A Bayesian Alternative to the Chi-Squared test of Association in a Two-Way Categorical Table Incorporating Intra-Class Correlation

Balgobin Nandram[1] and Jai Won Choi
(bbn9@cdc.gov, jwc7@cdc.gov)

Office of research and Methodology
National Center for Health Statistics, CDC
3311 Toledo Road, Hyattsville, MD 20782

June 4, 2004

Abstract

It is straight forward to analyze data from a single multinomial table. Specifically, for the analysis of a two-way categorical table, the common chi-squared test of independence between the two variables and maximum likelihood estimators are readily available. When the counts in the two-way categorical table are formed from familial data (clusters of correlated data), the common chi-squared test no longer applies. We note that there are several approximate adjustments to the common chi-squared test. For example, Choi and McHugh (1989, Biometrics **45**, 979-996) showed how to adjust the chi-squared statistic for clustered and weighted data. However, our main contribution is the construction and analysis of a Bayesian model which removes all analytical approximations. This is an extension of a standard multinomial-Dirichlet model to include the intra-class correlation associated with the individuals within a cluster. We have used a key formula described by Altham (1976, Biometrika **63**, 263-269) to incorporate the intra-class correlation. This intra-class correlation varies with the size of the cluster, but we assume that it is the same for all clusters of the same size for the same variable. We use Markov chain Monte Carlo methods to fit our model, and to make posterior inference about the intra-class correlations and the cell probabilities. Also, using Monte Carlo integration with a binomial importance function, we obtain the Bayes factor for a test of no association. To demonstrate the performance of the alternative test and estimation procedure, we have used data on activity limitation status and age from the National Health Interview Survey and a simulation study.

Key Words: Bayes factor, Gibbs sampler, Monte Carlo integration, Multinomial-Dirichlet.

[1]Balgobin Nandram (balnan@wpi.edu) is on sabbatical leave from the Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute road, Worcester, MA 01609.

1. Introduction

It is a common practice to use two-way categorical tables to present survey data. In this situation it is assumed that the cell counts in the $r \times c$ table follow a multinomial distribution. However, because of stratification and clustering the joint distribution of the cell counts is no longer multinomial. Thus, the standard chi-squared statistic no longer has a chi-squared distribution, and therefore the test based on the multinomial distribution may be inadequate. It is standard practice to make an adjustment to the standard chi-squared statistic, but in general the accuracy of this adjustment is not well understood, and one can not estimate the cell probabilities based on this adjustment. We propose a Bayesian alternative which is based on the Bayes factor to obtain a test for association between the two categorical variables. Our Bayesian method also provides posterior distributions for the cell probabilities.

Several authors have recognized inaccuracy in the analysis when the usual chi-squared test is applied to correlated "multinomial" data. Efforts to correct for spurious inflation in such tests have been based 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). For example, Rao and Scott (1981) investigate the effects of stratification and clustering on the asymptotic distribution of Pearson's chi-squared statistic for goodness of fit and independence. They propose new measures called generalized design effects. See also Rao and Scott (1984) who generalized the results of Rao and Scott (1981) to multi-way categorical tables. 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), shows how to adjust the standard chi-squared test statistic when there is an intra-class correlation.

The National Center for Health Statistics (NCHS) uses the National Health Interview Survey (NHIS) to collect data on chronic and acute conditions, doctor visits, hospital episodes, disability,

household and personal information, and other special aspects of health of the U.S. population. One of the variables we use in the NHIS is activity limitation status (ALS), a measure of long-term disability resulting from chronic conditions since 1957. ALS is defined as inability to carry out the major activity for one's age-sex group such as working, keeping house or going to school; restriction in the amount or kind of major activity; or restriction in relation to other activities such as recreational, church and civic interests. ALS is typically classified into three categories: "unable to perform major activity", "limited in kind/amount major activity and in other activities" and "not limited (includes unknowns)" ranging from severe individuals to individuals unnecessary to classify. The relation between age and activity limitation status is of interest. In the health interview survey, information (i.e., chronic disease and impairment) for each household member about the major activity she/he usually performed during the 12 months prior to interview is requested by the interviewer. There is possibly a positive association between ALS and age, and to study this association three age groups (under 56 years, 56-70 years and more than 70 years) are used. The analysis is complex because one can expect an intra-class correlation within households.

Let n_{jk} denote the number of individuals in the j^{th} row and k^{th} column of the $r \times c$ categorical table. Also let $n_{j\cdot} = \sum_{k=1}^{c} n_{jk},\ j = 1, \ldots, r,\ n_{\cdot k} = \sum_{j=1}^{r} n_{jk},\ k = 1, \ldots, c,\ n = \sum_{j=1}^{r} \sum_{k=1}^{c} n_{jk}$ and $e_{jk} = n_{j\cdot} n_{\cdot k}/n,\ j = 1, \ldots, r,\ k = 1, \ldots, c.$ Then, Pearson's chi-squared statistic, under independence of the row and column classification, 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 asymptotically (as $n \to \infty$) $X_u \to \chi^2_{(r-1)(c-1)}$, 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 larger than 5, and 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., familial 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 problem about the asymptotic distribution has received much attention, but the problem about the inaccuracy of the estimates of the cell proportions has received virtually no attention. In this paper we address both problems simultaneously within a Bayesian framework when there are familial count data.

We describe one solution that has been proposed for the problem about the asymptotic distribution. 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 ($\theta_1 \equiv 0$). Motivated by Rao and Scott (1981), Choi and McHugh (1989) derive 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 $\hat{\theta}$ is the maximum likelihood estimator of θ under their model. The statistic X_a is 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 further adjust $\chi^2_{(r-1)(c-1)}$ approximately by the average weight.

We provide a Bayesian analysis of this problem. This is a direct extension of the probabilistic development in Altham (1976) which is used to provide a likelihood function. Then proper but noninformative priors are assigned to the parameters to provide a full Bayesian approach. The model includes a nonnegative intra-class correlation which varies according to the number of individuals in a cluster (i.e., all clusters of the same size have the same intra-class correlation). In this framework we can provide (a) the posterior densities of the cell probabilities and (b) a test of association between the two categorical variables. In (b) we use the Bayes factor to quantify the difference between a model with association and one without. This is the ratio of the prior odds of one model to the other to their posterior odds (obtained through the use of Bayes' theorem), 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. If two models, M_0 and M_1, are fitted to data $\underset{\sim}{y}$, 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

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

where $\underset{\sim}{\theta_k}$ is the parameter vector under M_k, $p(\underset{\sim}{y}|\underset{\sim}{\theta_k}, \mathrm{M}_k)$ is the probability density (or mass) function and $p(\underset{\sim}{\theta_k} \mid \mathrm{M}_k)$ is the prior density. For example, in application M_0 is the model of no association and M_1 is the model of association. 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 $0 \leq \log(B_{10}) < 1$, the evidence against M_0 is "not worth more than a bare mention"; if $1 \leq \log(B_{10}) < 3$, the evidence against M_0 is "positive"; if $3 \leq \log(B_{10}) < 5$, the evidence against M_0 is "strong"; and if $\log(B_{10}) \geq 5$, the evidence against M_0 is "very strong". There are several methods to compute the marginal likelihood (e.g., see Section 1 of Chib and Jeliazkov 2001), and we note that one standard method is Monte Carlo integration with an importance function.

