The Annals of Statistics
1982, Vol. 10, No. 4, 1261–1268

MIXTURES OF DIRICHLET DISTRIBUTIONS AND ESTIMATION IN CONTINGENCY TABLES

BY JAMES H. ALBERT[1] AND ARJUN K. GUPTA

Bowling Green State University

Assuming a multinomial sampling model, prior distributions are developed which can accept prior information about symmetry and independence in a two-way contingency table. Bayesian estimates for the cell probabilities are obtained from the posterior distributions which are attractive alternatives to the usual classical estimates when vague prior information about symmetry or independence is available.

1. Introduction. Consider an $I \times J$ contingency table $\{X_{ij}, i = 1, \cdots, I, j = 1, \cdots, J\}$ where X_{ij} denotes the observed frequency in the (i, j) cell and let θ_{ij} denote the probability that an observation falls in that particular cell. Assume a sample of size N is taken from an infinite population and therefore $\{X_{ij}\}$ has a multinomial distribution with parameters N and $\boldsymbol{\theta} = (\theta_{11}, \cdots, \theta_{IJ})$.

Under this sampling model, consider the problem of estimating the vector of cell probabilities $\boldsymbol{\theta}$ using prior information about the cross-classification structure of the table. Specifically, say a user believes, before sampling, that there exists symmetry or independence in the table. That is, the cell probabilities are believed a priori to satisfy the relationships $\theta_{ij} = \theta_{ji}$ for $i < j$ (in the case of symmetry) or $\theta_{ij} = \theta_{i\cdot}\theta_{\cdot j}$, where $\theta_{i\cdot}$ and $\theta_{\cdot j}$ are marginal probabilities (in the case of independence). It is of interest to develop prior distributions for $\boldsymbol{\theta}$ which can reflect beliefs about symmetry and independence and lead to posterior estimates for $\boldsymbol{\theta}$.

The usual method of inputting prior information assumes that $\boldsymbol{\theta}$ possesses the conjugate Dirichlet distribution with density proportional to $\Pi_{i,j}\theta_{ij}^{K\eta_{ij}+x_{ij}-1}$, where $K > 0$, $\eta_{ij} > 0$, and $\sum_{i,j}\eta_{ij} = 1$. The posterior density of $\boldsymbol{\theta}$ is proportional to $\Pi_{i,j}\theta_{ij}^{K\eta_{ij}+x_{ij}-1}$, and a cell probability θ_{ij} can be estimated by its posterior mean, given by

$$E(\theta_{ij} \mid \mathbf{x}) = \frac{N}{N+K}\frac{x_{ij}}{N} + \frac{K}{N+K}\eta_{ij}. \tag{1.1}$$

The parameter η_{ij} is the prior mean of θ_{ij} and represents the user's prior guess at the probability. The parameter K is related to the prior variance of θ_{ij} and reflects the precision of the set of prior guesses $\{\eta_{ij}\}$. A natural way of enlarging the class of Dirichlet distributions is to consider a two-stage prior, introduced by Fienberg and Holland (1973) and Good (1967). The first stage of the prior assumes that $\boldsymbol{\theta}$ given $(K, \{\eta_{ij}\})$ is Dirichlet; the second stage gives K and $\boldsymbol{\eta} = (\eta_{11}, \cdots, \eta_{IJ})$ the prior density $\phi(K, \boldsymbol{\eta})$. The resulting prior distribution is a mixture of Dirichlets and the posterior mean of θ_{ij} is given by

$$E(\theta_{ij} \mid \mathbf{x}) = E\left(\frac{N}{N+K} \,\middle|\, \mathbf{x}\right)\frac{x_{ij}}{N} + E\left(\frac{K}{N+K}\eta_{ij} \,\middle|\, \mathbf{x}\right), \tag{1.2}$$

where the expectations on the right hand side of (1.2) are taken with respect to the posterior distribution of $(K, \boldsymbol{\eta})$. See Good (1976) and Crook and Good (1980) for discussions on the use of Dirichlet mixtures in Bayesian testing.

Received February 1981; revised April 1982.

[1] Research partially supported by a Faculty Research Fellowship at Bowling Green State University.

AMS 1979 *subject classification.* Primary 62F15; secondary 62C10.

Key words and phrases. Categorical data, symmetry, independence, mixture.

