Sankhyā: The Indian Journal of Statistics
2001, Volume 63, Series B, Pt. 3, pp 270-285

BAYESIAN ANALYSIS OF CORRELATED PROPORTIONS

By MARIA KATERI,
University of Ioannina
TAKIS PAPAIOANNOU
University of Pireaus
and
PETROS DELLAPORTAS
Athens University of Economics and Business

SUMMARY. In this paper we present a Bayesian analysis of 2×2 contingency tables, corresponding to matched pairs designs. We provide Bayes and empirical Bayes estimates for the cell probabilities of these tables as well as the Bayes factor for testing the equality of correlated proportions. The approximate highest posterior density (HPD) region for the difference of the correlated proportions is also obtained. Finally, a Bayesian variable selection approach is applied to a hierarchical logistic regression model and posterior model probabilities for the equality of the correlated proportions are estimated. This latter approach has the feature that the posterior model probabilities depend on the main-diagonal cells.

1. **Introduction**

Correlated proportions are usually expressed in the form of a 2×2 contingency table and their standard treatment consists of testing the null hypothesis of equality of proportions and evaluating confidence intervals for their difference. In the frequentists' approach these aspects have been treated by a number of authors (cf. Agresti 1990, Lloyd 1990) beginning with the well-known test of McNemar (1947). An issue of interest is the fact that McNemar's test is based only on the discordant pairs of the table. The

Paper received September 2001.
AMS (1991) *subject classifications.* Primary 62F15; secondary 62H17.
Key words and phrases. Bayes factor, empirical bayes, Gibbs variable selection, hierarchical logistic regression, highest posterior density region, matched pairs, Markov chain Monte Carlo.

concordant pairs play no role as one would expect since the off-diagonal cell counts alone are sufficient for the difference of the two marginal proportions. This lack of contribution of concordant counts, however, runs against the intuition of many data analysts and has been a point of discussion since Cochran (1950). Attempts to derive inference for matched pairs taking into consideration concordant and discordant pairs have been made by May and Johnson (1997) and Liang and Zeger (1988), to mention the most recent relevant references.

Although Bayesian analysis of two independent proportions has been in the center of attention several times (Nurminen and Mutanen 1987, Walters 1986, Kass and Vaidyanathan 1992), as far as we know, correlated proportions have not been examined as yet from the Bayesian point of view with the exception of Altham (1971). Altham mainly focused on the comparison of different available models for correlated proportions (logistic model, random effects model, etc.) and their Bayesian analogues. She also further extended these methods to the case where the order of application of the two treatments (corresponding to the two classification variables) may matter. However, for the basic simple problem in its standard formulation, she only dealt with hypothesis testing and developed the corresponding Bayes test.

This paper is an attempt to provide an extensive Bayes and empirical Bayes analysis of matched pairs with particular emphasis on the comparison of the two proportions assuming either no pair effect or a random pair effect. In Section 2 we provide Bayes estimates of the cell probabilities based on the symmetry hypothesis and we compute the corresponding Bayes factor (BF) for testing the equality of correlated proportions. An alternative approach based on empirical Bayes estimation of prior parameters is described in Section 3. The Highest Posterior Density (HPD) region for the difference of correlated proportions is developed in Section 4 noting that the classical confidence interval is derived for a particular choice of a Dirichlet prior. In Section 5 we investigate a hierarchical logistic regression model which provides, through a Markov Chain Monte Carlo variable selection methodology, a way to determine the symmetry hypothesis taking into account the main-diagonal cells. The issues which arise are illustrated with the help of two simulated data sets in Section 6 while final conclusions are provided in Section 7.

2. Bayesian Analysis

2.1 *Estimation.* Let $(n_{ij})_{2\times 2}$ be the contingency table of observed frequencies corresponding to a matched pairs design of two correlated proportions $\pi_{1.}$ and $\pi_{.1.}$ We assume that the sampling scheme is multinomial, so that

$$\mathbf{n} = (n_{11}, n_{12}, n_{21}, n_{22}) \sim Mult(N, \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}), \Sigma_{ij}\pi_{ij} = 1. \qquad (2.1)$$

In standard contingency tables analysis the cell probabilities $\pi_{ij}(i,j = 1,2)$ are estimated by the corresponding sample proportions p_{ij}. Under the symmetry hypothesis, the off-diagonal probabilities are estimated by $(p_{12}+p_{21})/2$ and the correlated proportions $\pi_{1.}$ and $\pi_{.1}$ by $p_{11}+(p_{12}+p_{21})/2$. Bayesian conjugate analysis proceeds by imposing a Dirichlet prior on the vector of probability parameters

$$(\pi_{11}, \pi_{12}, \pi_{21}) \sim Di(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}), \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22} > 0, \qquad (2.2)$$

resulting to a posterior probability of

$$(\pi_{11}, \pi_{12}, \pi_{21})|\mathbf{n} \sim Di(n_{11} + \alpha_{11}, n_{12} + \alpha_{12}, n_{21} + \alpha_{21}, n_{22} + \alpha_{22}),$$

with the posterior mean of π_{ij} equal to

$$\tilde{\pi}_{ij} = E(\pi_{ij}|\mathbf{n}) = w \cdot p_{ij} + (1-w) \cdot k_{ij}, \qquad i,j = 1,2, \qquad (2.3)$$

