Kuwait J. Sci. **40** (2) pp. 103-113, 2013

A conditional Bayesian approach for testing independence in two-way contingency tables

Z. SABERI* AND M. GANJALI*

**Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran.*

ABSTRACT

Bayesian methods for exact small-sample analysis with categorical data in $I \times J$ contingency tables are considered. Point null hypotheses versus two-sided hypothesis are tested concerning log odds ratios in these tables with fixed row margins. The conditional distribution of the sufficient statistics for interesting parameters conditional on sufficient statistics of other nuisance parameters in the model is obtained and used to eliminate the effect of nuisance parameters. This distribution is Fisher's multivariate noncentral hypergeometric distribution. Three Bayesian approaches, hierarchical Bayes, empirical Bayes, and noninformative Bayes are considered and compared by simulation studies.

Keywords: Bayes Factor; Bayesian P-value; hierarchical Bayes; important sampling; noncentral hypergeometric distribution.

INTRODUCTION

To test for independence in two-way contingency tables by use of classical methods, one may use the Pearson's Chi-squared test, the likelihood ratio test or tests based on a divergence measure (see for instance Cressie & Read, 1984; Pardo, 2006). But when the sample sizes are small, these approximate methods are not valid and other methods for exact inference should be used (Cochran, 1954). For some discussion about the robustness of tests see Casella & Moreno (2009). In this paper, we shall use a conditional Bayesian approach. Previously Altham (1969,1971) presented Bayesian analogs of small-sample frequentist tests for 2×2 tables using prior information. An alternative approach using normal priors for logit received considerable attention in the 1970s by Leonard (1972). For a review of Bayesian inference for categorical data see Agresti & Hitchcock (2005).

Here, we will condition on all row totals of $I \times J$ tables, i.e., the experimental design will be the same as that of Fisher in his famous example for 2×2 tables. Under independence, conditioning as well on column totals to eliminate the nuisance parameter yields Fisher's multivariate noncentral hypergeometric distribution. This is a member of the exponential family. Here, we consider different Bayesian approaches such as, empirical and hierarchical Bayes to make inference about the parameters of interest [log odds ratios, a matrix of dimension $(I-1)\times(J-1)$]. Then we present a Bayesian test of independence. For testing hypothesis the Bayes factor and Bayesian P-value will be used as Bayesian evidence.

Consider testing independence (H_0) against having correlation (H_1) between two variables of an $I\times J$ contingency table with given row margins. Suppose N_{ij}, $i=1,2,...,I; j=1,2,....,J$, is the variable which shows the number of events in the i th row and jth column, and $N_i=(N_{i1},N_{i2},....,N_{iJ})'$, for $i=1,2,...,I$, denote independent multinomial random variables with parameters $N_{i+}=\sum_j N_{ij}=n_{i+}$ and probability vector $\pi_i=(\pi_{i1},\pi_{i2},....,\pi_{iJ})'$ ($\sum_{j=1}^J \pi_{ij}=1$ for all i). The conditional distribution of $N=\{N_{ij}\}$ given $N_{+j}=n_{+j}$ (where $N_{+j}=\sum_{i=1}^I N_{ij}$) for $j=1,2,....,J$ under H_0 is a multivariate hypergeometric distribution with probability mass function,

$$Pr(N=n|n_{1+},n_{2+},...,n_{I+},n_{+1},n_{+2},...,n_{+J})=\frac{\prod_{j=1}^J n_{+j}!\prod_{i=1}^I n_{i+}!}{M!\prod_i\prod_j n_{ij}!}, \tag{1}$$

where $n=\{n_{ij}\}$, $M=\sum_i n_{i+}=\sum_j n_{+j}$ and n_{ij}'s are the observed counts of the table (Agresti, 2002). The non-null (dependence) conditional distribution of $N=\{N_{ij}\}$ is given by:

$$f(N_{11},...,N_{1J},....,N_{I1},...,N_{IJ}|n_{1+},...,n_{I+},n_{+1},n_{+2},...,n_{+J},M;\Theta)=$$