First, consider the use of the "natural" Dirichlet prior to reflect prior beliefs about symmetry or independence. (For simplicity, we will restrict discussion to the 2×2 table, where $I = J = 2$.) The user first will select the vector of prior means $\boldsymbol{\eta}$ which represents a guess at the vector of cell probabilities $\boldsymbol{\theta}$. If the user believes a priori in a particular cross-classification structure, then clearly $\boldsymbol{\eta}$ should satisfy the same structure. Thus, in the symmetry case, $\eta_{12} = \eta_{21}$, and in the independence case, $\eta_{11} = \eta_a \eta_b$, where $\eta_a = \eta_{11} + \eta_{12}$, $\eta_b = \eta_{11} + \eta_{21}$. These prior beliefs can be represented by the following configurations of prior means:

(1.3)

Symmetry

η_{11}	η_{12}
η_{12}	$1 - \eta_{11} - 2\eta_{12}$

Independence

$\eta_a \eta_b$	$\eta_a(1 - \eta_b)$
$(1 - \eta_a)\eta_b$	$(1 - \eta_a)(1 - \eta_b)$

Note that the dimension of the space of $\boldsymbol{\eta}$ is reduced to two, because of the restriction $\sum_{i,j} \eta_{ij} = 1$ and the additional assumption of symmetry or independence. If a user could specify values of η_{11}, η_{12} (in the symmetry case) or η_a, η_b (in the independence case), then $\boldsymbol{\eta}$ would be completely specified and the Dirichlet prior could be used. However, in this situation, the user believes solely in the symmetry (independence) configuration, and the actual values of the parameters η_{11} and η_{12} (η_a and η_b) are unknown.

The Dirichlet distribution appears unsuitable for use in situations where the user cannot obtain an exact value of the prior mean vector $\boldsymbol{\eta}$. But the Dirichlet can be used as the first stage of a two-stage prior which reflects both the belief in the configuration of prior means (1.3) and the ignorance of the independent parameters in this configuration. In the first stage, the user will select the configuration (1.3) and a value of the Dirichlet parameter K. The parameter K reflects the sureness of the configuration of prior means, and, in the two-stage prior, it will be shown to reflect the strength of the user's belief in symmetry or independence. The second stage of the prior assigns to the independent parameters (η_{11}, η_{12} or η_a, η_b) a noninformative uniform distribution. This stage reflects ignorance about the location of these independent parameters.

In Sections 2 and 3, we first discuss how these two-stage priors reflect the beliefs of symmetry and independence, and then give posterior mean estimates for the cell probabilities. In the symmetry case, the posterior means are shown to approach "symmetry" estimates as the value of the prior parameter K approaches infinity, or equivalently, as the user's prior belief in symmetry grows. In Section 3, a generalized version of the independence prior is developed which can accept additional prior information about the marginal cell probabilities in the table. In an example, the posterior means appear to shrink the classical estimates towards "independence" estimates, the degree of shrinkage depending on the value of the parameter K. These Bayesian estimates of the cell probabilities will be appropriate for use when prior information about symmetry or independence exists, but this information is not sufficiently strong to use symmetry or independence estimates. By using the Dirichlet parameter K, the user can state the precision of his knowledge about the cross-classification structure, and the resulting posterior estimates compromise between the usual estimates and the estimates assuming the particular cross-classification structure.

2. Estimation using prior knowledge of symmetry. Assume that a square contingency table is given ($I = J$) and it is of interest to estimate the cell probabilities using prior knowledge about symmetry. As discussed in Section 1, this prior belief can be represented by the following two-stage prior distribution.

Stage I: $\boldsymbol{\theta}$ is distributed Dirichlet (K, $\boldsymbol{\eta}$), where the elements in the prior mean vector $\boldsymbol{\eta}$ satisfy the relationships $\eta_{ij} = \eta_{ji}$ for $i < j$. Let $\boldsymbol{\eta}^* = (\eta_{11}, \eta_{22}, \cdots, \eta_{II}, \eta_{12}, \cdots, \eta_{1I}, \eta_{23}, \cdots, \eta_{I-1,I})$ denote the vector of distinct prior means.

