Alternative Tests of Independence in Two-Way Categorical Tables

Balgobin Nandram[1] and Jai Won Choi[2]
[1]*Worcester Polytechnic Institute and* [2]*National Center for Health Statistics*

Abstract: The chi-squared test for independence in two-way categorical tables depends on the assumptions that the data follow the multinomial distribution. Thus, we suggest alternatives when the assumptions of multinomial distribution do not hold. First, we consider the Bayes factor which is used for hypothesis testing in Bayesian statistics. Unfortunately, this has the problem that it is sensitive to the choice of prior distributions. We note here that the intrinsic Bayes factor is not appropriate because the prior distributions under consideration are all proper. Thus, we propose using Bayesian estimation which is generally not as sensitive to prior specifications as the Bayes factor. Our approach is to construct a 95% simultaneous credible region (i.e., a hyper-rectangle) for the interactions. A test that all interactions are zero is equivalent to a test of independence in two-way categorical tables. Thus, a 95% simultaneous credible region of the interactions provides a test of independence by inversion.

Key words: Bayes factor, chi-squared statistic, interactions, intra-class correlation, prior distribution, simultaneous inference.

1. Introduction

There are many occasions when we need to understand the extent of the association of two attributes. For example, the National Center for Health Statistics has been collecting data on obesity in the U.S. for many years. The data have been used to establish the relation between obesity and socio-demographic variables. These data are typically presented in two-way categorical tables. Scientists routinely use the chi-squared test to analyze such tables. However, in many applications the chi-squared test can be defective; one example is when there is an intra-class correlation violating the assumptions in the multinomial distribution. This paper reviews shortcomings of methods based on adjusted chi-squared statistic and the Bayes factor, an alternative in Bayesian hypothesis testing to the chi-squared statistic, and we propose a simple method based on estimation rather than hypothesis testing to "test" for independence in two-way categorical tables.

We consider the analysis of data summarized in a $r \times c$ categorical table (i.e., there are two attributes, the first with r levels and the second with c levels). The dataset is a sample from a population, and the individuals in the sample are categorized according to the two attributes. Let π_{jk} denote the probability that an individual falls in the jth level of the first attribute and k^{th} level of the second attribute. Here $\sum_{j=1}^{r} \sum_{k=1}^{c} \pi_{jk} = 1$. The two attributes are not associated if $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}$, $j = 1, \ldots, r$; $k = 1, \ldots, c$; otherwise they are associated. When a simple random sample of individuals is taken from the population, the sampled individuals can be allocated to the cells of the $r \times c$ categorical table (multinomial sampling) to obtain a chi-squared test of association between the two categories based on the data collected.

Traditionally, there are two ways to test for association (or no association) in a $r \times c$ categorical table. First, we can use the well-known Pearson chi-squared statistic which works for simple random sampling (i.e., multinomial sampling). The second approach is to use the Bayes factor (Kass and Raftery 1995), an alternative to the chi-squared test. In multinomial sampling the individuals fall in the cells independently, and in this case the standard chi-squared test for association between the two categories of $r \times c$ table is correct asymptotically. There is a correlation because the sample size is fixed. However, if the counts in the cells are formed from members in a cluster, the assumptions of multinomial sampling (i.e., independence) no longer hold because there is an additional correlation among the members in the cluster (i.e., the intra-class correlation).

Thus, the standard chi-squared test is inappropriate, especially for borderline cases of significance, when there is an intra-class correlation. For members within the same cluster (e.g., children in the same family, rats in the same litter, students in the same class), one can expect a positive correlation among the members of the cluster because the members will tend to have similar trait or share the same effect. This is the intra-class correlation (i.e., correlation among the members inside the cluster). For data with an intra-class correlation, the information available from all the members within the cluster is effectively less than the information when there is independence among the members within the cluster. Thus, with an intra-class correlation the effective sample size is smaller than the cluster size, so that there is an increase in variability when there is an intra-class correlation. For a simple nonparametric method to calculate the intra-class correlation see Rao (1965, p. 159).

Several authors have recognized inaccuracy in the analysis when the usual chi-squared test is applied to correlated data; efforts to correct for spurious inflation in such test statistics have been based mainly on two approaches. The design-based approach provides inference with respect to the asymptotic sampling distribution of estimates over repetitions of the sample design (Fellegi 1980,

Holt, Scott and Ewings 1980, Rao and Scott 1981, 1984, Bedrick 1983, and Fay 1985). They use design effect which is the ratio of two variances of an appropriate estimator, one from a more complex design than simple random sampling and the other from simple random sampling. For example, Rao and Scott (1981, 1984) investigate the effects of stratification and clustering on the asymptotic distribution of Pearson's chi-squared statistic for goodness of fit and independence in multi-way categorical tables. They propose generalized design effects which are used to adjust the standard chi-squared statistic. The model-based approach postulates a probability distribution to model the sample data (Altham 1976, Cohen 1976, Brier 1980, Fienberg 1979, and Choi and McHugh 1989). For example, Choi and McHugh (1989), applying the probabilistic development in Altham (1976) and Cohen (1976), shows how to adjust the standard chi-squared test statistic when there is an intra-class correlation. In the Bayesian approach these adjustments are not necessary because appropriate Bayesian models can be constructed to capture unusual features in the data.

For $r \times c$ categorical tables the Bayes factor is used to quantify the difference between a model with association and one without. This is the ratio of the posterior odds of one model to the other to their prior odds, and it is the same as the ratio of the marginal likelihoods of the data under two models, one without association and the other with association. The Bayes factor has recently been of much scientific interest. However, there are two important difficulties with the use of the Bayes factor. First, it is sensitive to the prior specifications, especially when there are not enough data to estimate the parameters under test; see Sinharay and Stern (2002) for an interesting discussion on nested models. Second, but less important is that we need to be careful with its interpretation; see Lavine and Schervish (1999). We discuss these issues in detail in Appendix A.

It is natural to think about alternative approaches when the cell counts do not follow a multinomial distribution, or some cell counts are small or zero. A Bayesian approach to the problem is desirable especially when the assumption of multinomial distribution does not hold. In this case, one can use the Bayes factor for testing association versus no association. The purpose of this paper is to alert practitioners of the improper routine use of the chi-squared test or Bayes factor, and to suggest alternatives. Thus, the reasons for writing this paper are:

a. To review some of the defects of standard chi-squared testing;

b. To show that the Bayes factor is defective as an alternative to the chi-squared test because it is sensitive to prior specifications;

c. To construct a Bayesian estimative alternative to the Bayes factor based on the interactions in the $r \times c$ table.

This paper has five more sections. Section 2 has a review on chi-squared test and Bayesian test. In Section 3 we present the Bayesian estimative alternative to the test of independence. Section 4 has a review of chi-squared testing and intra-class correlation and a simple implementation of the estimative procedure. Section 5 shows several examples which are related to activity limitation and age, forms of material and heat in fire incidences, and body mass index and family income. Section 6 has a discussion.