$$\frac{\prod_{i=1}^I\binom{n_{i+}}{n_{i1},n_{i2},...,n_{iJ}}\prod_{i=1}^{I-1}\prod_{j=1}^{J-1}\theta_{ij}^{n_{ij}}}{\sum\limits_{\{n_{ij};\ \sum_i n_{ij}=n_{+j};\ j=1,2,...,J\}}[\prod_{i=1}^I\binom{n_{i+}}{n_{i1},n_{i2},...,n_{iJ}}\prod_{i=1}^{I-1}\prod_{j=1}^{J-1}\theta_{ij}^{n_{ij}}]}, \tag{2}$$

where Θ is the vector of all odds ratios $\theta_{ij}\geq 0$ and nonnegative integers n_{ij} are the observed values of random variables N_{ij}'s consistent with the marginal totals. Also in (2) n_{i+}'s, $i=1,2,...,I$, are row margins, n_{+j}'s, $j=1,2,....,J$, are the column margins, M is the sample size, and $\theta_{i,j}=\frac{\pi_{i,j}\pi_{I,J}}{\pi_{i,J}\pi_{I,j}}$ and take values 1 under the null hypothesis of independence. This is the *multivariate Fisher's noncentral hypergeometric distribution* (McCullagh & Nelder, 1989).

2. Bayesian Approaches in $I\times J$ Contingency Tables

In this subsection empirical, noninformative and hierarchical priors for $I\times J$ contingency tables are presented for testing independence.

Let us reparametrize the conditional distribution of the vector N using $\delta_{ij} = \ln(\theta_{ij}),\ i = 1, 2, ..., I-1;\ j = 1, 2, ..., J-1$, in order to provide a natural prior distribution for $\Delta = (\delta_{11}, ..., \delta_{1(J-1)}, ..., \delta_{(I-1)(J-1)})$. When the two variables are independent the vector of log odds ratios Δ is 0. In general Δ is symmetric about 0.

It is also known that the empirical estimates of the log odds ratios based on the observed data is approximately normally distributed in studies of even moderate sample sizes. This reparametrization helps us because we can use the multivariate normal distribution as a natural prior distribution for the vector of log odds ratios (McCullagh & Nelder, 1989).

In this paper, we consider the test of independence against any kind of association which is the test of $H_0 : \Delta = 0$ against $H_1 : \delta_{ij} \neq 0$; at least for one $(i,j),\ i = 1, 2, ..., I-1; j = 1, 2, ..., J-1$.

2.1 Empirical Bayes approach

Empirical Bayes prior approximates the prior distribution by *frequentist methods* when the prior information is too vague. For Fisher's exact test in $I \times J$ contingency tables let us assume a multivariate normal distribution with mean $\underline{\mu}$ and covariance matrix Σ, i.e.,

$$\pi(\Delta|\underline{\mu}, \Sigma) \sim N_{(I-1)(J-1)}(\underline{\mu}, \Sigma).$$

to compute the empirical estimate of $(\underline{\mu}, \Sigma)$, we have to obtain the marginal density of $(N|\underline{\mu}, \Sigma)$ which is

$$f(N|\underline{\mu}, \Sigma) = \int_{\Delta^*} f(N|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta)\pi(\Delta|\underline{\mu}, \Sigma)d\Delta,$$

where Δ^* is $\mathrm{R}^{(I-1)(J-1)}$. In computing $f(N|\underline{\mu}, \Sigma)$, we shall use importance sampling (Srinivasan, 2002). So we estimate $(\underline{\mu}, \Sigma)$ by the classical methods such as maximum likelihood or other estimation methods by using $f(N|\underline{\mu}, \Sigma)$. This prior is denoted by $\pi_{EB}(.)$ where in this situation is $N(\underline{\mu},\ \hat{\Sigma})$.

