BAYESIAN ESTIMATION METHODS FOR INCOMPLETE TWO-WAY CONTINGENCY TABLES USING PRIOR BELIEFS OF ASSOCIATION

James H. Albert, Bowling Green State University

ABSTRACT. Suppose that a user possesses vague prior information about the association structure of a 2×2 contingency table and consider the problem of estimating the cell probabilities of the table using this prior information together with sample counts from an incomplete table. It is shown that a special mixture of Dirichlet distributions can reflect vague prior beliefs about an odds ratio and this prior information is used in the development of a posterior credible region for the vector of cell probabilities. The computation of this region is illustrated in the special case when the classification variables are believed independent.

1. INTRODUCTION. In this paper, Bayesian estimation methods are proposed for the cell probabilities of a 2×2 contingency table, when both completely and partially cross-classified data are collected. To illustrate the sampling scheme, consider data on 456 premature live births, given in Chen and Fienberg (1974) and presented in Table 1. The classification variables in this example are the infants' health index score (low, high) and their serum bilirium reading (low, high). Of the entire sample, 279 infants are completely classified with respect to both variables; 24 are partially classified with respect to their serum bilirium reading and the remaining 153 are classified only with respect to their health index. It is a trivial problem to estimate the cell probabilities of the 2×2 table using solely the completely classified counts. A nontrivial problem is how to use the counts in the two partially classified tables together with the completely classified counts to estimate the cell probabilities.

TABLE 1

DATA OF PREMATURE INFANTS CLASSIFIED WITH RESPECT TO HEALTH INDEX AND SERUM BILIRIUM LEVEL (from Chen and Fienberg (1974))

		Health Index: Low	Health Index: High		
Serum Bilirium Level	Low	35	75		11
	High	57	112		13
				279	24
		117	36	153	

Many authors, e.g. Hocking and Oxspring (1974), Chen and Fienberg (1974), and Fuchs (1982), have found maximum likelihood estimates (MLE's) of the cell probabilities. As will be shown in Section 3.1, these estimates allocate the partially classified counts to the 2×2 tables using proportions that are obtained from the completely classified table. The manner in which the partially classified counts are allocated to the complete table depends primarily on the association structure in the table. In the MLE procedure, the association structure is "estimated" by the completely classified counts.

Consider the situation where only a small portion of the total number of counts are completely classified. In this situation, the completely classified counts provide little information about the manner in which the partially classified counts are allocated to the table. In the extreme case where all of the counts are partially classified, the cell probabilities are not even estimable by the data. However, if prior information exists about the association structure in the table, then this information can be used (together with the completely classified counts) to allocate the partially classified counts to the table and give estimates for the cell probabilities. As Antelman (1972) explains, this Bayesian approach is necessary when all the data collected is partially classified.

To use the Bayesian method, the main task is to find a prior distribution which can reflect the typical vague form of prior information about the association structure of the table. To this end, Albert and Gupta (1982) introduced a class of priors, a mixture of Dirichlet distributions, which is designed to reflect vague prior beliefs about the cross-product ratio α, a common measure of association in a 2×2 table. (The rationale for the use of this class versus the use of the conjugate class is given in Albert and Gupta (1982).) One advantage of this class is that only two parameters are elicited from the user; basically these parameters reflect a guess at the association structure of the table and a statement of the precision of this guess.

Before we proceed, some notation will be given. Suppose that n observations are completely classified with respect to classification variables A and B and n_1 (n_2) observations are partially classified with respect to variable A (B), resulting in the observed counts below.