where $w = N/(\alpha_0 + N), k_{ij} = \alpha_{ij}/\alpha_0$ and $\alpha_0 = \Sigma_{i,j}\alpha_{ij}$. Equation (2.3) indicates that the π_{ij}'s are estimated as weighted averages of the sampling proportion and the prior mean. The weights depend on the sample size N and on α_0, which expresses the degree of uncertainty associated with the prior mean. If α_{ij} tends to zero then $\tilde{\pi}_{ij} \to p_{ij}$ while if $\alpha_0 \to \infty$ then $\tilde{\pi}_{ij} \to k_{ij}$. The Bayes estimators (with square error loss function) of the correlated proportions are $\tilde{\pi}_{i.} = w \cdot p_{i.} + (1-w).k_{i.}$, and $\tilde{\pi}_{.j} = w \cdot p_{.j} + (1-w).k_{.j}$

2.2 *Hypothesis testing.* The hypothesis of equality of two correlated proportions ($\pi_{1.} = \pi_{.1}$) can equivalently be expressed as

$$H_s : \pi_{12} = \pi_{21} \quad \text{versus} \quad H_A : \pi_{12} \neq \pi_{21}, \qquad (2.4)$$

where $\mathbf{n}$ is distributed as defined by (2.1). Hypothesis (2.4) can be tested using the Bayes Factor $BF(H_s, H_A)$ which is defined as the ratio of the marginal likelihood under model H_s to the marginal likelihood under model H_A

$$\begin{aligned} BF(H_s, H_A) &= \frac{MLD(H_s)}{MLD(H_A)} \\ &= \frac{\int\int pr(x|\pi_{12},\pi_{11},H_s).pr(\pi_{12},\pi_{11}|H_s)d\pi_{12}d\pi_{11}}{\int\int\int pr(x|\pi_{12},\pi_{21},\pi_{11},H_A).pr(\pi_{12},\pi_{21},\pi_{11}|H_A)d\pi_{12}d\pi_{21}d\pi_{11}} \end{aligned} \tag{2.5}$$

(see, for example, Bernardo and Smith 1993). Kass and Raftery (1995) provided a table, based on one suggested by Jeffreys in 1961, to be used as a guide in interpreting the BF value.

Assuming that the prior distribution of $(\pi_{11}, \pi_{12}, \pi_{21})$ under H_A is provided by the Dirichlet distribution given in (2.2), under H_A, the marginal likelihood is given by

$$MLD(H_A) = C \cdot \frac{\Gamma(\alpha_0) \cdot \Gamma(n_{11} + \alpha_{11}) \cdot \Gamma(n_{12} + \alpha_{12}) \cdot \Gamma(n_{21} + \alpha_{21}) \cdot \Gamma(n_{22} + \alpha_{22})}{\Gamma(\alpha_{11}) \cdot \Gamma(\alpha_{12}) \cdot \Gamma(\alpha_{21}) \cdot \Gamma(\alpha_{22}) \cdot \Gamma(N + \alpha_0)}, \tag{2.6}$$

where

$$C = \frac{N!}{n_{11}!n_{12}!n_{21}!n_{22}!}.$$

Under H_s let π_{12}^s denote the common value of π_{12} and π_{21} (i.e., $\pi_{12}^s = \pi_{12} = \pi_{21}$). From (2.2) it follows that $(\pi_{11}, \pi_{12} + \pi_{21})$ is distributed as a Dirichlet variable with parameters $\alpha_{11}, \alpha_{12} + \alpha_{21}$ and α_{22}. In other words

$$(\pi_{11}, 2\pi_{12}^s) - Di(\alpha_{11}, \alpha_{12} + \alpha_{21}, \alpha_{22}),$$

which leads to

$$MLD(H_s) = C \cdot \frac{\Gamma(\alpha_0) \cdot \Gamma(n_{11} + \alpha_{11}) \cdot \Gamma(n + \alpha_{12} + \alpha_{21}) \cdot \Gamma(n_{22} + \alpha_{22})}{2^n \Gamma(\alpha_{11}) \cdot \Gamma(\alpha_{12} + \alpha_{21}) \cdot \Gamma(\alpha_{22}) \cdot \Gamma(N + \alpha_0)}, \tag{2.7}$$

where $n = n_{12} + n_{21}$. Finally, substituting (2.6) and (2.7) in (2.5) we obtain

$$BF(H_s, H_A) = \frac{\Gamma(\alpha_{12}) \cdot \Gamma(\alpha_{21}) \cdot \Gamma(n + \alpha_{12} + \alpha_{21})}{2^n \Gamma(\alpha_{12} + \alpha_{21}) \cdot \Gamma(n_{12} + \alpha_{12}) \cdot \Gamma(n_{21} + \alpha_{21})}, \tag{2.8}$$

which is not affected by main-diagonal cells or the corresponding parameters α_{11} and α_{22}. This is in agreement with the classical McNemar test, where the total sample size and the main-diagonal cells do not contribute to the inference about the difference $\pi_{i.} - \pi_{.i.}$ Consequently, in testing (2.4) is irrelevant whether we use $\mathbf{n} = (n_{11}, n_{12}, n_{21}, n_{22})$ or $\mathbf{n}^* = (n_{12}, n_{21}, n_{11} + n_{22})$ following a trinomial distribution with the Dirichlet prior on the vector of cell probabilities having parameters α_{12}, α_{21} and $\alpha_1 = \alpha_{11} + \alpha_{22}$ respectively. For reasons of simplicity from now on we shall use the $\mathbf{n}^*$ vector for the data representation.

