A Bayesian Test for a Two-Way Contingency Table Using Independence Priors
Author(s): James H. Albert
Source: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 18, No. 4 (Dec., 1990), pp. 347-363
Published by: Statistical Society of Canada
Stable URL: http://www.jstor.org/stable/3315841

The Canadian Journal of Statistics
Vol. 18, No. 4, 1990, Pages 347–363
La Revue Canadienne de Statistique

A Bayesian test for a two-way contingency table using independence priors

James H. ALBERT

Bowling Green State University

Key words and phrases: Bayes factor, independence hypothesis, mixtures, p-values, weight of evidence.
AMS 1985 subject classifications: Primary 62F15; secondary 62F05.

ABSTRACT

One method of testing for independence in a two-way table is based on the Bayes factor, the ratio of the likelihoods under the independence hypothesis $\mathcal{H}$ and the alternative hypothesis $\bar{\mathcal{H}}$. The main difficulty in this approach is the specification of prior distributions on the composite hypotheses $\mathcal{H}$ and $\bar{\mathcal{H}}$. A new Bayesian test statistic is constructed by using a prior distribution on $\bar{\mathcal{H}}$ that is concentrated about the "independence surface" $\mathcal{H}$. Approximations are proposed which simplify the computation of the test statistic. The values of the Bayes factor are compared with values of statistics proposed by Gunel and Dickey (1974), Good and Crook (1987), and Spiegelhalter and Smith (1982) for a number of two-way tables. This investigation suggests a strong relationship between the new statistic and the p-value.

RÉSUMÉ

Une des méthodes pour tester l'indépendance dans un tableau 2×2 est basée sur le facteur de Bayes, c'est-à-dire le rapport des vraisemblances sous l'hypothèse d'indépendance, $\mathcal{H}$, et l'hypothèse alternative $\bar{\mathcal{H}}$. Suivant cette approche, on doit spécifier des lois a priori sous les hypothèses composées $\mathcal{H}$ et $\bar{\mathcal{H}}$, cela représente une difficulté importante. Un nouveau test bayesien est construit. On utilise une loi a priori sous $\bar{\mathcal{H}}$ qui est concentrée autour de la "surface d'indépendance" $\mathcal{H}$. Des approximations permettant de simplifier le calcul de la statistique sont proposées. Pour divers tableaux 2×2, les valeurs du facteur de Bayes sont comparées aux valeurs de statistiques proposées par Gunel et Dickey (1974), Good et Crook (1987) et Spiegelhalter et Smith (1982). Cette comparaison suggère qu'il y une forte relation entre la nouvelle statistique et la probabilité de dépassement.

1. INTRODUCTION

1.1. The Problem.

Consider the problem of testing for independence in an $I \times J$ contingency table. Suppose that the data are sampled so that the total count in the table is fixed. Then the observed counts $\{X_{ij}\}$ possess a multinomial distribution with total sample size n and probabilities $\{p_{ij}\}$, where X_{ij} and p_{ij} denote the count and probability, respectively, in the ith row and jth column of the table. Let $f(\mathbf{x}|\mathbf{p})$ denote the multinomial density

$$\binom{n}{\mathbf{x}} \prod_{ij} p_{ij}^{x_{ij}},$$

where $\mathbf{x} = (x_{11}, \ldots, x_{IJ})$ and $\mathbf{p} = (p_{11}, \ldots, p_{IJ})$. In this sampling model, the hypothesis of independence can be written as

$$\mathcal{H} \; : p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, \ldots I, \quad j = 1, \ldots, J,$$

where $p_{i\cdot} = \Sigma_{j=1}^{J} p_{ij}$ and $p_{\cdot j} = \Sigma_{i=1}^{I} p_{ij}$ denote the marginal probabilities in the ith row and jth column, respectively. Generally a dot will refer to summation over a given subscript.

Two classical test statistics for this problem are the Pearson statistic $X^2 = \Sigma_{ij} (X_{ij} - n\hat{p}_{ij})^2 / n\hat{p}_{ij}$ and the log-likelihood statistic $G^2 = \Sigma_{ij} X_{ij} \log(X_{ij}/n\hat{p}_{ij})$, where $\hat{p}_{ij} = X_{i\cdot}X_{\cdot j}/n^2$ is the maximum-likelihood estimator of p_{ij} under independence. Under the hypothesis $\mathcal{H}$, each statistic is asymptotically distributed chi-squared with $(I-1)(J-1)$ degrees of freedom. If y denotes the observed value of the test statistic (either X^2 or G^2), then a decision is typically made to accept or reject $\mathcal{H}$ based on the p-value $P[\chi^2(I-1)(j-1) > y]$, the probability that a chi-squared $(I-1)(J-1)$ variate exceeds y. One difficulty with this approach is the interpretation of the p-value. Diaconis and Efron (1985) note that, when the hypothesis is strongly rejected, the p-value conveys no additional information. More generally, Edwards *et al.* (1963) and Berger and Sellke (1987) have noted that in testing a sharp null hypothesis (where the dimension of $\mathcal{H}$ is smaller than the dimension of the unrestricted parameter space), the p-value significantly overstates the evidence against $\mathcal{H}$. This discussion suggests that one should only reject when the p-value is much smaller than the usual nominal level of 0.05.