Stage II: The vector $\boldsymbol{\eta}^*$ is given a uniform distribution over the set $A = \{\boldsymbol{\eta}^*: \eta_{ij} > 0, \sum_{i=1}^{I} \eta_{ii} + 2 \sum_{i<j} \eta_{ij} = 1\}$. This stage of the prior distribution expresses ignorance about the location of $\boldsymbol{\eta}^*$.

The resulting prior density for $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}) = \int_A \pi_1(\boldsymbol{\theta} \mid \boldsymbol{\eta}^*) \pi_2(\boldsymbol{\eta}^*) \, d\boldsymbol{\eta}^*, \tag{2.1}$$

where π_1 is the Dirichlet density given by

$$\pi_1(\boldsymbol{\theta} \mid \boldsymbol{\eta}) = \frac{\Gamma(K)}{\prod_{i=1}^I \Gamma(K\eta_{ii}) \prod_{i<j} \Gamma^2(K\eta_{ij})} \prod_{i=1}^I \theta_{ii}^{K\eta_{ii}-1} \prod_{i<j} (\theta_{ij}\theta_{ji})^{K\eta_{ij}-1},$$

and π_2 is the uniform density. Several facts can be shown about this prior. First, the correlation between two cell probabilities θ_{ij} and θ_{ji} (for $i \neq j$) is given by

$$\rho(\theta_{ij}, \theta_{ji}) = \frac{2K\{I(I+1)+2\}^{-1} - (K+1)\{I(I+1)\}^{-1}}{2K\{I(I+1)+2\}^{-1} - (K+1)\{I(I+1)\}^{-1} + 1}.$$

Note that $\rho(\theta_{ij}, \theta_{ji})$ is an increasing function of K and approaches 1 as K approaches infinity. A user specifying a value of the prior parameter K is specifying a value of $\rho(\theta_{ij}, \theta_{ji})$ and is indicating a degree of belief in symmetry in the table. Second, it can be shown that, as K approaches infinity, the density (2.1) approaches a uniform density on the "space of symmetry" $\{\boldsymbol{\theta}\colon \theta_{ij} \geq 0,\ \theta_{ij} = \theta_{ji},\ i<j,\ \sum_{i,j} \theta_{ij} = 1\}$. Thus, for a large selected value of K, the prior is placing most of its weight on symmetric values of $\boldsymbol{\theta}$ and is reflecting a strong prior belief in symmetry.

In practice, it will be difficult for a user to choose a value of K by guessing at a value of the correlation coefficient $\rho(\theta_{ij}, \theta_{ji})$. One simple interpretation of the precision parameter K of a Dirichlet distribution is that it represents the sample size of a "preliminary" contingency table which reflects one's prior beliefs about $\boldsymbol{\theta}$ (Good, 1965). Thus, if a user's prior belief is based on a previously observed contingency table where the cell frequencies approximately satisfy $x_{ij}/N = x_{ji}/N$, for $i \neq j$, then K could be chosen to be the sample size of this table. A value of the parameter K can also be chosen through specifying subjective probabilities. First, for K large, the marginal distribution of $(\theta_{ij}, \theta_{ji})$ is approximately bivariate normal with correlation coefficient ρ. Then, in this situation, Gokhale and Press (1980) discuss methods of indirectly assessing a value of ρ by assessing the probability of "concordance" $\tau(\rho)$. In our notation, if two pairs of independent observations $(\theta_{ij1}, \theta_{ji1})$, $(\theta_{ij2}, \theta_{ji2})$ are drawn from a bivariate normal density, the probability of "concordance" is given by

$$\tau(\rho) = P(\theta_{ij2} > \theta_{ij1} \mid \theta_{ji2} > \theta_{ji1}) \cong \frac{1}{2} + \frac{1}{\pi} \arcsin \rho.$$

A value of K can then be indirectly chosen by specifying the probability $\tau(\rho)$.

To estimate a cell probability θ_{ij}, we will use its posterior mean $E(\theta_{ij} \mid x)$. The following theorem gives a general expression for the posterior mean, and, additionally, evaluates this expression in the limiting cases $K = 0$ and $K = \infty$. These two cases correspond, respectively, to a very weak belief and a very strong belief in symmetry in the table.

