Journal of the Korean Data & Information Science Society
2017, **28**(1), 207–215

http://dx.doi.org/10.7465/jkdi.2017.28.1.207
한국데이터정보과학회지

Bayes tests of independence for contingency tables from small areas

Aejung Jo[1] · Dal Ho Kim[2]

[12]Department of Statistics, Kyungpook National University

Received 9 October 2016, revised 30 October 2016, accepted 1 November 2016

Abstract

In this paper we study pooling effects in Bayesian testing procedures of independence for contingency tables from small areas. In small area estimation setup, we typically use a hierarchical Bayesian model for borrowing strength across small areas. This techniques of borrowing strength in small area estimation is used to construct a Bayes test of independence for contingency tables from small areas. In specific, we consider the methods of direct or indirect pooling in multinomial models through Dirichlet priors. We use the Bayes factor (or equivalently the ratio of the marginal likelihoods) to construct the Bayes test, and the marginal density is obtained by integrating the joint density function over all parameters. The Bayes test is computed by performing a Monte Carlo integration based on the method proposed by Nandram and Kim (2002).

Keywords: Bayes factor, Dirichlet priors, Gibbs sampler, pooling, small areas.

1. Introduction

In many surveys, there are several small areas and a contingency table is constructed for each area. We consider a hierarchical Dirichlet-multinomial model to analyze the counts from these small areas. Our concern is to perform a test of independence which is competitive to the chi-square test for a single table. We follow a Bayesian inferential procedure so that appropriate priors are needed.

Statistical inference for small areas requires considerable care because the sample sizes for small areas are usually very small. To solve this problem, we use a hierarchical model which is to borrow strength some information from the neighboring areas. In this paper, we use the hierarchical Bayesian model to study the pooling effects in Bayesian tests of independence for contingency tables from small areas.

There are several literatures on methods for pooling of data. Malec and Sedransk (1992) developed a Bayesian procedure for estimation of the means for the specified experiments among a set of seemingly similar experiments. The proposed flexible prior distribution allows the intensity and nature of the pooling to be influenced by the sample data. Evans and Sedransk (1999) proposed an alternative Bayesian model with covariates that is more flexible.

[1] Ph.D. candidate, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

Evans and Sedransk (2003) provided a fully Bayesian justification for the results in Malec and Sedransk.

There is a wide literature on Bayesian methods for analyzing data with contingency tables. Agresti and Hitchcock (2005) surveyed Bayesian methods for categorical data analysis, with emphasis on contingency table analysis. The general concern with hierarchical Bayesian approach to contingency table analysis is how to handle the hyperparameters. In Dirichlet-multinomial model, Leonard (1977) made approximations when deriving the posterior to account for hyperparameter uncertainty. By contrast, Nandram (1998) used the Metropolis-Hastings algorithm to sample from the posterior distribution, rendering Leonard's approximation unnecessary. Recently hierarchical Bayesian models in the contingency tables from small areas with nonresponses have been studied in Woo and Kim (2015, 2016).

In this paper, we will construct Bayesian tests of independence using a hierarchical multinomial model with Dirichlet priors. We will investigate the pooling effects in Bayes factors through the three different types of pooling strategies for Dirichlet priors; no pooling, complete pooling and adaptive pooling. In Section 2, we introduce the hierarchical Bayesian models under the three different types of pooling strategies for the test of independence. Then we obtain the corresponding three Bayes factors using the marginal likelihoods. In Section 3, we present the results of numerical study with some simulated data. Finally, we provide some discussion and concluding remarks in Section 4.

2. Hierarchical Bayesian models

2.1. General models

For the sth area of S small areas, we consider the $r \times c$ contingency tables with cell counts, n_{sjk}, which are the responses for the kth column and jth row in the sth area. Let π_{sjk} denote the corresponding probabilities of each unit cell in the sth area. When p_{sj} and q_{sk} are marginal probabilities for each column and each row in the sth area, the independence models have $\pi_{sjk} = p_{sj}q_{sk}$, $j = 1, \cdots, r$, $k = 1, \cdots, c$, where $\sum_{j=1}^{r} p_{sj} = 1$ and $\sum_{k=1}^{c} p_{sk} = 1$ for $s = 1, \cdots, S$. Let n_{si}, $i = 1, \cdots, I\ (= rc)$ denote the cell counts for the sth area and π_{si} denote the corresponding probabilities of each area. We assume that