An alternative testing approach is based on the Bayes factor

$$F = \frac{m(\mathbf{x} \mid \bar{\mathcal{H}})}{m(\mathbf{x} \mid \bar{\mathcal{H}})},$$

the ratio of the weighted averages of the likelihoods under the unrestricted hypothesis $\bar{\mathcal{H}}$ and the independence hypothesis $\mathcal{H}$. This approach requires the specification of two prior distributions, one on the space $\bar{\mathcal{H}}$ and a second on the space $\mathcal{H}$. Specifically, suppose that p, conditional on $\bar{\mathcal{H}}$, is assigned a prior π_1 and, under the hypothesis of independence, the marginal vectors of probabilities $\mathbf{p}_a = (p_{1\cdot}, \ldots, p_{I\cdot})$ and $\mathbf{p}_b = (p_{\cdot 1}, \ldots p_{\cdot J})$ are assigned the joint density π_2. Then the Bayes factor is given by

$$F = \frac{\displaystyle\int f(\mathbf{x} \mid \mathbf{p})\pi_1(\mathbf{p})\, d\mathbf{p}}{\displaystyle\int f(\mathbf{x} \mid \mathbf{p})\pi_2(\mathbf{p}_a, \mathbf{p}_b)\, d\mathbf{p}_a\, d\mathbf{p}_b}. \tag{1.1}$$

In this paper, a Bayesian test of independence is developed which is similar in spirit to Jeffreys's (1961) test of a normal mean. Suppose X is distributed $\mathrm{N}(\mu, \sigma^2)$ and one is testing the hypothesis that $\mu = \mu_0$. To compute a Bayes factor, a prior needs to be assigned to the space $\mu \neq \mu_0$. On this space, Jeffreys suggested the use of a Cauchy prior centered about μ_0, the rationale being that values of μ near μ_0 are believed more likely a priori than values of μ far from μ_0. Similarly, in this testing problem, suppose that the independence hypothesis is believed *a priori* to be a plausible model. Then values of p located near the independence surface $\{p_{ij} = p_{i\cdot}p_{\cdot j}, i = 1, \ldots, I, j = 1, \ldots, J\}$ will be believed more likely than values of p far from this surface. Under these assumptions, it will then be appropriate to choose a prior on the space $\bar{\mathcal{H}}$ which is concentrated about the surface of independence.

In Albert and Gupta (1982), such a class of "independence priors" was developed by mixing Dirichlet distributions. In Section 2, we review the construction of this family of distributions and describe some of its properties. The class is indexed by a hyperparameter K which reflects the tightness of the distribution about the independence surface. Some methods for assessing K are described in that section. As K approaches infinity, the distribution places its entire mass on the $(I+J-2)$-dimensional independence surface.

Suppose a user is able to assess a value for the hyperparameter K, say K_0. Then a test for $\mathcal{H}$ against $\bar{\mathcal{H}}$ is equivalent to testing $K = \infty$ against $K = K_0$. In Section 3, we discuss the computation of the Bayes factor. This computation is difficult for large tables, since the evaluation of $m(\mathbf{x} \mid \bar{\mathcal{H}})$ requires an $(I+J-2)$-dimensional integration. One approximation is discussed which greatly simplifies the calculations.

In situations where it is difficult to assess K, we use the reporting philosophy of Dickey (1973) and compute the Bayes factor for a sequence of K-values. One can summarize this sequence of Bayes factors by the computation of $F_{\max}$, the maximum Bayes factor over values of K (Good 1967).

In Section 4 we briefly describe the alternative Bayesian testing methods of Good and Crook (1987), Gunel and Dickey (1974), and Speigelhalter and Smith (1982). In Section 5, we compare the maximum Bayes factor $F_{\max}$ with values of these alternative Bayes factors for a large number of two-way tables. The aim of this section is to compare the methods and to see the influence of the sample size, dimensions (I and J), and table margins on the test statistics. In addition, this section investigates the relationship between the p-value and the values of the different Bayes factors. In Section 6, we summarize our findings.

1.2. Notation.

Since all the prior distributions considered will be Dirichlet distributions or mixtures of them, it will be helpful to define some notation. The conjugate prior for a multinomial probability vector $\mathbf{p} = (p_1, \ldots, p_N)$ with N cells in the Dirichlet density with hyperparameter vector $\mathbf{a} = (a_1, \ldots, a_N)$:

$$\pi_d(\mathbf{p}; N, \mathbf{a}) = D_N^{-1}(a) \prod_i p_i^{a_i - 1},$$

where $D_N(\mathbf{a})$ denotes the Dirichlet function $\int \Pi_i p_i^{a_i-1}\, d\mathbf{p}$. This is denoted $\mathbf{p} \sim \text{Dirichlet}_N(\mathbf{a})$.

