

Bayesian test of independence and conditional independence of two ordinal variables

Zahra Saberi*

*Department of Mathematical Sciences, Isfahan University of Technology,
Isfahan 84156-83111, Iran.
z_saberi@cc.iut.ac.ir*

Mojtab Ganjali†

*Department of Statistics, Faculty of Mathematical Sciences,
Shahid Beheshti University, Tehran, Iran.
m-ganjali@sbu.ac.ir*

Received 14 August 2014

Accepted 12 January 2015

For analysis of contingency tables with large sample size, classical approaches using approximate methods have high power. However, when the sample size is small or some cells have frequencies less than 5, classical approaches are so conservative. Also asymptotic behavior may be poor when the table contains small counts. So, Bayesian test of independence for two-way contingency tables with ordinal variables is considered. The conditional independence of two ordinal variables given values of a controlling variable is also considered. To do these tests, gamma and partial gamma are used as association parameters for ordinal variables. Since gamma has a complex posterior form, it is intractable to compute directly the necessary inferential measures. So, a Dirichlet distribution is used as a prior distribution for the vector of cell probabilities, then the use of computational methods such as the Monte Carlo algorithm is introduced to generate samples from posterior distribution of gamma. Also, the Bayesian P-value and Bayes factor are obtained. In a simulation study, the choice of appropriate prior distribution for gamma is discussed and also the performance of gamma is compared to that of kappa. It is shown that, in contingency tables with ordinal variables, it is better to apply gamma as a measure of association. Some sensitivity analysis to the choice of prior are also performed on real applications.

Keywords: Association parameters; Bayesian P-value; Gamma; Sensitivity Analysis.

2000 Mathematics Subject Classification: 62F03, 62F15, 62H15, 62H17.

1. Introduction

Many different methods have been proposed for testing independence and conditional independence of two ordinal variables. For systematic reviews, see the books of Agresti (1984, 2002) and Clogg and Shihadeh (1994) and also McCullagh and Nelder (1989, pp.151-155). However, despite recent intensive interest in using the notion of conditional independence (see Wermuth and Cox, 1998a

*Assistant Professor of Statistics.

†Professor of Statistics.

and 1998b) to simplify multivariate systems (Edwards, 1995; Cox and Wermuth, 1996; Lauritzen, 1996; Wermuth, 1998) such possibilities have not been set out in detail for Bayesian approach of testing independence of ordinal variables.

A commonly used measure of association in two-way contingency tables where both the row and column variables are ordinal is Goodman and Kruskal's γ . It was introduced in the first paper of a series of four papers by Goodman and Kruskal (1954, 1959, 1963, 1972), which were later published together as a book (Goodman and Kruskal, 1979). The measure γ has since become a familiar measure of ordinal association for social, behavioral and medical scientists; see, for example, Freeman (1986) and Gonzalez and Nelson (1996). Also partial γ is a measure of partial association often used for analyzing a three-way contingency table with two ordinal responses and a controlling variable.

The measure γ is constructed based on the number of concordant and discordant pairs of observations in a sample. Denoting the ordinal row and column variables in a contingency table by X and Y , respectively, a pair of observations is concordant if the member that ranks higher on X also ranks higher on Y , and discordant if the member that ranks higher on X ranks lower on Y . Ties, namely pairs of observations that have equal values of X or equal values of Y , are ignored by γ .

In contingency tables when sample sizes are small, approximate methods, such as chi-squared approximation, are not valid, and other methods such as exact inference should be used. These methods use distributions determined exactly rather than as large-sample approximations. The P-value for exact test contingency tables is defined as the set of tables (with the same sample size) that are no more likely to occur than the table observed. So computing P-value for these tables when the sample sizes and/or dimensions of tables increased, is intractable and time-consuming.

Special algorithms and software for computing exact tests for these tables are widely available (e.g., Mehta and Patel, 1983). We recommend these tests when asymptotic approximations may be invalid. However, computing time increases exponentially as n or number of levels of variables increases. One can use Monte Carlo to sample randomly from the set of tables with the given sample size to reduce the time. The estimated P-value is then the sample proportion of tables having test statistic value at least as large as the value observed. As dimensions of table increase, the number of possible values for any test statistic tends to increase. Thus, the conservativeness issue for testing conditional independence or independence becomes problematic (Agresti, 2002).

Problems arising for calculating P-value and also its conservative nature of rejecting H_0 lead us to think about a Bayesian approach which solves these problems.

In this paper we like to introduce Bayesian approaches for testing conditional independence (H_0 : all partial association parameters are equal to 0) versus existence of conditional association in a three-way contingency table with two ordinal responses, X and Y , and a controlling variable, Z , with K categories. Our approaches utilize Bayesian criteria such as Bayes factor or Bayesian P-value (Agresti and Hitchcock, 2005) for testing this hypothesis. These approaches are not related to sample size and dimension of table, and so Bayesian approaches, considered in this paper, may be a better option than classical approaches.

In a simulation study we investigate how to choose, in an objective view, the prior distribution. Also in another simulation study, we compare the performance of two measures of association, γ (which considers the ordinality of the variables) and κ (which ignores the ordinality of the variables and considers them as nominal) and show that the performance of γ based on the posterior probabilities is better than κ when variables are ordinal (for a Bayesian approach using κ for measuring agreement between raters, see Broomeling, 2009, page 42).

Section 2 gives a review of the concept of conditional independence in three-way contingency tables with ordinal responses. The details of the Bayesian approach for testing conditional independence in three-way contingency tables with two ordinal responses are given in Section 3. In section 4, some simulation studies are given in order to investigate the behavior of the prior distribution of γ . In section 5, some real applications are analyzed. In the end, some conclusions are given.