THEOREM 1. *Let the prior distribution for* $\boldsymbol{\theta}$ be given by (2.1). Then

$$E(\theta_{ij} \mid x) = \frac{N}{N+K} \frac{x_{ij}}{N} + \frac{K}{N+K} E(\eta_{ij} \mid \mathbf{x}), \tag{2.2}$$

where the expectation $E(\eta_{ij} \mid \mathbf{x})$ *is taken with respect to the posterior distribution of* $\boldsymbol{\eta}^*$, *with density given by*

$$\pi_3(\boldsymbol{\eta}^* \mid \mathbf{x}) = C_1 \prod_{i=1}^I (K\eta_{ii})^{(x_{ii})} \cdot \prod_{i<j} \{(K\eta_{ij})^{(x_{ij})} (K\eta_{ij})^{(x_{ji})}\}, \quad \boldsymbol{\eta}^* \in A,$$

C_1 *being the proportionality constant and* $a^{(r)} = a(a+1) \cdots (a+r-1)$, $a^{(0)} = 1$. *In the limiting situations where K approaches zero and infinity,* $E(\theta_{ij} \mid \mathbf{x})$ *approaches, respectively,*

$$\lim_{K \to 0} E(\theta_{ij} \mid \mathbf{x}) = x_{ij}/N, \qquad \text{for } i, j = 1, \cdots, I \tag{2.3}$$

and

$$\lim_{K\to\infty} E(\theta_{ij} \mid \mathbf{x}) = \begin{cases} \dfrac{x_{ij}+1}{N+I(I+1)/2} & \text{for } i=j=1, \cdots, I \quad (2.4) \\[2ex] \dfrac{1}{2}\dfrac{x_{ij}+x_{ji}+1}{N+I(I+1)/2} & \text{for } i \neq j. \quad (2.5) \end{cases}$$

PROOF. See Albert and Gupta (1981).

Note that the posterior mean of θ_{ij} in the limiting case $K = 0$ is simply the maximum likelihood estimator under an unrestricted model. In addition, note that in the case $K = \infty$ (a strong prior belief of symmetry), the posterior mean of an off-diagonal probability is a type of "symmetry" estimator which pools the counts in cells (i, j) and (j, i) to estimate θ_{ij}. Thus in the usual situation where a finite positive value of K is used, it is expected that the posterior mean (2.2) will shrink the MLE x_{ij}/N towards a symmetry estimator, the degree of shrinkage depending on the value of K.

3. Estimation using prior knowledge of independence. Consider the related problem of estimating the vector of cell probabilities using prior information about independence in the table. That is, the cell probabilities are believed to satisfy the relationships $\theta_{ij} = \theta_{i.}\theta_{.j}$ for all i, j, where $\theta_{i.} = \sum_{j=1}^{J} \theta_{ij}$, $i = 1, \cdots, I$ and $\theta_{.j} = \sum_{i=1}^{I} \theta_{ij}$, $j = 1, \cdots, J$ are the marginal probabilities of the two variables in the contingency table. In addition to the belief about independence, the user may possess prior beliefs about the location of the two sets of marginal probabilities. Beliefs about independence and the marginal probabilities can be reflected by means of the following two-stage prior distribution.

Stage I: $\boldsymbol{\theta}$ is distributed Dirichlet $(K, \boldsymbol{\eta})$, where $\eta_{ij} = \eta_{ai}\eta_{bj}$, $i = 1, \cdots, I$, $j = 1, \cdots, J$, $\eta_{ai} \geq 0$, $\eta_{bj} \geq 0$ for all i, j and $\sum_{i=1}^{I} \eta_{ai} = \sum_{j=1}^{J} \eta_{bj} = 1$.

Stage II: The vectors $\boldsymbol{\eta}_a = (\eta_{a1}, \cdots, \eta_{aI})$ and $\boldsymbol{\eta}_b = (\eta_{b1}, \cdots, \eta_{bJ})$ are independent, $\boldsymbol{\eta}_a$ possesses a Dirichlet $(L_a, \boldsymbol{\lambda}_a)$ distribution and $\boldsymbol{\eta}_b$ possesses a Dirichlet $(L_b, \boldsymbol{\lambda}_b)$ distribution, where $\boldsymbol{\lambda}_a = (\lambda_{a1}, \cdots, \lambda_{aI})$, $\boldsymbol{\lambda}_b = (\lambda_{b1}, \cdots, \lambda_{bJ})$. The prior density for $\boldsymbol{\theta}$ is then given by