2. Chi-squared Test and Bayesian Test: A Review

In this section we give a detailed quantitative review of the chi-squared statistic and the Bayes factor for tests of association between the two categorical variables in a $r \times c$ categorical table.

2.1 Standard chi-squared test

Let n_{jk} denote the number of individuals in the jth row and k^{th} column of the $r \times c$ categorical table. Also let $n_{j.} = \sum_{k=1}^{c} n_{jk}$, $j = 1, \ldots, r$; $n_{.k} = \sum_{j=1}^{r} n_{jk}$, $k = 1, \ldots, c$; $n = \sum_{j=1}^{r} \sum_{k=1}^{c} n_{jk}$ and $e_{jk} = n_{j.} n_{.k} / n$, $j = 1, \ldots, r$; $k = 1, \ldots, c$. Then, Pearson's chi-squared statistic, under independence of the row and column classifications, is

$$X_u = \sum_{j=1}^{r} \sum_{k=1}^{c} (n_{jk} - e_{jk})^2 / e_{jk}.$$

If the responses from the individual members are independent and identically distributed, then as n goes to infinity, X_u converges in law to a chi-squared random variable with $(r-1)(c-1)$ degrees of freedom. In practice, the validity of the chi-squared test depends on (a) the magnitude of the expected values e_{jk}, and (b) whether the cell counts $(n_{jk}, j = 1, \ldots, r; k = 1, \ldots, c)$ follow a multinomial distribution given the sample size n (i.e., the individual responses are independent and identically distributed). In (a) the test is valid if the e_{jk} are all larger than 5; see Greenwood and Nikulin (1996, Chapter 1, Section 2). Clearly the only way to achieve this is to increase the sample size subject to cost. In (b) when there is correlation among the members (e.g., intra-class correlation), the asymptotic distribution of X_u is no longer $\chi^2_{(r-1)(c-1)}$, and the estimates of the cell proportions can be inaccurate. The Pearson chi-squared test has received much attention. See Mirkin (2001) for a review of interpretations of the chi-squared statistic as a measure of association or independence.

We describe one solution that has been proposed for the problem about the asymptotic distribution when sampling is not simple random sampling. Let n_t denote the number of members in all families of the same size $t = 1, \ldots, T$, and

let θ_t denote the intra-class correlation for clusters of size t ($\theta_1 \equiv 0$). First, we describe the method of Rao and Scott (1981). Omitting redundancy, let V and W be the covariance matrices of the estimators of the cell proportions under the assumption of simple random sampling and a "complex" sample design respectively.

The design effects are the eigen-values of the matrix $V^{-1}W$, and the adjusted chi-squared statistic is χ_u^2/a, where a is the average of the eigen-values. Motivated by Rao and Scott (1981), for cluster sampling Choi and McHugh (1989) derives the following adjusted chi-squared statistic

$$X_a = X_u\{1 + n^{-1}\sum_{t=1}^{T}(t-1)n_t\hat{\theta}_t\}^{-1},$$

where $\hat{\theta}_t$ is the maximum likelihood estimator of θ_t under their model. The statistic $X_a \sim \chi^2_{(r-1)(c-1)}$ asymptotically, an improvement over X_u (i.e., more accurately $\chi^2_{(r-1)(c-1)}$). The p-value corresponding to the adjusted chi-squared statistic will be larger. For weighted data they divide X_a by the average weight. We note that sometimes X_a is difficult to compute because W itself is difficult to compute.

2.2 Bayes factor

We now discuss the Bayes factor as an alternative to the chi-squared test. If two models, M_0 and M_1, are fit to data y, the Bayes factor for comparing models M_1 and M_0 is defined as the ratio of the marginal likelihoods of the data $\underset{\sim}{y}$ as

$$B_{10} = \frac{p(\underset{\sim}{y}|M_1)}{p(\underset{\sim}{y}|M_0)} \quad \text{with} \quad p(\underset{\sim}{y} \mid M_k) = \int p(\underset{\sim}{y}|\underset{\sim}{\theta}_k,M_k)p(\underset{\sim}{\theta}_k \mid M_k)d\underset{\sim}{\theta}_k, \ k = 0, 1,$$

where $\underset{\sim}{\theta}_k$ is the parameter vector under M_k, $p(y|\underset{\sim}{\theta}_k,M_k)$ is the probability density (or mass) function and $p(\underset{\sim}{\theta}_k \mid M_k)$ is the prior density. For example, in our application M_0 is the model of no association and M_1 is the model of association, or vice versa. The Bayes factor summarizes the evidence provided by the data in favor of one scientific hypothesis M_1 relative to another M_0. Kass and Raftery (1995) gave a comprehensive description of Bayes factors including their interpretation. For example, if $0 \leq \log_e(B_{10}) < 1$, the evidence against M_0 is "not worth more than a bare mention"; if $1 \leq \log_e(B_{10}) < 3$, the evidence against M_0 is "positive"; if $3 \leq \log_e(B_{10}) < 5$, the evidence against M_0 is "strong"; and if $\log_e(B_{10}) \geq 5$, the evidence against M_0 is "very strong".

For the $r \times c$ categorical table, we can consider two multinomial-Dirichlet models, one with association and the other with no association. The model with

association is $n \mid \pi \sim$ Multinomial(n, π) and $\pi \sim$ Dirichlet$(1, \ldots, 1)$. Thus, for the model with no association, with $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}$, $j = 1, \ldots, r$; $k = 1, \ldots, c$; $n \mid \pi^{(1)}, \pi^{(2)} \sim$ Multinomial(n, π), $\pi^{(1)} \sim$ Dirichlet$(1, \ldots, 1)$, and independently $\pi^{(2)} \sim$ Dirichlet$(1, \ldots, 1)$, where $\pi^{(1)}$ and $\pi^{(2)}$ have r and c components respectively. It is easy to show that the marginal likelihood with association (as) is $p_{as}(n) = (rc-1)!n!/(n+rc-1)!$ and with no association (nas) is

$$p_{nas}(n) = p_{as}(n) \frac{(r-1)!(c-1)!}{(rc-1)!} \frac{(n+rc-1)!}{(n+r-1)!(n+c-1)!} \frac{\prod_{j=1}^{r} n_{j\cdot}! \prod_{k=1}^{c} n_{\cdot k}!}{\prod_{j=1}^{r} \prod_{k=1}^{c} n_{jk}!}.$$

Unfortunately, with this approach, the Bayes factor is sensitive to the prior specification (e.g., $\pi \sim$ Dirichlet(1)). It is well-known that $\pi \sim$ Dirichlet$(1/2, 1/2, \ldots, 1/2)$ will give different inference from $\pi \sim$ Dirichlet(1); e.g., see Nandram, Cox and Choi (2005), Nandram and Choi (2006) and the examples in Section 5. Thus, we consider estimation theory (credible intervals) to form a test (i.e., an inverted interval).