2. Conditional independence in three-way contingency tables with ordinal responses

An important part of most studies is the choice of controlling variables. In studying the relation between responses X and Y , one should control for any covariate that can influence that relationship. This involves using some mechanism to hold the covariate constant. Otherwise, an observed relationship between X and Y may actually reflect effects of that covariate on both X and Y . The relationship between X and Y then shows confounding.

In this section we discuss test of conditional independence between ordinal responses X and Y in each category of a controlling variable, say Z . For simplicity, we consider a single control variable, but generalization to more controlling variables may be done in the same manner.

We first define $\gamma_{(k)}$ as a measure of partial ordinal association. We illustrate it for $I \times J \times K$ tables, where K denotes the number of categories of a control variable, Z . Let $\{\pi_{ijk}\}$ denote the joint probability distribution in this table. Within a fixed category k of Z , in order to define partial association, $\gamma_{(k)}$, we must describe the partial probabilities of concordance and discordance in each partial table k . For two independent observations from the k th partial table, the partial probabilities of concordance and discordance are denoted by $\Pi_{C(k)}$ and $\Pi_{D(k)}$, respectively, and are defined as:

$$\Pi_{C(k)} = 2 \sum_i \sum_j \pi_{ijk} \left(\sum_{h>i} \sum_{t>j} \pi_{htk} \right), \text{ and } \Pi_{D(k)} = 2 \sum_i \sum_j \pi_{ijk} \left(\sum_{h>i} \sum_{t<j} \pi_{htk} \right).$$

Conditional on the two independent observations not being tied on both ordinal variables in each category of control variable (Z), the partial probabilities of concordance and discordance are given by $\frac{\Pi_{C(k)}}{\Pi_{C(k)} + \Pi_{D(k)}}$ and $\frac{\Pi_{D(k)}}{\Pi_{C(k)} + \Pi_{D(k)}}$, respectively. So, $\gamma_{(k)}$ as a measure of partial association in table k is the difference between these partial probabilities, i.e.,

$$\gamma_{(k)} = \frac{\Pi_{C(k)} - \Pi_{D(k)}}{\Pi_{C(k)} + \Pi_{D(k)}}; \quad k = 1, 2, \dots, K.$$

If there are no control variables, we define γ as a measure of association between two ordinal variables. In this case, the index of k is removed from the above formula.

2.1. Conditional Independence

Consider an $I \times J \times K$ table that describes the relationship between ordinal responses, X and Y , controlling for Z . If X and Y are independent in partial table k , then X and Y are called *conditionally independent at level k* of Z .

More generally, X and Y are said to be conditionally independent given Z when they are conditionally independent at every level of Z .

Suppose $\pi_{ijk} = P(X = i, Y = j, Z = k)$ denotes the joint probability distribution for three variables, X, Y and Z . Then

$$\pi_{ijk} = P(X = i, Z = k)P(Y = j|X = i, Z = k),$$

which under conditional independence of X and Y , given Z , equals

$$\pi_{ijk} = \pi_{i.k} \frac{P(Y = j, Z = k)}{P(Z = k)} = \frac{\pi_{i.k} \pi_{.jk}}{\pi_{.k}}; \text{ for all } i, j, \text{ and } k.$$

where $\pi_{i.k}$ is the marginal distribution of X and Z , $\pi_{.jk}$ are marginal probabilities of Y and Z and $\pi_{.k}$ are marginal probabilities of Z . Also conditional independence implies that $\gamma_{(k)} = 0$, for $k = 1, 2, \dots, K$, but the converse of this statement may not be true.

3. The Bayesian approach for testing conditional independence

First, consider a three-way contingency table having two ordinal responses, X with I categories and Y with J categories, and a controlling variable Z with K categories. The number of subjects in all $I \times J$ subtables is known. Our interest is on testing conditional independence ($H_0 : \gamma_{(1)} = \dots = \gamma_{(K)} = 0$) against $H_1 : \gamma_{(k)} \neq 0$, for at least one $k, k = 1, \dots, K$. Suppose cell counts from different partial tables are independent. Let $N_{ijk}, i = 1, 2, \dots, I; j = 1, 2, \dots, J$, be the number of events in the cell at the intersection of i th row and j th column in table $k, k = 1, 2, \dots, K$, and $\mathbf{N}_k = (N_{11k}, N_{12k}, \dots, N_{1Jk}, \dots, N_{IJk})'$ denotes vector of multinomial random variables with respective indexes $n_k = \sum_i \sum_j n_{ijk}$ (where n_{ijk} is the observed value of N_{ijk} and n_k is a known value) and probability vector parameters $\Pi_k = (\pi_{11k}, \pi_{12k}, \dots, \pi_{1Jk}, \dots, \pi_{IJk})'$ (where $\sum_i \sum_j \pi_{ijk} = 1$). The parametric vector of interest is the vector of partial associations between two ordinal responses (X and Y), $\Gamma = (\gamma_{(1)}, \dots, \gamma_{(K)})$, each element of which was defined in the previous section.