2. THE PRIOR DISTRIBUTION

2.1. Construction.

The Dirichlet mixture class of Albert and Gupta (1982) is constructed in two stages. At the first stage, $\mathbf{p}$, conditional on the hyperparameters $K, \mathbf{q}_a = (q_{ai}, \ldots, q_{aI})$, and $\mathbf{q}_b = (q_{b1}, \ldots, q_{bj})$, is assigned a $\text{Dirichlet}_{IJ}(K\mathbf{q})$ distribution, where the vector of prior means $\mathbf{q} = (q_{11}, \ldots, q_{IJ})$ satisfies the independence configuration $q_{ij} = q_{ai}q_{bj}$ for all i and j. Then, at the second stage, the hyperparameter vectors $\mathbf{q}_a$ and $\mathbf{q}_b$ are assumed independent and assigned the symmetric Dirichlet densities

$$\pi_d(\mathbf{q}_a;\, I,\, w\mathbf{1}) = D_I^{-1}(w\mathbf{1}) \prod_i q_{ai}^{w-1}, \qquad \pi_d(\mathbf{q}_b;\, J,\, w\mathbf{1}) = D_J^{-1}(w\mathbf{1}) \prod_j q_{bj}^{w-1},$$

where $\mathbf{1}$ is the vector of ones. The resulting prior for $\mathbf{p}$, conditional on the hyperparameter K, is given by

$$\pi_A(\mathbf{p} \mid K) = \int \pi_d(\mathbf{p}; IJ, (Kq_{ai}q_{bj}))\, \pi_d(\mathbf{q}_a; I, w\mathbf{1})\pi_d(\mathbf{q}_b; J, w\mathbf{1})\, d\mathbf{q}_a\, d\mathbf{q}_b. \tag{2.1}$$

In this paper, we assign $\mathbf{q}_a$ and $\mathbf{q}_b$ uniform densities ($w = 1$). This choice is somewhat arbitrary; other possible "noninformative" choices for w are 0 and $\frac{1}{2}$. One reason for the choice $w = 1$ (instead of the improper choice $w = 0$) is that the predictive density $m(\mathbf{x} \mid K) = \int f(\mathbf{x} \mid \mathbf{p})\pi_A(\mathbf{p} \mid K)\, d_{\mathbf{p}}$ will be finite for all $\mathbf{x}$. We will briefly investigate in Section 3.4 the sensitivity of the Bayes factor with respect to w.

2.2. *Properties.*

The prior (2.1) induces a conjugate analysis on the independence hypothesis $\mathcal{H}$. To see this, note that, as $K \longrightarrow \infty$, the prior (2.1) places all of its mass on the surface $p_{ij} = p_{i\cdot}p_{\cdot j}$, where the marginal probability vectors $\mathbf{p}_a$ and $\mathbf{p}_b$ are assigned the independent priors Dirichlet$_I(w\mathbf{1})$ and Dirichlet$_J(w\mathbf{1})$, respectively.

The distribution (2.1) is only weakly informative about the marginal probability vectors $\mathbf{p}_a$ and $\mathbf{p}_b$. This can be demonstrated by noting that the marginal density of the single-row probability p_{a1} is given by

$$\pi_A(p_{a1} \mid K) = \int_0^1 \pi_d\big(p_{a1};\ 2,\ \big(Kq_{a1},\ K(1 - q_{a1})\big)\big)\, dq_{a1}. \tag{2.2}$$

Suppose K is an integer ≥ 2. Then one can approximate the uniform prior on q_{a1} by a discrete prior which assigns uniform mass on the $K - 1$ points $1/K, 2/K, \ldots, (K-1)/K$. Then, letting $\beta = Kq_{a1}$, (2.2) is approximated by

$$\begin{aligned}
\pi_A(p_{a1} \mid K) &\simeq (K-1)^{-1} \sum_{\beta=1}^{K-1} \frac{\Gamma(K)}{\Gamma(\beta)\Gamma(K-\beta)}\, p_{a1}^{\beta-1}(1 - p_{a1})^{K-\beta-1} \\
&= \sum_{\beta=0}^{K-2} \binom{K-2}{\beta}\, p_{a1}^{\beta}(1 - p_{a1})^{K-2-\beta} \\
&= 1.
\end{aligned} \tag{2.3}$$

Thus the marginal density of p_{a1} is approximately uniform. To check the accuracy of this approximation for small K, Figure 1 gives the exact marginal prior density of p_{a1} for the hyperparameter values $K = 2, 8$, and 32. Note that this density is nonuniform only near the boundaries for p_{a1} and approaches the uniform density for larger values of K.