$$\pi(\boldsymbol{\theta}) = \int_B \psi(\boldsymbol{\theta} \mid \boldsymbol{\eta}^+, K)\psi(\boldsymbol{\eta}_a \mid \boldsymbol{\lambda}_a, L_a)\psi(\boldsymbol{\eta}_b \mid \boldsymbol{\lambda}_b, L_b)\, d\boldsymbol{\eta}_a\, d\boldsymbol{\eta}_b \tag{3.1}$$

where

$$B = \{(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b): \eta_{ai} \geq 0, \eta_{bj} \geq 0, i = 1, \cdots, I, j = 1, \cdots, J, \textstyle\sum_{i=1}^{I} \eta_{ai} = \sum_{j=1}^{J} \eta_{bj} = 1\},$$

$\boldsymbol{\eta}^+ = (\eta_{a1}\eta_{b1}, \cdots, \eta_{aI}\eta_{bJ})$, and

$$\psi(\boldsymbol{\theta} \mid \boldsymbol{\eta}, K) = \frac{\Gamma(K)}{\Pi_{i,j}\,\Gamma(K\eta_{ij})} \Pi_{i,j}\, \theta_{ij}^{K\eta_{ij}-1}. \tag{3.2}$$

This prior distribution implies the following knowledge about the marginal probabilities and the interaction in the table.

(i) The means and variances of the marginal probabilities $\theta_{i.}$ and $\theta_{.j}$ are given by

$$E(\theta_{i.}) = \lambda_{ai}, \ \mathrm{Var}(\theta_{i.}) = \frac{(K + L_a + 1)}{(K + 1)(L_a + 1)}\lambda_{ai}(1 - \lambda_{ai}), \qquad i = 1, \cdots, I,$$

$$E(\theta_{.j}) = \lambda_{bj}, \ \mathrm{Var}(\theta_{.j}) = \frac{(K + L_b + 1)}{(K + 1)(L_b + 1)}\lambda_{bj}(1 - \lambda_{bj}), \qquad j = 1, \cdots, J.$$

The parameter λ_{ai} represents a prior guess at the probability $\theta_{i.}$ and (for a fixed value of the parameter K) the parameter L_a reflects the precision of this prior guess.

(ii) Consider the vector of parameters $(\gamma_{11}, \gamma_{12}, \cdots, \gamma_{IJ})$, where $\gamma_{ij} = \theta_{ij} - \theta_{i.}\theta_{.j}$, $i = 1, \cdots, I$, $j = 1, \cdots, J$. This vector can represent the interaction structure in the table. The mean and variance of γ_{ij} from the above prior (3.1) are

$$E(\gamma_{ij}) = 0 \tag{3.3}$$

and

$$\text{Var}(\gamma_{ij}) = \frac{K}{(K+1)(K+3)} \frac{\lambda_{ai}(1-\lambda_{ai})L_a}{L_a+1} \frac{\lambda_{bj}(1-\lambda_{bj})L_b}{L_b+1}. \tag{3.4}$$

From observing (3.3), it is clear that the user believes that the two variables in the contingency table are independent and from (3.4) the strength of the user's belief in independence is indicated through the parameter K. To better understand the role of K in the prior, consider the special case where $I = J = 2$. In this 2×2 table, one common measure of association is the correlation coefficient, defined by

$$\rho_c = (\theta_{11} - \theta_{1.}\theta_{.1})/(\theta_{1.}\theta_{2.}\theta_{.1}\theta_{.2})^{1/2}.$$

Using (3.3), the prior mean of ρ_c can be approximated by

$$E(\theta_{11} - \theta_{1.}\theta_{.1})/\{E(\theta_{1.}\theta_{2.})E(\theta_{.1}\theta_{.2})\}^{1/2} = 0.$$

Similarly, the prior variance of ρ_c can be approximated by

$$\frac{E(\theta_{11} - \theta_{1.}\theta_{.1})^2}{E(\theta_{1.}\theta_{2.})E(\theta_{.1}\theta_{.2})} = \frac{K+1}{K(K+3)}. \tag{3.5}$$

Note from (3.5) that the prior variance of ρ_c is a decreasing function of K. The user who specifies a large value of K believes that the parameter ρ_c lies within a small interval about zero and is indicating a strong belief in independence.