$$\mathbf{n}_s|\boldsymbol{\pi}_s \overset{ind}{\sim} \text{Multinimial}(n_s, \boldsymbol{\pi}_s),\ s = 1, \cdots, S \tag{2.1}$$

where $\mathbf{n}_s = (n_{s1}, \cdots, n_{sI})$ for $s = 1, \cdots, S$ is the vector of responses with $n_s = \sum_{i=1}^{I} n_{si}$, total sum of responses, and $\boldsymbol{\pi}_s = (\pi_{s1}, \cdots, \pi_{sI})$ is the corresponding probability vector of each area with $\sum_{i=1}^{I} \pi_{si} = 1$. Here I is denoted by the number of cells for the table corresponding to each area.

Now we consider three types of pooling strategies for the general model (2.1).

1) No pooling, $\boldsymbol{\pi}_s \overset{iid}{\sim} \text{Dirichlet}(\mathbf{1}),\ s = 1, \cdots, S$;

2) Complete pooling, $\tilde{\boldsymbol{\pi}} \sim \text{Dirichlet}(\mathbf{1})$ with $\boldsymbol{\pi}_1 = \cdots = \boldsymbol{\pi}_S = \tilde{\boldsymbol{\pi}}$;

3) Adaptive pooling, $\boldsymbol{\pi}_s \overset{iid}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau)$,

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K)$, $0 \le \mu_k \le 1$, $\sum_{k=1}^{K} \mu_k = 1$ and $\tau > 0$ are hyperparameters for Dirichlet distribution and are assumed to have the noninformative and proper prior

$\pi(\boldsymbol{\mu}, \tau) = (K-1)!/(1+\tau)^2$. This prior is very similar to a half-Cauchy prior and can prevent overestimation of scale parameters from our models. Recall that $\mathbf{x}|\boldsymbol{\mu}, \tau \sim \text{Dirichlet}(\boldsymbol{\mu}\tau)$ has the density $f(\mathbf{x}|\boldsymbol{\mu}, \tau) = \prod_{i=1}^{k} x_i^{\mu_i \tau - 1}/D(\boldsymbol{\mu}\tau)$, $0 < x_i < 1$, $\sum_{i=1}^{k} x_i = 1$ where $D(\boldsymbol{\mu}\tau) = \prod_{i=1}^{k} \Gamma(\mu_i \tau)/\Gamma(\tau)$, $0 < \mu_i < 1$, $\tau > 0$, is the Dirichlet function, also known as the multivariate Beta function.

Under no pooling, the joint density function for all variables is

$$\pi(\mathbf{n}, \boldsymbol{\pi}) = \prod_{s=1}^{S} \{f(\mathbf{n}_s|\boldsymbol{\pi}_s)\pi(\boldsymbol{\pi}_s)\} = \prod_{s=1}^{S} \Big\{ \frac{n_s!}{\prod_{i=1}^{I} n_{si}!} \prod_{i=1}^{I} \pi_{si}^{n_{si}} (I-1)! \Big\},$$

where $\mathbf{n} = (\mathbf{n}_1, \cdots, \mathbf{n}_S)$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_S)$. In no pooling, we could not obtain any information from prior distribution. For the marginal likelihood, we can use the posterior density of $\boldsymbol{\pi}_s$ given $\mathbf{n}_s$,

$$\boldsymbol{\pi}_s \mid \mathbf{n}_s \overset{ind}{\sim} \text{Dirichlet}(\mathbf{n}_s + \mathbf{1}),\ s = 1, \cdots, S.$$

So we obtain the marginal likelihood for the sth area,

$$f(\mathbf{n}_s) \quad = \quad \frac{n_s!(I-1)!}{\prod_{i=1}^{I} n_{si}!} \frac{\prod_{i=1}^{I} \Gamma(n_{si}+1)}{\Gamma(\sum_{i=1}^{I} n_{si} + I)}.$$