	B			
A	x_{11}	x_{12}	$x_{1\cdot}$	y_1^A
	x_{21}	x_{22}	$x_{2\cdot}$	y_2^A
	$x_{\cdot 1}$	$x_{\cdot 2}$	n	n_1
	y_1^B	y_2^B	n_2	

(The dot notation represents summation over the appropriate index.) It is of interest to estimate $\underset{\sim}{p} = (p_{11}, p_{12}, p_{21}, p_{22})$, where p_{ij} denotes the probability of falling in the (i,j) cell. If observations are classified from an infinite population, then the likelihood is given by

$$\pi p_{ij}^{x_{ij}} \pi p_{i\cdot}^{y_i^A} p_{\cdot i}^{y_i^B} . \tag{1.1}$$

Since the prior used is a mixture of Dirichlet distributions, it will be convenient to define

$$f_D(\underset{\sim}{p}|K\underset{\sim}{\eta}) = \Gamma(K) \prod_{i,j} p_{ij}^{Kf_{ij}-1} / \Gamma(Kf_{ij}), \tag{1.2}$$

the Dirichlet density with prior mean vector $\underset{\sim}{\eta} = (f_{11}, f_{12}, f_{21}, f_{22})$ and precision parameter K.

In Section 2, the prior distribution on $\underset{\sim}{p}$ is defined and expressions are given for the posterior moments in the special case where partially classified counts exist for only one variable (the general case is considered in Albert (1983)). In Section 3, simple approximations are developed for the posterior means which show how the partially classified counts are allocated in the complete table. In Section 4, we conclude our discussion by illustrating the computation of the posterior means and variances for the data in Chen and Fienberg (1974) in the situation where the user believes that the two classification variables are independent.

2. PRIOR TO POSTERIOR ANALYSIS.

2.1. THE PRIOR DISTRIBUTION.

Albert and Gupta (1982) introduced the following two-stage prior distribution to reflect prior beliefs about association in a 2×2 table.

Stage I: The vector $\underset{\sim}{p}$ is given the Dirichlet distribution (1.2), where the components of $\underset{\sim}{\eta}$ have row margins η_a, $1 - \eta_a$, column margins η_b, $1 - \eta_b$, and cross-product ratio α_0. Equivalently, the set of prior means satisfy the configuration

$f_{11}(\eta_a, \eta_b)$	$\eta_a - f_{11}(\eta_a, \eta_b)$	η_a
$\eta_b - f_{11}(\eta_a, \eta_b)$	$1-\eta_a-\eta_b+f_{11}(\eta_a, \eta_b)$	$1 - \eta_a$
η_b	$1 - \eta_b$	

(2.1)

where $\alpha_0 = [f_{11}(\cdot,\cdot)(1-\eta_a-\eta_b+f_{11}(\cdot,\cdot)]/[(\eta_a-f_{11}(\cdot,\cdot))(\eta_b-f_{11}(\cdot,\cdot))]$.

Stage II: The vector of hyperparameters (η_a, η_b) is given a uniform distribution on the unit square.

The resulting prior density on $\underset{\sim}{p}$ is given by

$$\pi_m(p) = \int\int f_D(\underset{\sim}{p}|K\underset{\sim}{\eta}^*) d\eta_a d\eta_b, \tag{2.2}$$

where $\underset{\sim}{\eta}^* = (f_{11}, f_{12}, f_{21}, f_{22})$ is the vector of prior means with configuration (2.1) (for ease of notation, we will write f_{11} instead of $f_{11}(\eta_a, \eta_b)$, although it is understood that the prior mean is a function of the parameters η_a and η_b).

The prior distribution (2.2) is designed to accept the typical form of vague prior information about the association structure in the table. Two parameters are elicited from the user; the parameter α_0 is a guess at the cross-product ratio α and the parameter K reflects the sureness of one's guess at α_0. It is illustrated in Albert and Gupta (1982) that the induced prior on $\ell n\ \alpha$ is approximately bell-shaped and symmetric about $\ell n\ \alpha_0$. Therefore, by the specification of an interval which is thought to contain α with probability .9 and the use of Figure 2 in Albert and Gupta (1982), one can obtain values of α_0 and K.