Adopting a conditional approach, hypothesis (2.4) is equivalent to testing the hypothesis

$$H_0: \quad \pi^*_{12} = \frac{1}{2} \text{ towards } \quad H_1: \pi^*_{12} \neq \frac{1}{2},$$

where $\pi^*_{12} = \pi_{12}/(\pi_{12} + \pi_{21})$ is the conditional probability of an observation falling in cell (1,2) given that it will fall in the off-diagonal cells. A well-known test, conditional on the sum of the off-diagonal frequencies n, is based on the fact that n_{12} is distributed as binomial with parameters (n, π^*_{12}). It is straightforward to see that when (2.2) holds then the prior on π^*_{12} under H_1 is a Beta distribution with parameters α_{12} and α_{21}, denoted $Be(\alpha_{12}, \alpha_{21})$. In this context, the Bayes Factor $BF(H_0, H_1)$ is equal to the Bayes Factor $BF(H_s, H_A)$, which will be shortly denoted as BF_{01}. This is an anticipated result due, for example, to the results of Günel and Dickey (1974) on factorized likelihoods. Also it coincides with the standard frequentist result, that for large samples the conditional on n and the unconditional approaches lead to the same test.

For $\alpha_{12} = \alpha_{21} = 1$ (uniform prior on π^*_{12}), BF_{01} reduces to

$$BF_{01} = \frac{\Gamma(n+2)}{2^n \cdot \Gamma(n_{12}+1) \cdot \Gamma(n_{21}+1)} \tag{2.11}$$

which is the usual Bayes Factor for testing the null hypothesis that a binomial parameter equals to $\frac{1}{2}$; see, for example, Bernardo and Smith (1993, p.414).

2.3 *Prior choice.* We now discuss the choice of parameter values for the Dirichlet prior. Strong prior belief that $\pi_{1.} > \pi_{.1}$ is expressed by $\alpha_{12} >> \alpha_{21}$ while great prior uncertainty by $\alpha_{12} = \alpha_{21} = q, q \in [0, 1]$. The consideration of $\alpha_{12} = \alpha_{21}$ corresponds to strong prior preference for $\pi_{1.} = \pi_{.1}$, i.e., to prior belief for the symmetry model.

A rather more interesting issue is the quantification of prior parameters. If there exists prior information for the mean (μ_0) and the variance (σ_0^2)

of π_{12}^* i.e., for π_{12} conditional on the off-diagonal cells, then since $\pi_{12}^* \sim Be(\alpha_{12}, \alpha_{21})$, α_{12} and α_{21} can be computed from the beta mean and variance as follows

$$\alpha_{12} = \mu_0[\mu_0(1-\mu_0) - \sigma_0^2]/\sigma_0^2, \tag{2.12a}$$

and

$$\alpha_{21} = (1-\mu_0)[\mu_0(1-\mu_0) - \sigma_0^2]/\sigma_0^2. \tag{2.12b}$$

Since the initial problem is the comparison of the marginal proportions $\pi_{1.}$ and $\pi_{.1}$, it is realistic to assume that prior knowledge, if available, refers to these marginals rather than the off-diagonal proportions π_{12} and π_{21}. In order to incorporate information of this kind we need to transform it to information for the cells of the corresponding 2×2 table. From (2.2) we have that

$$\pi_{1.} \sim Be(\alpha_{1.}, \alpha_{2.}), \text{ and } \pi_{.1} \sim Be(\alpha_{.1}, \alpha_{.2}).$$

Thus, if the means and standard deviations of $\pi_{1.}$ and $\pi_{.1}$ are available, we can assign values to the beta parameters $\alpha_{1.}$, $\alpha_{2.}$, $\alpha_{.1}$ and $\alpha_{.2}$ as before, i.e., by (2.12). Now we need to express the α_{ij}'s, i,j=1,2 in terms of their marginals. In order to do this, besides the means and standard deviations of the correlated proportions $\pi_{1.}$ and $\pi_{.1}$, we also need their covariance, since $\alpha_{1.}$, $\alpha_{2.}$, $\alpha_{.1}$ and $\alpha_{.2}$ are linearly dependent. Let c be the value of $cov(\pi_{1.}, \pi_{.1})$ known in advance. We have

$$cov(\pi_{1.}, \pi_{.1}) = cov(\pi_{11}, \pi_{12}) + cov(\pi_{11}, \pi_{21}) + cov(\pi_{12}, \pi_{21}) + var(\pi_{11}).$$

Expressing the covariances in the second term of the above equation in terms of the α_{ij}'s (from the Dirichlet density) and after some algebra we obtain

$$\alpha_{11} = \alpha_0(\alpha_0 + 1)c + (\alpha_{1.}\alpha_{.1})/\alpha_0, \tag{2.16}$$

where $\alpha_0 = \alpha_{1.} + \alpha_{2.} = \alpha_{.1} + \alpha_{.2}$. The remaining α_{ij}'s are computed easily from their corresponding marginals. This implementation of prior knowledge on marginal proportions has some drawbacks. For example, if the two marginals have the same prior mean, then the beta parameters $(\alpha_{1.}, \alpha_{2.})$ and $(\alpha_{.1}, \alpha_{.2})$ are proportional, i.e. $\alpha_{1.}/\alpha_{2.} = \alpha_{.1}/\alpha_{.2} = k$ and this imposed a limitation on the relationship between the variances, namely $var(\pi_{1.})/var(\pi_{.1}) = (\alpha_0 + k^{-1})/(\alpha_0 + 1)$.