For this prior, the Bayes factor for testing independence is given by

$$B_{01_{EB}} = \frac{f(N = n|N_{1+}, ..., N_{I+}, N_{+1}, ..., N_{+J}, M, \Delta = 0)}{\int_{\Delta^*} f(N = n|N_{1+}, ..., N_{I+}, N_{+1}, ..., N_{+J}, M, \Delta)\pi_{EB}(\Delta)d\Delta}$$

$$= \frac{f(N = n|N_{1+}, ..., N_{I+}, N_{+1}, ..., N_{+J}, M, \Delta = 0)}{\int_{\Delta^*} |\hat{\Sigma}^{-12}| f(N = n|N_{1+}, ..., N_{I+}, N_{+1}, ..., N_{+J}, M, \Delta)\varphi(\hat{\Sigma}^{-12}(\Delta - \underline{\mu}))d\Delta},$$

where $\varphi(.)$ is the density function of the standard multivariate normal distribution. This form helps us to approximate the mean of $f(N = n|N_{1+}, ..., N_{I+}, N_{+1}, ..., N_{+J}, M, \Delta)$ by Monte Carlo method, using sample means of the simulated values of the multivariate normal distribution with mean $\underline{\mu}$ and covariance matrix $\hat{\Sigma}$. So the denominator of B_{01EB} can be calculated and the value of B_{01EB} can be obtained.

2.2 Hierarchical Bayes approach

For the hierarchical Bayesian approach, as a prior distribution at the first stage we consider $\pi(\Delta|\underline{\mu}, \Sigma)$ as the multivariate normal distribution with unknown mean and unknown covariance matrix and at the second stage, we consider the following priors for $\underline{\mu}$ and Σ.

$$\underline{\mu} \sim \quad N(0, \sigma_{\mu}^2 I) \ ,$$

$$\Sigma \sim W_{(I-1)\times(J-1)}(\upsilon_0 I, (I-1)(J-1)),$$

where σ_{μ}^2 and υ_0 are known constants and these values may be chosen to reflect noninformativeness of priors for $\underline{\mu}$ and Σ.

The resulting Bayes factor is given by

$$B_{01_{HB}} = \frac{f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta = 0)}{\int_{\Delta} f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta)\pi(\Delta)d\Delta} \tag{3}$$

where

$$\pi(\Delta) = \int_{\underline{\mu}} \int_{\Sigma} \pi(\Delta|\underline{\mu}, \Sigma)\pi(\underline{\mu})\pi(\Sigma) \ d\Sigma d\underline{\mu}.$$

The denominator of (3) is computationally intractable. We shall use numerical methods such as Monte Carlo to approximate it. This form helps us to approximate the mean of $f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta)$ with respect to the joint distribution of $(\Delta, \underline{\mu}, \Sigma)$.

2.3 Noninformative prior

For a noninformative Bayesian approach, δ (the vector of log odds ratios) is assigned the Jeffreys' prior; this is one of the popular noninformative priors for parameters of interest. So for determining Jeffreys' prior (by definition), at first we find the Fisher's information matrix for the multivariate Fisher's noncentral hypergeometric distribution.

Fisher's information matrix assuming the multivariate Fisher's noncentral hypergeometric distribution is

$$I(\Delta) = Cov(N).$$

Then Jeffreys' prior for Δ in a 2×2 contingency table is

$$\pi_{NJ}(\Delta) \propto |Cov(N)|^{\frac{1}{2}}.$$

The resulting Bayes factor is given by

$$B_{01_{NJ}} = \frac{f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta = 0)}{\int_{\Delta^*} f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, M; \Delta)\pi_{NJ}(\Delta)d\Delta}. \tag{4}$$

In (4), the denominator is computationally intractable, in which case we shall use numerical methods such as Monte Carlo to approximate it. This form helps us to approximate the mean of $f(n|n_{1+}, ..., n_{I+}, n_{+1}, ..., n_{+J}, N; \Delta)$ with respect to $\pi_{NJ}(\Delta)$. As generating random values from π_{NJ} is difficult, one may use the following acceptance-rejection algorithm (Robert \& Casella, 2004). In this algorithm, suppose there exists a density function $g(\Delta)$ and a constant $c \geq 1$ such that $\pi_{NJ}(\Delta) \leq cg(\Delta)$ for all $\Delta \in \mathrm{R}^{(I-1)(J-1)}$. Typically, g is a density such that Monte Carlo sampling from g is easy, for instance, a multivariate uniform distribution.