2.2 POSTERIOR ANALYSIS.

If $\underset{\sim}{p}$ is given the prior (2.2), then the posterior density of $\underset{\sim}{p}$ is proportional to

$$\int\int \pi[p_{ij}^{x_{ij}+Kf_{ij}-1}/\Gamma(Kf_{ij})] \pi p_{i\cdot}^{y_i^A} p_{\cdot i}^{y_i^B} d\eta_a d\eta_b. \tag{2.3}$$

This density can be seen to be a mixture of densities with kernel

$$\pi p_{ij}^{a_{ij}-1} \pi p_{i\cdot}^{y_i^A} p_{\cdot i}^{y_i^B}, \tag{2.4}$$

where $a_{ij} = x_{ij} + Kf_{ij}$. Here attention is restricted to the special case $y_1^B = y_2^B = 0$. In this case, the family of distributions with kernel (2.4) is called by Antleman (1972) the simple Dirichlet-beta ($D\beta$) family. First some facts about the simple $D\beta$ distribution are summarized in Section 2.2.1 and these results are applied in obtaining expressions for the posterior moments in Section 2.2.2.

2.2.1. The Simple Dirichlet-Beta Distribution.

After some manipulation, it can be shown that the simple $D\beta$ density can be represented as

$$\begin{aligned}\pi_s(\underset{\sim}{p}|\underset{\sim}{a},\underset{\sim}{y}^A) = \sum_{i=0}^{y_1^A} \sum_{j=0}^{y_2^A} & f_{Bb}(i|y_1^A, a_{11}, a_{12}) \\ & \cdot f_{Bb}(j|y_2^A, a_{21}, a_{22}) \\ & \cdot f_D(\underset{\sim}{p}|a_{11}+i,\ a_{12}+y_1^A-i, \\ & a_{21}+j,\ a_{22}+y_2^A-j),\end{aligned} \tag{2.5}$$

where $\underset{\sim}{a} = (a_{11}, a_{12}, a_{21}, a_{22})$, $\underset{\sim}{y}^A = (y_1^A, y_2^A)$ and f_{Bb} denotes the beta-binomial density given by

$$f_{Bb}(k|m,b,c) = \binom{m}{k} B(b+k, c+m-k)/B(b,c).$$

Using the representation (2.5), Albert (1983) showed that the mean vector of $\underset{\sim}{p}$ is given by $\underset{\sim}{\lambda} = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$, where

$$\lambda_{ij} = E(p_{ij}|\underset{\sim}{a},\underset{\sim}{y}^A) = \frac{a_{ij}}{a_{i\cdot}} \frac{a_{i\cdot} + y_i^A}{a_{\cdot\cdot} + n_1} . \tag{2.6}$$

The posterior covariance matrix is given by

$$\mathrm{cov}(\underset{\sim}{p}|\underset{\sim}{a},\underset{\sim}{y}^A) = (a_{..}+n_1+1)^{-1}$$

$$(\mathrm{diag}\{\lambda_{11},\lambda_{12},\lambda_{21},\lambda_{22}\}-\underset{\sim}{\lambda}\,\underset{\sim}{\lambda}'$$

$$+\begin{bmatrix} \underset{\sim}{A}_1 & 0 \\ 0 & \underset{\sim}{A}_2 \end{bmatrix}), \qquad (2.7)$$

where

$$\underset{\sim}{A}_i = \frac{y_i^A\lambda_{i1}a_{i2}}{(a_{i.}+1)a_{i.}} \cdot \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad i = 1,2.$$

2.2.2. Posterior Moments in the Case $y_1^B = y_2^B = 0$.

In the special case $y_1^B = y_2^B = 0$, the posterior density (2.3) can be represented by

$$\pi(\underset{\sim}{p}|\underset{\sim}{x},\underset{\sim}{y}^A) = \int\int \pi_1(\eta_a,\eta_b|\underset{\sim}{x},\underset{\sim}{y}^A) \cdot\pi_s(p|\underset{\sim}{\eta},\underset{\sim}{x},\underset{\sim}{y}^A)d\eta_a d\eta_b, \qquad (2.8)$$

where

$$\pi_1(\eta_a,\eta_b|\underset{\sim}{x},\underset{\sim}{y}^A) \propto \prod \frac{\Gamma(Kf_{ij}+x_{ij})}{\Gamma(Kf_{ij})} \frac{\Gamma(K\eta_a+x_{1.}+y_1^A)}{\Gamma(K\eta_a+x_{1.})} \cdot \frac{\Gamma(K(1-\eta_a)+x_{2.}+y_2^A)}{\Gamma(K(1-\eta_a)+x_{2.})} \qquad (2.9)$$