3. Empirical Bayes Analysis

In the case that no prior information is available, a different approach to the use of non-informative priors is to adopt an empirical Bayes analysis. For example, we may assume that the prior cell means k_{ij} in (2.3) follow a specific model and estimate them from the data. One approach is to specify α_0 by minimizing the total mean square error $MSE = E\left\{\Sigma_{i,j=1,2}(\tilde{\pi}_{ij} - \pi_{ij})^2\right\}$ (Fienberg and Holland, 1973, referred to this approach as pseudo-Bayes).

In the framework of $R \times C$ contingency tables, Agresti and Chuang (1989) estimated the k_{ij}'s by the uniform association model. In our context we could estimate them under the symmetry model. Thus we have

$$\hat{k}_{ij} = \frac{p_{ij} + p_{ji}}{2}, \quad i, j = 1, 2, \tag{3.1}$$

and minimization of the MSE given above, leads to

$$\alpha_0 = \frac{1 - \pi_{12}^2 - \pi_{21}^2}{(\pi_{12} - k_{12})^2 + (\pi_{21} - k_{21})^2} \tag{3.2}$$

which is estimated by

$$\hat{\alpha}_0 = \frac{2(1 - p_{12}^2 - p_{21}^2)}{(p_{12} - p_{21})^2}, \quad p_{12} \neq p_{21} \tag{3.3}$$

[for $\hat{\pi}_{ij} = p_{ij}$ and $\hat{k}_{ij}$ given by (3.1)], i.e., provided that the symmetry model does not fit the data perfectly. From (3.1) and (3.3), the parameters of the Dirichlet distribution are estimated by

$$\hat{\alpha}_{12} = \hat{\alpha}_{21} = \frac{(p_{12} + p_{21})(1 - p_{12}^2 - p_{21}^2)}{(p_{12} - p_{21})^2}, \quad p_{12} \neq p_{21} \tag{3.4a}$$

and

$$\hat{\alpha}_{ii} = p_{ii}\frac{2(1 - p_{12}^2 - p_{21}^2)}{(p_{12} - p_{21})^2}, \quad i = 1, 2, \quad p_{12} \neq p_{21}. \tag{3.4b}$$

If the symmetry model fits the data perfectly, (i.e. $p_{12} = p_{21}$), then $\hat{k}_{ij} = p_{ij}$ and

$$\hat{\alpha}_{ij} = p_{ij}, \quad i, j = 1, 2. \tag{3.5}$$

Finally, the empirical Bayes estimates of the cell probabilities are given by

$$\hat{\hat{\pi}}_{ij} = \frac{1}{2(\hat{\alpha}_0 + N)} \cdot [(\hat{\alpha}_0 + 2N)p_{ij} + \hat{\alpha}_0 p_{ji}]. \tag{3.6}$$

Of course, it is possible to model the diagonal cells as well, e.g. $\pi_{11} = 2\pi_{22}$, etc. In this case (3.1) through (3.5) would change accordingly. Note that the empirical Bayes approach takes into account the total sample size and therefore the size of the main-diagonal cells.

4. Approximate Highest Posterior Density Regions for the Difference of Two Correlated Proportions

Since the difference of the two correlated proportions $\pi_{1.} - \pi_{.1}$ is equal to $\pi_{12} - \pi_{21}$, we shall work on the latter difference. The construction of the approximate Highest Posterior Density (HPD) region is based on approximating the posterior density of (π_{12}, π_{21}) by a bivariate normal distribution. This approximation will be good if $(a_{12}+n_{12}), (a_{21}+n_{21})$ and (a_1+N-n) are sufficiently large. For other approximations see Lindley (1964) and Blotch and Watson (1967). For small sample sizes a Monte Carlo approach is preferable, since the it is easy to simulate from the Dirichlet distribution.

The fitted normal density will be centered at the mode of the posterior and its variance-covariance matrix will be the inverse of the curvature of the log-posterior density at the mode (cf. Gelman et al. 1998, p.100-101, 274). Hence, we need to evaluate the strict local maximum $\mathbf{m}_2 = (\theta_{12}, \theta_{21})^T$ and the variance-covariance matrix Σ at $\mathbf{m}_2$ of the log-posterior density of (π_{12}, π_{21}). Alternatively, we could approximate the Dirichlet posterior by a multivariate normal distribution having the same first and second order moments. This approximation is straightforward and is close to the one proposed here. But it does not lead to the classical Confidence Interval (CI) for the uniform choice on priors, as it will be shown to be true for the approximation adopted.