2.3. *Assessing the Hyperparameter K.*

The hyperparameter K reflects the strength of one's prior belief about independence. To understand how to assess this hyperparameter, consider the 2×2 table where one parameter describes the association structure. Suppose that the user has prior information about the log odds ratio $\lambda = \ln(p_{11}p_{22}/p_{12}p_{21})$. The Dirichlet mixture prior (2.1) induces a prior for λ which is symmetric about 0. Table 1, which gives quantiles of the distribution of $|\lambda|$, can be used to assess a value for K. To illustrate how this table is used, suppose

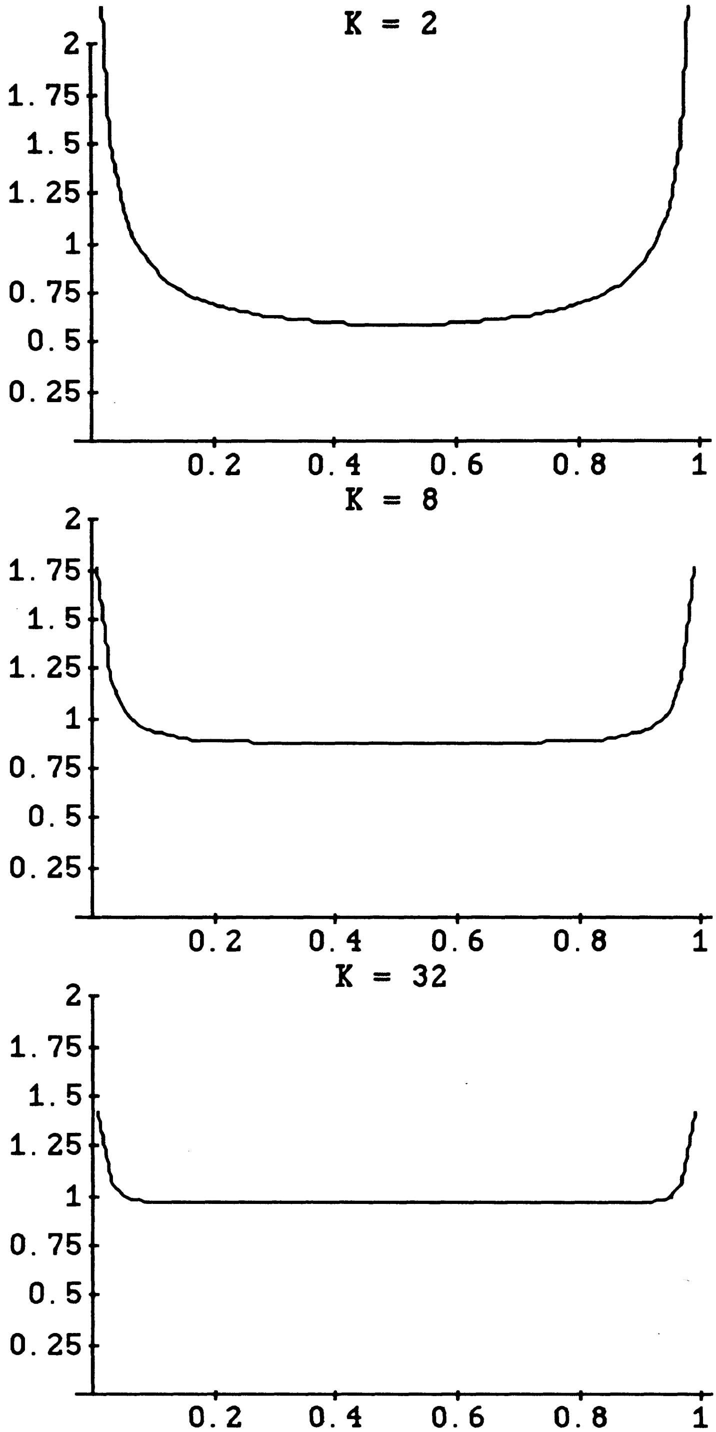


FIGURE 1: Prior density of the marginal probability p_{a1} for three values of the Dirichlet mixture hyperparameter K.

TABLE 1: Quantiles of $|\lambda|$ for various values of the hyperparameter K for a 2×2 table.

	$\vert\lambda\vert$			
K	Prob. = 0.6	0.7	0.8	0.9
15	1.75	2.31	3.17	5.41
25	1.22	1.57	2.08	3.25
50	0.81	1.03	1.33	1.91
100	0.55	0.70	0.90	1.25
200	0.39	0.49	0.62	0.86
500	0.24	0.30	0.39	0.53
1000	0.17	0.22	0.27	0.37

the user believes that λ lies between -1.25 and 1.25 with probability 0.9. This statement is matched with $K = 100$ using Table 1.

For a general $I \times J$ table, the association in the table is described by $(I-1)(J-1)$ parameters (Altham 1970). For large I and J, the hyperparameter K will be difficult to assess. In these situations, it may be more useful to assess quantiles of the predictive distribution of a standard test statistic such as X^2. It will be shown in Section 6 that $\{(K+1)/(K+n)\}X^2$ has approximately a chi-squared distribution with $(I-1)(J-1)$ degrees of freedom. To use this result, suppose that the user can assess the $\gamma \times 100\%$ quantile c for X^2 for an $I \times J$ table with given sample size n. Since $P[X^2 < c] = P[X^2(I-1)(j-1) < (K+1)c/(K+n)] = \gamma$, a value of K can be found by solving the equation $(K+1)c/(K+n) = y_\gamma$, where y_γ is the chi-square $(I-1)(J-1)$ quantile. Note that this assessed value for K depends on n and the size of the table. Thus it may be wise to assess K for a number of tables of varying n and size before deciding on a final value.