Acceptance-rejection sampling may be performed as follows:

1. Simulate Δ from $g(\Delta)$, and U uniformly on (0,1).
2. If $U < \dfrac{\pi_{NJ}(\Delta)}{cg(\Delta)}$, then accept Δ as a draw from the $\pi_{NJ}(\Delta)$. If not, reject Δ and try again.

The algorithm is repeated until the desired sample size is obtained.

3. Simulation studies

In this section, we present a simulation study to consider and compare the performance of the three Bayesian approaches (empirical, hierarchical and noninformative Bayes) for independence test, in a 2×3 contingency table with given margins. Four different values of $\Theta = (\theta_{11}, \theta_{12})$ and different values of m with assumption $n_{1+} = 2m$ and $n_{2+} = n_{+2} = n_{+1} = m$ are chosen. The results of these simulation studies are given in Tables 1 and 2.

In Table 1, to compare the three Bayesian approaches (empirical, noninformative and hierarchical Bayes), posterior probabilities of null hypothesis, $P(\Theta = 1|N = n)$ are computed for each contingency table generated in each simulation iteration. Table 1 reports the mean of these posterior probabilities for different values of m and $\Theta = (\theta_{11}, \theta_{12})$ over all simulations. In this study we consider $\theta_{11} = \theta_{12}$. The results of Table 1 show that, when $\Theta = 1$, for all values of m, the empirical Bayes is the most conservative approach (although its posterior probability for $m = 2$ is near 0.52). When sample size increases and the true values of Θ are equal to 1, noninformative and empirical Bayesian approaches perform in a similar manner, and these are better than empirical Bayes. But, when the true values of Θ are far from 1 as the posterior probability of $\Theta = 1$ is the smallest for empirical Bayes, this method performs the best for all values of m. In general, one may say that empirical Bayes performs well for any value of Θ and any value of m.

In Table 2 the percentage of times the various two-sided tests reject H_0 (or power) is recorded for all observed simulations from the noncentral ($\Theta \neq 1$) or central $[\Theta = (\theta_{11}, \theta_{12}) = (1, 1)]$ multivariate Fisher's distribution, assuming various values of the true value Θ.

The guidelines discussed in Jeffreys (1961) and in Kass & Raftery (1995) concerning the use of Bayes factors in testing hypotheses will be used here, which is an upper bound cutoff for the Bayes factor of at most 0.1 for rejecting H_0 (The former states that B_{01} between 0.1 and 0.3162 indicates moderate or substantial evidence against H_0, B_{01} between 0.01 and 0.1 indicates strong evidence against H_0, and B_{01} less than 0.01 indicates decisive evidence against H_0).

The results of Table 2 show that, for small samples $m = 2$ ($N = 6$, M is sample size), the three Bayesian approaches are very conservative. When the sample sizes are increased, for $m = 4$, the hierarchical and noninformative Bayes methods are more conservative than empirical Bayes. But for $m = 8, 16$ and 32, when Θ is far from 1, the empirical Bayes is the best because the powers of rejecting H_0 in favour of H_1 is greatest for this approach. Also in this table for all values of Θ in Θ_1 and large m, noninformative Bayes is the most conservative approach. In general, Table 2 shows that empirical Bayes performs well for any value of Θ and any value of $m > 2$. This is consistent with the results of Table 1.

Table 1. Mean of simulated posterior probabilities of H_0 for 3000 generated contingency tables for different values of m and $\Theta = (\theta_{11}, \theta_{12}) = (\theta, \theta)$
[*: Value of Θ in which H_0 is true.]