How sensitive is the Bayes factor to the choice of the prior distributions? First, note that the prior density, any reasonable person might use in this problem, is the Dirichlet distribution. For the model with association we have selected the prior distributions to be $\pi \sim$ Dirichlet(κ), and for the model with no association $\pi^{(1)} \sim$ Dirichlet$(\kappa^{(1)})$ and independently $\pi^{(2)} \sim$ Dirichlet$(\kappa^{(2)})$. Let $n_{j\cdot}^{(1)} = \sum_{k=1}^{c} n_{jk}$, $j = 1, \ldots, r$ and $n_{\cdot k}^{(2)} = \sum_{j=1}^{r} n_{jk}$, $k = 1, \ldots, c$. Then, it is easy to show that the Bayes factor for a test of association versus no association is

$$BF = \frac{D_{rc}(n + \kappa)/D_r(n_{\cdot}^{(1)} + \kappa^{(1)})D_c(n_{\cdot}^{(2)} + \kappa^{(2)})}{D_{rc}(\kappa)/D_r(\kappa^{(1)})D_c(\kappa^{(2)})},$$

where generically $D_k(ct) = \{\prod_{j=1}^{k} \Gamma(c_j t)\}/\Gamma(t)$ is the Dirichlet function with $t > 0$, $c_j > 0$, $\sum_{j=1}^{k} c_j = 1$. Then, we can choose each of the components of κ, $\kappa^{(1)}$ and $\kappa^{(2)}$ to be δ [e.g., in $p_{as}(n)$ and $p_{nas}(n)$, $\delta = 1$]. Sensitivity to the choice of prior distributions can be studied in terms of δ. Here $\delta = 1$ corresponds to the uniform prior distribution and $\delta = .50$, Jeffreys' prior; these are "noninformative" priors usually used in the multinomial-Dirichlet model. We will call a study of BF as a function of δ a *sensitivity analysis*. In a small sensitivity analysis we can study the behavior of BF at $\delta = 0.1, 0.5, 1., 1.5, 2, 3$ with illustrative example.

3. The Bayesian Estimative Alternative to the Test of Independence

Our test for independence is simple, and our discussion consists of two parts. First, we state and prove a theorem which shows that the test of association is equivalent to a test that the $(r-1)(c-1)$ two-way interactions are all 0.

Second, rather than using the Bayes factor for hypothesis testing, we obtain a 95% simultaneous credible region, a hyper-rectangle, for the $(r-1)(c-1)$ interaction effects in the two-way table, and then we check to see whether each interval contains 0. Thus, we consider the problem of estimating the effects in a $r \times c$ categorical table.

Our basic model is

$$\underset{\sim}{n} \mid \underset{\sim}{\pi} \sim \text{Multinomial}(n, \underset{\sim}{\pi}),$$

where $\underset{\sim}{n} = \{n_{jk}, j = 1, ..., r, \ k = 1, ..., c\}$ is the vector of cell counts. A priori we take $\underset{\sim}{\pi} \sim \text{Dirichlet}(\underset{\sim}{1})$, where $\underset{\sim}{1}$ is a rc vector of ones. It is well-known that a posteriori

$$\underset{\sim}{\pi} \mid \underset{\sim}{n} \sim \text{Dirichlet}(\underset{\sim}{n} + \underset{\sim}{1}),$$

where $\underset{\sim}{1}$ is a vector of ones.

Now, we include the effects in the $r \times c$ table by taking

$$\pi_{jk} = A e^{\alpha_j + \beta_k + \gamma_{jk}}, \ j = 1, \ldots, r; \ k = 1, \ldots, c,$$

where $A^{-1} = \sum_{j=1}^{r} \sum_{k=1}^{c} e^{\alpha_j + \beta_k + \gamma_{jk}}$ with the corner-point restrictions $\alpha_r = \beta_c = \gamma_{jc} = \gamma_{rk} = 0, \ j = 1, \ldots, r; \ k = 1, \ldots, c$, instead of $\sum_{j=1}^{r} \alpha_j = \sum_{k=1}^{c} \beta_k = 0$, $\sum_{j=1}^{r} \gamma_{jk} = 0, \ k = 1, \ldots, c; \ \sum_{k=1}^{c} \gamma_{jk} = 0, \ j = 1, \ldots, r$. Note that in the general log-linear model for $r \times c$ tables $\log(n\pi_{jk}) = \mu + \alpha_j + \beta_k + \gamma_{jk}$, and the grand mean μ gets absorbed into A; see Agresti (1990, Section 5.1) for a discussion of the saturated model.

It is intuitive that $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}, \ j = 1, \ldots, r; \ k = 1, \ldots, c$ is equivalent to $\gamma_{jk} = 0, j = 1, \ldots, r, \ k = 1, \ldots, c$. Although there are related results elsewhere (e.g., Agresti 1990), we state the following theorem, which will be proved in Appendix B.

Theorem   Suppose that $\alpha_r = \beta_c = 0, \ \gamma_{jc} = 0, \ j = 1, \ldots, r$ and $\gamma_{rk} = 0; \ k = 1, \ldots, c$. Then, $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}, \ j = 1, \ldots, r; \ k = 1, \ldots, c$ is equivalent to $\gamma_{jk} = 0, j = 1, \ldots, r, \ k = 1, \ldots, c$.

That is, to test for independence, one can use the interactions γ_{jk} to provide a simultaneous inference on $\gamma_{jk}, \ j = 1, \ldots, r-1, \ k = 1, \ldots, c-1$.

Let $q_{jk}, \ j = 1, \ldots, r; \ k = 1, \ldots, c$ denote the log contrasts. Specifically, $q_{jk} = \log \pi_{jk} - \log \pi_{rc}$. Then,

$$
\begin{aligned}
q_{jk} &= \alpha_j + \beta_k + \gamma_{jk}, \ j = 1, \ldots, r-1; \ k = 1, \ldots, c-1, \\
q_{jc} &= \alpha_j, \ j = 1, \ldots, r-1, \\
q_{rk} &= \beta_k, \ k = 1, \ldots, c-1.
\end{aligned}
\tag{3.1}
$$

Note that $q_{rc} = 0$. It follows easily from (3.1) that

$$
\begin{aligned}
\alpha_j &= \log \pi_{jc} - \log \pi_{rc}, \\
\beta_k &= \log \pi_{rk} - \log \pi_{rc}, \\
\gamma_{jk} &= [\log \pi_{jk} - \log \pi_{rc}] - [\log \pi_{jc} - \log \pi_{rc}] - [\log \pi_{rk} - \log \pi_{rc}], \quad (3.2)
\end{aligned}
$$

$j = 1, \ldots, r - 1$, $k = 1, \ldots, c - 1$. Note that the effects in (3.2) are simple linear functions of the log contrasts.

We need a 95% simultaneous credible region for γ_{jk}, $j = 1, \ldots, r - 1$; $k = 1, \ldots, c - 1$. We can do so using the method of Besag et al. (1995). Note that we do not need to specify the prior distributions of the α_j, β_k and γ_{jk}. These parameters inherit their prior and posterior densities from the π_{jk}.