Expressing the data in the form $\mathbf{n}^* = (n_{12}, n_{21}, n_{11} + n_{22})$ and assuming that the prior of (π_{12}, π_{21}) is

$$(\pi_{12}, \pi_{21}) \;\sim\; Di(\alpha_{12}, \alpha_{21}, \alpha_1), \quad \alpha_{12}, \alpha_{21}, \alpha_1 > 0,$$

derived from (2.2) for $\alpha_1 = \alpha_{11} + \alpha_{22}$, the posterior density of (π_{12}, π_{21}) is

$$(\pi_{12}, \pi_{21})|\mathbf{n}^* \;\sim\; Di(\alpha_{12} + n_{12}, \alpha_{21} + n_{21}, \alpha_1 + N - n).$$

Setting the first partial derivatives of the log-posterior density $L(\pi_{12}, \pi_{21})$ with respect to π_{12} and π_{21} equal to zero, we obtain $\mathbf{m}_2 = (\theta_{12}, \theta_{21})^T$. The matrix of second order derivatives of $L(\pi_{12}, \pi_{21})$ with respect to π_{12} and π_{21} is Σ^{-1} and by matrix inversion we end up to the variance-covariance matrix

$$\Sigma = \Sigma(\theta_{12}, \theta_{21}) = d^2 \cdot \begin{pmatrix} \frac{B_{12}(1+cB_{21})}{1+c(B_{12}+B_{21})} & -\frac{cB_{12}B_{21}}{1+c(B_{12}+B_{21})} \\ -\frac{cB_{12}B_{21}}{1+c(B_{12}+B_{21})} & \frac{B_{21}(1+cB_{12})}{1+c(B_{12}+B_{21})} \end{pmatrix},$$

where

$$\begin{aligned} \theta_{ij} &= d \cdot B_{ij} \\ B_{ij} &= \alpha_{ij} + n_{ij} - 1 \quad , i, j = 1, 2; i \neq j \\ A_{ij} &= \alpha_1 + \alpha_{ij} + N - n_{ji} - 2 \\ c &= \frac{(\alpha_1 + N - n - 1)^3}{[A_{12}A_{21} - B_{12}B_{21} - (\alpha_1 + N - n - 1)(B_{12} + B_{21})]^2} \\ d &= \frac{\alpha_1 + N - n - 1}{A_{12}A_{21} - B_{12}B_{21}}. \end{aligned}$$

Hence, *the normal approximation of the posterior distribution of* (π_{12}, π_{21}) *given* $\mathbf{n}^*$, *based on its mode, is given by* $N(\mathbf{m}_2, \Sigma)$.

¿From this approximation it follows that $(\pi_{12} - \pi_{21})|\mathbf{n}^* \sim N(\theta_{12} - \theta_{21}, \sigma^2)$, where $\sigma^2 = \frac{(B_{12}+B_{21}+4cB_{12}B_{21})d^2}{1+c(B_{12}+B_{21})}$. Consequently, the approximate HPD region for the difference $\pi_{12} - \pi_{21}$ is

$$d \cdot \left\{ (B_{12} - B_{21}) \pm z_{a/2} \sqrt{\frac{B_{12} + B_{21} + 4cB_{12}B_{21}}{1 + c(B_{12} + B_{21})}} \right\}.$$

In the case of $\alpha_{12} = \alpha_{21} = \alpha_1 = 1$ (uniform prior) the strict local maximum is simplified to $\mathbf{m}_2^0 = (n_{12}/N, n_{21}/N)^T$ and the variance-covariance matrix equals

$$\Sigma^{(0)}(\theta_{12}, \theta_{21}) = N^{-2} \cdot \begin{pmatrix} \frac{n_{12}(N-n_{12})}{N} & -\frac{n_{12}n_{21}}{N} \\ -\frac{n_{12}n_{21}}{N} & \frac{n_{21}(N-n_{21})}{N} \end{pmatrix}.$$

Thus $(\pi_{12} - \pi_{21}) \sim N\left(\frac{n_{12}-n_{21}}{N}, \sigma_0^2\right)$ with $\sigma_0^2 = \frac{1}{N^2} \cdot \left[n - \frac{(n_{12}-n_{21})^2}{N}\right]$, which is the standard approximation of the frequentists' approach and leads to the classical $(1-\alpha)\%$ CI for the difference of correlated proportions [cf. Snedecor and Cochran 1971, Llyod 1990].

5. A Logistic Regression Approach

The Bayesian approach for the models described in the previous sections has been used to smooth the estimates of the cell probabilities. It is evident that this shrinking of the observed probabilities towards some prespecified model does not affect the symmetry hypothesis testing. In this section we propose the use of logistic regression to model correlated proportions assuming a subject (pair) effect as well. We are driven by the desire to propose a Bayesian model uncertainty methodology which investigates the existence of a symmetry structure without ignoring the main-diagonal cell counts. From a classical perspective, this has been achieved in the work of Lesperance and Kalbfleisch (1994) and Tibshirani and Redelmeier (1997). The idea is that instead of using the usual logistic regression model for modeling correlated proportions (see, for example, Cox and Snell 1989), we use a mixture model, similar to one of the models suggested by Altham (1971), as follows.