Under complete pooling, the joint density function is

$$\pi(\mathbf{n}, \tilde{\boldsymbol{\pi}}) = \prod_{s=1}^{S} \{f(\mathbf{n}_s|\tilde{\boldsymbol{\pi}})\}\pi(\tilde{\boldsymbol{\pi}}) = \prod_{s=1}^{S} \Big\{ \frac{n_s!}{\prod_{i=1}^{I} n_{si}!} \prod_{i=1}^{I} \tilde{\pi}_i^{n_{si}} \Big\} (I-1)!$$

where $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \cdots, \tilde{\pi}_I)$. The parameter $\tilde{\boldsymbol{\pi}}$ is shared by the data in entire areas. Then we can calculate the marginal likelihood using the posterior density of $\tilde{\boldsymbol{\pi}}$ which is the Dirichlet distribution with parameter $\sum_{s=1}^{S} \mathbf{n}_s + \mathbf{1}$. Our marginal likelihood under complete-pooling is

$$f(\mathbf{n}) \quad = \quad \frac{\prod_{s=1}^{S} n_s!(I-1)!}{\prod_{s=1}^{S} \prod_{i=1}^{I} n_{si}!} \frac{\prod_{s=1}^{S} \prod_{i=1}^{I} \Gamma(n_{si}+1)}{\Gamma(\sum_{s=1}^{S} \sum_{i=1}^{I} n_{si} + I)}$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the gamma function. In complete pooling, the data are tied by interested parameter $\tilde{\boldsymbol{\pi}}$, so our marginal likelihood for entire areas is expressed in just one equation. The idea of complete-pooling can be a contrast to no-pooling case.

Under adaptive pooling, the joint density function for all variables is

$$\pi(\mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\mu}, \tau) \quad = \quad \prod_{s=1}^{S} \Big\{ \frac{n_s!}{\prod_{i=1}^{I} n_{si}!} \prod_{i=1}^{I} \pi_{si}^{n_{si}} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{i=1}^{I} \pi_{si}^{\mu_i \tau - 1} \Big\} \frac{(I-1)!}{(1+\tau)^2}.$$

Here we need to know the posteriors for all parameters to be integrated out in the joint density function. The posterior density of $\boldsymbol{\pi}_s$, $s = 1 \cdots, S$ under adaptive-pooling is

$$\boldsymbol{\pi}_s \mid \mathbf{n}_s, \boldsymbol{\mu}, \tau \overset{ind}{\sim} \text{Dirichlet}(\mathbf{n}_s + \boldsymbol{\mu}\tau),\ s = 1, \cdots, S;$$

$$\pi(\boldsymbol{\mu},\tau \mid \mathbf{n}) \propto \prod_{s=1}^{S}\Big\{\prod_{i=1}^{I}\frac{D(\mathbf{n}_s+\boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau)}\Big\}\frac{1}{(1+\tau)^2}.$$

Each area has separate parameter vector $\boldsymbol{\pi}_s$ for $s=1,\cdots,S$. However, the data in all areas is indirectly pooled by hyper-parameters $\boldsymbol{\mu}$ and τ. Using these posteriors, we obtain the marginal likelihood

$$f(\mathbf{n}_s) \;=\; \frac{(I-1)!n_s!}{\prod_{i=1}^{I} n_{si}!}\int_{\boldsymbol{\mu}}\int_{\tau}\frac{D(\boldsymbol{\mu}\tau+n_s)}{D(\boldsymbol{\mu}\tau)(1+\tau)^2}d\boldsymbol{\mu}d\tau,\ s=1\cdots,S.$$

For the computation of this marginal likelihood, we can use the method developed by Nandram and Kim (2002). They use an importance function which exploits the multiplication rule of probability, and is appropriate for any hierarchical model.

2.2. Independence models

Let n_{sjk}, $j=1,\cdots,r$, $k=1,\cdots,c$, be the cell counts for jth row and kth column in sth area, $s=1,\cdots,S$ with corresponding cell probability $\pi_{sjk}=p_{sj}q_{sk}$ where $p_{sj}=\sum_{k=1}^{c}\pi_{sjk}$ and $q_{sk}=\sum_{j=1}^{r}\pi_{sjk}$. We assume that

$$\mathbf{n}_s|\boldsymbol{p}_s,\boldsymbol{q}_s \overset{ind}{\sim} \text{Multinomial}(n_s, vec(\boldsymbol{p}_s\boldsymbol{q}_s')),\ s=1,\ldots,S \tag{2.2}$$

where $\mathbf{n}_s=(n_{s11},\cdots,n_{s1c},\cdots,n_{sr1},\cdots,n_{src})$, $n_s=\sum_{j=1}^{r}\sum_{k=1}^{c}n_{sjk}$, $\boldsymbol{p}_s=(p_{s1},\cdots,p_{sr})$, $\boldsymbol{q}_s=(q_{s1},\cdots,q_{sc})$, $\sum_{j=1}^{r}p_{sj}=1$, and $\sum_{k=1}^{c}q_{sk}=1$.