It is easy to draw a random sample from $\underset{\sim}{\pi} \mid \underset{\sim}{n} \sim \text{Dirichlet}(\underset{\sim}{1})$. Let $\underset{\sim}{\pi}^{(1)}, \ldots, \underset{\sim}{\pi}^{(M)}$ denote a large random sample (M=10,000). Then we have automatically a large random sample from the posterior distribution of $\underset{\sim}{\gamma}$ by taking

$$
\gamma_{jk}^{(h)} = [\log \pi_{jk}^{(h)} - \log \pi_{rc}^{(h)}] - [\log \pi_{jc}^{(h)} - \log \pi_{rc}^{(h)}] - [\log \pi_{rk}^{(h)} - \log \pi_{rc}^{(h)}], \quad h = 1, ..., M.
$$
$$(3.3)$$

Now, we can find a 95% simultaneous credible region for the $\{\gamma_{jk}\}$, obtained by using the sample $\{\gamma_{jk}^{(h)}\}$, $h = 1, \ldots, M$; see our Appendix C for the method of Besag et al. (1995).

Note that for a 2×2 table, there is just a single γ, so that we can have a credible interval for γ. Here $\gamma = [\log \pi_{11} - \log \pi_{12}] - [\log \pi_{21} - \log \pi_{22}] = \log[\pi_{11}\pi_{22}/\pi_{12}\pi_{21}]$, the log of the odds ratio. So the odds ratio is $\exp(\gamma)$, and we can find a 95% credible interval by using $\exp(\gamma^{(h)})$, $h = 1, ..., M$.

Note that the 95% simultaneous credible region formed by this procedure (if $r, c > 2$) is a hyper-rectangle. We can form a test that $\gamma_{jk} = 0$, $j = 1, \ldots, r - 1$; $k = 1, \ldots, c - 1$, easily. This is done by simply checking that the hyper-rectangle contains $\underset{\sim}{\gamma} = \underset{\sim}{0}$. If each of the components of $\underset{\sim}{\gamma}$ contains 0, then the hyper-rectangle does. That is, if all components of $\underset{\sim}{\gamma}$ are 0, there is independence (i.e., no association) between the two attributes, and if at least one component of $\underset{\sim}{\gamma}$ does not contain 0, there is dependence (i.e., association) between the two attributes. This is analogous to the F-test of the regression coefficients in a normal-theory linear regression model.

In a $r \times c$ table there are $(r - 1)(c - 1)$ interactions. The theorem shows that under independence all these interactions must be zero. One should not compute $(r - 1)(c - 1)$ individual 95% credible intervals, because the overall coverage will be much less than 95%. Thus, a 95% simultaneous credible region (a hyper-rectangle in $(r - 1)(c - 1)$ dimensional Euclidean space) is needed; see Miller (1981) for a general discussion of simultaneous inference. Simultaneous regions of the regression parameters in the standard normal theory linear model

are optimal ellipsoid (Box and Tiao, 1973). However, because r and c are both generally small, the individual intervals which form the hyper-rectangle are not much wider than those from the individual 95% credible intervals.

4. Chi-Squared Testing and Intra-Class Correlation

Nandram and Choi (2006) used results from Altham (1976) to describe a Bayesian methodology to fit "multinomial" data when there is an intra-class correlation. We first review this model, and we show how to obtain the 95% simultaneous credible intervals for the interactions. We show more explicitly the difficulties using the chi-squared test in the situation when there is an intra-class correlation.

Suppose there are s_i individuals in the ith cluster, $i = 1, \ldots, \ell$, and s_{ijk} individuals fall in the jth row and k^{th} column in the $r \times c$ table, $j = 1, \ldots r; k = 1, \ldots, c$. Here $\sum_{j=1}^{r} \sum_{k=1}^{c} s_{ijk} = s_i$, $s_{ijk} \geq 0$. Also, let $\mathcal{C}$ denote the set of clusters in which all individuals fall in a single cell of the $r \times c$ table. Then, using formulas in Altham (1976) and letting $\underset{\sim}{s}_i = (s_{i11}, \ldots, s_{irc})$, $i = 1, \ldots, \ell$,

$$p(\underset{\sim}{s}_i \mid \theta_{s_i}, \underset{\sim}{\pi}) = \begin{cases} \theta_{s_i} \pi_{jk} + (1 - \theta_{s_i}) \pi_{jk}^{s_i}, & i \in \mathcal{C} \\ \\ (1 - \theta_{s_i}) s_i! \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{jk}^{s_{ijk}} / s_{ijk}!, & i \notin \mathcal{C}, \end{cases} \tag{4.1}$$

where we take $\theta_1 = 0$ for one-member family. Let $\mathbf{s} = (\underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_\ell)$.

Suppose each cluster has size t, $t = 1, \ldots, T$; in applications T is 2 to 5 or so. Also, let g_{tjk} denote the number of clusters in $\mathcal{C}$ of size t with all individuals in cell (j, k), and $\tilde{g}_t$ the number of clusters of size t outside $\mathcal{C}$. Then, assuming independence over the clusters,

$$\begin{aligned} p(\mathbf{s} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) &= [\prod_{t=1}^{T} \prod_{j=1}^{r} \prod_{k=1}^{c} \{\theta_t \pi_{jk} + (1 - \theta_t) \pi_{jk}^t\}^{g_{tjk}}] \\ &\quad \times [\prod_{t=1}^{T} (1 - \theta_t)^{\tilde{g}_t} \{\prod_{i \notin \mathcal{C}} s_i! \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{jk}^{s_{ijk}} / s_{ijk}!\}]. \end{aligned} \tag{4.2}$$

Note that when $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}$, we simply replace π_{jk} in (4.2) by $\pi_j^{(1)} \pi_k^{(2)}$ to form $p(\underset{\sim}{s} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}^{(1)}, \underset{\sim}{\pi}^{(2)})$. Note also that there are two parts in (4.2), one corresponds to the clusters in $\mathcal{C}$ and the other to those clusters outside $\mathcal{C}$.

Finally, for a full Bayesian approach, noting that $\theta_1 \equiv 0$, we assume

$$\theta_t \overset{iid}{\sim} \text{Uniform}(0, 1), \ t = 2, \ldots, T \tag{4.3}$$

and independently

$$\underset{\sim}{\pi} \sim \text{Dirichlet}(1,\ldots,1). \tag{4.4}$$

The prior densities in (4.3) and (4.4) are relatively noninformative but proper. Thus, the posterior density is

$$p(\theta, \underset{\sim}{\pi} \mid \mathbf{s}) \propto p(\underset{\sim}{s} \mid \theta, \underset{\sim}{\pi}).$$

Nandram and Choi (2006) show how to use the Gibbs sampler to fit the posterior density.

The estimative procedure is easy to implement. The procedure to obtain the 95% simultaneous credible region for $\underset{\sim}{\gamma}$ is similar to the one in Section 3. Here we obtain sample $\{\pi_{jk}^{(h)}\}$, $h = 1, \ldots, M$ from the posterior density $p(\theta, \underset{\sim}{\pi} \mid \mathbf{s})$; here $\underset{\sim}{\theta}$ is a nuisance parameter. We obtain samples from the posterior density of $\underset{\sim}{\gamma}$. That is, we use (3.3) to obtain $\{\gamma_{jk}^{(h)}\}$, $h = 1, \ldots, M = 10,000$.