Assume that we have N binary pairs of observations, $Y_{i1}, Y_{i2}, i = 1, \ldots, N$, and that

$$p_{ij} = P(Y_{ij} = 1|\beta_i, \theta) = [1 + \ \exp\ (-\beta_i - (j-1)\theta]^{-1}, j = 1, 2,$$

or equivalently

$$\text{logit}\ (p_{ij}) = \beta_i + (j-1)\theta, \quad j = 1, 2. \tag{5.1}$$

By testing the hypothesis $\theta = 0$ the above model formulation leads to the McNemar test. A Bayesian hierarchical model that incorporates a random effect component is

$$\text{logit}\ (p_{ij}) = \mu + (j-1)\theta + \epsilon_i, \quad j = 1, 2, \tag{5.2a}$$

$$\begin{bmatrix} \mu \\ \theta \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right) \tag{5.2b}$$

$$\epsilon_i \sim N(0, \sigma^2), \quad \sigma^2 \sim IG(\lambda_1, \lambda_2), \tag{5.2c}$$

where IG denotes the Inverse Gamma density.

Note that (5.2) differs from (5.1) in that in (5.2) the estimates of p_{ij} borrow strength from all observations through the random effect expressed in (5.2c).

Implementation of model (5.2) is only feasible by using numerical approximation techniques. In particular, Markov Chain Monte Carlo (MCMC) algorithms can converge very rapidly by using Gibbs sampling steps and exploiting the log-concavity of the resulting full conditional posterior densities; see Dellaportas and Smith (1993). Moreover, it is clear that (5.2) allows us to determine our uncertainty about the symmetry model via the posterior model probabilities derived from the MCMC output; see, for example, Dellaportas et al. (2000, 2001). In particular, as pointed out by these authors, a simple way to produce posterior model probabilities based on a Gibbs sampling Markov chain is to formulate, when possible, the model choice as a variable selection problem. Thus, (5.2) can be written as

$$\text{logit } (p_{ij}) = \mu + \gamma(j-1)\theta + \epsilon_i, \quad j = 1, 2. \tag{5.3a}$$

$$\begin{bmatrix} \mu \\ \theta \end{bmatrix} \mid \gamma \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & k^{(1-\gamma)}\sigma_2^2 \end{pmatrix} \right) \tag{5.3b}$$

$$\epsilon_i \sim N(0, \sigma^2), \quad \sigma^2 \sim IG(\lambda_1, \lambda_2), \quad \gamma \sim Bernoulli(0.5) \tag{5.3c}$$

for some constant k. We have replaced $(j-1)$ in (5.2a) with $\gamma(j-1)$ in (5.3a). The idea is that the posterior distribution of γ represents the posterior probability that $\theta \neq 0$. Note that Gibbs variable selection is a special case of reversible jump MCMC (Green, 1995). For details, see Appendix and Ntzoufras (1999).

In (5.3), the prior for θ consists of a mixture of two densities. The first ($\gamma = 0$) is a pseudoprior, which does not affect the posterior density. Its role is simply to facilitate the convergence of the MCMC algorithm. A choice of k that has found to be rather good in practice is $k = 10$. For related ideas for the use of pseudopriors in MCMC see Carlin and Chib (1995). The hierarchical setup (5.3) allows us to obtain the relevant information about our uncertainty for the symmetry model through the marginal posterior density of γ. Essentially, this posterior marginal updates our prior belief ($Pr[\gamma = 1] = 1/2$)) through the data.

We shall assume that no strong prior information is available so a diffuse prior distribution for μ and θ is needed. This choice may have a considerable impact on the posterior model probabilities so we proceed to discuss it in some detail. Clearly, for the models of interest in this section, the key element is the variance of θ, but it is more convenient to specify the joint prior of μ and θ given that $\gamma = 1$. We adopt ideas from Kass and Wasserman (1995) who suggest the use of prior variances which correspond to unit prior information for a 'reference' Bayesian comparison of nested models.

In model (5.3) the information matrix for (μ, θ) is

$$I(\mu, \theta) = X^T W X \tag{5.4}$$

where X is a design matrix of dimension $2N \times 2$ and $W = diag(p_{ij}(1-p_{ij}))$. If we represent our data in the usual form of a 2×2 table presented in Section 2.1, then the second column of X contains $n_{12}+n_{11}$ ones and $2N-n_{12}-n_{11}$ zeroes. The unit information matrix is given by $I(\mu,\theta)/N$, and hence a maximum unit information of the experiment can be estimated by

$$\frac{I(\mu,\theta)}{N} \leq \frac{1}{4N} X^T X = \frac{1}{4N}\begin{bmatrix} 2N & n_{11}+n_{12} \\ n_{11}+n_{12} & n_{11}+n_{12} \end{bmatrix} \leq \frac{1}{4}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

so that we propose to use in (5.2)

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = 4\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

We also use a vague prior for σ^2 with $\lambda_1 = \lambda_2 = 10^{-4}$ and $k = 10$ which turned out to be satisfactory for our models.

6. Examples

To test our logistic regression model we applied it in two sets of simulated data appearing in Table 1. We chose two different total sample sizes N=20 and N=100, keeping the main-diagonal cells equal and the off-diagonal cells constant.

Table 1. Simulated data

Example 1		Example 2	
2	3	42	3
13	2	13	42

Gibbs sampler had a rapid convergence in both examples. The results, presented in Table 2, are based on ergodic averages of a sample of 20,000 iterations after discarding the first 10,000 iterations as a burn-in. The program ran with BUGS software and it took 4 and 19 seconds per 1000 iterations in a Pentium II PC for Examples 1 and 2 respectively.