For the independence model (2.2), we consider no pooling

$$\boldsymbol{p}_s \overset{iid}{\sim} \text{Dirichlet}(\mathbf{1}_p);$$
$$\boldsymbol{q}_s \overset{iid}{\sim} \text{Dirichlet}(\mathbf{1}_q).$$

Under no pooling, the joint density function for all variables is

$$\pi(\mathbf{n}_s,\boldsymbol{p}_s,\boldsymbol{q}_s)=\frac{n_s!}{\prod_{j=1}^{r}\prod_{k=1}^{c}n_{sjk}!}\prod_{j=1}^{r}\prod_{k=1}^{c}(p_{sj}q_{sk})^{n_{sjk}}(r-1)!(c-1)!,\ s=1,\ldots,S.$$

Then the marginal likelihood for the sth area is

$$f(\mathbf{n}_s) \;=\; (r-1)!(c-1)!\frac{n_s!}{\prod_{j=1}^{r}\prod_{k=1}^{c}n_{sjk}!}\int_{\boldsymbol{p}_s}\prod_{j=1}^{r}p_{sj}^{n_{sjk}}d\boldsymbol{p}_s\int_{\boldsymbol{q}_s}\prod_{k=1}^{c}q_{sk}^{n_{sjk}}d\boldsymbol{q}_s.$$

Using the posterior distributions,

$$\boldsymbol{p}_s \mid \mathbf{n}_s^{(1)} \overset{ind}{\sim} Dirichlet(\mathbf{n}_s^{(1)}+\mathbf{1}) \text{ and } \boldsymbol{q}_s \mid \mathbf{n}_s^{(2)} \overset{ind}{\sim} Dirichlet(\mathbf{n}_s^{(2)}+\mathbf{1}),$$

where $\mathbf{n}_s^{(1)}=(n_{s1}^{(1)},\cdots,n_{sr}^{(1)})$, $n_{sj}^{(1)}=\sum_{k=1}^{c}n_{sjk}$, $j=1,\cdots,r$, $\mathbf{n}_s^{(2)}=(n_{s1}^{(2)},\cdots,n_{sc}^{(2)})$, $n_{sk}^{(2)}=\sum_{j=1}^{r}n_{sjk}$, $k=1,\cdots,c$, we obtain the marginal likelihood

$$f(\mathbf{n}_s)=(r-1)!(c-1)!\frac{n_s!}{\prod_{j=1}^{r}\prod_{k=1}^{c}n_{sjk}!}\frac{\prod_{j=1}^{r}\Gamma(n_{sj}^{(1)}+1)\prod_{k=1}^{c}\Gamma(n_{sk}^{(2)}+1)}{\Gamma(\sum_{j=1}^{r}n_{sj}^{(1)}+r)\Gamma(\sum_{k=1}^{c}n_{sk}^{(2)}+c)}.$$

Under no pooling, the Bayes factor (BF) of the general model versus the independence model for the sth area is

$$BF_s^1 = \frac{(rc-1)!\prod_{j=1}^r \prod_{k=1}^c \Gamma(n_{sjk}+1)\Gamma(\sum_{j=1}^r n_{sj}^{(1)}+r)\Gamma(\sum_{k=1}^c n_{sk}^{(2)}+c)}{(r-1)!(c-1)!\Gamma(\sum_{j=1}^r \sum_{k=1}^c n_{sjk}+I)\prod_{j=1}^r \Gamma(n_{sj}^{(1)}+1)\prod_{k=1}^c \Gamma(n_{sk}^{(2)}+1)}.$$

Next, we consider complete pooling for the independence model,

$$\boldsymbol{p}_1 = \cdots = \boldsymbol{p}_S = \boldsymbol{p} \sim \text{Dirichlet}(\mathbf{1}_p);$$
$$\boldsymbol{q}_1 = \cdots = \boldsymbol{q}_S = \boldsymbol{q} \sim \text{Dirichlet}(\mathbf{1}_q),$$

where $\boldsymbol{p} = (p_1, \cdots, p_r)$, $\boldsymbol{q} = (q_1, \cdots, q_c)$. Under complete pooling, the joint density function for all variables is

$$\pi(\mathbf{n}, \boldsymbol{p}, \boldsymbol{q}) = \prod_{s=1}^S \frac{n_s!}{\prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} \prod_{j=1}^r \prod_{k=1}^c (p_j q_k)^{n_{sjk}} (r-1)!(c-1)!,\ s = 1, \ldots, S.$$