For testing conditional independence using a Bayesian approach, the values of π_0 and π_1 , prior probabilities of H_0 and H_1 , are chosen such that $\pi_0 + \pi_1 = 1$. A proper prior, denoted by $g(\Gamma)$ defined on $\Gamma_1 = [-1, 1]^K = [-1, 1] \times \dots \times [-1, 1]$ (Robert, 2007), should be also chosen. We consider $g(\Gamma) = g_1(\gamma_{(1)}) \times \dots \times g_K(\gamma_{(K)})$, where $g_k(\gamma_{(k)}); k = 1, \dots, K$, are proper distributions on $[-1, 1]$. So, the Bayes factor in favor of H_0 and Bayesian P-value for this test are defined as,

$$B_{01} = \frac{\pi_0 f(\mathbf{N}|\Gamma=0)}{\pi_1 m(\mathbf{N})}, \text{ and Bayesian } P\text{-value} = \frac{\pi_0 f(\mathbf{N}|\Gamma=0)}{\pi_0 f(\mathbf{N}|\Gamma=0) + \pi_1 m(\mathbf{N})},$$

respectively, where $f(\mathbf{N}|\Gamma=0) = \prod_{k=1}^K f(\mathbf{N}_k|\gamma_{(k)}=0)$ and

$$\begin{aligned} m(\mathbf{N}) &= \int_{\Gamma_1} f(\mathbf{N}|\Gamma) g(\Gamma) d\Gamma = \int \left[\prod_{k=1}^K f(\mathbf{N}_k|\gamma_{(k)}) g_k(\gamma_{(k)}) \right] d\Gamma \\ &= \prod_{k=1}^K \int_{-1}^1 f(\mathbf{N}_k|\gamma_{(k)}) g_k(\gamma_{(k)}) d\gamma_{(k)} = \prod_{k=1}^K m_k(\mathbf{N}_k). \end{aligned}$$

By considering $\pi_0 = \frac{1}{2}$, the Bayes factor and Bayesian P-value for this test are given by the following formula, respectively,

$$B_{01} = \frac{\prod_{k=1}^K f(\mathbf{N}_k|\gamma_{(k)}=0)}{\prod_{k=1}^K m_k(\mathbf{N}_k)}, \quad (3.1)$$

$$\text{Bayesian } P\text{-value} = \frac{\prod_{k=1}^K f(\mathbf{N}_k | \gamma_{(k)} = 0)}{\prod_{k=1}^K f(\mathbf{N}_k | \gamma_{(k)} = 0) + \prod_{k=1}^K m_k(\mathbf{N}_k)} = \frac{1}{1 + \frac{1}{B_{01}}}. \quad (3.2)$$

For testing conditional independence in this table, by use of a Bayesian approach, the posterior probability of H_0 and Bayes factor as some evidence must be computed, so we need to determine the posterior distribution of Γ . For computing the posterior distribution of Γ , it is enough to determine the posterior distribution of each $\gamma_{(k)}$. Computing this distribution ($\pi(\gamma_{(k)} | \mathbf{N}_k)$) is very intractable due to the complex form of $\gamma_{(k)}$. So we first obtain the posterior distribution of Π_k 's and then by use of it and the relationship between Π_k and $\gamma_{(k)}$, the posterior probability of $\gamma_{(k)}$ will be computed.

For determining the posterior probability of Π_k , we consider a Dirichlet distribution with $\mathbf{a}_k = (\alpha_{11k}, \dots, \alpha_{IJk})$ as a prior distribution for Π_k , i.e.,

$$\pi((\pi_{11k}, \pi_{12k}, \dots, \pi_{IJk}) | (\alpha_{11k}, \dots, \alpha_{IJk})) \sim \text{Dirich}(\mathbf{a}_k).$$

Since the Dirichlet distribution is a conjugate prior for multinomial distribution, the posterior distribution of Π_k is a Dirichlet distribution with vector of parameters $\mathbf{a}_k^* = (\alpha_{11k}^*, \dots, \alpha_{IJk}^*)$, where $\alpha_{ijk}^* = \alpha_{ijk} + N_{ijk}$; $i = 1, \dots, I, j = 1, \dots, J$ and $k = 1, \dots, K$. So the posterior distribution has the form,

$$\pi((\pi_{11k}, \pi_{12k}, \dots, \pi_{IJk}) | \mathbf{N}_k, (\alpha_{11k}, \dots, \alpha_{IJk})) = \frac{1}{B(\mathbf{a}_k^*)} \prod_i \prod_j \pi_{ijk}^{\alpha_{ijk}^* + N_{ijk}},$$

where $B(\mathbf{a}_k^*)$ is the multinomial beta function, which can be expressed in terms of the gamma function as,

$$B(\mathbf{a}_k^*) = \frac{\prod_i \prod_j \Gamma(a_{ijk}^*)}{\Gamma(\sum_i \sum_j a_{ijk}^*)}.$$

Now in order to determine the posterior distribution of $\gamma_{(k)}$, $\pi(\gamma_{(k)} | \mathbf{N}_k)$, we can use the following algorithm to achieve the posterior value of $\gamma_{(k)}$.

- (1) Simulate Π_k from a Dirichlet distribution with known vector of parameters $\mathbf{a}_k = (\alpha_{11k}, \dots, \alpha_{IJk})$. Then the prior values of $\gamma_{(k)}$ using Π_k can be computed.
- (2) Then simulate Π_k^* from a Dirichlet distribution with vector of parameters $\mathbf{a}_k^* = (\alpha_{11k}^*, \dots, \alpha_{IJk}^*)$, where $\alpha_{ijk}^* = \alpha_{ijk} + N_{ijk}$. Then the posterior values of $\gamma_{(k)}$ using Π_k^* can be computed.

The algorithm is repeated until the desired number of samples is obtained.

The values of $\gamma_{(k)}$ that were generated in the first stage, give us information about the prior distribution of $\gamma_{(k)}$. By generating the posterior value of $\gamma_{(k)}$, the posterior distribution of $\gamma_{(k)}$ may be obtained empirically.

3.1. Computing Bayes factor and Bayesian P-value

Consider a test of the conditional independence hypothesis in a three-way contingency table with two ordinal responses, i.e., $H_0 : \gamma_{(1)} = \dots = \gamma_{(K)} = 0$. Let $m_{ijk} = \frac{n_{i.k} n_{.jk}}{n_{..k}}$ be the maximum likelihood estimates of π_{ijk} under the conditional independence model, where $n_{i.k}$ and $n_{.jk}$, respectively are row and column marginal counts in table k and $n_{..k}$ is the total number of subjects in each subtable.