Finally, for comparison to test for association versus no association between the two attributes, we can use the Bayes factor. Retaining the prior distribution in (4.3), the other two prior distributions are

$$\underset{\sim}{\pi}^{(1)} \sim \text{Dirichlet}(1,\ldots,1) \ \text{ and independently } \ \underset{\sim}{\pi}^{(2)} \sim \text{Dirichlet}(1,\ldots,1). \tag{4.5}$$

Recalling that the vector $\pi^{(1)}$ has r components and $\pi^{(2)}$ has c components, the marginal likelihoods with association is

$$p_{as}(\mathbf{s}) = (rc - 1)! \int \int p(\underset{\sim}{s} \mid \theta, \underset{\sim}{\pi}) d\theta d\underset{\sim}{\pi}$$

and the marginal likelihood with no association is

$$p_{nas}(\mathbf{s}) = (r - 1)!(c - 1)! \int \int \int p(\underset{\sim}{s} \mid \theta, \underset{\sim}{\pi}^{(1)}, \underset{\sim}{\pi}^{(2)}) d\theta d\underset{\sim}{\pi}^{(1)} d\underset{\sim}{\pi}^{(2)}.$$

Nandram and Choi (2006) construct Monte Carlo consistent estimators of the marginal likelihoods.

5. Examples

In this section we discuss four examples to compare the methods for testing association in $r \times c$ categorical tables, and to show sensitivity of the Bayes factor to prior specifications. In Example 1 we show that inference from the chi-squared test and the simultaneous credible region is similar, but this differs from that of the Bayes factor. In Example 2 we show that when there is a strong association, the three methods give the same inference.

While in Examples 1 and 2 there is mild sensitivity of the Bayes factor to the choice in the prior specifications, in Example 3 we show that there could be strong sensitivity to the prior specifications and to the stability of a cell size. In Example 4 we show how the intra-class correlation affects inference by the three methods.

5.1 Example 1: Bone mineral density and family income

We consider the 3×3 categorical table of bone mineral density (BMD) and family income (FI) for a probability sample drawn from the U.S. population. FI is a discrete variable, and there are three levels: low, medium and high. While BMD is a continuous variable, the World Health Organization has classified BMD into three levels: normal, osteopenia and osteoporosis. BMD is used to diagnose osteoporosis, a disease of elderly females, and in the National Health and Nutrition Examination Survey (NHANES III) it is measured for individuals at least twenty years old. In Table 1 we summarize the data on white females with chronic conditions.

Table 1: Classification of bone mineral density (BMD) and family income (FI) for 1,844 white females, at least 20 years old (20+)

	FI		
BMD	1	2	3
1	621	290	284
2	260	131	117
3	93	30	18

Note: BMD: $1(> 0.82g/cm^2$; normal), $2(> 0.64, \leq 0.82g/cm^2$; osteopenia), $3(\leq 0.64g/cm^2$; osteoporosis); FI: $1(< \$20,000)$, $2(\geq \$20,000, < \$45,000)$, $3(\geq \$45,000)$; BMD is only measured for age 20+.

Under independence (i.e., no association) the observed chi-squared statistic is 12.7 on 4 degrees of freedom with a p-value of .013, and no association is rejected. The log Bayes factor is 3.40 for evidence of no association relative to association. Therefore, while the chi-squared test provides strong evidence *against* no association, the log Bayes factor provides strong evidence *for* no association. This is the contradictory evidence because the chi-squared test *rejects* independence and the Bayes factor *accepts* independence.

Our new test, the 95% simultaneous credible region, for $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ is $(-1.47, -0.24)$, $(-1.21, 0.23)$, $(-1.49, -0.19)$, $(-1.15, 0.36)$. Thus, this region does not contain $(0, 0, 0, 0)$, (i.e., only two contain the 0), and like the Pearson chi-squared test, it rejects independence (or no association); this is different from the Bayes factor.

We have performed the sensitivity test of Bayes factor for the prior parameter $\delta = .1, .5, 1., 2., 3$. The corresponding values of the log Bayes factor are -7.0, -3.6, -3.4, -4.7, -6.6. As expected, as δ increases, the independence hypothesis becomes stronger. The log Bayes factor is sensitive to the choices of the prior δ, but around $\delta = 1$ (i.e., the uniform prior), there is little sensitivity of the log Bayes factor. Here the evidence for independence from both prior distributions (Jeffreys and uniform) is "strong", and this evidence increases on either side of $\delta = 1$. We note that over these values of δ the 95% simultaneous region changes very little as seen above (i.e., independent from the prior sensitivity), and inference is unaffected.

5.2 Example 2: Material and Heat

The U.S. Consumer Product Safety Commission (CPSC) staff analyzes fire incident data to determine patterns and losses (e.g., fire deaths, injuries and property damage) associated with fires that involve household products. The purposes of this analysis are (1) to support or evaluate standards that would make household products less likely to ignite and (2) to identify products that pose new hazards and the patterns of usage associated with such hazards. Attributing fire losses to fire causes is an important part of this task (Greene et al. 2002). In Table 2 we present a 4×6 table of forms of materials by forms of heat of 468 fires completely classified, a scaled down version of a typical problem at CPSC (Greene et al. 2002). Note the sparseness of the data in Table 2 makes the standard asymptotic chi-squared test untrustworthy.

Table 2: Classification of fire deaths by form of materials and form of heat for 468 fires

Form of Materials	Form of heat					
	1	2	3	4	5	6
1	55	20	13	14	51	20
2	21	4	18	16	74	20
3	4	21	18	12	15	30
4	8	7	1	2	21	3

Note: As reported by Greene *et al.* (2002) the values in the table do not represent actual data which they described as a scaled down hypothetical CPSC raking problem. Forms of materials are 1(Not furniture), 2(Furniture not in scope), 3(Upholstered furniture), 4(Unknown furniture), and forms of heat are 1(Fuel fired in scope), 2(Fuel fire not in scope), 3(Fuel fire unknown if in scope), 4(Smoking materials in scope), 5(Smoking materials not in scope) and 6(Smoking materials unknown if in scope).

Under independence condition (i.e., no association), both chi-squared test and Bayes factor reject the independence (or no association) (i.e., the observed chi-squared statistic is 102.4 on 15 degrees of freedom with a p-value of .000 and the log Bayes factor is 32.1).

The new method produces the same result as the other two tests (i.e., the 95% simultaneous credible region for $(\alpha_{11}, \ldots, \alpha_{15}, \ldots \ldots, \alpha_{35})$ is $(-1.99, 1.88)$, $(-2.88, 1.06)$, $(-2.24, 3.74)$, $(-2.44, 2.72)$, $(-2.88, 0.68)$, $(-2.89, 1.02)$, $(-4.68, -0.16)$, $(-1.95, 4.05)$, $(-2.33, 2.83)$, $(-2.45, 1.10)$, $(-5.15, -0.74)$, $(-3.30, 0.72)$, $(-2.26, 3.63)$, $(-3.09, 2.19)$, and $(-4.54, -0.87)$). Note that the intervals corresponding to α_{22}, α_{31} and α_{35} do not contain 0. Thus, this region does not contain the origin in the 15-dimensional space, and also rejects independence. Although the data are sparse for at least unknown furniture (4), all three procedures give similar results, partly because the nature of data has a very strong association. Also, note that the Monte Carlo errors associated with estimating the γ_{jk} are all smaller than 0.01 (i.e., the estimative procedure is highly reproducible).