3. COMPUTATION OF THE BAYES FACTOR

3.1. Basic Formula.

The Bayes factor using the prior (2.1) is given by

$$F_K = \frac{m(\mathbf{x} \mid K)}{m(\mathbf{x} \mid \infty)}. \tag{3.1}$$

The numerator of (3.1) is given by

$$m(\mathbf{x} \mid K) = \int m(\mathbf{x} \mid K, \mathbf{q}_a, \mathbf{q}_b)\, \pi_d(\mathbf{q}_a;\, I,\, w\mathbf{1})\, \pi_d(\mathbf{q}_b;\, J,\, w\mathbf{1})\, d\mathbf{q}_a\, d\mathbf{q}_b, \tag{3.2}$$

where $m(\mathbf{x} \mid K, \mathbf{q}_a, \mathbf{q}_b)$ is the Dirichlet-multinomial density

$$m(\mathbf{x} \mid K, \mathbf{q}_a, \mathbf{q}_b) = \binom{n}{\mathbf{x}} \frac{D_{IJ}\left((Kq_{ai}q_{bj} + x_{ij})\right)}{D_{IJ}\left((Kq_{ai}q_{bj})\right)}. \tag{3.3}$$

If one takes the limit of (3.2) as $K \longrightarrow \infty$, one obtains the denominator of (3.1):

$$\begin{aligned} m(\mathbf{x} \mid \infty) &= \int \binom{n}{\mathbf{x}} \prod_{ij} (q_{ai}q_{bj})^{x_{ij}} \pi_d(\mathbf{q}_a;\, I,\, w\mathbf{1}) \pi_d(\mathbf{q}_b;\, J,\, w\mathbf{1})\, d\mathbf{q}_a\, d\mathbf{q}_b \\ &= \binom{n}{\mathbf{x}} \frac{D_I(\mathbf{x}_a + w\mathbf{1}) D_J(\mathbf{x}_b + w\mathbf{1})}{D_I(w\mathbf{1}) D_J(w\mathbf{1})}. \end{aligned} \tag{3.4}$$

3.2. Approximations.

Since the computation of $m(\mathbf{x} \mid K)$ requires the evaluation of a $(I+J-2)$-dimensional integral, it is necessary to develop some approximate methods. First note that, if we define the random variable $\mathbf{Y} = \tau\mathbf{X} = (\tau X_{11}, \dots, \tau X_{IJ})$, where $\tau = (K+1)/(n+K)$, then $\mathcal{E}(\mathbf{Y} \mid K, \mathbf{q}) = \tau n\mathbf{q}$ and $Cov(\mathbf{Y} \mid K, \mathbf{q}) = \tau n\{\text{diag}(\mathbf{q}) - \mathbf{q}^T\mathbf{q}\}$, where diag($\mathbf{q}$) is the diagonal matrix with elements (q_{ij}). Since $\mathbf{Y}$ has the same first and second moments as a multinomial random variable with parameters $n\tau$ and $\mathbf{q}$, we can approximate the Dirichlet-multinomial density of $\mathbf{X}$ by the quasimultinomial density

$$m_{\mathbf{q}}(x \mid K, \mathbf{q}_a, \mathbf{q}_b) = \tau^{IJ-1} \frac{\Gamma(n\tau+1)}{\prod_{ij} \Gamma(x_{ij}\tau+1)} \prod_{ij} (q_{ai}q_{bj})^{\tau x_{ij}}. \tag{3.5}$$

[This approximation was proposed by Leonard (1977).] Then, by substitution into (3.2), we obtain the approximation

$$m_{\mathbf{q}}(\mathbf{x} \mid K) = \tau^{IJ-1} \frac{\Gamma(n\tau+1)D_I(\tau\mathbf{x}_a + w\mathbf{1})D_J(\tau\mathbf{x}_b + w\mathbf{1})}{\prod_{ij} \Gamma(x_{ij}\tau+1)D_I(w\mathbf{1})D_J(w\mathbf{1})}. \tag{3.6}$$

The approximation (3.6) is simple to compute and approaches the exact density (3.4) as K approaches infinity. However, from our experience, $m_{\mathbf{q}}(\mathbf{x} \mid K)$ is significantly larger than $m(\mathbf{x} \mid K)$ for small values of K. Thus in practice a second approximation will be used which appears to be more accurate for small K. As in the above approach, replace the Dirichlet-multinomial density by the quasimultinomial likelihood in (3.2), obtaining

$$m(\mathbf{x} \mid K) \simeq \frac{\Gamma(n\tau+1)\tau^{IJ-1}}{\prod_{ij} \Gamma(x_{ij}\tau+1)D_1(w\mathbf{1})D_J(w\mathbf{1})} \times \int \prod_i q_{ai}^{\tau x_{i\cdot}+w-1}\, d\mathbf{q}_a \int \prod_j q_{bj}^{\tau x_{\cdot j}+w-1}\, d\mathbf{q}_b. \tag{3.7}$$