In this paper, we introduce a Bayesian method to analyze data from an $r \times c$ categorical table. We consider the situation in which there are no missing data, but one in which the table is built up by aggregating clustered multinomial data. In Section 2, we describe the methodology to obtain estimates of the cell probabilities, and to obtain the Bayes factor for a test of no association between the two categorical variables. We also show how to use Markov chain Monte Carlo methods to fit the models. We show how to use Monte Carlo integration with an importance function to compute the marginal likelihoods under different models. In Section 3 we illustrate our method using data from the National Health Interview Survey. In Section 4, we perform several simulated examples to compare inference using our model with another model which does not incorporate the intra-class correlation. Finally, Section 5 has concluding remarks.

2. Bayesian Model and Computation

We describe the methodology to fit "multinomial" data when there is an intra-class correlation. We build our model based on the work of Altham (1976).

2.1 Model

Suppose there are s_i individuals in the i^{th} cluster, $i = 1, \ldots, \ell$, and s_{ijk} individuals fall in the j^{th} row and k^{th} column in the $r \times c$ table, $j = 1, ...r, k = 1, ..., c$. Here $\sum_{j=1}^{r} \sum_{k=1}^{c} s_{ijk} = s_i$, $s_{ijk} \geq 0$. Altham (1976) shows that the probability that all s_i individuals fall in the j^{th} row and k^{th} column is

$$\theta_{s_i} \pi_{jk} + (1 - \theta_{s_i}) \pi_{jk}^{s_i}, \ 0 \leq \theta_{s_i} \leq 1. \tag{1}$$

There is exactly one sequence in (**??**). Note that (**??**) can be interpreted as a mixture of two distributions. Let w_{s_i} be the latent variable such that $w_{s_i} = 1$ for perfect dependence and $w_{s_i} = 0$, for perfect independence, where dependence/independence refers to the intra-class correlation. Then $p(w_{s_i} = 1 \mid \theta_{s_i}) = 1 - p(w_{s_i} = 0 \mid \theta_{s_i}) = \theta_{s_i}$.

Also, the probability that the individuals are in different *specified* cells is

$$(1 - \theta_{s_i}) \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{jk}^{s_{ijk}} \tag{2}$$

where we allow the intraclass correlation θ_{s_i}, $0 \leq \theta_{s_i} \leq 1$, to depend on the cluster size s_i. Note that there is at least one sequence in (**??**).

This model of clustering permits only positive association or independence among the individuals within a cluster, and this is typically the case for many demographic, social and economic characteristics.

Note that $\theta_{s_i} \pi_{jk} + (1 - \theta_{s_i}) \pi_{jk}^{s_i}$ is strictly increasing in θ_{s_i}. When $\theta_{s_i} = 0$, the probability that all individuals in the i^{th} cluster belong to cell (j, k) is $\pi_{jk}^{s_i}$, and when $\theta_{s_i} = 1$, the probability that all individuals in the i^{th} cluster belong to cell (j, k) is π_{jk}, which can be much larger. In addition, $(1 - \theta_{s_i}) \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{jk}^{s_{ijk}}$ is a strictly decreasing function in θ_{s_i}. When $\theta_{s_i} = 0$, the probability that the individuals in the i^{th} cluster belong to different specified cells is $\pi_{jk}^{s_i}$, and when $\theta_{s_i} = 1$, the probability that the individuals in the i^{th} cluster belong to different specified cells is 0. Thus, the intra-class correlation has an important role when inference is made about the π_{jk} and the association between the two categorical variables. Henceforth, $s_1, \ldots, s_\ell$ are assumed known.

Let $\mathcal{C}$ denote the set of clusters in which all individuals fall in a single cell of the $r \times c$ table. Then, letting $\underset{\sim}{s}_i = (s_{i11}, \ldots, s_{irc}),\ i = 1, \ldots, \ell$,

$$p(\underset{\sim}{s}_i \mid \theta_{s_i}, \underset{\sim}{\pi}) = \begin{cases} \theta_{s_i}\pi_{jk} + (1-\theta_{s_i})\pi_{jk}^{s_i}, & i \in \mathcal{C} \\ (1-\theta_{s_i})s_i! \prod_{j=1}^{r}\prod_{k=1}^{c} \pi_{jk}^{s_{ijk}}/s_{ijk}! & i \notin \mathcal{C}. \end{cases} \tag{3}$$

Assuming independence over clusters and letting $\underset{\sim}{s} = (\underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_\ell)$, we have

$$p(\underset{\sim}{s} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \prod_{i\in\mathcal{C}}\prod_{j=1}^{r}\prod_{k=1}^{c}\{\theta_{s_i}\pi_{jk} + (1-\theta_{s_i})\pi_{jk}^{s_i}\} \prod_{i\notin\mathcal{C}}\{(1-\theta_{s_i})s_i! \prod_{j=1}^{r}\prod_{k=1}^{c}\pi_{jk}^{s_{ijk}}/s_{ijk}!\}. \tag{4}$$

Observe that if $\theta_{s_i} = 0,\ i = 1, \ldots, \ell$,

$$p(\underset{\sim}{s} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \prod_{i=1}^{\ell}\{s_i! \prod_{j=1}^{r}\prod_{k=1}^{c}\pi_{jk}^{s_{ijk}}/s_{ijk}!\}, \tag{5}$$

which is a product of multinomial probability functions. In (**??**) the statistics $\sum_{i=1}^{\ell} s_{ijk} = n_{jk}$ are sufficient as in regular multinomial sampling (i.e., observations are from a simple random sample) and each individual belongs to cell (j,k) with probability $\pi_{jk} \geq 0$, $\sum_{j=1}^{r}\sum_{k=1}^{c}\pi_{jk} = 1$.

Suppose that each cluster has size $t,\ t = 1, ..., T$; in applications T is 2 to 5 or so. Then letting 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 in $\tilde{\mathcal{C}}$ (i.e., outside $\mathcal{C}$),

$$\begin{aligned} p(\underset{\sim}{s} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) \;\; = \;\; & \prod_{t=1}^{T}\prod_{j=1}^{r}\prod_{k=1}^{c}(\theta_t\pi_{jk} + (1-\theta_t)\pi_{jk}^{t})^{g_{tjk}} \\ & \times \{\prod_{t=1}^{T}(1-\theta_t)^{\tilde{g}_t}\} \prod_{i\notin\mathcal{C}}\{s_i! \prod_{j=1}^{r}\prod_{k=1}^{c}\pi_{jk}^{s_{ijk}}/s_{ijk}!\}. \end{aligned} \tag{6}$$

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

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

and independently

$$\underset{\sim}{\pi} \sim \text{Dirichlet}(\underset{\sim}{1}).$$

These are noninformative but proper prior densities. Thus, the joint posterior density of $(\underset{\sim}{\theta}, \underset{\sim}{\pi})$ is proper.