In order to conduct a test of independence via the Bayesian approach, we need to compute Bayes factor (3.1). For computing Bayes factor, we must determine denominator of equation (3.1).

Under null hypothesis, the numerator of B_{01} is $\prod_{k=1}^K f(\mathbf{N}_k = \mathbf{n}_k | \hat{\pi}_{ijk} = m_{ijk})$. For determining the denominator, we must compute,

$$m(\mathbf{N}) = \int_{\Gamma_1} f(\mathbf{N}|\Gamma)g(\Gamma)d\Gamma, \quad (3.3)$$

where $g(\cdot)$ is the prior distribution of Γ and $\Gamma_1 = [-1, 1]^K$. In order to determine the prior distribution of Γ , the prior distribution of each $\gamma_{(k)}$ must be obtained. So, we compute it using prior distribution of Π_k as follows. By use of the simulated values of Π_k , from a Dirichlet distribution with vector of parameters \mathbf{a}_k , and relationship between Π_k and $\gamma_{(k)}$, the prior values of $\gamma_{(k)}$ are computed. This is equivalent to the fact that prior distribution for $\gamma_{(k)}$ is obtained by a transformation of the assumed prior distribution of Π_k which is a Dirichlet Distribution. With these values of $\gamma_{(k)}$, prior distribution of $\gamma_{(k)}$ are determined empirically. Since $\gamma_{(k)}$'s are independent, then the prior distribution of Γ can be obtained empirically.

This form of (3.3) helps us to approximate the mean of $f(\mathbf{N}_k|\gamma_{(k)})$ by Monte Carlo method, using sample means of the computed prior values of $\gamma_{(k)}$, i.e.,

$$m_k(\mathbf{N}_k) \simeq \frac{1}{M} \sum_{i=1}^M f(\mathbf{N}_k|\gamma_{(k)}^{(i)}),$$

where M is the number of simulated Π_k . So the value of B_{01} can be obtained and by use of (3.2), we can obtain the Bayesian P-value.

3.2. The Bayesian approach for testing independence

In a similar manner, the test of independence ($H_0 : \gamma = 0$) against having association ($H_1 : \gamma \neq 0$) between two ordinal variables of a $I \times J$ contingency table with given sample size can be performed. Suppose N_{ij} , $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$, is the number of events in the cell at the intersection of i th row and j th column, and $\mathbf{N} = (N_{11}, N_{12}, \dots, N_{1J}, \dots, N_{IJ})'$ denotes multinomial random variables with respective indexes $n = \sum_i \sum_j n_{ij}$ (where n_{ij} denotes the cell frequency in the i th row and j th column of the table) and probability vector parameters $\Pi = (\pi_{11}, \pi_{12}, \dots, \pi_{1J}, \dots, \pi_{IJ})'$ (where $\sum_i \sum_j \pi_{ij} = 1$).

For testing independence (H_0) versus some association using a Bayesian approach, Bayes factor and Bayesian P-value are defined as,

$$B_{01} = \frac{f(\mathbf{N}|\gamma = 0)}{m(\mathbf{N})}, \quad (3.4)$$

and,

$$\text{Bayesian P-value} = \frac{f(\mathbf{N}|\gamma = 0)}{f(\mathbf{N}|\gamma = 0) + m(\mathbf{N})} = \frac{1}{1 + \frac{1}{B_{01}}}, \quad (3.5)$$

respectively, where $m(\mathbf{N}) = \int_{\Gamma_1} f(\mathbf{N}|\gamma)g(\gamma)d\gamma$.

By generating γ values by an algorithm similar to that outlined in the previous subsection, the posterior distribution of γ may be obtained empirically. Also the Bayes factor and Bayesian P-value for this test may be computed in a similar manner to (3.1) and (3.2) where the index k is removed.

4. Simulation studies

In this section, we present some simulation studies to investigate the behavior of extracted prior distribution for γ . Also the performance of γ is compared with that of κ (Cohen, 1960) for measuring association of two ordinal variables (We have used R software for simulation studies).

At first we shall try to assess the behavior of prior distribution of γ in a 2×3 contingency table. So, we consider a Dirichlet distribution as a prior distribution for vector of probabilities $\Pi = (\pi_{11}, \pi_{12}, \dots, \pi_{23})$ with vector of parameters $\mathbf{a} = (a_{11}, a_{12}, \dots, a_{23})$. We have chosen different vectors of parameters (different \mathbf{a}) in order to study the sensitivity of results with respect to different priors. (all elements of \mathbf{a} equal to α , Jeffrey's prior when $\alpha = 0.05$). Symmetric Dirichlet distributions are often used when a Dirichlet prior has no prior knowledge favoring one component over another. When $\alpha = 1$, the symmetric Dirichlet distribution is equivalent to a uniform distribution over all points in its support. Values of α above 1 refer to distributions that are dense, evenly-distributed distributions, i.e. all probabilities returned are similar to each other. Values of α below 1 refer to sparse distributions, i.e. most of the probabilities returned are close to 0, and most of the mass is concentrated in a few of the probabilities. If we have some prior knowledge about one or more components over another, then we may choose a Dirichlet distribution with different parameters. In this case we allocate large values of α_i 's to components that are more informative or have high frequencies.

To see the effect of the above decisions on the prior densities of γ , Figure 1 shows the prior distribution of γ for different values of the vector \mathbf{a} . This figure indicates that symmetric Dirichlet distribution with $\alpha = 1$ is a low informative prior with highly dispersed distribution for γ . Also in symmetric Dirichlet distribution, when $\alpha > 1$ the prior distribution of γ is less dispersed and generally as $\alpha (> 1)$ increases the prior distribution of γ is going to have low variance. For $\alpha < 1$, the prior values of γ are concentrated around 1 and -1 .