Then the marginal likelihood is

$$f(\mathbf{n}) = \int_{\boldsymbol{p}} \int_{\boldsymbol{q}} \prod_{s=1}^S \frac{n_s!}{\prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} \prod_{j=1}^r \prod_{k=1}^c (p_j q_k)^{n_{sjk}} (r-1)!(c-1)! d\boldsymbol{p} d\boldsymbol{q}.$$

Using the posterior distributions,

$$\boldsymbol{p} \mid \mathbf{n}^{(1)} \sim \text{Dirichlet}(\sum_{s=1}^S \mathbf{n}_s^{(1)} + \mathbf{1}) \text{ and } \boldsymbol{q} \mid \mathbf{n}^{(2)} \sim \text{Dirichlet}(\sum_{s=1}^S \mathbf{n}_s^{(2)} + \mathbf{1}),$$

where $\mathbf{n}^{(1)} = (\mathbf{n}_1^{(1)}, \cdots, \mathbf{n}_S^{(1)})$, $\mathbf{n}_s^{(1)} = \sum_{k=1}^c n_{sjk}$, $\mathbf{n}^{(2)} = (\mathbf{n}_1^{(2)}, \cdots, \mathbf{n}_S^{(2)})$, $\mathbf{n}_s^{(2)} = \sum_{j=1}^r n_{sjk}$, we have the marginal density as follows.

$$f(\mathbf{n}) = \frac{(r-1)!(c-1)!\prod_{s=1}^S s_s!}{\prod_{s=1}^S \prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} \frac{\prod_{j=1}^r \Gamma(\sum_{s=1}^S n_{sj}^{(1)}+1)\prod_{k=1}^c \Gamma(\sum_{s=1}^S n_{sk}^{(2)}+1)}{\Gamma(\sum_{j=1}^r \sum_{s=1}^S n_{sj}^{(1)}+r)\Gamma(\sum_{k=1}^c \sum_{s=1}^S n_{sk}^{(2)}+c)}.$$

Under complete-pooling, we obtain the Bayes factor of the general model versus the independence model for all areas,

$$BF^2 = \frac{(rc-1)!\prod_{s=1}^S \prod_{i=1}^I \Gamma(n_{si}+1)\Gamma(\sum_{j=1}^r \sum_{s=1}^S n_{sj}^{(1)}+r)\Gamma(\sum_{k=1}^c \sum_{s=1}^S n_{sk}^{(2)}+c)}{(r-1)!(c-1)!\Gamma(\sum_{s=1}^S \sum_{i=1}^I n_{si}+I)\prod_{j=1}^r \Gamma(\sum_{s=1}^S n_{sj}^{(1)}+1)\prod_{k=1}^c \Gamma(\sum_{s=1}^S n_{sk}^{(2)}+1)},$$

where $n_{sj}^{(1)} = \sum_{k=1}^c n_{sjk}$, $n_{sk}^{(2)} = \sum_{j=1}^r n_{sjk}$. This Bayes factor is the same for all areas. It means that the all areas are regarded as strata with same characteristics in one grand area.

Under two pooling strategies, the Bayes factors are too much influenced by observed sample data because the priors are noninformative. To complement this concern, we consider the hierarchical Bayesian model under adaptive-pooling. In specific, adaptive pooling for the independence model is given by

$$\begin{aligned}&\boldsymbol{p}_s \overset{iid}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_1, \tau_1);\\ &\boldsymbol{q}_s \overset{iid}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_2, \tau_2);\\ &\pi(\boldsymbol{\mu}_1, \tau_1) = \frac{(r-1)!}{(\tau_1+1)^2},\ \pi(\boldsymbol{\mu}_2, \tau_2) = \frac{(c-1)!}{(\tau_2+1)^2},\end{aligned}$$

where $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1r})$, $\sum_{j=1}^r \mu_{1j} = 1$, $0 < \mu_{1j} < 1$, $j = 1, \cdots, r$, $\boldsymbol{\mu}_2 = (\mu_{21}, \cdots, \mu_{2c})$, $\sum_{k=1}^c \mu_{2k} = 1$, $0 < \mu_{2k} < 1$, $k = 1, \cdots, c$, $\tau_1 > 0$, and $\tau_2 > 0$. Under adaptive pooling, the joint density function for the sth area is