The likelihood function is described in Appendix A where we introduce latent variables z_{tjk} (i.e., an augmented likelihood function used primarily to simplify the computations). Specifically, in (**??**) we have described the joint probability mass function of $(\underset{\sim}{s}, \underset{\sim}{z})$ given $(\underset{\sim}{\theta}, \underset{\sim}{\pi})$. Then, letting $\underset{\sim}{z} = \{z_{tjk},\ t = 2, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c\}$ and using Bayes' theorem the joint posterior density is

$$p(\underset{\sim}{\theta}, \underset{\sim}{\pi}, \underset{\sim}{z} \mid \underset{\sim}{s}) \propto \left\{ \prod_{t=2}^{T} (1-\theta_t)^{\tilde{g}_t} \right\}$$

$$\times \prod_{j=1}^{r} \prod_{k=1}^{c} \left\{ \pi_{jk}^{g_{1jk}+\tilde{s}_{jk}} \prod_{t=2}^{T} \left(\begin{array}{c} g_{tjk} \\ z_{tjk} \end{array} \right) (\theta_t \pi_{jk})^{z_{tjk}} ((1-\theta_t)\pi_{jk}^t)^{g_{tjk}-z_{tjk}} \right\}. \tag{7}$$

The joint posterior density in (7) is complex, so we use the Gibbs sampler to draw samples which are used to make inference about π_{jk} and θ_t.

2.2 Computation

To run the Gibbs sampler we need starting values for $\underset{\sim}{\theta}$ and $\underset{\sim}{\pi}$, and these are easy to obtain. Letting $n_{jk} = \sum_{i=1}^{\ell} s_{ijk}$ and $n = \sum_{j=1}^{r} \sum_{k=1}^{c} n_{jk}$, we take $\hat{\pi}_{jk} = n_{jk}/n$ and $\theta_t = 1/t,\ t = 2, ..., T$. Note also that we estimate z_{tjk} by $z_{tjk} = g_{tjk}[\theta_t \pi_{jk} / \{\theta_t \pi_{jk} + (1-\theta_t)\pi_{jk}^t\}]$.

The conditional posterior densities (cpd's) of each parameter given the others are needed to implement the Gibbs sampler. Note that $z_{1jk} = \theta_1 = 0$. Specifically, the cpd for $\underset{\sim}{\theta}$ is

$$\theta_t \mid \underset{\sim}{\pi}, \underset{\sim}{z}, \underset{\sim}{s} \overset{ind}{\sim} \text{Beta} \left\{ 1 + \sum_{j=1}^{r} \sum_{k=1}^{c} z_{tjk}, 1 + \tilde{g}_t + \sum_{j=1}^{r} \sum_{k=1}^{c} (g_{tjk} - z_{tjk}) \right\},\ t = 2, \ldots, T,$$

the cpd for $\underset{\sim}{\pi}$ is

$$\underset{\sim}{\pi} \mid \underset{\sim}{\theta}, \underset{\sim}{z}, \underset{\sim}{s} \sim \text{Dirichlet} \left\{ 1 + g_{1jk} + \tilde{s}_{jk} + \sum_{t=1}^{T} [z_{tjk} + t(g_{tjk} - z_{tjk})],\ j = 1, \ldots, r,\ k = 1, \ldots, c \right\}$$

and the cpd for $\underset{\sim}{z}$ is

$$z_{tjk} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}, \underset{\sim}{s} \overset{ind}{\sim} \text{Binomial} \left\{ g_{tjk}, \frac{\theta_t \pi_{jk}}{\theta_t \pi_{jk} + (1-\theta_t)\pi_{jk}^t} \right\},\ t = 2, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c.$$

We "burn in" 1000 iterates, and took every tenth to get 1000 iterates which we use for inference. These choices are very conservative, and the algorithm runs very quickly.

3. Alternative Test for Association and Computation

To test for association versus no association between the two categorical variables, we use the Bayes factor, the ratio of the two marginal likelihoods. By no association we mean that $\pi_{jk} = q_j^{(1)} q_k^{(2)}$ where $\sum_{j=1}^{r} q_j^{(1)} = \sum_{k=1}^{c} q_k^{(2)} = 1$. A problem of the slightly less interest is to test for no intra-class correlation. The model without intra-class correlation is given in Appendix B.

3.1 Bayes Factor

Consider our problem with intra-class correlation. Letting $\underset{\sim}{z} = \{z_{ijk},\ t = 2, ..., T,\ j = 1, ..., r,\ k = 1, ..., c\}$, we define $\mathcal{Z} = \{\underset{\sim}{z} :\ 0 \le z_{tjk} \le g_{tjk},\ t = 2, ..., T,\ j = 1, ..., r,\ k = 1, ..., c\}$. For the model with association, taking $\theta_1 = 0$ and letting $A(\underset{\sim}{g}, \underset{\sim}{z}) = \prod_{t=1}^{T} \prod_{j=1}^{r} \prod_{k=1}^{c} \begin{pmatrix} g_{tjk} \\ z_{tjk} \end{pmatrix}$, the marginal likelihood is

$$p_{\text{as}}(\underset{\sim}{s}) = (rc-1)! \sum_{\underset{\sim}{z} \in \mathcal{Z}} [A(\underset{\sim}{g}, \underset{\sim}{z}) \int \int \prod_{t=1}^{T} \prod_{j=1}^{r} \prod_{k=1}^{c} \{(\theta_t \pi_{jk})^{z_{tjk}} \{(1-\theta_t)\pi_{jk}^{t}\}^{g_{tjk}-z_{tjk}}$$

$$\times \{\prod_{t=1}^{T} (1-\theta_t)^{\tilde{g}_t}\} \prod_{i \notin \mathcal{C}} \{s_i! \prod_{j=1}^{r} \prod_{k=1}^{c} \frac{\pi_{jk}^{s_{ijk}}}{s_{ijk}!}\} \mathrm{d}\underset{\sim}{\theta} \mathrm{d}\underset{\sim}{\pi}]$$

and for the model without association the marginal likelihood is

$$p_{\text{nas}}(\underset{\sim}{s}) = (r-1)!(c-1)! \sum_{\underset{\sim}{z} \in \mathcal{Z}} [A(\underset{\sim}{g}, \underset{\sim}{z}) \int \int \int \prod_{t=1}^{T} \prod_{j=1}^{r} \prod_{k=1}^{c} (\theta_t q_j^{(1)} q_k^{(2)})^{z_{tjk}} \{(1-\theta_t)(q_j^{(1)} q_k^{(2)})^{t}\}^{g_{tjk}-z_{tjk}}$$

$$\times \{\prod_{t=1}^{T} (1-\theta_t)^{\tilde{g}_t}\} \prod_{i \notin \mathcal{C}} \{s_i! \prod_{j=1}^{r} \prod_{k=1}^{c} \frac{(q_j^{(1)} q_k^{(2)})^{s_{ijk}}}{s_{ijk}!}\} \mathrm{d}\underset{\sim}{q}^{(1)}_{j} \mathrm{d}\underset{\sim}{q}^{(2)}_{k} \mathrm{d}\underset{\sim}{\theta}] .$$