Table 2 indicates that the posterior probability of $\theta \neq 0$ is 0.99 and 0.87 for Examples 1 and 2 respectively. The posterior probability of γ decreased in Example 2 because there is more support for the symmetry model due to the increased sample size. We also calculated the Bayes Factor (2.8) using the data above. For representative choices for the Dirichlet prior parameters $(\alpha_{12}, \alpha_{21}, \alpha_1) = (1, 1, 1), (0.5, 0.5, 0.5)$ and $(0.25, 0.25, 0.25)$, the resulting values of BF_{01}, which are equal for both data sets, are $0.15, 0.18$ and 0.26 respectively and indicate rejection of $\theta = 0$ based on the Kass and Raftery table. Note that for $N = 20$ the estimated correlated proportions are $\pi_{1\cdot} = 0.25$ and $\pi_{\cdot 1} = 0.75$, having a difference of -.50, while for $N = 100$ they change to $\pi_{1\cdot} = 0.45$ and $\pi_{\cdot 1} = 0.55$, with difference -.10, whereas the BF remains unaltered. The estimated difference $\pi_{1\cdot} - \pi_{\cdot 1}$ would reduce further if we further increased the main-diagonal cell counts. This (possibly undesirable) feature does not appear in the hierarchical logistic regression model.

Table 2. Estimates of posterior mean, standard deviation, 2.5% and 97.5% quantiles

	Example 1				Example 2			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
θ	- 2.19	0.75	-3.70	-0.79	-1.35	0.84	-0.78	2.33
γ	0.99	0.07	–	–	0.87	0.34	–	–

7. Conclusions

We have presented a thorough investigation of the analysis of correlated proportions under a Bayesian perspective. Our investigation contained conjugate Bayesian analysis, empirical Bayes analysis and determination of model uncertainty through Bayes factors, evaluation of asymptotic posterior intervals, and numerical evaluation of posterior model probabilities through a hierarchical logistic setup. For testing the equality of correlated proportions, the latter approach as well as the empirical Bayes approach take into account the total sample size and, therefore, the main-diagonal cell counts and could be considered preferable to the classical McNemar test or the Dirichlet based Bayes Factors.

Appendix

We show here that the Gibbs variable selection algorithm (Dellaportas et al., 2001) is a special case of the reversible jump algorithm suggested by Green (1995). First, note that Gibbs variable selection consists of two Gibbs steps that update (μ, θ) and γ. If we denote all data by Y, the required conditional densities of (μ, θ) given γ are:

$$\begin{aligned} f(\mu,\theta|Y,\gamma=1) &\propto f(Y|\mu,\theta,\gamma=1)\cdot f(\mu,\theta|Y,\gamma=1) \\ f(\mu,\theta|Y,\gamma=0) &\propto f(Y|\mu,\gamma=0)\cdot f(\mu,\theta|\gamma=0) \end{aligned}$$

and γ is updated through

$$\frac{f(\gamma=1|\mu,\theta,Y)}{f(\gamma=0|\mu,\theta,Y)} = \frac{f(Y\gamma|\mu,\theta,\gamma=1)}{f(Y|\mu,\gamma=0)} \cdot \frac{f(\mu,\theta|\gamma=1)}{f(\mu,\theta|\gamma=0)} \cdot \frac{f(\gamma=1)}{f(\gamma=0)}. \qquad (A1)$$

Assume now that we wish to use a reversible jump algorithm to obtain posterior model probabilities between models m_0 (corresponding to $\gamma = 0$) and m_1 (corresponding to $\gamma = 1$). If we are, for example, in model m_0 with current parameter value μ, the required steps are:

1. Propose model m_1 with probability $j(m_0, m_1)$.
2. Generate u (of dimension 1) from a proposal density $q(u|\mu, m_0, m_1)$.
3. Set $(\mu', \theta') = (\mu, u)$.
4. Accept the move with probability

$$\alpha = \min\left\{1, \frac{f(Y|\mu',\theta',m_1)}{f(Y|\mu,m_0)} \cdot \frac{f(\mu',\theta'|m_1)}{f(\mu|m_0)} \cdot \frac{f(m_1)}{f(m_0)} \cdot \frac{j(m_1,m_0)}{j(m_0,m_1)} \cdot \frac{1}{q(u|\mu,m_0,m_1)}\right\}.$$

Now first note that by setting $q(u|\mu, m_0, m_1) = f(\theta|\mu, \gamma = 0)$ we obtain

$$\alpha = \min\left\{1, \frac{f(Y|\mu',\theta',m_1)}{f(Y|\mu,m_0)} \cdot \frac{f(\mu',\theta'|m_1)}{f(\mu,\theta|m_0)} \cdot \frac{f(m_1)}{f(m_0)} \cdot \frac{j(m_1,m_0)}{j(m_0,m_1)}\right\}$$

with (μ', θ') being generated by $f(\mu, \theta|Y, \gamma = 0)$. This is clearly just a Metropolis step on the discrete random variable γ, of which a special case (by setting $j(m_1, m_0)$ and $j(m_0, m_1)$ equal to the full conditional densities) is the Gibbs sampler specified in $(A1)$. Thus, the Gibbs variable selection is obtained from the reversible jump algorithm by choosing an appropriate proposal $q(u|\mu, m_0, m_1)$ and by performing a Gibbs sampling rather than a Metropolis Hasting step for the model indicator. The proof for the reverse move is similar.