If the vector of parameters \mathbf{a} is chosen such that the large values of a_i 's are allocated to the probability of diagonal cells in the table, then the prior distribution of γ would be skewed. Existence of a positive association between two ordinal variables in a contingency table, indicates an increasing trend between the two variables. Then the main diagonal would have more frequencies than the other elements. So, if we want to determine prior distribution objectively, we shall allocate the large a_{ij} 's to cells about the main diagonal. This leads to a prior distribution for γ which is skewed to the left.

Where there exists a negative value of the measure of association, there is a decreasing trend between the two variables. This gives frequencies which are concentrated on the subordinate diagonal. So, we may allocate the large values of a_{ij} 's to cells about the subordinate diagonal to have a prior distribution for γ which is skewed to the right. In order to choose the Dirichlet distribution as a prior distribution of vector Π_k , one may use the symmetric Dirichlet distribution. One may choose α such that for every real number $\varepsilon > 0$,

$$|\hat{\pi}_{ijk(Bayes)} - \hat{\pi}_{ijk(classic)}| < \varepsilon,$$

where $\hat{\pi}_{ijk(Bayes)}$ is the Bayesian estimator of π_{ijk} and $\hat{\pi}_{ijk(classic)}$ is the classical estimator of π_{ijk} , i.e.,

$$\left| \frac{\alpha + N_{ijk}}{N_{..k} + IJ\alpha} - \frac{N_{ijk}}{N_{..k}} \right| < \varepsilon.$$

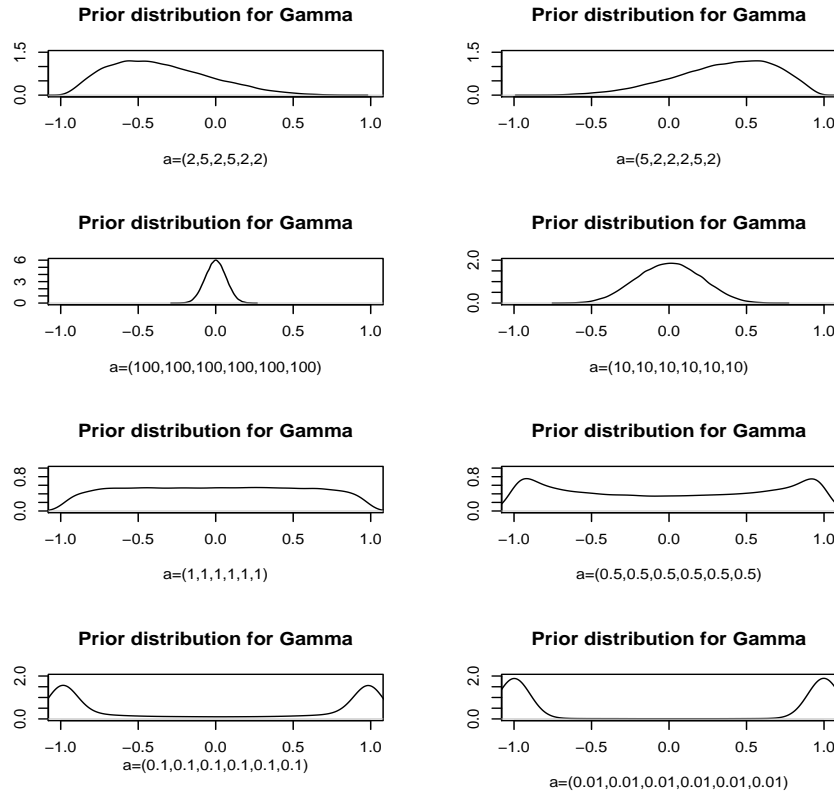


Fig. 1. Prior distribution of γ for different Dirichlet distributions with hyperparameters $\mathbf{a} = (a_{11}, a_{12}, \dots, a_{23})$ in a 2×3 contingency table.

This means to have a low-informative prior, we need to choose a very small value for ε . By considering $\varepsilon \leq \frac{1}{N_{..k}^2}$, we can have

$$\alpha \leq [|N_{..k} - N_{ijk}IJ| - \frac{IJ}{N_{..k}}]^{-1},$$

as $0 \leq \alpha$, then we must choose α such that,

$$0 \leq \alpha \leq \{||N_{..k} - N_{ijk}IJ| - \frac{IJ}{N_{..k}}|\}^{-1}.$$

In the second simulation study, we compare the performance of two measures of association γ and κ . So a 3×3 contingency table with $N = 100$ and different values of γ are chosen. In this study, 3000 3×3 contingency tables are simulated and for each table, the value of $\kappa (= \frac{\sum_{i=1}^3 (\pi_{ii} - \pi_i \pi_{.i})}{1 - \sum_{i=1}^3 \pi_i \pi_{.i}})$ is computed. In order to compare the performance of γ and κ , a Dirichlet prior distribution with vector of parameters $\mathbf{a} = (1, 1, \dots, 1)$ is considered for vector of probabilities Π . This leads to a noninformative prior for γ . Then the proportion of posterior events $(\gamma > 0|\mathbf{N})$, $(\gamma < 0|\mathbf{N})$, $(\kappa > 0|\mathbf{N})$ and $(\kappa < 0|\mathbf{N})$ are reported in Table 1.

The results of Table 1 indicate that in contingency tables with ordinal variables, when there is a positive association, both γ and κ detect the positive direction of association but κ underestimate

Table 1. Prior and posterior means of γ and κ and proportion of times that different posterior events occur for 3000 simulated 3×3 contingency tables with two ordinal variables.