Instead of evaluating the integrals in (3.7) directly, we use the following approximation to the Dirichlet function obtained by the substitution of Stirling's formula for the individual gamma functions:

$$D_I(a) = \int \prod_i y_i^{a_i-1} dy \simeq \prod_i \hat{y}_i^{a_i-\frac{1}{2}} (2\pi)^{(I-1)/2} a_{\cdot}^{-(I-1)/2},$$

where $\hat{y}_i = a_i/a_{\cdot}$. Applying this approximation twice in (3.7) and simplifying, we obtain

$$\begin{aligned} m(\mathbf{x} \mid K) &\simeq \frac{\Gamma(n\tau+1)\tau^{IJ-1}}{\prod_{ij} \Gamma(x_{ij}\tau+1)D_1(w\mathbf{1})D_J(w\mathbf{1})} (2\pi)^{(I+J)/2-1}(n\tau+Iw)^{-(I-1)/2} \\ &\quad \times (n\tau+Jw)^{-(J-1)/2} \prod_i \hat{q}_{ai}^{x_{i\cdot}\tau+w-\frac{1}{2}} \prod_j \hat{q}_{bj}^{x_{\cdot j}\tau+w-\frac{1}{2}} \\ &= m_{\mathbf{q}}(\mathbf{x} \mid K, \hat{\mathbf{q}}_a, \hat{\mathbf{q}}_b)(2\pi)^{(I+J)/2-1}(n\tau+Iw)^{-(I-1)/2} \\ &\quad \times (n\tau+Jw)^{-(J-1)/2} \prod_i \hat{q}_{ai}^{w-\frac{1}{2}} \prod_j \hat{q}_{bj}^{w-\frac{1}{2}} D_I^{-1}(w\mathbf{1})D_J^{-1}(w\mathbf{1}), \end{aligned} \tag{3.8}$$

TABLE 2: Computation of weight of evidence $\log_{10} F_K$ using exact and approximate methods for four tables.

Table; $-\log_{10} p$-value	Method	$\log_{10} F_K$: $\log_{10} K = 0.0$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
[8, 5; 5, 32]	Exact	0.55	1.23	1.40	1.17	0.70	0.30	0.11	0.03	0.01
3.16	Approx.	0.65	1.28	1.43	1.15	0.65	0.27	0.09	0.03	0.00
[7, 2; 18, 23]	Exact	−0.97	−0.23	0.14	0.23	0.15	0.07	0.02	0.01	0.00
1.18	Approx.	−0.84	−0.15	0.17	0.23	0.14	0.06	0.02	0.00	0.00
[10, 7; 15, 18]	Exact	−1.81	−0.92	−0.40	−0.15	−0.04	−0.01	0.00	0.00	0.00
0.43	Approx.	−1.65	−0.86	−0.39	−0.15	−0.05	−0.02	−0.01	0.00	0.00
[3, 16; 20, 11]	Exact	0.70	1.47	1.71	1.38	0.77	0.31	0.11	0.04	0.01
3.10	Approx.	0.86	1.53	1.69	1.33	0.73	0.30	0.10	0.03	0.01

where $\hat{q}_{ai} = (\tau x_{i\cdot} + w)/(n\tau + Iw)$, $\hat{q}_{bj} = (\tau x_{\cdot j} + w)/(n\tau + Jw)$, and $\hat{\mathbf{q}}_a = (\hat{q}_{a1}, \ldots, \hat{q}_{aI})$, $\hat{\mathbf{q}}_b = (\hat{q}_{b1}, \ldots, \hat{q}_{bJ})$.

Finally we replace the quasilikelihood density by the more accurate Dirichlet-multinomial density in (3.8), giving the recommended approximation

$$m_b(\mathbf{x} \mid K) = m(\mathbf{x} \mid K, \hat{\mathbf{q}}_a, \hat{\mathbf{q}}_b)(2\pi)^{(I+J)/2-1}(n\tau + Iw)^{-(I-1)/2} \times (n\tau + Jw)^{-(J-1)/2} \prod_i \hat{q}_{ai}^{\,w-\frac{1}{2}} \prod_j \hat{q}_{bj}^{\,w-\frac{1}{2}} D_I^{-1}(w\mathbf{1}) D_J^{-1}(w\mathbf{1}). \tag{3.9}$$

Brooks (1984) used a similar idea for constructing an approximate likelihood-ratio test for beta-binomial data.

In Table 2, we evaluate the goodness of the approximation for four 2×2 tables where $n = 50$. In this table, values of $\log_{10} F_K$ [called the weight of evidence by Good (1967)] are computed for nine values of $\log_{10} K$ using numerical integration (the row labelled "Exact") and the approximation. Recall that we have assigned the hyperparameter $w = 1$. Note that the largest absolute errors of approximation occur for the smallest values of K. The accuracy of the approximation has also been checked for a number of tables where I and J both exceed two. In all of the examples that we have studied, the approximation appears to be accurate.

3.3. *Behaviour of the Bayes Factor.*

To gain an understanding of the behaviour of the new Bayes factor, we test the hypothesis of independence on two well-known tables. Table 3 gives a two-way table from Yule (1900) on the height classification of husband and wife for 205 married couples. Table 4 gives a table from White and Eisenberg (1959) on the blood and site classification of 707 patients with cancer of the stomach. Figure 2 plots $\log_{10} F_K$ as a function of $\log_{10} K$ for both tables. Yule's data is close to an independence structure ($X^2 = 2.91$ on 4 degrees of freedom), and the Bayes factor is maximized at the independence hypothesis ($K = \infty$). In contrast, the second data set is relatively far from independence ($X^2 = 12.65$ on 6 d.f.), and the Bayes factor is maximized at the interior point $\log_{10} K = 2.8$.