We have performed the sensitivity analysis as in Example 1. At $\delta = .1, .5, 1., 2., 3.$ the values of the log Bayes factor are 16.9, 30.3, 32.1, 30.3, 27.0; also at $\delta = 10$, it is 7.4 and at $\delta = 20$, it is -7.1. As expected, as δ increases, the log Bayes factor decreases. Again while the log Bayes factor is sensitive to changes in δ, inference is not so sensitive around $\delta = 1$ ("very strong" evidence anyway). There is a large change from $\delta = .1$ to $\delta = .5$. Again, we note that over these values of δ the 95% simultaneous region changes very little, and inference is unaffected.

5.3 Example 3: Artificial example for sensitivity analysis

We have constructed a simple example to study the sensitivity of the Bayes factor to the prior specification and the stability of a small cell. This is a 3×3 table of two attributes; the cell counts are 50, 50, 25, 50, 50, 20, 40, 30, 1, starting from the (1, 1) cell going across the rows. Note that cell (3, 3) is small (i.e., 1).

At $\delta = .10, .15, .25, .50, 1.00, 2.00, 3.00, 4.00, 5.50$, the log Bayes factor are -0.56, 0.55, 1.73, 2.86, 3.24, 2.69, 1.84, 0.98, $-.21$. Note that at $\delta = .5$ the evidence against an association is "positive" (i.e., 2.86), at $\delta = 1$ the evidence against an association is "strong" (i.e., 3.24), and at $\delta = 2$ the evidence against an association is "positive" (i.e., 2.69), again. Thus, the log Bayes factor is sensitive to the choice of the prior distribution, but the change in the evidence against association is small. We note also that there is no change in the evidence against association using the 95% simultaneous credible region.

In Figure 1 we have presented a plot of the log Bayes factor by δ for five values of the (3, 3) cell (i.e., albf1 corresponds to $n_{33} = 1$; albf2 to $n_{33} = 2$;

albf3 to $n_{33} = 3$; albf4 to $n_{33} = 4$; and albf5 to $n_{33} = 5$). Observe how sensitive the Bayes factor is to the value of the (3, 3) cell, and to the choices of δ. As δ moves away from one on either side, the log Bayes factor decreases rapidly from a positive value to negative values. The chi-squared test of no independence versus independence is not better. At $n_{33} = 1, 2, 3, 4, 5$ the p-values are .006, .015, .031, .061, .106 respectively. Independence is rejected at the 5% significance level for the first three choices of n_{33} but not at the others, showing sensitivity to n_{33}. Inference from the estimative alternative remains unaltered.
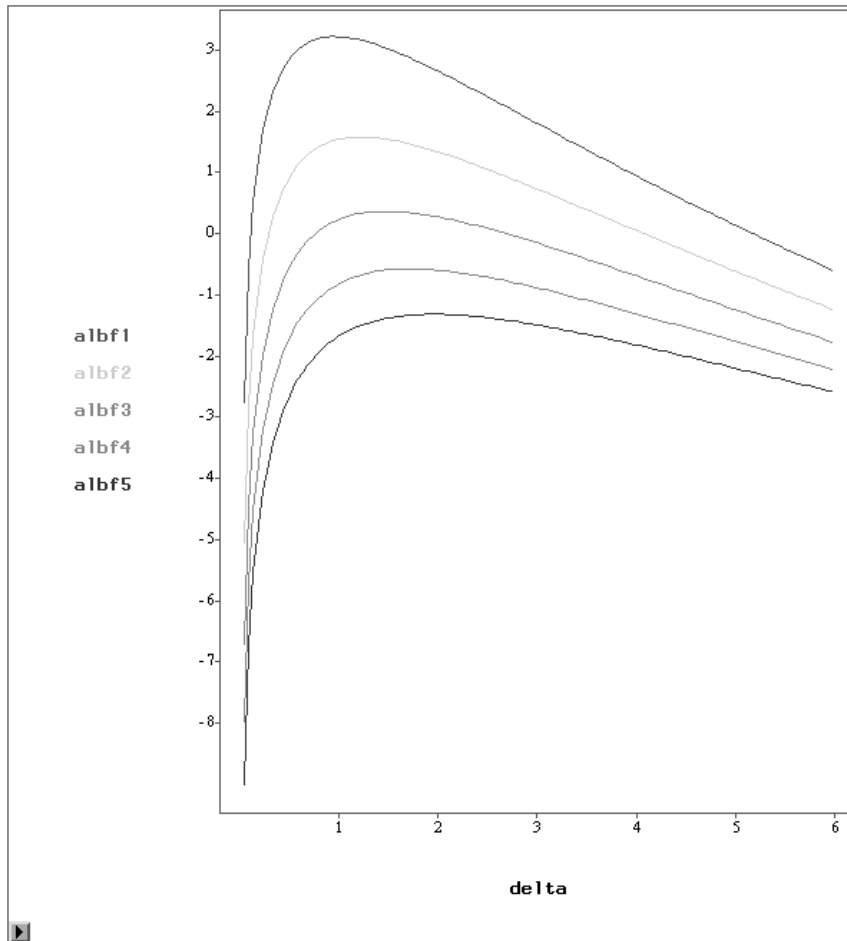

Figure 1: Scatterplots of log Bayes factor versus δ by size of the (3, 3) cell in artificial example; in order from top: albf1 - cell (3, 3) has 1 observation; albf2 - cell (3, 3) has 2 observations; albf3 - cell (3, 3) has 3 observations; albf4 - cell (3, 3) has 4 observations; albf5 - cell (3, 3) has 5 observations.

5.4 Example 4: Activity limitation status and age with intra-class correlation

We use data from the U.S. National Health Interview Survey on activity limitation status (ALS), a measure of long-term disability resulting from chronic conditions. In Table 3 we present a 3×3 categorical table of ALS by age for two states. In State 1 there are 164 sampled adults and in State 2 there are 153 sampled adults. Note that in both tables there are very sparse cell counts especially for the first category of ALS in State 1. We also note that in both tables, members are from 1-member to 5-member families.

Table 3: Classification of activity limitation status (ALS) and Age for two states

	Age					
	State 1			State 2		
ALS	1	2	3	1	2	3
1	2	3	2	9	8	0
2	9	4	2	16	6	6
3	107	17	18	78	15	15

Note: ALS (1: unable to perform major activity; 2: limited in kind/amount major activity and limited in other activities; 3: not limited (includes unknowns)) and Age (1: under 56 years; 2: 56-70 years; 3: over 70 years).