Then, letting $d = (rc-1)!S$ and $e = (r-1)!(c-1)!S$ with $S = \prod_{i \notin \mathcal{C}} \{s_i! / \prod_{j=1}^{r} \prod_{k=1}^{c} s_{ijk}!\}$, it is easy to show that

$$p_{\text{as}}(\underset{\sim}{s}) = d \sum_{\underset{\sim}{z} \in \mathcal{Z}} \left[A(\underset{\sim}{g}, \underset{\sim}{z}) \{\prod_{t=2}^{T} D(b_{1t}+1, b_{2t}+1)\} D(a_{11}+1, ..., a_{rc}+1) \right] \tag{8}$$

and

$$p_{\text{nas}}(\underset{\sim}{s}) = e \sum_{\underset{\sim}{z} \in \mathcal{Z}} \left[A(\underset{\sim}{g}, \underset{\sim}{z}) \{\prod_{t=2}^{T} D\left(b_{1t}+1, b_{2t}+1\right)\}\ D_1(\underset{\sim}{a})\ D_2(\underset{\sim}{a}) \right] \tag{9}$$

where $b_{1t} = \sum_{j=1}^{r} \sum_{k=1}^{c} z_{tjk}$, $b_{2t} = \tilde{g}_t + \sum_{j=1}^{r} \sum_{k=1}^{c} (g_{tjk} - z_{tjk})$, $a_{jk} = g_{1jk} + \tilde{s}_{jk} + \sum_{t=2}^{T} \{z_{tjk} + t(g_{tjk} - z_{tjk})\}$, $a_{j\cdot} = \sum_{k=1}^{c} a_{jk}$, $a_{\cdot k} = \sum_{j=1}^{r} a_{jk}$, $j = 1, \ldots, r,\ \ k = 1, \ldots, c$, $D_1(\underset{\sim}{a}) = D(a_{1\cdot}+1, \ldots, a_{r\cdot}+1)$ and $D_2(\underset{\sim}{a}) = D(a_{\cdot 1}+1, \ldots, a_{\cdot k}+1)$. In (**??**) and (**??**) $D(\cdot, \ldots, \cdot)$ is the Dirichlet function, where

for a κ-dimensional vector $\underset{\sim}{x}$, $D(\underset{\sim}{x}) = \prod_{s=1}^{\kappa}\Gamma(x_s)/\Gamma(\sum_{s=1}^{\kappa}x_s)$ (e.g., when $\kappa = 2$, $D(x_1, x_2) = \Gamma(x_1)\Gamma(x_2)/\Gamma(x_1+x_2)$ is the beta function).

3.2 Computation

To compute (**??**) and (**??**) we use Monte Carlo integration with the importance function,

$$z_{tjk} \overset{ind}{\sim} \text{Binomial}(g_{tjk}, q_{tjk}),\ t = 2, ..., T,\ k = 1, ..., c. \tag{10}$$

In (**??**), $q_{tjk} = \hat{\theta}_t\hat{\pi}_{jk}/\{\hat{\theta}_t\hat{\pi}_{jk} + (1-\hat{\theta}_t)\hat{\pi}_{jk}^{\kappa}\}$ where κ is a tuning constant and $\hat{\theta}_t$ and $\hat{\pi}_{jk}$ are respectively the posterior means of θ_t and π_{jk} obtained from the Gibbs sampler. We choose the tuning constant $\kappa = 2$.

Then, simulation consistent estimators of $p_{\text{as}}(\underset{\sim}{s})$ and $p_{\text{nas}}(\underset{\sim}{s})$ are

$$\widehat{p_{\text{as}}(\underset{\sim}{s})} = dM^{-1}\sum_{h=1}^{M}\frac{\{\prod_{t=2}^{T}D(b_{1t}^{(h)}+1, b_{2t}^{(h)}+1)\}D(a_{11}^{(h)}+1, ..., a_{rc}^{(h)}+1)}{\prod_{t=2}^{T}\prod_{j=1}^{r}\prod_{k=1}^{c}q_{tjk}^{z_{tjk}^{(h)}}(1-q_{tjk})^{g_{tjk}-z_{tjk}^{(h)}}}$$

and

$$\widehat{p_{\text{nas}}(\underset{\sim}{s})} = eM^{-1}\sum_{h=1}^{M}\frac{\{\prod_{t=2}^{T}D(b_{1t}^{(h)}+1, b_{2t}^{(h)}+1)\}D_1(\underset{\sim}{a}^{(h)})D_2(\underset{\sim}{a}^{(h)})}{\prod_{t=2}^{T}\prod_{j=1}^{r}\prod_{k=1}^{c}q_{tjk}^{z_{tjk}^{(h)}}(1-q_{tjk})^{g_{tjk}-z_{tjk}^{(h)}}}$$

where $b_{1t}^{(h)} = \sum_{j=1}^{r}\sum_{k=1}^{c}z_{tjk}^{(h)}$, $b_{2t}^{(h)} = \tilde{g}_t + \sum_{j=1}^{r}\sum_{k=1}^{c}(g_{tjk} - z_{tjk}^{(h)})$, $a_{jk}^{(h)} = g_{1jk} + \tilde{s}_{jk} + \sum_{t=2}^{T}\{z_{tjk}^{(h)} + t(g_{tjk} - z_{tjk}^{(h)})\}$, and $\underset{\sim}{z}^{(h)}$, $h = 1, ..., M$ is a random sample from (**??**). We have chosen M=10,000.

4. An Illustrative Example

In the NHIS the households are poststratified by states and there are data from all 51 states (including the District of Columbia). For some states there are extremely small numbers of sampled households (e.g., Iowa, Idaho, Wyoming) and for some states there are extremely large numbers of sampled households (e.g., California, New York, Texas). We have studied these states individually and to illustrate our procedure we use the data from Maryland (medium size state). In column 2 of Table **??** we present the cell counts of the 3×3 table of age and ALS for Maryland. It is of general interest to test the hypothesis that age and ALS are independent and to estimate the proportion of

individuals in each cell of the 3×3 table. We have compared our model with intra-class correlation with the simple (without intra-class correlation) multinomial model; see Appendix B for a brief discussion of the simple multinomial model.

Using the method of Rao (1965, p. 159) we have calculated the intra- class correlation coefficient for age and ALS separately for Maryland data. For age the estimates are: families of size 2 (.71), 3 (.45), 4 (.20) and for families of size 5 (-.03), and for ALS the estimates are: families of size 2 (.29), 3 (.26), 4 (.20 and for families of size 5 (.02). Note that there are three families of size 7 and 2 of size 8 that we have omitted from our data analysis, and the total number of individuals is 897 with 104 one-member families, 140 two-member families, 79 three-member families, 49 four-member families and 16 five-member families.

We have constructed 95% credible intervals for the intra-class correlations $(\theta_2, \ldots, \theta_5)$ and they are: θ_2 (.39, .58), θ_3 (.46, .70), θ_4 (.57, .80), and θ_5 (.28, .74). Thus, there is substantial intra-class correlation especially among families of size 4.