	γ				
	(-0.33)	(-0.12)	(0)	(0.23)	(0.44)
$E(\gamma)$	-0.4404	-0.1329	0.0016	0.2276	0.4385
$E(\kappa)$	0.0126	-0.0032	0.0004	0.1384	0.3118
$E(\gamma \mathbf{N})$	-0.4335	-0.1307	0.0015	0.2233	0.4321
$E(\kappa \mathbf{N})$	0.0125	-0.0030	0.0005	0.1360	0.3061
$p(\gamma > 0 \mathbf{N})$	0.001	0.154	0.492	0.962	0.999
$p(\gamma < 0 \mathbf{N})$	0.999	0.846	0.508	0.038	0.001
$p(\kappa > 0 \mathbf{N})$	0.578	0.478	0.506	0.944	0.998
$p(\kappa < 0 \mathbf{N})$	0.422	0.522	0.496	0.056	0.002

Table 2. Effect of smoking on self reported health status five years later [Source: Nanny Wermuth and D. R. Cox, 1998, On the Application of Conditional Independence to Ordinal Data. *International Statistical Review* 66: 181-199]

Health status	Smoking habits fi ve years ago					Total
	Never smoked	Quit smoking	1-9	10-20	> 20	
very good	16	15	13	10	1	55
fair	73	75	59	81	29	317
bad	6	6	7	17	3	39
very bad	1	0	1	3	1	6
Total	96	96	80	111	34	417

the amount of association compared with that of γ , i.e. $\hat{\kappa} < \hat{\gamma}$. When there is a negative association i.e. $\gamma < 0$ ($\gamma = -0.33$), the estimated values of κ indicate independence or positive association since κ ignores the ordinal nature of variables. In such cases that there is a negative trend between two ordinal variables, κ can not detect it and could be highly misleading for the test of independence. Hence, there are situations in which γ truly rejects the independence hypothesis, but κ accepts it. So, in contingency tables with variables that are ordinal in nature, it is more appropriate to apply γ as a measure of association. Figure 2 shows posterior distributions of γ and κ . This confirms the result of Table 1.

5. Real applications

We consider two real data sets.

Application 1: Table 2 shows a 4×5 contingency table for two ordinal variables obtained for 417 adults in an epidemiological cohort study in Denmark. The variables are self reported health status ($i = 1$, very good; $i = 2$, fair; $i = 3$, bad; $i = 4$, very bad), and habits concerning cigarette smoking five years earlier ($j = 1$, never smoked; $j = 2$, did not smoke then; $j = 3, 4$, and 5 smoked fewer than 10, between 10 and 20 and more than 20 cigarettes per day, respectively).

For this table the classical estimate of γ is $\hat{\gamma} = 0.2415$. In order to test for independence, $H_0 : \gamma = 0$, by use of classical methods, we must compute the P -value corresponding to this hypothesis. The P -value to test for independence in these data is the probability of the set of tables that are no more likely to occur than the table observed. This task is intractable in the classical frequentist context.

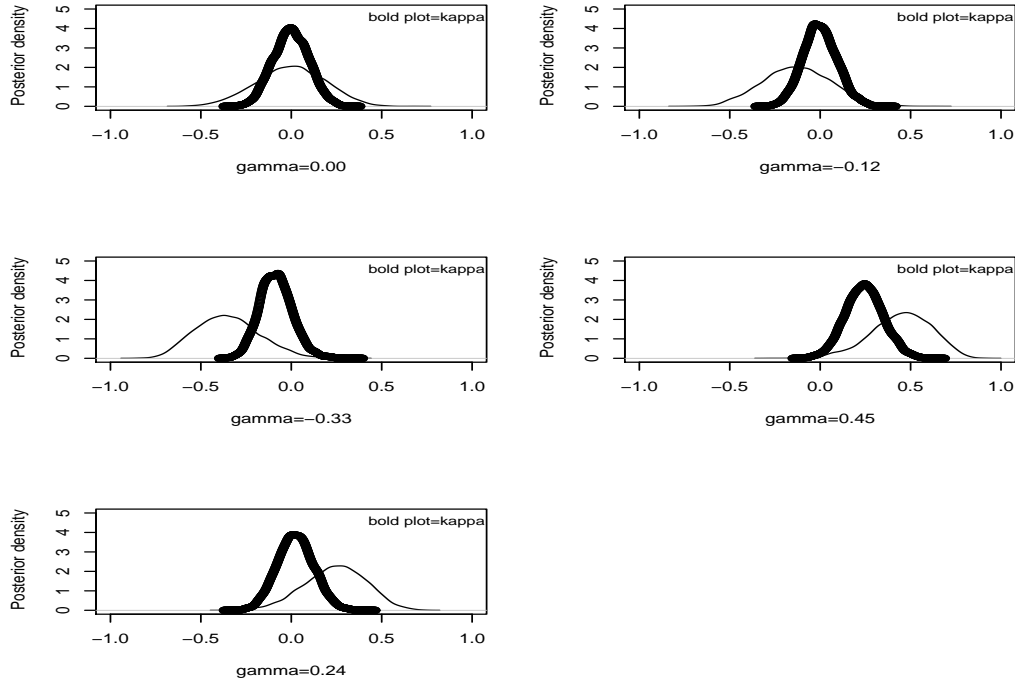


Fig. 2. Posterior distributions of γ and κ for $\alpha = 1$ and different values of γ (bold plot shows the distribution of κ and the other shows that for Γ , the real value of γ is given in the label on the x-axis).

Table 3. Prior and posterior means of γ , posterior probabilities and Bayes factor in favor of no association for different values of α .