Acknowledgements. The first two authors have been supported by the Hellenic Secretariat of Research and Technology [Project PENED-95, No.377].

The last author was supported by the European Union's research nertwork 'Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0095'.

References

AGRESTI, A. (1990). *Categorical Data Analysis*, Wiley, New York.

AGRESTI, A and CHUANG, C. (1989). Model-based Bayesian methods for estimating cell proportions in cross-classification tables having ordered categories, *Computational Statistics and Data Analysis*, **7**, 245-258.

ALTHAM, P.M.E. (1971). The analysis of matched proportions, *Biometrika*, **58**, 561-576.

BERNARDO, J.M. and SMITH, A.F.M. (1993). *Bayesian Theory*, Wiley, New York.

BLOTCH, D.A. and WATSON, G.S. (1967). A Bayesian study of the multinomial distribution. *Annals of Mathematical Statistics*, **38**, 1423-1435.

CARLIN, B.P. and CHIB, S. (1995). Bayesian model choice via Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society, Series B*, **157**, 473-484.

COCHRAN, W.G. (1950). The comparison of percentages in matched samples, *Biometrika*, **37**, 256-266.

COX, D.R. and SNELL, E.J. (1989). *Analysis of Binary Data*, Champman & Hall, London.

DELLAPORTAS P, FORSTER J.J. and NTZOUFRAS, I. (2000). Bayesian model and variable selection for generalised linear models. In *Generalised Linear models: A Bayesian perspective*, eds. D.K. Dey, S. Ghosh and B. Mallick, Marcel Dekker, New York, 273-286.

DELLAPORTAS, P., FORSTER, J.J., NTZOUFRAS, I. (2001). On Bayesian model and variable selection using MCMC, *Statistics and Computing*, **12**, 27-36.

DELLAPORTAS, P. and SMITH, A.F.M. (1993). Bayesian inference for generalised linear and proportional hazards models via Gibbs Sampler, *Applied Statistics*, **42**, 443-459.

FIENBERG, S.E. and HOLLAND, P.W. (1973). Simultaneous estimation of multinomial cell probabilities, *Journal of the American Statistical Association*, **68**, 683-691.

GELMAN, A., CARLIN, J.B., STERN, H.S. and RUBIN, D.B. (1998). *Bayesian Data Analysis*, Chapman & Hall, London.

GREEN, P. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination, *Biometerika*, **82**, 711-732.

GÜNEL, E. and DICKEY, J. (1974). Bayes factors for independence in contingency tables, *Biometrika*, **61**, 545-557.

KASS, R.E. and RAFTERY, A.E. (1995). Bayes Factors, *Journal of the American Statistical Association*, **90**, 773-795.

KASS, R.E. and VAIDYANATHAN, S.K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions, *Journal of the Royal Statistical Society, Series B*, **54**, 129-144.

KASS, R.E. and WASSERMAN, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwartz criterion, *Journal of the American Statistical Association*, **90**, 928-934.

LESPERANCE M. and KALBFLEISCH J. (1994). Mixture models for matched pairs, *Canadian Journal of Statistics*, **22**, 65-74.

LIANG, K.Y. and ZEGER, S.L. (1988). On the use of concordant pairs in matched case-control studies, *Biometrics*, **44**, 1145-1156.

LLYOD, C.J. (1990). Confidence intervals from the difference between two correlated proportions, *Journal of the American Statistical Association*, **85**, 1154-1158.

LINDLEY D.V. (1964). The Bayesian analysis of contingency tables, *Annals of Mathematical Statistics*, **35**, 1622-1643.

MAY, W.L. and JOHNSON, W.D. (1997). The validity and power of tests for equality of two correlated proportions, *Statistics in Medicine*, **16**, 1081-1096.

MCNEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika*, **12**, 153-157.

NTZOUFRAS, I. (1999). *Aspects of Bayesian model and variable selection using MCMC.* Ph.D. Thesis, Department of Statistics, Athens University of Economics and Business, Athens, Greece.

NURMINEN, M. and MUTANEN, P. (1987). Exact Bayesian analysis of two proportions, *Scandinavian Journal of Statistics*, **14**, 67-77.

SNEDECOR, G.W. and COCHRAN, W.G. (1971). *Statistical Methods*, Iowa State University Press, Iowa, USA.

TIBSHIRANI, R and REDELMEIER, D.A. (1997). Cellular telephones and motor-vehicle collisions: Some variations on matched-pairs analysis, *Canadian Journal of Statistics*, **25**, 581-591.

WALTERS, D.E. (1986). On the reliability of Bayesian confidence limits for a difference of two proportions, *Biometrical Journal*, **28**, 337-346.

MARIA KATERI
DEPT. OF PHILOSOPHY - EDUCATION - PSYCHOLOGY
UNIVERSITY OF IOANNINA
IOANNINA 45110, GREECE
E-mail: me00126@cc.uoi.gr

TAKIS PAPAIOANNOU
DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE
UNIVERSITY OF PIREAUS
PIREAUS-18535, GREECE
E-mail: takpap@unipi.gr

PETROS DELLAPORTAS
DEPARTMENT OF STATISTICS
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
76 PATISSION STR.
10434 ATHENS, GREECE
E-mail: petros@aueb.gr