$$\begin{aligned}\pi(\mathbf{n}_s, \boldsymbol{p}_s, \boldsymbol{q}_s, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau_1, \tau_2) &= \frac{n_s!}{\prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} \prod_{j=1}^r \prod_{k=1}^c (p_{sj} q_{sk})^{n_{sjk}} \frac{1}{D(\boldsymbol{\mu}_1 \tau_1)} \prod_{j=1}^r p_{sj}^{\mu_{1j}\tau_1 - 1}\\ &\times \frac{1}{D(\boldsymbol{\mu}_2 \tau_2)} \prod_{k=1}^c q_{sk}^{\mu_{2k}\tau_2 - 1} \frac{(r-1)!}{(1+\tau_1)^2} \frac{(c-1)!}{(1+\tau_2)^2}.\end{aligned}$$

In the joint density function, $f(\mathbf{n}_s \mid \boldsymbol{p}_s, \boldsymbol{q}_s)$ is also rewritten as

$$f(\mathbf{n}_s | \boldsymbol{p}_s \boldsymbol{q}_s) = \frac{\prod_{j=1}^r n_{sj}^{(1)} \prod_{k=1}^c n_{sk}^{(2)}}{n_s! \prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} f(\mathbf{n}_s^{(1)} | \boldsymbol{p}_s) f(\mathbf{n}_s^{(2)} | \boldsymbol{q}_2)$$

where $f(\mathbf{n}_s^{(1)} | \boldsymbol{p}_s) = \frac{n_s!}{\prod_{j=1}^r n_{sj}^{(1)}} \prod_{j=1}^r p_{sj}^{n_{sjk}}$ and $f(\mathbf{n}_s^{(2)} | \boldsymbol{q}_2) = \frac{n_s!}{\prod_{k=1}^c n_{sk}^{(2)}} \prod_{k=1}^c q_{sk}^{n_{sjk}}$ for $s = 1, \cdots, S$. Then we can calculate the marginal likelihood under adaptive pooling using the rewritten expression for $f(\mathbf{n}_s \mid \boldsymbol{p}_s, \boldsymbol{q}_s)$. The marginal likelihood is

$$f(\mathbf{n}_s) = \frac{(r-1)!(c-1)! n_s!}{\prod_{j=1}^r \prod_{k=1}^c n_{sjk}!} \int\int\int\int \frac{D(\boldsymbol{\mu}_1 \tau_1 + \mathbf{n}_s^{(1)}) D(\boldsymbol{\mu}_2 \tau_2 + \mathbf{n}_s^{(2)})}{D(\boldsymbol{\mu}_1 \tau_1) D(\boldsymbol{\mu}_2 \tau_2)(1+\tau_1)^2 (1+\tau_2)^2} d\boldsymbol{\mu}_1 d\tau_1 d\boldsymbol{\mu}_2 d\tau_2.$$

Hence the Bayes factor of the general model versus the independence model for the sth area under adaptive pooling is

$$BF_s^3 = \frac{(rc-1)! \int\int \frac{D(\boldsymbol{\mu}\tau + \mathbf{n}_s)}{D(\boldsymbol{\mu}\tau + 1)^2} d\boldsymbol{\mu} d\tau}{(r-1)!(c-1)! \int\int\int\int \frac{D(\boldsymbol{\mu}_1 \tau_1 + n_s^{(1)}) D(\boldsymbol{\mu}_2 \tau_2 + n_s^{(2)})}{D(\boldsymbol{\mu}_1 \tau_1) D(\boldsymbol{\mu}_2 \tau_2)(\tau_1+1)^2 (\tau_2+1)^2} d\boldsymbol{\mu}_1 d\tau_1 d\boldsymbol{\mu}_2 d\tau_2}.$$

3. Numerical study

Our interest is to investigate the pooling effects by comparing Bayes factors through the three different types of pooling strategies; no pooling, complete pooling and adaptive pooling. For the comparative study, we generate the simulated data which construct the two categorical variables. Specifically, we generated the data such that the one variable is divided into 3 levels, while the other is composed of 4 categories in 20 areas. So we have 3×4 contingency tables from 20 areas. Here the response probabilities are taken from uniform Dirichlet distribution.