and $\pi_s(\underset{\sim}{p}|\underset{\sim}{\eta},\underset{\sim}{x},\underset{\sim}{y}^A)$ is given by (2.5) with $a_{ij} = Kf_{ij} + x_{ij}$. Using expressions in Section 2.2.1 and rules of conditional expectation,

$$E(p_{ij}|\underset{\sim}{x},\underset{\sim}{y}^A) = E[E(p_{ij}|\underset{\sim}{\eta},\underset{\sim}{x},\underset{\sim}{y}^A)] = E[\lambda_{ij}|\underset{\sim}{x},\underset{\sim}{y}^A] \qquad (2.10)$$

$$\mathrm{Var}(p_{ij}|\underset{\sim}{x},\underset{\sim}{y}^A) = E[\mathrm{Var}(p_{ij}|\underset{\sim}{\eta},\underset{\sim}{x},\underset{\sim}{y}^A)] + \mathrm{Var}[E(p_{ij}|\underset{\sim}{\eta},\underset{\sim}{x},\underset{\sim}{y}^A)] = v_{ij}^1 + v_{ij}^2 + v_{ij}^3$$

where

$$v_{1j}^1 = (n+K+n_1+1)^{-1}E[\lambda_{ij}(1-\lambda_{ij})|\underset{\sim}{x},\underset{\sim}{y}^A]$$

$$v_{2j}^2 = (n+K+n_1+1)^{-1}E[\frac{y_i^A\lambda_{i1}(x_{i2}+Kf_{i2})}{u_i(u_i+1)}|\underset{\sim}{x},\underset{\sim}{y}^A] \qquad (2.11)$$

$$v_{3j}^3 = \mathrm{Var}[\lambda_{11}|\underset{\sim}{x},\underset{\sim}{y}^A],$$

$$\lambda_{ij} = (x_{ij}+Kf_{ij})(u_i+y_i^A)/[u_i(n+K+n_1)],$$

$u_1 = x_{1.} + K\eta_a$, $u_2 = x_{2.} + K(1-\eta_a)$, and the expectations and variance in (2.11) are taken with respect to the posterior distribution of (η_a,η_b) (2.9).

3. NUMERICAL STUDY. The posterior expressions in (2.10) and (2.11) are not written in closed form and, therefore, it is difficult to see how these moments incorporate the information contained in the prior and the sample counts. In this section, the computation of these expressions is discussed, and simple approximations are proposed in the independence case which illuminate the behavior of the posterior means and variances.

First note that the posterior quantities (2.10) and (2.11) involve expectations using the posterior density of (η_a,η_b) (2.9), which is not expressible in closed form. Thus it is necessary to compute expectations of the form

$$E[g(\eta_a,\eta_b)|\underset{\sim}{x},\underset{\sim}{y}^A] = \frac{\int_0^1\int_0^1 g(\eta_a,\eta_b)\pi_1(\eta_a,\eta_b)d\eta_a d\eta_b}{\int_0^1\int_0^1 \pi_1(\eta_a,\eta_b)d\eta_a d\eta_b}, \qquad (3.1)$$

where g is an arbitrary function of η_a and η_b. One efficient way of computing the integrals in (3.1) uses the notion of importance sampling. The first step of this simulation technique finds a simple approximation for the posterior density π_1. Since it can be shown for $\alpha_0 = 1$ that

$$\lim_{K\to\infty} \pi_1(\eta_a,\eta_b|\underset{\sim}{x},\underset{\sim}{y}^A) = \pi_L(\eta_a,\eta_b|\underset{\sim}{x},\underset{\sim}{y}^A) = f_\beta(\eta_a|x_{1.}+y_1^A+1,\ x_{2.}+y_2^A+1) \cdot f_\beta(\eta_b|x_{.1}+1,\ x_{.2}+1), \qquad (3.2)$$

the limiting distribution π_L can serve as a rough approximation to π_1 for values of α_0 near one. Next, rewrite the expectation (3.1) as

$$E[g(\eta_a,\eta_b)|\underset{\sim}{x},\underset{\sim}{y}^A] = \frac{\int\int g(\eta_a,\eta_b)\left[\frac{\pi_1(\pi_a,\pi_b)}{\pi_L(\eta_a,\eta_b)}\right]\pi_L(\eta_a,\eta_b)d\eta_a d\eta_b}{\int\int\left[\frac{\pi_1(\eta_a,\eta_b)}{\pi_L(\eta_a,\eta_b)}\right]\pi_L(\eta_a,\eta_b)d\eta_a d\eta_b}. \qquad (3.3)$$