	$\alpha_1 = \alpha_2 = \dots = \alpha_{20} = \alpha$				
	10^{-3}	10^{-2}	10^{-1}	1	10
$E(\gamma)$	-0.0034	-0.0011	-0.0102	-0.0091	-0.0081
$var(\gamma)$	0.9978	0.8861	0.3789	0.0533	0.0056
$E(\gamma \mathbf{N})$	0.2412	0.2402	0.2386	0.2147	0.1155
$\pi(\gamma \geq 0 \mathbf{N})$	0.9994	0.9995	0.9999	0.9986	0.9913
$\pi(\gamma = 0 \mathbf{N})$	0.00	0.00	0.00	0.00	0.00
B_{01}	0.00	0.00	0.00	0.00	0.00

In order to perform a Bayesian test of independence in Table 2, we choose the Dirichlet distribution as a prior distribution with several different values for $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_{20})$, in order to conduct a sensitivity analysis. Also we consider $\alpha_1 = \alpha_2, \dots = \alpha_{20} = \alpha$. The results of Table 3, show expectation and variance of prior distributions for different values of α . This table also shows the posterior mean of γ , posterior probability of $\gamma \geq 0$, Bayesian P-value and Bayes factor. These results indicate that for the values of α that are smaller than 1, the prior distributions are less informative than the priors corresponding to values of α that are greater than 1 (due to the prior variance of γ). So, the posterior means of γ corresponding to priors that are less informative, are similar to the classical estimate of γ . But for the priors that are more informative ($\alpha = 10$) than the others, the Bayesian

Table 4. HPD credible sets of γ and their lengths for different values of α .

	γ	Length(γ)
$\alpha = 0.001$	(0.1066, 0.3823)	0.2757
$\alpha = 0.01$	(0.0954, 0.3682)	0.2728
$\alpha = 0.1$	(0.1005, 0.3670)	0.2665
$\alpha = 1$	(0.0792, 0.3204)	0.2412
$\alpha = 10$	(0.0179, 0.2093)	0.1914

estimate of γ is far from the classical estimate. Also, the Bayesian methods indicate strong evidence against the null hypothesis, due to the calculated values of Bayesian P-values and Bayes factors, i.e., smoking and health status are highly associated. Also in order to check sensitivity of the choice of prior, we compute the highest posterior density (HPD) credible sets for different values of α and the results are given in Table 4. These results indicate that the length of HPD credible sets correspond to low informative priors ($\alpha \leq 1$) are larger than those of more informative priors ($\alpha > 1$). Also for $\alpha = 10$ (corresponding to the most informative prior in this example), the HPD credible set does not cover the classical estimate of γ (=0.2415).

Application 2: Table 5 shows the results of a randomized, double-blind clinical trial comparing an active hypnotic drug with a placebo in patients who have insomnia problems. The response is the patient's reported time in minutes to fall asleep after going to bed. Patients responded before and following a two-week treatment period. The subjects, receiving the two treatments, were regarded as forming two independent samples.

In this table the classical estimate of $\Gamma = (\gamma_1, \gamma_2)$ is $(\hat{\gamma}_1, \hat{\gamma}_2) = (0.453, 0.631)$ where γ_1 is the partial gamma for two responses of people who use active drug and γ_2 is the partial gamma for people who use placebo. In order to perform a Bayesian test of conditional independence in Table 5, we have chosen independence Dirichlet distributions as prior distributions for the vector of probabilities in each subtable. We have chosen different vectors of parameters (\mathbf{a}_1 and \mathbf{a}_2) in order to conduct a sensitivity analysis with respect to different priors (see, table 6).

These results indicate that for the values of a that are smaller than 1, the prior distributions are less informative than the priors corresponding to values of a that are greater than 1. So, the posterior means of Γ corresponding to priors that are noninformative, are similar to the classical estimate of Γ . But for priors that are more informative ($a = 10$), the Bayesian estimates of Γ are significantly different from those of the classical estimates. Also, the Bayesian methods indicate strong evidence against null hypothesis, due to the very low values of Bayesian P-values and Bayes factors, i.e., the time that patients reported to fall asleep after going to bed before and following a two week treatment period are not independent at both levels of treatment. Then, the time that patients reported to fall asleep after going to bed before and following a two week treatment period are not conditionally independent. So, to do any longitudinal analysis to study the effect of the hypnotic drug on insomnia, one has to take into account the association between two responses (see Ganjali and Rezaee, 2007). As this association is different for the two levels of treatment, this association should be itself modelled as a function of treatment. By considering a Dirichlet distribution with $a = 1$ for test of $H_0 : \gamma_{(2)} = \gamma_{(1)}$, we have $E(\gamma_{(2)} - \gamma_{(1)} | \mathbf{N}) = 0.18447$, $P(\gamma_{(2)} > \gamma_{(1)} | \mathbf{N}) = 0.9228$ and the Bayes factor in favor of H_0 is 0. Hence there is strong evidence against H_0 .

Table 5. Time to falling asleep, by treatment and occasion, [Source: From S. F. Francom, C.Chuang-Stein, and J. R. Landis, Statist. Med. **8**, 571-582, (1989).]

Time to falling Asleep					
Treatment	Initial	Follow up			
		< 20	[20, 30)	[30, 60)	≥ 60
Active	< 20	7	4	1	0
	[20, 30)	11	5	2	2
	[30, 60)	13	23	3	1
	≥ 60	9	17	13	8
Placebo	< 20	7	4	2	1
	[20, 30)	14	5	1	0
	[30, 60)	6	9	18	0
	≥ 60	4	11	14	22

Table 6. Posterior mean and standard error of $\Gamma = (\gamma_{(1)}, \gamma_{(2)})$, Bayesian P-value and Bayes factor in favor of $\Gamma = \mathbf{0}$ for different values of a .