We have also studied the tests (Bayes factor and adjusted chi-squared statistics). Working with logarithm, for the model without intra-class correlation the Bayes factor is 15.6; the value of the unadjusted chi-squared test statistic is 67.2 giving a p-value of virtually 0. For the model with intra-class correlation the Bayes factor is 7.6 with a NSE of 3.3; the value of the adjusted chi-squared test statistic is 45.8 giving a p-value of virtually 0 again. The Bayes factor gives very strong evidence for an association between age and ALS with or without the intra-class correlation; the same is true for the chi-squared test. Note that the count for cell $(2, 1)$ is only 3, showing a possible problem for the chi-squared test. We also note that in this example, because there is a strong assosciation between age and ALS, the difference among these tests is small. However, in cases where the association is not so large, there could be differences in these tests.

For the π_{jk} in Table **??** we present the posterior mean (PM), posterior standard deviation (PSD), numerical standard error (NSE), and the 95% credible intervals. First, the NSE's are small showing that the results can be reproduced. But note, as expected, the model without intra-class correlation has PM's close to the $\hat{\pi}_{jk}$, but there are some cases where the PM's from the model with

intra-class correlation differ (e.g., for cell (2,1) compare .067 with .027 and for cell (3, 1) compare .706 with .675). Also the PSD's under the model with intra-class correlation are larger than those under the model without intra-class correlation, as is expected. These differences are reflected in the 95% credible intervals (e.g., for cell (3,1) compare $(.62, .73)$ with $(.67, .73)$). Except for cell (1, 2) the intervals under the model with intra-class correlation contain those under the model without intra-class correlation.

It is interesting that inference about the π_{jk} can differ under the model with intra-class correlation and the one without, reflecting the presence of an intra-class correlation. The presence of a substantial intra-class correlation has the effect of reducing the number of observations. Thus, we can deduce that the absence of intra-class correlation in the simple multinomial model leads to an under estimation of variability.

5. A Simulation Study

We have simulated data from our model to assess the quality of our methodology. Specifically, we have studied how changes in the intra-class correlation affect inference about the π_{jk} and the association between the variables in a $r \times c$ categorical table.

We have chosen the π_{jk} to represent different degrees of association between the categorical variables in a 3×3 table. Specifically, we have chosen three different sets of π_{jk}: (a) low (or no) assiciation ($\pi_{jk} = 1/9,\ j = 1, \ldots r,\ k = 1, \ldots, c$) (b) medium association ($\pi_{11} = .220$, $\pi_{12} = .150$, $\pi_{13} = .100$, $\pi_{21} = .075$, $\pi_{22} = .100$, $\pi_{23} = .075$, $\pi_{31} = .050$, $\pi_{32} = .100$, and $\pi_{33} = .130$) and (c) strong association ($\pi_{11} = .250$, $\pi_{12} = .050$, $\pi_{13} = .010$, $\pi_{21} = .030$, $\pi_{22} = .250$, $\pi_{23} = .030$, $\pi_{31} = .050$, $\pi_{32} = .080$, and $\pi_{33} = .250$). We have taken $\theta_k = \theta,\ k = 2, \ldots, T$, and 5 values of θ $(.2,\ .4,\ .5,\ .6,\ .8)$. Thus, we study the effect of our choice of θ on inference about π_{jk} and the association between the two categorical variables. We note that when θ is small (large), there is a tendency for the simulated individuals to be in different (same) cell(s) of the $r \times c$ table. Letting c_k denote the number of clusters of size $k,\ k = 1, \ldots, T = 5$, we take $c_1 = 50$, $c_2 = 70$, $c_3 = 50$, $c_4 = 40$ and $c_5 = 20$ to get a total of 600 observations. These are held fixed for all simulation

experiments. Thus, we study how the posterior distributions of the intraclass correlation and the π_{jk} are affected by choices of the intra-class correlation and the degree of association between the two categorical variables.

We simulated the cell counts using the probabilities of allocation. Let $\underset{\sim}{s}_i$ denote the counts for the i^{th} cluster. Then, for $t = 1$, $\underset{\sim}{s}_i \mid \underset{\sim}{\pi} \sim$ Multinomial$(1, \underset{\sim}{\pi})$. For $t \geq 2$ the probability that all t individuals fall in the same cell is $\theta\pi_{jk} + (1-\theta)\pi_{jk}^t$ and the probability that they fall in different cells is $(1-\theta)(t!)\prod_{j=1}^{r}\prod_{k=1}^{c}\pi_{jk}^{s_{ijk}}/s_{ijk}!$, $\sum_{j=1}^{r}\sum_{k=1}^{c}s_{ijk} = t$. Thus, it is straight forward to draw the cell counts.

We have simulated 1000 datasets of size 600 for each value of θ and the degree of association (i.e., there are 15000 datasets). We fit our model incorporating the intra-class correlation to each data set using the Gibbs sampler as described in Section 2.2. We compute (a) the posterior mean, posterior standard deviation, and 95% credible intervals for the π_{jk} for each data set, (b) the Bayes factor to test for association in the 3×3 table and (c) we have also compared inference using the model with intra-class correlation and the standrad multinomial-Dirichlet model, one that ignores the intra-class correlation (see Appendix B).

We have presented results for our simulated examples in Tables **??**, **??**,**??** and **??**, taking averages over the 1000 datasets for each quantity.

[Need to discuss!] In Tables **??** we have presented posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval for the θ_k and the π_{jk}. Apart from $\theta = .10$ the posterior summaries are concordant with the design values for θ_2 to θ_5.

[Need to discuss!] In Table **??** we have presented posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval for the and the π_{jk}. The posterior summaries indicate that the model with the intra-class correlation is more in concordant with the design values than the regular multinomial, and the regular multinomial degrades as θ increases.

[Need to discuss!] The log-Bayes factors in Table **??** for both the model with intra-class correlation and the regular multinomial are bigger than 5, indicating very strong evidence for no

association. But while for the model with intra-class correlation the Bayes factor decreases rapidly as θ increases, the decrease under the regular multinomial model is not so steep.

We have studied the datasets generated from the design level of weak association ($\pi_{jk} = 1/9,\ j = 1, 2, 3,\ k = 1, 2, 3$) to compare the adjusted ch-squared test and the Bayes factor from the model with intra-class correlation; see Table **??**. First, using the the model with intra-class correlation all the datasets show some degree of evidence for no association; of the 5000 datasets 3812 show strong evidence and 1188 show weak evidence for no association. Observe that the number of datasets showing weak (strong) evidence increase (decrease) with the intra-class correlation. The situation is different when the adjusted chi-squared test is used; of the 5000 datasets 3948 do not reject (accept in Table **??**) and 1052 (much too large) reject no association at the 5% significance level. With aberrations the effect of increasing intra-class correlation is the same as for the Bayes factor. Thus, inference using the adjusted chi-squared test can be incorrect.

6. Concluding Remarks

We have shown how to analyze multinomial data from $r \times c$ categorical tables when there is an intra-class correlation. We have also shown that by using the Bayes factor (ratio of the marginally likelihoods of two models) we can test for association between the two categories.

We have analyzed 3×3 categorical data of age and activity limitation status from the 1996 National Health Interview Survey. We have found moderately large intra-class correlations, and these correlations have small effects on tests of hypothesis (both the standard chi-squared test and our Bayesian alternative). While we have reported results for Maryland, we have found similar results for many of the other states.