There is significant correlation between the two members of the 2-member families. A 95% confidence interval for θ_2 in State 1 is (.36, .71) and in State 2 it is (.50, .81) (θ_2 is the intra-class correlation between the two members in the 2-member families). The intra-class correlations among the members of the 3-member, 4-member and 5-member families are not substantial. Thus, we need to adjust the chi-squared test for intra-class correlation. We discuss each state separately.

First, consider State 1. For a test of no association versus association, the log Bayes factor is 5.40 for the model with intra-class correlation and 2.56 for the model with no intra-class correlation. This difference is important because the evidence in the former is "very strong" (i.e., 5.40) whereas in the latter it is "positive" (i.e., 2.56).

The unadjusted chi-squared statistic is 9.60 on 4 degrees of freedom with a p-value of .048, and the adjusted chi-squared statistic is 7.75 on 4 degrees of freedom with a p-value of .101. Thus, the *adjusted* test does not reject the hypothesis of independence (i.e., p=0.048) whereas the unadjusted test rejects independence

(i.e., p=0.101) at the 5% significance level, with the *unadjusted* chi-squared test leading to an erroneous conclusion.

Using the model with intra-class correlation the 95% simultaneous credible region is formed by the intervals (-2.93, 4.49), (-2.66, 5.59), (-2.56, 2.43), (-2.12, 3.29); all the intervals contain 0 and the estimative procedure does not reject independence (same conclusion as the *adjusted* chi-squared test). We note that, using the model with no intra-class correlation, inference from the 95% simultaneous credible region remains unchanged. This implies that the two different correlation models (intra-class and no intra-class) does not give different results for the border line situations (seen for both Bayes factor and chi-squared test).

Second, consider State 2. For a test of association versus no association, the log Bayes factor is 3.09 for the model with intra-class correlation and 0.38 for the model with no intra-class correlation. This difference is important because the evidence of the difference between association and no association in the former is "strong" whereas in the latter it is "not worth more than a bare mention".

The unadjusted chi-squared statistic is 13.52 on 4 degrees of freedom with a p-value of .009, and the adjusted chi-squared statistic is 11.18 on 4 degrees of freedom with a p-value of .025. Thus, there is strong evidence against independence with the unadjusted test, but for the test adjusted for correlation, the evidence is marginal at the 2.5% significance level.

Using the model with *intra-class correlation*, the 95% simultaneous credible region is formed by the intervals (-2.39, 5.09), (-1.35, 5.95), (-4.49, -0.26), (-2.39, 1.16); only the interval for γ_{21} does not contain 0, and the estimative procedure rejects independence (or no association). Using the model with *no intra-class correlation* inference from the 95% simultaneous credible region remains unchanged, but now it is only the credible interval for γ_{12}, not γ_{21}, that does not contain 0.

6. Discussion

We have reviewed two current tools, the chi-squared test and the Bayes factor, to test for association in two-way tables. We have demonstrated the difficulties associated with the standard chi-squared test in two-way categorical tables when the multinomial assumptions are violated. We have also demonstrated difficulties associated with the use of the Bayes factor, especially its sensitivity to prior specifications. To overcome these problems, we propose a new method for "testing" association, capitalizing on the fact that Bayesian hypothesis testing is sensitive to prior specifications whereas estimation is not that sensitive.

Thus, responding to the important issue of the sensitivity of Bayes factors to the choice of the prior distributions in Bayesian hypothesis testing, we introduce

a Bayesian estimative procedure for the analysis of two-way categorical tables. In a two-way categorical table, we have utilized a standard relation between the cell probabilities and the main effects and the interactions. A 95% simultaneous credible region for the interactions provides a "test" of independence. This is easy to implement, and it requires samples from the posterior distribution of the interactions.

We also demonstrated that the chi-squared test is inaccurate in nonstandard situations such as when there is intra-class correlation in the data (e.g., familial data). These situations require a model more elaborate than the simple multinomial-Dirichlet model. Thus, the chi-squared test needs special adjustment, and the Bayes factor is also sensitive to prior specifications. Our new method about independence via the 95% simultaneous credible region is simple and not sensitive to prior specifications. When there exists strong association between the two attributes, the three methods give the same conclusions. When there are borderline cases, adjustment is necessary, and our method should to be preferred.

We recommend the 95% simultaneous credible region to make inference about independence in two-way categorical tables in non-standard situations where both the chi-squared test and Bayes factor can fail. One can possibly quantify the strength of the evidence by counting the number of credible intervals containing 0; this is a topic for further research. There are other situations where the 95% simultaneous credible region will be useful. For example, when missing data differ from observed data, the chi-squared test will fail. It is possible to extend our Bayesian estimative procedure to higher dimensional categorical tables. We have not addressed model averaging in this paper because a test is performed using a selected model.

Appendix A: A Discussion of the Defects of the Bayes Factor

We discuss some defects of the Bayes factor in hypothesis testing. The discussion here holds for any Bayesian hypothesis testing problem in which the Bayes fastor is used. Thus, specifically it applies directly to the test of association in a $r \times c$ categorical table.

Kass and Raftery (1995) have popularized the use of the Bayes factor in scientific problems. However, they have discussed controversies associated with the use of the Bayes factor. As is well known, Bayes factor requires the specification of a prior distribution. They stated, "This may be considered both good and bad. Good, because it is a way of including other information about the values of the parameters. Bad, because these prior densities may be hard to set when there is no such information." Indeed, a serious problem with the Bayes factor is the sensitivity to the prior specification as Kass and Raftery (1995) wrote "An

important issue is the sensitivity of the Bayes factor to the choices of priors."

With respect to the sensitivity of the Bayes factor to the choices of priors, Kass and Raftery (1999) wrote "Also in contrast with Bayesian point estimates such as the posterior mean, the Bayes factor does tend to be sensitive to the choices of priors on the model parameters," and they also stated that "Bayes factor tends to be more sensitive to the choice of prior than the posterior probability of an interval." These statements are now well known (e.g., see Kass 1993). Berger and Pericchi (1996) agrue that one should operate in strict accordance with two basic premises, selection should have a Bayesian basis and it should be automatic, and they stated that "The reason is that Bayes factors in hypothesis testing and model selection typically depend rather strongly on the prior distributions, much more so than in, say, estimation." We take advantage of these difficulties of the Bayes factor to construct a Bayesian estimative alternative to test of association in a two-way categorical table.

For a sensible calibration of the Bayes factor proper priors are needed (i.e., they must integrate to 1). Berger and Pericchi (1996) wrote "For most model selection problems, one cannot use standard improper noninformative priors; such priors are defined only up to a constant multiple, and the Bayes factor is itself a multiple of this arbitrary constant." Of course, this difficulty can be overcome by using proper priors, but then there is the issue of the sensitivity to prior specification. To overcome this problem, Berger and Pericchi (1996) introduced the *intrinsic* Bayes factors. But these also have their own problems. Clearly, intrinsic Bayes factors are not Bayes factors, they require part of the data to be treated as a *training* sample, they require enormous computation even if a minimal training sample or a small random sample of the set of training samples is used, and they are not calibrated with respect to the strength of the evidence as in Kass and Raftery (1995). Moreover, for our problem in the $r \times c$ categorical table, the set of minimal training samples is essentially empty; both Jeffreys' prior and the uniform prior are proper. The use of the intrinsic Bayes factor fixes the calibration problem, but not the sensitivity problem, and therefore, they are not appropriate in our context.