Finally, to approximate the integrals in (3.3) using simulation, N_0 values of (η_a,η_b) are randomly generated from the beta densities in (3.2). Call the randomly generated values (e_{ai},e_{bi}), $i = 1,\ldots,N_0$. Then (3.3) is approximated by

$$\frac{\sum_{i=1}^{N_0} g(e_{ai},e_{bi})\pi_1(e_{ai},e_{bi})/\pi_L(e_{ai},e_{bi})}{\sum_{i=1}^{N_0} \pi_1(e_{ai},e_{bi})/\pi_L(e_{ai},e_{bi})}. \qquad (3.4)$$

In the example which follows, we will consider the situation where the user believes a priori that the two classification variables are independent. The prior parameter α_0 will be set to one (reflecting a belief in independence) and the value of the parameter K selected will reflect the precision of a user's belief in independence.

Since the posterior means are, in some sense, a compromise between estimates from an unrestricted model and estimates from an independence model, we will first discuss the computation of these "traditional" estimates. Consider the hypothetical sample counts presented in Table 2. Under the unrestricted model, the MLE estimates a cell probability by allocating the partially classified counts according to the counts in the completely classified table. In this example the 30 counts partially classified in category one are allocated into the (1,1), (1,2) cells in the complete table by the proportions 100/(100 + 50), 50/(100 + 50), respectively. In general, the MLE of p_{ij} is given by

$$\hat{p}_{ij} = \frac{x_{ij} + y_i^A \cdot x_{ij}/x_i}{n + n_1}, \qquad (3.5)$$

and the values of these estimates are given in Table 3. To understand the computation of the MLE under an independence model, first note that if the partially classified counts are ignored, then the expected cell count in cell (1,1) is $x_{1\cdot}x_{\cdot 1}$. Then the 30 partially classified counts are allocated into the (1,1), (1,2) cells by the "pooled" proportions 175/300, 125/300, respectively. The independence MLE of p_{ij} is given by

$$\tilde{p}_{ij} = \frac{x_{i\cdot}x_{\cdot j} + y_i^A x_{\cdot j}/n}{n + n_1}. \qquad (3.6)$$

TABLE 2

SOME HYPOTHETICAL SAMPLE COUNTS

		B			
A	100	50	150		30
	75	75	150		60
	175	125	300		90

TABLE 3

COMPUTED VALUES OF MLE'S, EXACT AND APPROXIMATE POSTERIOR MEANS FOR DATA OF TABLE 2

	Expected cell counts	Probability estimates	
MLE, unrestricted	100 + 30(.667)	.308	
	50 + 30(.333)		.154
	75 + 60(.500)	.269	
	75 + 60(.500)		.269
MLE, independence	87.5 + 30(.583)	.269	
	62.5 + 30(.417)		.192
	87.5 + 60(.583)	.314	
	62.5 + 60(.417)		.224
Exact posterior means, K = 100	100+100(.266)+30(.646)	.298	
	50+100(.192)+30(.354)		.163
	75+100(.310)+60(.520)	.280	
	75+100(.226)+60(.480)		.258
Approximate posterior means, K = 100	100+100(.269)+30(.646)	.299	
	50+100(.192)+30(.354)		.163
	75+100(.314)+60(.521)	.281	
	75+100(.224)+60(.479)		.257

The posterior mean (2.10) can be rewritten as

$$E(p_{ij}|\underset{\sim}{x},\underset{\sim}{y}^A) = (n+K+n_1)^{-1}[x_{ij}+KE(f_{ij}|\underset{\sim}{x},y^A) + y_i^A E((x_{ij}+Kf_{ij})/u_i|\underset{\sim}{x},y^A)]. \qquad (3.7)$$