We have also performed a small simulation study to assess the impact of the intra-class correlations on the alternative to the chi-squared test and posterior inference of the cell probabilities. It appears that the Bayes factor decreases as the intra-class correlation increases (further investigation is required), but for the examples we have not found much difference in inference between

the model with intra-class correlation and the one without. There are also small differences for inference about the π_{jk}.

In future we can extend our methodology to accommodate (a) small areas (b) nonresponse and (c) an intra-class correlation coefficient corresponding to each categorical variable. In (a) we can consider the states (including the District of Columbia) as small areas. There are very sparse data from some of the states (e.g., Iowa, Idaho, Wyoming), and to make reliable inference about one of these states, one needs to "borrow strength" across the states. In (b) there is a non-negligible number of nonrespondents from each state, and one would need to construct a model that can adjust for nonignorable nonresponse. Finally, in (c) for two categorical variables one intra-class correlation is θ_{s_i} and the other is $\kappa_{s_i}\theta_{s_i},\ 0 \le \kappa_{s_i} \le \theta_{s_i}^{-1}$. Then we can replace θ_{s_i} in Altham's formula by $\frac{1}{2}(\kappa_{s_i}+1)\theta_{s_i},\ 0 \le \theta_{s_i} \le 1,\ 0 < \kappa_{s_i} \le \theta_{s_i}^{-1},\ i = 1, \ldots, \ell$ with an appropriate joint prior density on $(\theta_{s_i},\ \kappa_{s_i})$.

APPENDIX A: Augmented Likelihood Function

We derive the likelihood function of $\underset{\sim}{\theta}$ and $\underset{\sim}{\pi}$, augmented with latent variables. Let the cell counts for the i^{th} cluster be $s_{ijk},\ i = 1, \ldots, \ell,\ j = 1, \ldots, c,\ k = 1, \ldots, c$, and $s_i = \sum_j^r \sum_k^c s_{ijk}$. Also let $\underset{\sim}{s_i} = (s_{i11}, \ldots, s_{irc})$. Assuming that the s_i are known, we have $\underset{\sim}{s_i}$ given $\underset{\sim}{\theta}$and $\underset{\sim}{\pi}$ are independent. We separate the derivation into two parts.

First, for each $i \epsilon C$ (i.e., the first part in which all members fall in the same cell),

$$p(\underset{\sim}{s_i} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = (\theta_{s_i}\pi_{jk}) + (1-\theta_{s_i})\pi_{jk}^{s_i} = \sum_{\omega_{ijk}=0}^{1} (\theta_{s_i}\pi_{jk})^{\omega_{ijk}} \{(1-\theta_{s_i})\pi_{jk}^{s_i}\}^{1-\omega_{ijk}},$$

$j = 1, \ldots, r,\ k = 1, \ldots, c$ where ω_{ijk} is an indicator variable (i.e., $\omega_{ijk} = 0, 1$). Then, by de-marginalization over ω_{ijk}, we have

$$p(\omega_{ijk}, \underset{\sim}{s_i} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = (\theta_{s_i}\pi_{jk})^{\omega_{ijk}} \{(1-\theta_{s_i})\pi_{jk}^{s_i}\}^{1-\omega_{ijk}},\ \omega_{ijk} = 0, 1.$$

That is, $\omega_{ijk} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi} \sim$ Bernoulli$\{\frac{\theta_{s_i}\pi_{jk}}{\theta_{s_i}\pi_{jk}+(1-\theta_{s_i})\pi_{jk}^{s_i}}\},\ i \in C$. Note that for each $i \in C$ there is contribution from only one of the cells $j = 1, \ldots, r,\ k = 1, \ldots, c$ (i.e., $s_{ijk} \equiv s_i$). Note also that

for $i\epsilon C$ and $s_i = 1$, $p(\underset{\sim}{s}_i \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \pi_{jk}$. Thus, for $i \in C$ letting $\underset{\sim}{\omega}$ denote the vector of all ω_{ijk}, by independence the joint probability mass function of $(\underset{\sim}{s}_i,\ i \in C, \underset{\sim}{\omega})$ is

$$p_1(\underset{\sim}{s}_i, i\epsilon C, \underset{\sim}{\omega} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \left[\prod_{i\epsilon C, s_i=1} \pi_{jk}^{s_i}\right]\left[\prod_{i\epsilon C, s_i\geq 2} \{(\theta_{s_i}\pi_{jk})^{\omega_{ijk}}\{(1-\theta_{s_i})\pi_{jk}^{s_i}\}^{1-\omega_{ijk}}\right]. \tag{A.1}$$

Second, for $i \notin C$ (i.e., the second part in which members fall in the different cells),

$$p(\underset{\sim}{s}_i \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = (1-\theta_{s_i})\{s_i! \prod_{j=1}^{r}\prod_{k=1}^{c} \pi_{jk}^{s_{ijk}}/s_{ijk}!\}$$

and the joint probability mass function of $\underset{\sim}{s}_i,\ i \notin C$ is

$$p_2(\underset{\sim}{s}_i, i \notin C \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \prod_{i\notin C} p(\underset{\sim}{s}_i \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \prod_{i\notin C}(1-\theta_{s_i})\{s_i! \prod_{j=1}^{r}\prod_{k=1}^{c} \pi_{jk}^{s_{ijk}}/s_{ijk}!\}. \tag{A.2}$$

Thus, letting $\underset{\sim}{s} = (\underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_\ell)$, by independence the joint probability mass function of $(\underset{\sim}{s}, \underset{\sim}{\omega})$ is

$$p(\underset{\sim}{s}, \underset{\sim}{\omega} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = p_1(\underset{\sim}{s}_i, i\epsilon C, \underset{\sim}{\omega} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) p_2(\underset{\sim}{s}_i, i \notin C \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) \tag{A.3}$$

where $p_1(\underset{\sim}{s}_i, \underset{\sim}{\omega}_i, i\epsilon C \mid \underset{\sim}{\theta}, \underset{\sim}{\pi})$ is given in (**??**) and $p_2(\underset{\sim}{s}_i, i \notin C \mid \underset{\sim}{\theta}, \underset{\sim}{\pi})$ is given in (**??**).

Now, let T denote the largest cluster size and $\theta_1 = 0$. For the clusters in C, let g_{tjk} denote the number of clusters of size t with all their members in cell (j,k), and

$$\underset{\sim}{z} = \{z_{tjk} : z_{tjk} = \sum_{i\epsilon C, s_i=t} \omega_{ijk},\ t = 1, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c\}. \tag{A.4}$$

Note that in (**??**) for all $i \in C$ whenever $s_{ijk} = 0$, $\omega_{ijk} = 0$. For the clusters outside C, let $\tilde{g}_t$ denote the number of clusters of size t and $\tilde{s}_{jk} = \sum_{i\notin C} s_{ijk}$, the total number of individuals in cell $(j,\ k)$ outside C. Then, using the assumption that the intraclass correlation coefficient depends only on the cluster size with $\theta_{s_i} = \theta_t$, $t = 1, \ldots, T$ and (**??**), the joint probability mass function of $(\underset{\sim}{s}, \underset{\sim}{z})$ is

$$p(\underset{\sim}{s}, \underset{\sim}{z} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi}) = \prod_{i\notin C}\left\{s_i! / \prod_{j=1}^{r}\prod_{k=1}^{c} s_{ijk}!\right\}$$

$$\times \left\{\prod_{t=2}^{T}\prod_{j=1}^{r}\prod_{k=1}^{c}\begin{pmatrix} g_{tjk} \\ z_{tjk}\end{pmatrix}(\theta_t\pi_{jk})^{z_{ijk}}\{(1-\theta_t)\pi_{jk}^{t}\}^{g_{tjk}-z_{tjk}}\right\}\left\{\prod_{t=2}^{T}(1-\theta_t)^{\tilde{g}_t}\prod_{j=1}^{r}\prod_{k=1}^{c}\pi_{jk}^{\tilde{s}_{jk}}\right\}. \tag{A.5}$$