As in Section 2, a value of the prior parameter K can be chosen by regarding it as the sample size of a preliminary table in which there exists independence between the two variables. In the special 2×2 table, an alternative way to indirectly choose K is to specify an interval of the form $(-a, a)$ which is felt to contain ρ_c with high probability, say, 0.91. By using tables of incomplete beta integrals, a translated beta density (symmetric about 0) can be matched to this prior belief. The variance of this translated beta density can then be set equal to (3.5) to obtain the value of K.

As in Section 2, a cell probability θ_{ij} is estimated by its posterior mean. Theorem 2 below gives an expression for the posterior mean and gives simple expressions for this estimator in the limiting cases where the prior parameter K approaches zero and infinity.

THEOREM 2. *Let the prior density for* $\boldsymbol{\theta}$ *be given by* (3.1). *The posterior mean of* θ_{ij} *is given by*

$$E(\theta_{ij} \mid \mathbf{x}) = \frac{N}{N+K}\frac{x_{ij}}{N} + \frac{K}{N+K}E(\eta_{ai}\eta_{bj} \mid \mathbf{x}), \tag{3.6}$$

where the expectation in (3.6) *is taken with respect to the posterior distribution of* $(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b)$, *with density given by*

$$\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x}) = C\,\Pi_{i,j}(K\eta_{ai}\eta_{bj})^{(x_{ij})}\,\Pi_{i=1}^{I}\eta_{ai}^{L_a\lambda_{ai}-1}\,\Pi_{j=1}^{J}\eta_{bj}^{L_b\lambda_{bj}-1}, \quad (\boldsymbol{\eta}_a, \boldsymbol{\eta}_b) \in B,$$

C being the proportionality constant. Furthermore

$$\lim_{K\to 0}E(\theta_{ij} \mid \mathbf{x}) = \frac{x_{ij}}{N}, \tag{3.7}$$

$$\lim_{K\to\infty}E(\theta_{ij} \mid \mathbf{x}) = \frac{(x_{i.} + L_a\lambda_{ai})(x_{.j} + L_b\lambda_{bj})}{(N+L_a)(N+L_b)}, \tag{3.8}$$

where $x_{i.} = \sum_{j=1}^{J} x_{ij}$, $x_{.j} = \sum_{i=1}^{I} x_{ij}$.

PROOF. See Albert and Gupta (1981).

TABLE 1
Parental decision-making and political affiliation. Source: Braungart (1971)

		Political Affiliation	
		SDS	**YAF**
Parental Decision Making	**Authoritarian**	29	33
	Democratic	131	78

Recall that a large value of K implies a strong belief in independence in the table. Note that one can write (3.8) as the product $\tilde{\theta}_{i.}\tilde{\theta}_{.j}$, where $\tilde{\theta}_{i.} = (x_{i.} + L_a\lambda_{ai})/(N + L_a)$ and $\tilde{\theta}_{.j} = (x_{.j} + L_b\lambda_{bj})/(N + L_b)$. The posterior mean, in the limiting case $K = \infty$, first combines the prior information and the sample information to estimate the marginal probabilities $\theta_{i.}$ and $\theta_{.j}$ by $\tilde{\theta}_{i.}$ and $\tilde{\theta}_{.j}$ respectively. Then the estimators $\tilde{\theta}_{i.}$ and $\tilde{\theta}_{.j}$ are multiplied to estimate θ_{ij}, reflecting the strong belief in independence.

To illustrate the behavior of the posterior mean (3.6) for finite values of K, consider the 2×2 table (from Braungart, 1971, and analyzed in Bishop, Fienberg and Holland, 1975) given in Table 1 which classifies college students with respect to their political affiliation and their family structure. Consider the estimation of the cell probabilities using the prior belief that the two variables of the study are independent. To use the prior density (3.1), first one specifies the vectors $(L_a, \lambda_{a1}, \lambda_{a2})$ and $(L_b, \lambda_{b1}, \lambda_{b2})$ which reflect prior knowledge about the proportions of students in the two political affiliations and the two family structures respectively. In this example, we will set $L_a\lambda_{a1} = L_a\lambda_{a2} = L_b\lambda_{b1} = L_b\lambda_{b2} = 1$, reflecting ignorance about the location of these proportions (η_a and η_b each will be assigned uniform distributions). Next, one specifies a value of the prior parameter K, which indicates the strength of the belief of independence.

