inversion of Bayes formula: explicit formulae for unconditional pdf. In *Advance in the Theory and Practice in Statistics* (Edited by N. L. Johnson and N. Balakrishnan), 571-584. Wiley, New York.
- Park, T. and Brown, M. B. (1994). Models for categorical data with nonignorable nonresponse. *J. Amer. Statist. Assoc.* **89**, 44-52.
- Paulino, C. D. M. and Pereira, C. A. de B. (1992). Bayesian analysis of categorical data informatively censored. *Comm. Statist. Theory Methods* **21**, 2689-2705.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate data*. Chapman and Hall, London.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82**, 528-540.
- Tian, G. L. and Tan, M. (2001). Exact statistical solutions using the inverse Bayes formulae. Technical Report #2001-02, Dept. of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81**, 82-86.

Division of Biostatistics, University of Maryland Greenebaum Cancer Center, Baltimore, MD 21201, U.S.A.

E-mail: gtian2@umm.edu

Department of Statistics and Actuarial Science, the University of Hong Kong, P. R. China.

E-mail: kaing@hku.hk

Department of Probability and Statistics, Peking University, P. R. China.

E-mail: zgeng@math.pku.edu.cn

(Received November 2000; accepted July 2002)