Note that the joint probability mass fucntion in (**??**) as a function of $\underset{\sim}{\theta}$, $\underset{\sim}{\pi}$ and $\underset{\sim}{z}$ is the augmented likelihood function (i.e., $\underset{\sim}{\theta}$ and $\underset{\sim}{\pi}$ are augmented with $\underset{\sim}{z}$). Note also that from (**??**)

$$z_{tjk} \mid \underset{\sim}{\theta}, \underset{\sim}{\pi} \overset{ind}{\sim} \text{Binomial}\{g_{tjk}, \frac{\theta_t \pi_{jk}}{\theta_t \pi_{jk} + (1-\theta_t)\pi_{jk}^t}\},$$

$t = 2, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c.$

APPENDIX B: Product Multinomial-Dirichlet Model

Letting $\pi_{jk},\ j = 1, \ldots, r,\ k = 1, \ldots, c$ denote the cell probabilities and $s_i = \sum_{j=1}^{r} \sum_{k=1}^{c} s_{ijk}$, the multinomial-Dirichlet model for the cell counts s_{ijk} in a $r \times c$ categorical table is

$$\underset{\sim}{s_i} \mid \underset{\sim}{\pi} \overset{ind}{\sim} \text{Multinomial}(s_i,\ \underset{\sim}{\pi}),\ i = 1, \ldots, \ell \text{ and } \underset{\sim}{\pi} \sim \text{Dirichlet}(1, \ldots, 1), \tag{B.1}$$

where the s_i are assumed known.

Let $n_{jk} = \sum_{i=1}^{\ell} s_{ijk}$ denote the cell counts over all clusters, $j = 1, \ldots, r,\ k = 1, \ldots, c.$ Then, a posteriori

$$\underset{\sim}{\pi} \mid \underset{\sim}{s_1}, \ldots, \underset{\sim}{s_\ell} \sim \text{Dirichlet}(n_{11} + 1, \ldots, n_{rc} + 1).$$

Because the posterior density is in closed form, one can obtain inference about the π_{jk} in a straight forward manner.

The corresponding marginal likelihoods (association: as, no association: nas) are

$$p_{\text{as}}(\underset{\sim}{n}) = \frac{(rc-1)! \prod_{j=1}^{r} \prod_{k=1}^{c} n_{jk}!}{(n+rc-1)!} \left\{ \prod_{i=1}^{\ell} \frac{s_i!}{\prod_{j=1}^{r} \prod_{k=1}^{c} s_{ijk}!} \right\} \tag{B.2}$$

and

$$p_{\text{nas}}(\underset{\sim}{n}) = p_{\text{as}}(\underset{\sim}{n})\ \frac{(r-1)!(c-1)!}{(rc-1)!} \frac{(n+rc-1)!}{(n+r-1)!(n+c-1)!} \frac{\prod_{j=1}^{r} n_{j.}! \prod_{k=1}^{c} n_{.k}!}{\prod_{j=1}^{r} \prod_{k=1}^{c} n_{jk}!} \tag{B.3}$$

where $n_{j.} = \sum_{k=1,}^{c} n_{jk},\ j = 1, \ldots, r,\ n_{.k} = \sum_{j=1,}^{r} n_{jk},\ k = 1, \ldots, c$ and $\sum_{j=1}^{r} \sum_{k=1}^{c} n_{jk} = n.$

References

Adams, P. F. and Marano, M. A. (1995). Current estimate from the National Health Interview Survey, 1994. *National Center for Health Statistics, CDC*, Series **10**, (193).

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.

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

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

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96**, 270-281.

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 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.

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

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: 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.

Table 1: Comparison of the posterior means (PM), posterior standard deviations (PSD), numerical standard deviation (NSE), and 95% credible intervals (CI) for π_{jk} with and without intra-class correlation for Maryland

		Intra-class				No intra-class			
Cell	$\hat{\pi}$	PM	PSD	NSE	CI	PM	PSD	NSE	CI
(1, 1)	.014	.024	.010	.002	(.008, .045)	.015	.004	.001	(.008, .024)
(1, 2)	.014	.027	.010	.002	(.011, .048)	.016	.004	.001	(.008, .025)
(1, 3)	.003	.008	.006	.001	(.001, .021)	.004	.002	.000	(.001, .010)
(2, 1)	.067	.027	.010	.002	(.012, .050)	.067	.008	.002	(.053, .084)
(2, 2)	.022	.028	.010	.002	(.011, .051)	.023	.005	.001	(.015, .035)
(2, 3)	.026	.024	.010	.002	(.009, .045)	.027	.005	.001	(.017, .038)
(3, 1)	.706	.675	.029	.005	(.616, .730)	.699	.016	.004	(.669, .730)
(3, 2)	.090	.099	.019	.004	(.067, .138)	.091	.010	.002	(.071, .110)
(3, 3)	.057	.088	.018	.003	(.056, .125)	.057	.008	.001	(.043, .074)

NOTE: The total number of observations from Maryland is 897, and $\hat{\pi}_{jk}$ is the observed proportion of observations in cell (j, k). The row and column variables are age and activity limitation status (ALS) respectively; Age (1: under 56 years; 2: 56-70 years; 3: over 70 years) and ALS (1: unable to perform major activity; 2: limited in kind/amount major activity and limited in other activities; 3: not limited (includes unknowns). The numerical standard errors are obtained using the batch-means method with batches of 25 for the selected sample from the Gibbs sampler.

Table 2: Ratio (R) and the probability content (C) of the 95% credible intervals for the intra-class correlation (θ) over the 1000 simulated datasets by the degree of association (low, medium, high) and five values of the intra-class correlations (.2, .4, .5, .6, .8)

	Low		Medium		High	
θ	R	C	R	C	R	C
	1.04	0.94	1.06	0.95	1.45	0.68
0.2	1.05	0.95	1.06	0.96	1.16	0.92
	1.07	0.94	1.05	0.95	1.10	0.94
	1.12	0.96	1.12	0.97	1.12	0.96
	0.99	0.94	1.01	0.95	1.15	0.86
0.4	1.02	0.94	1.00	0.97	1.05	0.92
	0.99	0.96	1.01	0.95	1.01	0.95
	1.03	0.95	1.01	0.96	1.02	0.96
	0.99	0.95	1.00	0.94	1.10	0.87
0.5	1.00	0.94	1.00	0.95	1.02	0.93
	1.00	0.95	1.00	0.96	1.00	0.95
	1.00	0.97	0.99	0.96	1.00	0.96
	0.99	0.94	0.99	0.95	1.05	0.92
0.6	0.99	0.95	0.99	0.96	1.02	0.94
	1.00	0.96	1.00	0.94	1.00	0.96
	0.98	0.96	0.98	0.95	0.99	0.96
	0.99	0.96	0.99	0.95	1.01	0.95
0.8	0.99	0.95	0.99	0.94	0.99	0.96
	0.99	0.95	0.98	0.95	0.98	0.96
	0.96	0.95	0.97	0.95	0.96	0.95

NOTE: For the i^{th} dataset $R_i = PM_i/DV_i$, where PM_i is the posterior maen of θ and DV_i is the design value of θ; R is the average over the 1000 simulated datasets.