This behaviour is qualitatively similar to the behaviour of the Bayes factor of Good (1967) in the multinomial-testing situation. From our experience, F_K is maximized at

TABLE 3: Yule's data on the classification of 205 married couples.

		Wife height		
		Tall	Med.	Short
Husband height	Tall	18	28	14
	Med.	20	51	28
	Short	12	25	9

TABLE 4: White and Eisenberg's data on the classification of cancer patients.

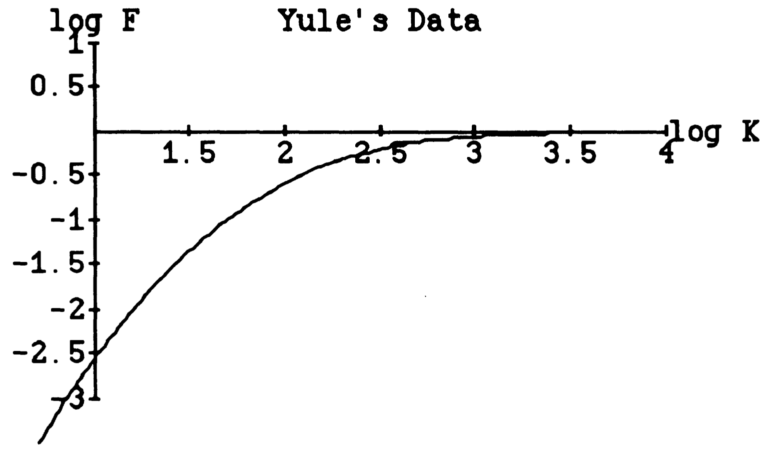
		Blood group		
		O	A	B or AB
Site	Pylorus and antrum	104	140	52
	Body and fondus	116	117	52
	Cardia	28	39	11
	Extensive	28	12	8

$K = \infty$ for tables where, approximately, $X^2 < (I-1)(J-1)$; otherwise it possesses one local maximum at $K = K_0 < \infty$.

Also, it is of interest to see the behaviour of the Bayes factor for different sample sizes. Consider the three 2×2 tables (32, 18; 18, 32), (272, 228, 228, 272), (2570, 2430, 2430, 2570) of respective sample sizes 100, 1000, and 10,000. All three tables have equal row and column margins and have a p-value approximately equal to 0.005. Figure 3 plots $\log_{10} F_K$ (as a function of $\log_{10} K$) for the three tables. Note that, for small values of $\log_{10} K$ (say smaller than one), the Bayes factor is a decreasing function of the sample size n. This behaviour is consistent with the well-known conflict between Bayes factors and p-values for large samples (see Berger 1985, Chapter 4). Also, note that the value of $\log_{10} K$ that maximizes the Bayes factor is an increasing value of n. This indicates that the table with $n = 10{,}000$ is "closest" to the independence hypothesis. Finally, note that the maximum value of $\log_{10} F_K$ is approximately equal to one for all three tables. In Section 4, we will further investigate the relationship between this maximum Bayes factor and the p-value.

3.4. Sensitivity of the Bayes Factor with Respect to Hyperparameters.

Here we briefly investigate the sensitivity of the Bayes factor F_K with respect to the second-stage hyperparameter w. In Table 5, values of $\log_{10} F_K$ and $\log_{10} F_{\max}$ are computed for three 2×2 tables and for the three "noninformative choices" $w = 0, \frac{1}{2}$, and 1. Note from this table that the values of $\log_{10} F_K$ are insensitive to the value of w for $\log_{10} K \geq 1.5$. In addition, since the value where F_K is maximized lies in this region for these examples, $F_{\max}$ is also insensitive to the choice of w.



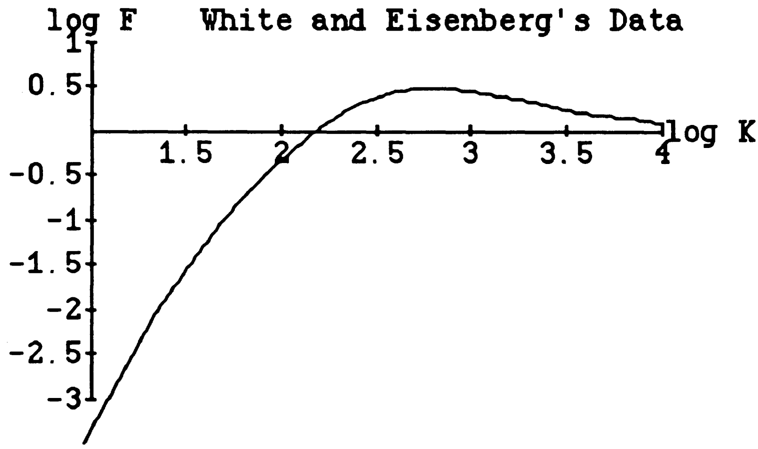


FIGURE 2: Graphs of the Bayes factor $\log_{10} F_K$ for the two data sets.