Using techniques similar to those discussed in Albert and Gupta (1982), the following approximation to (3.7) is proposed:

$$E(p_{ij}|\underset{\sim}{x},\underset{\sim}{y}^A) \cong (n+K+n_1)^{-1}[x_{ij}+K\tilde{p}_{ij}+y_i^A\hat{\gamma}_{ij}] = p_{ij}^* \qquad (3.8)$$

where

$$\gamma_{ij} = \frac{n}{n+K}\frac{x_{ij}}{x_{i\cdot}} + \frac{K}{n+K}\frac{x_{\cdot j}}{n}.$$

Values of the exact posterior mean together with the approximate values are also given in Table 3. To illustrate the computation of (3.7), note that for the (1,1) cell the observed count 100 is first added to the count 26.6, reflecting a shift of the observed count towards an expected count assuming an independence model. Then the 30 partially classified counts are allocated to the (1,1), (1,2) cells by the probabilities .646, .354, respectively. The probability .646 is a compromise between the allocation probabilities assuming an unrestricted model and an independence model. Thus the posterior means allocate the partially classified counts to the complete table in a way which reflects the vague prior beliefs in independence.

4. AN EXAMPLE. To illustrate the application of the Bayesian estimation procedures proposed in this paper, consider the Chen and Fienberg (1974) data discussed in Section 1. Suppose the user believes a priori that an infant's health index score is unrelated to his/her serum bilirium reading. Equivalently, the odds of a low health index score infant having a high serum bilirium reading are believed to be equal to the odds of a high health index score infant having a high serum bilirium reading. In addition, suppose that the user is 90 per cent confident that the ratio of the above odds is between .2 and 5. Using the Albert and Gupta (1982) table, this prior belief is translated to the values of the prior parameters $\alpha_0 = 1$, $K = 150$. Using the prior (2.2) with this prior knowledge, Table 4 gives the (approximate) posterior means and variances. (Expressions for these moments are given in Albert (1983).) These moments can be used to construct approximate credible intervals for the components of $\underset{\sim}{p}$. For

example, by assuming that the marginal posterior distribution of p_{11} is approximately normal, the interval

$$E(p_{11}|\underset{\sim}{x},\underset{\sim}{y}) \pm 2(\mathrm{Var}(p_{11}|\underset{\sim}{x},\underset{\sim}{y}))^{\frac{1}{2}}$$
$$= .186 \pm 2(395 \cdot 10^{-6})^{\frac{1}{2}}$$
$$= .186 \pm .040$$

is an approximate 95 per cent credible interval. These procedures are attractive alternatives to the usual classical procedures when vague prior beliefs exist about the association structure in the 2×2 table. For future research, we plan to identify situations where vague prior beliefs exist and suggest ways of eliciting these beliefs so they can be used in the estimation process.

TABLE 4

APPROXIMATE POSTERIOR MEANS AND VARIANCES FOR CHEN AND FIENBERG DATA, $\alpha_0 = 1$, $K = 150$

Posterior Means

.186	.211
.291	.313

Posterior Variances (unit = 10^{-6})

395	416
550	510

REFERENCES

Albert, J. H. (1983). Bayesian estimation methods for incomplete two-way contingency tables using mixtures of Dirichlet distributions. Technical Report No. 83-10, Department of Mathematics and Statistics, Bowling Green State University.

Albert, J. H., and A. K. Gupta (1982). Bayesian estimation methods for 2×2 contingency tables using mixtures of Dirichlet distributions. *J. Amer. Statist. Assoc.*, to appear.

Antleman, G. R. (1972). Interrelated Bernoulli processes. *J. Amer. Statist. Assoc.*, 67, 831-841.

Chen, T., and S. E. Fienberg (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*, 30, 629-642.

Dempster, A. P., N. M. Laird, and D. E. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Statist. Soc.*, *B*, 1, 1-38.

Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *J. Amer. Statist. Assoc.*, 77, 270-278.

Hocking, R. R., and H. H. Oxspring (1974). The analysis of partially categorized contingency data. *Biometrics*, 30, 469-483.