Table 3: Comparison of the two models via the ratio (R) and the probability content (C) of the 95% credible intervals for the π_{jk} over the 1000 simulated datasets by the degree of association (low, medium, high) and three values of the intra-class correlation (.2, .5, .8)

	Low				Medium				High			
θ	R_I	R_N	C_I	C_N	R_I	R_N	C_I	C_N	R_I	R_N	C_I	C_N
	1.00	1.00	0.95	0.91	0.99	0.99	0.94	0.89	1.00	1.00	0.96	0.88
	1.01	1.01	0.96	0.90	1.00	1.00	0.95	0.90	1.04	1.02	0.96	0.91
	1.00	1.00	0.96	0.91	1.00	0.99	0.95	0.91	1.19	1.13	0.96	0.91
	1.00	0.99	0.95	0.91	1.01	1.01	0.96	0.91	1.06	1.04	0.95	0.89
0.2	1.00	1.01	0.95	0.90	1.00	1.00	0.96	0.90	0.97	0.99	0.95	0.88
	1.00	1.00	0.96	0.92	1.01	1.01	0.95	0.90	1.07	1.05	0.95	0.90
	1.00	1.01	0.94	0.88	1.03	1.02	0.93	0.88	1.05	1.02	0.95	0.91
	0.99	0.99	0.96	0.89	1.00	1.00	0.95	0.90	1.02	1.00	0.95	0.89
	1.00	1.00	0.95	0.90	0.99	1.00	0.96	0.89	0.98	0.99	0.94	0.90
	1.01	1.00	0.94	0.81	0.99	1.00	0.95	0.93	1.00	1.00	0.96	0.83
	1.00	1.00	0.96	0.82	0.99	1.00	0.95	0.83	1.02	1.00	0.96	0.82
	1.00	1.00	0.95	0.82	1.00	1.00	0.94	0.81	1.25	1.15	0.94	0.82
	0.99	0.99	0.94	0.81	1.02	1.02	0.95	0.84	1.07	1.03	0.95	0.81
0.5	1.00	1.00	0.95	0.82	1.00	1.00	0.94	0.82	0.97	0.99	0.96	0.84
	1.00	1.00	0.95	0.83	1.01	1.01	0.96	0.84	1.08	1.04	0.94	0.83
	1.01	1.01	0.94	0.82	1.02	1.01	0.94	0.83	1.05	1.02	0.95	0.82
	1.00	1.00	0.94	0.81	1.01	1.00	0.96	0.82	1.02	1.01	0.97	0.84
	1.00	1.00	0.95	0.83	0.99	0.99	0.95	0.82	0.98	0.99	0.95	0.83
	0.99	1.00	0.95	0.74	0.99	0.99	0.94	0.75	0.99	0.99	0.94	0.75
	1.00	1.00	0.96	0.76	0.99	0.99	0.96	0.75	1.06	1.04	0.95	0.75
	0.99	0.99	0.95	0.75	1.00	1.00	0.95	0.77	1.30	1.15	0.96	0.75
	1.01	1.00	0.96	0.78	1.00	0.99	0.94	0.75	1.09	1.05	0.95	0.77
0.8	1.00	1.00	0.95	0.77	1.00	1.00	0.95	0.75	0.98	1.00	0.95	0.77
	1.01	1.01	0.94	0.76	1.02	1.02	0.94	0.75	1.10	1.05	0.95	0.76
	1.00	1.01	0.96	0.77	1.04	1.02	0.95	0.78	1.03	1.01	0.95	0.75
	1.00	1.00	0.95	0.77	1.01	1.01	0.95	0.75	1.01	1.00	0.97	0.79
	1.01	1.00	0.94	0.76	1.00	1.00	0.94	0.75	0.97	0.99	0.94	0.77

NOTE: The two models are the model with intra-class (I) correlation and the model with no (N) intra-class correlation (see Appendix B). For the i^{th} dataset $R_i = PM_i/DV_i$, where PM_i is the posterior mean of π_{jk} and DV_i is the design value of π_{jk}; R is the average over the 1000 simulated datasets.

Table 4: Comparison of the test based on the logarithm of Bayes factor and the test based on the chi-squared statistic by the degree of association (low, medium, high) and five values of intra-class correlation (.2, .4, .5, .6, .8) averaged over the 1000 simulated datasets

	Low				Medium				High			
θ	N	Y	χ^2_u	χ^2_a	N	Y	χ^2_u	χ^2_a	N	Y	χ^2_u	χ^2_a
0.2	-3.6	-4.2	5.8	5.2	21.8	17.8	56.5	49.9	226.3	194.3	489.0	426.6
0.4	-2.7	-3.9	7.8	6.2	22.7	14.8	58.0	46.5	227.0	167.6	490.4	389.7
0.5	-2.5	-3.8	8.1	6.2	22.3	13.1	57.0	43.7	228.6	153.6	493.0	374.5
0.6	-2.0	-3.6	9.2	6.7	23.3	12.1	59.2	43.3	228.5	138.5	492.0	358.1
0.8	-1.0	-3.2	11.2	7.5	24.7	9.5	62.0	41.8	227.3	110.6	489.1	329.2

NOTE: The Bayes factor is the ratio of the marginal likelihood for a model with (Y) association (i.e., no restriction on π_{jk}, $\sum_{j=1}^{r}\sum_{k=1}^{c}\pi_{jk}=1$) to the marginal likelihood for a model with no (N) association (i.e., $\pi_{jk}=\pi_j^{(1)}\pi_k^{(2)}$, $\sum_{j=1}^{r}\pi_j^{(1)}=1$, $\sum_{k=1}^{c}\pi_k^{(2)}=1$). Also, χ^2_u and χ^2_a are repectively the unadjusted and adjusted chi-squared statistic.

Table 5: Classification of the simulated data sets by statistical significance using the adjusted chi-squared statistic (reject: pvalue$<$ .05; accept: pvalue$\geq$ .05) and the strength of evidence using the logarithm of the Bayes factor (weak: log-Bayes factor <3; strong: log-Bayes factor ≥ 3) for five values of intra-class correlation (.2, .4, .5, .6, .8)

	weak			strong		
θ	reject	accept	total	reject	accept	total
.2	83	77	160	25	815	840
.4	141	71	212	75	713	788
.5	134	99	233	71	696	767
.6	144	104	248	95	657	752
.8	209	126	335	75	590	665
total	711	477	1188	341	3471	3812

NOTE: Inference using the model with intra-class correlation is compared with inference from the adjusted chi-squared test.