Table 3.1 Contingency tables from the simulated data

Areas	n	Two categorical variables											
		1			2			3			4		
		1	2	3	1	2	3	1	2	3	1	2	3
1	156	11	2	2	40	4	3	11	8	16	29	0	30
2	100	5	0	2	4	21	28	2	18	14	2	4	0
3	167	23	10	18	65	7	17	4	6	0	9	8	0
4	103	14	9	0	7	3	8	1	7	6	1	15	32
5	141	3	0	23	45	5	4	4	11	4	7	15	20
6	193	4	66	13	16	19	3	3	25	17	4	9	14
7	146	7	58	1	1	11	14	7	3	20	11	0	13
8	141	8	2	2	0	40	26	10	6	1	3	30	13
9	182	6	53	8	27	5	33	5	23	2	8	4	8
10	190	18	4	1	33	11	11	44	7	0	20	12	29
11	103	3	6	7	11	10	11	8	18	2	10	2	15
12	168	10	5	8	39	42	28	0	5	6	12	6	7
13	147	9	3	15	4	2	17	1	66	3	16	11	0
14	157	25	8	24	7	9	19	10	1	17	11	5	21
15	109	11	4	3	48	2	1	12	10	0	8	9	1
16	164	25	1	15	5	10	11	5	15	7	56	2	12
17	174	33	11	2	17	12	22	3	4	21	13	18	18
18	162	0	65	7	10	12	4	4	2	11	2	33	12
19	176	28	8	14	11	1	34	32	2	8	14	11	13
20	109	28	5	1	25	17	6	6	1	1	1	12	6

We consider three Bayes tests of independence as well as classical χ^2 test. The χ^2 test is done for the single table in each area and so it does not allow for any pooling of data from different areas. Bayes test under no pooling is a direct Bayes test which uses only data from each area. However, Bayes test under complete pooling regards all areas as one grand area. Moreover, Bayes test under adaptive pooling is a indirect Bayes test which uses the data in all areas indirectly pooled by hyperparameters. Table 3.1 shows that the simulated counts in each area (a row of the table) are formed into a 3×4 contingency table.

Strictly speaking while the p-value can be used to provide evidence for the alternative, the Bayes factor can be used to provide evidence of either the null hypothesis or the alternative hypothesis over the other. If the $log(BF)$ is in $(0, 1)$ or the p-value is in $(.05, .10)$ we get borderline evidence against the null hypothesis. If the $log(BF)$ is in $(1, 3)$ or the p-value is in $(.025, .05)$, we get positive evidence against the null hypothesis. If the $log(BF)$ is in $(3, 5)$ or the p-value is in $(.01, .025)$, we get strong evidence against the null hypothesis. If the $log(BF)$ is greater than 5 or the p-value is in $(.000, .010)$, we get very strong evidence against the null hypothesis; see Kass and Raftery (1995).

In Table 3.2, we compare the $log(BF)$ values of three Bayes tests with the p-values of the classical χ^2 test. The classical χ^2 test is comparable to both Bayes test under no pooling and Bayes test under adaptive pooling. But Bayes test under complete pooling provides very strong evidence against the null hypothesis in all areas which might lead misleading in inference. There are some areas in which the χ^2 test and Bayes test under no pooling provide same inferences, but Bayes test under no pooling and Bayes test under adaptive pooling lead to different inference.

Table 3.2 Comparisons of the χ^2 test and three Bayes tests

Areas	χ^2test	$log(BF^1)$	$log(BF^2)$	$log(BF^3)$
1	0.0000	16.8213	37.4837	17.1179
2	0.0000	5.2395	37.4837	9.5066
3	0.0000	9.5482	37.4837	14.0984
4	0.0000	20.7604	37.4837	23.8498
5	0.0000	40.4786	37.4837	39.4519
6	0.0000	17.0117	37.4837	19.5477
7	0.0000	48.9853	37.4837	49.6256
8	0.0000	21.3097	37.4837	21.6755
9	0.0000	40.3248	37.4837	43.1446
10	0.0000	19.1170	37.4837	21.1589
11	0.0005	8.5958	37.4837	9.967
12	0.0827	-0.9977	37.4837	3.877
13	0.0000	57.6440	37.4837	59.1607
14	0.0915	-0.8139	37.4837	0.0185
15	0.0000	8.5715	37.4837	9.277
16	0.0000	27.4994	37.4837	28.0636
17	0.0000	20.3411	37.4837	21.729
18	0.0000	23.4972	37.4837	25.3115
19	0.0000	16.4594	37.4837	16.1591
20	0.0000	10.7442	37.4837	14.2011

We have monitored the convergence of Gibbs sampler using Geweke's test and trace plots. It turns out that our Gibbs sampler is enough to have convergence to get random samples based on some diagnostic measures.