In practice, the posterior mean (3.6) is difficult to compute numerically, due to the posterior expectation

$$E(\eta_{ai}\eta_{bj} \mid \mathbf{x}) = \frac{\int_B \eta_{ai}\eta_{bj}\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})\, d\boldsymbol{\eta}_a\, d\boldsymbol{\eta}_b}{\int_B \pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})\, d\boldsymbol{\eta}_a\, d\boldsymbol{\eta}_b}, \tag{3.9}$$

a ratio of two multidimensional integrals each of dimension $I + J - 2$. One efficient method of numerically computing these integrals uses the notion of importance sampling. First, it can be shown that as the prior parameter K approaches zero and infinity,

$$\lim_{K\to 0}\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x}) = \pi_1^1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x}) = \psi(\boldsymbol{\eta}_a \mid \mathbf{1}/I, I(J+1))\psi(\boldsymbol{\eta}_b \mid \mathbf{1}/J, J(I+1)) \tag{3.10}$$

and

$$\lim_{K\to \infty}\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x}) = \pi_1^2(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x}) = \psi(\boldsymbol{\eta}_a \mid \tilde{\boldsymbol{\theta}}_a, N + L_a)\psi(\boldsymbol{\eta}_b \mid \tilde{\boldsymbol{\theta}}_b, N + L_b), \tag{3.11}$$

where $\mathbf{1}$ is the vector containing all ones, $\tilde{\boldsymbol{\theta}}_a = (\tilde{\theta}_{1.}, \cdots, \tilde{\theta}_{I.})$ and $\tilde{\theta}_b = (\tilde{\theta}_{.1}, \cdots, \tilde{\theta}_{.J})$. Thus for small or large values of K, the posterior density π_1 can be approximated by the simpler asymptotic densities π_1^1 and π_1^2 respectively. Next, rewrite the posterior expectation (3.9) as

$$E(\eta_{ai}\eta_{bj} \mid \mathbf{x}) = \frac{\int_B \eta_{ai}\eta_{bj}\left\{\dfrac{\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})}{\pi_1^q(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})}\right\}\pi_1^q(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})\, d\boldsymbol{\eta}_a\, d\boldsymbol{\eta}_b}{\int_B \left\{\dfrac{\pi_1(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})}{\pi_1^q(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})}\right\}\pi_1^q(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \mid \mathbf{x})\, d\boldsymbol{\eta}_a\, d\boldsymbol{\eta}_b}, \tag{3.12}$$

where π_1^q is one of the two limiting densities. Finally, to approximate the integrals in (3.12)

TABLE 2
Computed values of the maximum likelihood estimates (K = 0), "independence" estimates (K = ∞) and posterior means for different values of K

K	0	200	400	600	1000	2000	∞
$\hat{\theta}_{11}$	.107	.120	.124	.127	.130	.133	.136
$\hat{\theta}_{12}$	.122	.111	.106	.103	.100	.098	.095
$\hat{\theta}_{21}$	.483	.470	.466	.463	.460	.457	.454
$\hat{\theta}_{22}$	.288	.300	.304	.307	.310	.312	.316

using simulation, n_0 sets of values of $(\boldsymbol{\eta}_a, \boldsymbol{\eta}_b)$ are randomly generated from the Dirichlet densities of π_1^q. Call these randomly generated values $(d_{1m}, \cdots, d_{Im}, e_{1m}, \cdots, e_{Jm})$, $m = 1, \cdots, n_0$. Then (3.12) is approximated by

$$\frac{\sum_{m=1}^{n_0} d_{im} e_{jm} \pi_1(\mathbf{d}_m, \mathbf{e}_m \mid \mathbf{x}) / \pi_1^q(\mathbf{d}_m, \mathbf{e}_m \mid \mathbf{x})}{\sum_{m=1}^{n_0} \pi_1(\mathbf{d}_m, \mathbf{e}_m \mid \mathbf{x}) / \pi_1^q(\mathbf{d}_m, \mathbf{e}_m \mid \mathbf{x})}, \tag{3.13}$$