Value of m		$\theta_{11} = \theta_{12} = \theta$				
	Value of θ	(1*)	(5)	(10)	(20)	(50)
		$\delta = ln(\theta)$				
		(0*)	(1.6)	(2.3)	(2.99)	(3.91)
$m = 2$	Empirical Bayes	0.5256	0.4292	0.3455	0.2404	0.1970
	Noninformative (Jeffery's) Bayes	0.5276	0.4995	0.4032	0.3296	0.2934
	Hierarchical Bayes	0.5304	0.4862	0.3792	0.2907	0.2793
$m = 4$	Empirical Bayes	0.5621	0.4045	0.2840	0.1989	0.0883
	Noninformative (Jeffery's) Bayes	0.5638	0.4726	0.3481	0.2901	0.2006
	Hierarchical Bayes	0.5698	0.4855	0.4529	0.3576	0.2494
$m = 8$	Empirical Bayes	0.5862	0.3073	0.1805	0.0691	0.0211
	Noninformative (Jeffery's) Bayes	0.5912	0.4326	0.2741	0.1972	0.0847
	Hierarchical Bayes	0.6365	0.4502	0.2969	0.1667	0.0767
$m = 16$	Empirical Bayes	0.5979	0.1623	0.0422	0.0147	0.0008
	Noninformative (Jeffery's) Bayes	0.6023	0.3423	0.1813	0.0894	0.0046
	Hierarchical Bayes	0.7385	0.3295	0.1114	0.0399	0.0037
$m = 32$	Empirical Bayes	0.6009	0.0495	0.0040	5.01×10^{-5}	1.42×10^{-7}
	Noninformative (Jeffery's) Bayes	0.6307	0.1732	0.0873	0.0079	4.23×10^{-5}
	Hierarchical Bayes	0.7675	0.1314	0.0112	0.0003	2.88×10^{-6}

Table 2. Percentage of simulations out of 3000 in which $H_0 : \Theta = \Theta_0$ is rejected in favor of $H_1 : \Theta \neq \Theta_0$ when Bayes factor is less than 0.1 for different values of m and $\Theta = (\theta_{11}, \theta_{12}) = (\theta, \theta)$ [*: Value of Θ for which H_0 is true.].

Value of m	$\theta_{11} = \theta_{12} = \theta$					
	Value of θ	(1*)	(5)	(10)	(20)	(50)
		$\delta = ln(\theta)$				
		(0*)	(1.6)	(2.3)	(2.99)	(3.91)
$m = 2$	Empirical Bayes	0	0	0	0	0
	Noninformative (Jeffery's) Bayes	0	0	0	0	0
	Hierarchical Bayes	0	0	0	0	0
$m = 4$	Empirical Bayes	0.81	5.12	17.00	33.62	61.75
	Noninformative (Jeffery's) Bayes	0	0	0	0	0
	Hierarchical Bayes	0	0	0	0	0
$m = 8$	Empirical Bayes	0.70	27.12	57.20	77.62	95.00
	Noninformative (Jeffery's) Bayes	1.11	6.48	13.41	28.93	56.76
	Hierarchical Bayes	1.81	6.04	10.66	34.66	72.02
$m = 16$	Empirical Bayes	0.40	46.9	82.87	98.12	99.70
	Noninformative (Jeffery's) Bayes	0.98	10.78	41.31	60.12	78.19
	Hierarchical Bayes	1.53	18.03	67.09	82.06	98.10
$m = 32$	Empirical Bayes	0.20	85.12	99.01	99.99	99.99
	Noninformative (Jeffery's) Bayes	0.87	59.53	95.046	98.92	99.99
	Hierarchical Bayes	1.10	61.02	98.01	99.99	99.99

4. Real application

Table 3, taken from Graubard & Korn (1987) which refers to a prospective study of maternal drinking and congenital malformations, has 32574 observations. For testing independence of alcohol consumption and malformation, corrected Pearson statistic gives $\chi^2 = 12.0821$ (P − value =0.0377), likelihood ratio statistic gives $G^2 = 6.2019$ (P − value =0.18456) and corrected likelihood ratio statistic gives $G_c^2 = 4.0835$ (P − value =0.3948). These values illustrate that different statistics and approximations can give quite different results.

Table 3. Maternal drinking and congenital malformations (Graubard & Korn, 1987). Choice of column scores for testing independence in ordered $2 \times k$ tables. *Biometrics* **43**: 471-476.