a	$\alpha_1 = \alpha_2 = \dots = \alpha_{16} = a$				
	10^{-3}	10^{-2}	10^{-1}	1	10
$E(\gamma_{(1)} \mathbf{N}_1)$	0.4555	0.4505	0.4418	0.3456	0.0745
$E(\gamma_{(2)} \mathbf{N}_2)$	0.63187	0.6291	0.6213	0.5393	0.2307
Posterior SE($\gamma_{(1)}$)	0.1044	0.1041	0.1021	0.1009	0.0684
Posterior SE($\gamma_{(2)}$)	0.0759	0.0753	0.0750	0.0711	0.0668
$\pi(\gamma_{(1)} = 0 \mathbf{N}_1)$	0.0000	0.0000	0.0000	0.0000	0.0000
$\pi(\gamma_{(2)} = 0 \mathbf{N}_2)$	0.0000	0.0000	0.0000	0.0000	0.0000
B_{01}	0.0000	0.0000	0.0000	0.0000	0.0000

Conclusion

Although gamma, as a measure of association for two ordinal variables, is commonly used and it is easy to compute, for tables with small observed values, where one prefers to adopt a Bayesian approach, posterior distribution of gamma can not be characterized. In this paper we presented a method and an algorithm where one is able to generate values of posterior distribution of gamma. We performed Bayesian test of independence against association by Bayes factor and Bayesian P-value. We also extended this approach to test for conditional independence of two ordinal variables, given a controlling nominal variable. For considering more control variables, one may extend $\Pi_{C(k)}$ and $\Pi_{D(k)}$ to $\Pi_{C(kw)}$ and $\Pi_{D(kw)}$, respectively, where k and w may be used for given categories of two control variables.

In this paper, we investigated the behavior of the prior distribution for γ in a simulation study. Also, the performance of γ is compared with that of κ . So, in contingency tables with ordinal variables, if there is a positive trend, γ and κ perform in a similar manner, although $\hat{\gamma} > \hat{\kappa}$. When

there is a negative trend, κ can not detect it and then can not reject the independence hypothesis. So, in contingency tables with ordinal variables, it is better to apply γ as a measure of association.

References

- [1] A. Agresti, (*Analysis of Ordinal Categorical Data*, Wiley, New York, 1984).
- [2] A. Agresti, (*Categorical Data Analysis*, Wiley, New York, 2002).
- [3] A. Agresti and D.B. Hitchcock, Bayesian inference for categorical data analysis, *Stat. Methods Appl.*, **14**(2005) 297–330.
- [4] L. D. Broemeling, (*Bayesian Methods for Measures of Agreement*, Chapman & Hall/CRC press, New York, 2009).
- [5] C. C. Clogg and E. I. Shihadeh, *Statistical models for ordinal data*, (London: Sage, 1994).
- [6] J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, **20**(1960) 37–46.
- [7] D.R. Cox and N. Wermuth, (*Multivariate dependencies models, analysis and interpretation*, London: Chapman & Hall, 1996).
- [8] D. Edwards, (*Introduction to graphical modelling*. New York: Springer, 1995).
- [9] S. F. Francom, C. Chuang-Stein and J. R. Landis, A log-linear model for ordinal data to characterize differential change among treatments *Statist. Med.*, **8**(1998) 571–582.
- [10] L. C. Freeman, Order-Based Statistics and Monotonicity: A Family of Ordinal Measures of Association. *Journal of Mathematical Sociology*, **12** (1986) 49-69.
- [11] M. Ganjali and Z. Rezaee, A transition model for analysis of repeated measure ordinal response data to identify the effects of different treatments, *Drug information journal*, **41** (2007) 527–534.
- [12] R. Gonzalez and T. O. Nelson, Measuring Ordinal Association in Situations That Contain Tied Scores, *Psychological Bulletin*, **119** (1996) 159–165.
- [13] L.A. Goodman and W.H. Kruskal, Measures of Association for Cross Classifications, *Journal of the American Statistical Association*, **49** (1954) 732-764.
- [14] L.A. Goodman and W.H. Kruskal, Measures of Association for Cross Classifications II: Further Discussion and References, *Journal of the American Statistical Association*, **54**(1959) 123-163.
- [15] L.A. Goodman and W.H. Kruskal, Measures of Association for Cross Classifications III: Approximate Sampling Theory, *Journal of the American Statistical Association*, **58** (1963) 310-364.
- [16] L.A. Goodman and W.H. Kruskal, Measures of Association for Cross Classifications IV: Simplification of Asymptotic Variances, *Journal of the American Statistical Association*, **67** (1972) 415-421.
- [17] L.A. Goodman and W.H. Kruskal, (*Measures of Association for Cross Classifications*, Springer-Verlag, New York, 1979).
- [18] S.L. Lauritzen, (*Graphical models*. Oxford, United Kingdom: Clarendon Press, 1996).
- [19] P. McCullagh and J. A. Nelder, (*Generalized Linear Models*, **2nd**, London: Chapman and Hall, 1989).
- [20] C. R. Mehta and N. R. Patel. A network algorithm for performing Fishers exact test in $r \times c$ contingency tables, *J. Amer. Statist. Assoc.*, **78** (1983) 427-434.
- [21] C. P. Robert, (*The Bayesian Choice*, **2ed**, Springer-Verlag, New York, 2007).
- [22] Statistical dependence and independence. *Encyclopedia of Biostatistics*, P. Armitage and T. Colton (eds), New York: Wiley, (1998a)4260-4264.
- [23] N. Wermuth and D. R. Cox, On association models defined over independence graphs. *Bernoulli*, **4** (1998b) 477-495.
- [24] N. Wermuth and D. R. Cox, On the application of conditional independence to ordinal data, *International Statistical Review*, **66** (1998) 181-199.
- [25] N. Wermuth, Graphical Markov models. In S. Kotz, C. Read and D. Banks (eds), *Encyclopedia of Statistical Sciences*, **2**(1998) 284-300, New York: Wiley.