where $\mathbf{d}_m = (d_{1m}, \cdots, d_{Im})$ and $\mathbf{e}_m = (e_{1m}, \cdots, e_{Jm})$. In the situation where K is large and the asymptotic density π_1^2 is used, the approximation (3.13) reduces (in the case where $L_a \lambda_{ai} = L_b \lambda_{bj} = 1$ for all i, j) to

$$\frac{\sum_{m=1}^{n_0} d_{im} e_{jm} \Pi_{k,\ell} (K d_{km} e_{\ell m})^{(x_{k\ell})} / (\Pi_k d_{km}^{x_{k\cdot}} \Pi_\ell e_{\ell m}^{x_{\cdot \ell}})}{\sum_{m=1}^{n_0} \Pi_{k,\ell} (K d_{km} e_{\ell m})^{(x_{k\ell})} / (\Pi_k d_{km}^{x_{k\cdot}} \Pi_\ell e_{\ell m}^{x_{\cdot \ell}})}. \tag{3.14}$$

Several additional comments should be made concerning this computational method. First, in the large K situation, each summand in (3.14) was computed by first computing the natural log of each individual term (the log of $(K d_{km} e_{\ell m})^{(x_{kl})}$ was computed by first expressing it as a ratio of gamma functions and then using the IMSL procedure DLGAMA), combining the logs of the terms and then taking the exponential of the result. Second, this method can also be used to compute the posterior means in the symmetry situation, with the densities π_1^1 and π_1^2 replaced by the corresponding limiting distributions of $\boldsymbol{\eta}^*$.

In Table 2, values of the posterior means, given in (3.6), are calculated (numerically using the approximation (3.14)) for different values of the prior parameter K. These estimates were computed in each example using $n_0 = 5000$ iterations and requiring 42 seconds of CPU time. When the estimates were recomputed using a new set of 5000 values of $(\mathbf{d}_m, \mathbf{e}_m)$, the new estimates deviated no more than .001 (in each cell) from the estimates presented in Table 2. It appears that 5000 iterations are adequate in obtaining reasonably accurate approximations to the posterior means (3.6) for tables as large as 10 × 10, though a larger number of iterations may be required to estimate small cell probabilities. (One 10 × 10 example required 7 minutes and 20 seconds of CPU time.) Since this computational method simulates values from the limiting density π_1^2, this method appears to work best for large values of K. However, in many examples, this procedure appears to give stable values for even moderate values of K and we would recommend the use of the alternative density π_1^1 only when K is chosen small.

In addition to the numerically computed posterior means, Table 2 gives the values of the estimators in the limiting cases $K = 0$ and $K = \infty$. Note that as the value of K increases from 200 to 2000, the posterior means shrink the maximum likelihood estimates $\{x_{ij}/N\}$ towards the independence estimates $\{(x_{i\cdot} + 1)(x_{\cdot j} + 1)/(N + 2)^2\}$. These posterior estimates reflect the imprecise prior belief in independence by compromising between estimates assuming an unsaturated model and estimates assuming an independence model.

REFERENCES

ALBERT, J. H., and GUPTA, A. K. (1981). Mixtures of Dirichlet distributions and estimation in contingency tables. Bowling Green State University, Department of Mathematics and Statistics, T.R. No. 81-01.

BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis.* MIT Press, Cambridge, Mass.

BRAUNGART, R. G. (1971). Family status, socialization and student politics: a multivariate analysis. *Amer. J. Sociol.* **77** 108–130.

CROOK, J. F., and GOOD, I. J. (1980). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, part II. *Ann. Statist.* **8** 1198–1218.

FIENBERG, S. E., and HOLLAND, P. W. (1973). Simultaneous estimation of multinomial cell probabilities. *J. Amer. Statist. Assoc.* **68** 683–691.

GOKHALE, D. V., and PRESS, S. J. (1980). Assessment of a correlation coefficient in a bivariate normal distribution. University of California, Riverside, Department of Statistics technical report.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, Cambridge, Mass.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions (with discussion). *J. Roy. Statist. Soc.*, Ser. B, **29** 399–431.

GOOD, I. J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.* **4** 1159–1189.

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOWLING GREEN STATE UNIVERSITY
BOWLING GREEN, OHIO 43403