	Alcohol consumption (average no. of drinks/day)					
Malformation	0	< 1	1-2	3-5	≥ 6	Total
Absent	17066	14464	788	126	37	32481
Present	48	38	5	1	1	93
Total	17114	14502	793	127	38	32574

Using empirical, noninformative and hierarchical priors for $[\delta_{11} = ln(\theta_{11}),\ \delta_{12} = ln(\theta_{12}),\ \delta_{13} = ln(\theta_{13}),\ \delta_{14} = ln(\theta_{14})]$ in testing independence $(H_0 : \delta_{11} = \delta_{12} = \delta_{13} = \delta_{14} = 0)$, we found $B_{01_{EM}} = 2.28363$, $B_{01_{HB}} = 9.04001$ and $B_{01_{NJ}} = 7.4637$. The results of Bayesian analysis show that there is not any evidence against the null hypothesis. Between these, empirical Bayes gives more support to H_0 in a manner like the corrected likelihood ratio test. So consumption of alcohol and malformations are independent. So, our Bayesian final result agrees with that of likelihood ratio test statistic.

REFERENCES

Agresti, A. 2002. Categorical Data Analysis, Wiley, New York.

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

Altham, P. M. E. 1969. Exact Bayesian analysis of a 2 × 2 contingency table, and Fisher's exact significance test. Journal of the Royal Statistical Society, Series B., Methodological, **31**: 261-269.

Altham, P. M. E. 1971. The analysis of matched proportions. Biometrika **58**: 561-576.

Casella, G. & Moreno, E. 2009. Assessing robustness of intrinsic tests of independence in two-way contingency tables. JASA, **104**(487): 1261-1271.

Cochran, W. G. 1954. Some methods of strengthening common χ^2 tests. Biometrics, **10**: 417-451.

Cressie, N. & Read, T. R. C. 1984. Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society, Series B, **46**: 440-464.

Graubard, B. I., & Korn, E. I. 1987. Choice of column scores for testing independence in ordered $2 \times k$ tables. Biometrics, **43**: 471-476.

Jeffreys, H. 1961. Theory of Probability. Oxford University Press.

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

Leonard, T. 1972. Bayesian methods for binomial data. Biometrika, **59**: 581-589.

McCullagh, P. & Nelder, J. A. 1989. Generalized Linear Models, 2nd edition. London: Chapman & Hall.

Pardo, L. 2006. Statistical Inference Based on Divergence Measures. Chapman & Hall/CRC.

Robert, C.P. & Casella, G. 2004. Monte Carlo Statistical Methods, 2nd edition, New York: Springer-Verlag.

Srinivasan, R. 2002. Importance sampling - Applications in communications and detection, Springer-Verlag, Berlin.

Submitted : *1/11/2011*
Revised : *9/12/2012*
Accepted : *19/12/2012*

نهج بايزي شرطي لاختبار الاستقلال في جداول الطوارئ ذات الاتجاهين

***ز. صبيري، **م. كَنجعلى**
قسم الاحصاء - كلية علوم الاحصاء - جامعة شهيد بهشتي - إيران

خلاصة

أخذ بعين الاعتبار أساليب بايزي لتحليل العينة الصغيرة الدقيقة مع المعطيات الفئوية في جداول الطوارئ. وتم اختبار فرضية نقطة العدم مقابل الفرضية ذات الجانبين فيما يتعلق بسجل نسب الأرجحية في هذة الجداول مع هوامش صف ثابت. وتم الحصول على التوزيع الشرطي من إحصاءات كافية لعوامل مثيرة للاهتمام مشروطة بإحصاءات كافية من عوامل الإزعاج الأخرى في النموذج استخدمت للقضاء على تأثير عوامل الأزعاج. هذا التوزيع هو توزيع فيشر الهندسي الزائدي غير المحوري متعدد المتغيرات. وقد تم اعتماد ثلاثة أنواع من نهج بايزي ومقارنتها بدراسات المحاكاة وهي بايز الهرمية، وبايز التجريبية وبايز غير المفيدة.