Table 3.3 Geweke's statistic for convergence diagnostic in Gibbs sampler

General model						
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Statistics	-0.5503	0.4695	1.9387	-0.0118	0.4447	-1.4995
p-value	0.5821	0.6387	0.0525	0.9906	0.6565	0.1337
	μ_7	μ_8	μ_9	μ_{10}	μ_{11}	μ_{12}
Statistics	-0.6491	0.5523	2.1683	-0.2583	-1.8313	0.103
p-value	0.5163	0.5807	0.0301	0.7962	0.0671	0.9179
Independence model						
	μ_{11}	μ_{12}	μ_{13}			
Statistics	-0.6962	-0.1775	1.3423			
p-value	0.4863	0.8591	0.1795			
	μ_{21}	μ_{22}	μ_{23}	μ_{24}		
Statistics	2.2492	-0.6979	-0.9235	-1.1826		
p-value	0.0245	0.4853	0.3557	0.2370		

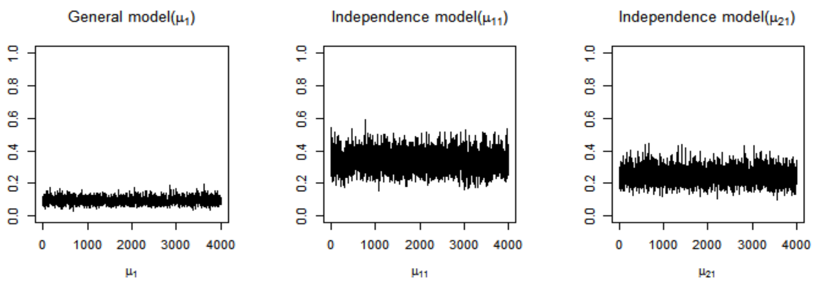

Figure 3.1 Trace plots of μ_1 in Gibbs sampler

4. Concluding remarks

This article has presented Bayes factors for the test of independence for contingency tables from small areas. To obtain the test statistics, we have constructed the hierarchical Bayesian model with Dirichlet priors. In both no pooling and complete pooling, the data are directly pooled by probability parameters of Multinomial distribution. On the contrary to this, the adaptive pooling is assumed that the data are pooled indirectly by hyper-parameters in Dirichlet prior. Then we compare the Bayes factors with the χ^2 test. We generate the data from uniform Dirichlet distribution and analyze the simulated data. As a result, the adaptive pooling seems to provide the better inference among three pooling strategies. By using indirect pooling method, we can calculate better test statistic which allows borrowing information from similar areas like in small area estimation.

References

Agresti, A. and Hitchcock, D. B. (2005). Bayesian inference for categorical data analysis. *Statistical Methods and Applications*, **14**, 297-330.

Evans, R. and Sedransk, J. (1999). Methodoloty for pooling subpopulation regressions when sample sizes are small and there is uncertainty about which subpopulations are similar. *Statistica Sinica*, **9**, 345-359.

Evans, R. and Sedransk, J. (2003). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain: II. *Journal of Statiatical Planning and Inference*, **111**, 95-100.

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

Leonard, T. (1977). Bayes simultaneous estimation for several multinomial distributions. *Communications in Statistics: Theory and Methods*, **6**, 619-630.

Malec, D. and Sedransk, J. (1992). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain. *Biometrika*, **79**, 593-601.

Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, **61**, 97-112.

Nandram, B. and Kim, H. (2002). Marginal likelihood for a class of Bayesian generalized linear models. *Journal of Statistical Computation and Simulation*, **72**, 319-340.

Woo, N. and Kim, D. H. (2015). A Bayesian uncertainty analysis for nonignorable nonresponse in two-way contingency table. *Journal of the Korean Data & Information Science Society*, **26**, 1547-1555.

Woo, N. and Kim, D. H. (2016). A Bayesian model for two-way contingency tables with nonignorable nonresponse from small areas. *Journal of the Korean Data & Information Science Society*, **27**, 245-254.