There is also a difficulty in interpreting the Bayes factor as pointed out by Lavine and Schervish (1999), "The removal of the prior odds from the posterior odds to produce the Bayes factor has consequences that affect the interpretation of the resulting ratio." As pointed out by Lavine and Schervish (1999), "Just because the data increase the support for a hypothesis H relative to its complement does not necessarily make H more likely than its complement, it only makes H more likely than it was a priori." See Lavine and Schervish (1999) for further discussion. We do not address this issue further in this paper.

Appendix B: Proof of the Theorem

Suppose $\gamma_{jk} = 0$, $j = 1, \ldots, r-1$; $k = 1, \ldots, c-1$. Then, it is easy to show that

$$\pi_{jk} = [e^{\alpha_j} / \sum_{j'=1}^{r} e^{\alpha_{j'}}][e^{\beta_k} / \sum_{k'=1}^{c} e^{\beta_{k'}}], \ j = 1, \ldots, r; \ k = 1, \ldots, c.$$

Now, set $\pi_j^{(1)} = e^{\alpha_j} / \sum_{j'=1}^{r} e^{\alpha_{j'}}$, $j = 1, \ldots, r$ and $\pi_j^{(2)} = e^{\beta_k} / \sum_{k'=1}^{c} e^{\beta_{k'}}$, $k = 1, \ldots, c$.

Suppose $\pi_{jk} = \pi_j^{(1)} \pi_k^{(2)}$, $j = 1, \ldots, r$; $k = 1, \ldots, c$. Then, using $\gamma_{jc} = 0$, $j = 1, \ldots, r$ and $\gamma_{rk} = 0$, $k = 1, \ldots, c$, it follows that

$$\gamma_{jk} = \text{constant}, \ j = 1, \ldots, r-1; \ k = 1, \ldots, c-1,$$

$$\sum_{k=1}^{c-1} e^{\beta_k}(e^{\gamma_{jk}} - 1) = 0, \ j = 1, \ldots, r-1.$$

Thus, the constant is 0, and $\gamma_{jk} = 0$, $j = 1, \ldots, r-1$; $k = 1, \ldots, c-1$.

Appendix C: A Nonparametric Simultaneous Credible Region

We describe the nonparametric method of Besag et al. (1995) to obtain a simultaneous credible region for $\gamma_{j,k}$, $j = 1, \ldots, r-1$; $k = 1, \ldots, c-1$. For convenience, we let $i = k + (j-1)(c-1)$, and we rewrite γ_{jk} as γ_i, $i = 1, \ldots, (r-1)(c-1) = \ell$.

Denoting the stored sample by $\{\gamma_i^{(h)} : i = 1, \ldots, \ell; \ h = 1, \ldots, M\}$, order $\{\gamma_i^{(h)} : h = 1, \ldots, M\}$ separately for each component i, to obtain order statistics $\gamma_i^{[h]}$ and ranks $r_i^{(h)}$, $h = 1, \ldots, M$. For fixed $\kappa \in \{1, \ldots, M\}$, let h^* be the smallest integer such that $\gamma_i^{[M+1-h^*]} \leq \gamma_i^{(h)} \leq \gamma_i^{[h^*]}$, for all i, for at least κ values of h; h^* is equal to the κth order statistic from the set $a^{(h)} = \max\left\{\max_i r_i^{(h)}, M+1-\min_i r_i^{(h)}\right\}$, $h = 1, \ldots, M$, that is, $h^* = a^{[\kappa]}$.

Then $\left\{\left[\gamma_i^{[M+1-t^*]}, \gamma_i^{[t^*]}\right] : i = 1, \ldots, \ell\right\}$ form a simultaneous credible region which contains at least $100\kappa/M\%$ of the empirical distribution. Thus, if $M = 10000$ and $\kappa = 95$, a 95% simultaneous credible region is obtained.

References

Agresti, A. (1990). *Categorical Data Analysis*. Wiley.

Altham, P. M. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika* **63**, 263-269.

Bedrick, E. J. (1983). Adjusted chi-squared tests for cross-classified tables of survey data. *Biometrika* **70**, 591-595.

Berger, J. O., and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association* **91**, 109-122.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3-66.

Box, G. E. P and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Wiley.

Brier, S. S. (1980). Analysis of contingency tables under cluster sampling. *Biometrika* **67**, 591-596.

Choi, J. W. and McHugh, R. B. (1989). A reduction factor in goodness-of-fit and independence tests for clustered and weighted observations. *Biometrics* **45**, 979-996.

Cohen, J. E. (1976). The distribution of the chi-squared statistic under cluster sampling from contingency tables. *Journal of the American Statistical Association* **71**, 591-596.

Fellegi, I. P. (1980). Approximate tests of independence and goodness of fit based upon stratified multistage samples. *Journal of the American Statistical Association* **75**, 261-268.

Fay, R. E. (1985). Complex samples. *Journal of the American Statistical Association* **80**, 148-157.

Fienberg, S. E.(1979). The use of chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society*, Series B **41**, 54-64.

Greene, M. A., Smith, L. E., Levenson, S. M., Hiser, S. and Mah, J. C. (2002). Raking fire data. *U.S. Consumer Product Safety Commission*, 1-11.

Greenwood, P. E. and Nikulin, M. S. (1996). *A Guide to Chi-Squared Testing*. Wiley.

Holt, D., Scott, A. J. and Ewings, P. O. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society*, Series A **143**, 302-320.

Kass, R. E. (1993). Bayes factors in practice. *Statistician* **42**, 551-560.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Lavine, M. and Schervish, M. J. (1999). Bayes factors: What they are what they are not. *The American Statistician* **53**, 119-122.

Miller, R. G. (1981). *Simultaneous Statistical Inference*. Springer.

Mirkin, B. (2001). Eleven ways to look at the chi-squared coefficient for contingency tables. *The American Statistician* **55**, 111-120.

Nandram, B. and Choi, J. W. (2006). A Bayesian analysis of a two-way categorical table incorporating intra-class correlation. *Journal of Statistical Computation and Simulation* **76**, 233-249.

Nandram, B., Cox, L. H. and Choi, J. W. (2005) Bayesian analysis of non-ignorable missing categorical data: An application to bone mineral density and family income. *Survey Methodology* **31**, 213-225.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association* **76**, 221-230.

Rao, J. N. K. and Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics* **12**, 46-60.

Sinharay, S. and Stern, H. S. (2002). On the sensitivity of bayes factors to the prior distributions. *The American Statistician* **56**, 196-201.

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute road, Worcester
Massachusetts 01609, USA
balnan@wpi.edu

Jai Won Choi
Office of Research and Methodology
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782, USA
jwc7@cdc.gov