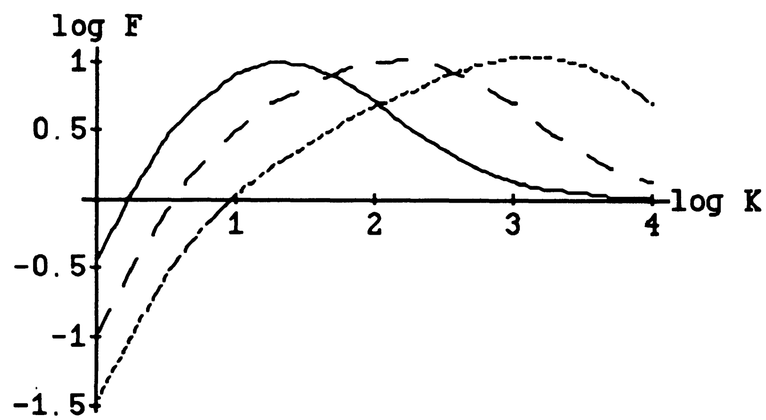


FIGURE 3: Plot of Bayes factor for three tables with equal margins and p-value = 0.005. Solid line is n = 100, dashed line is n = 1000, and dotted line is n = 10,000.

TABLE 5: Weight of evidence $\log_{10} F_K$ for three tables and three values of the hyperparameter w.

Table; $\log_{10}$ p-value	w	$-\log_{10} K = 0.0$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	max
		$\log_{10} F_K$									
[52, 38; 43, 67]	0	−0.37	0.32	0.70	0.85	0.76	0.48	0.21	0.07	0.02	0.85
2.1	0.5	−0.54	0.22	0.66	0.84	0.76	0.47	0.21	0.07	0.02	0.84
	1.0	−0.66	0.15	0.63	0.83	0.76	0.47	0.21	0.07	0.02	0.83
[23, 27; 27, 123]	0	1.07	1.81	2.20	2.25	1.87	1.14	0.50	0.18	0.06	2.25
4.1	0.5	1.13	1.83	2.20	2.24	1.86	1.13	0.50	0.18	0.06	2.24
	1.0	1.16	1.87	2.21	2.24	1.85	1.13	0.50	0.18	0.06	2.24
[42, 38; 48, 72]	0	−1.24	−0.54	−0.13	0.11	0.19	0.15	0.07	0.02	0.01	0.19
1.1	0.5	−1.40	−0.62	−0.16	0.10	0.19	0.14	0.07	0.02	0.01	0.19
	1.0	−1.51	−0.69	−0.19	0.09	0.18	0.14	0.07	0.02	0.01	0.18

4. ALTERNATIVE BAYESIAN STATISTICS FOR CONTINGENCY TABLES

4.1. The Approach of Gunel and Dickey.

Gunel and Dickey (1974) (henceforth referred to as GD) give a thorough discussion of Bayes factors for this problem, using the class of conjugate priors. Suppose that on $\bar{\mathcal{H}}$, the $(IJ-1)$-dimensional parameter $\mathbf{p}$ is assigned the Dirichlet density with hyperparameter vector $\alpha = (\alpha_{11}, \ldots, \alpha_{IJ})$. Then, by applying Savage's continuity condition, GD show that this conjugate prior induces on $\mathcal{H}$ the following prior on $\mathbf{p}_a$ and $\mathbf{p}_b$:

$$\mathbf{p}_a,\ \mathbf{p}_b \text{ independent,} \quad \mathbf{p}_a \sim D_I\big(\alpha_a - (J-1)\big), \quad \mathbf{p}_b \sim D_J\big(\alpha_b - (I-1)\big), \tag{4.1}$$

where $\alpha_a = (\alpha_{1\cdot}, \ldots, \alpha_{I\cdot})$ and $\alpha_b = (\alpha_{\cdot 1}, \ldots, \alpha_{\cdot J})$. One noninformative choice for α is $\mathbf{1}$, the vector of ones. This gives a uniform prior on $\bar{\mathcal{H}}$ and also induces independent uniform priors for $\mathbf{p}_a$ and $\mathbf{p}_b$ under $\mathcal{H}$.

4.2. The Approach of Spiegelhalter and Smith.

Spiegelhalter and Smith (1982) (henceforth referred to as SS) derive a Bayes factor for comparing two log-linear models. They develop a test for comparing the full multinomial model (with $IJ - 1$ parameters) and an unsaturated model defined by setting certain contrasts of the log probabilities equal to zero. Their Bayes factor, given in Equations (32) and (33) of their paper, is based on an assumption of approximate multivariate normality for the log counts, and the "device of imaginary observations" is used to completely specify the vague priors on the parameters in the saturated and unsaturated models.

4.3. The Approach of Good and Crook.

An alternative Bayesian approach to testing in contingency tables is described extensively in Good (1976), Crook and Good (1980), and Good and Crook (1987) (collectively referred to as GC in this paper). We will concentrate on the methodology described in the 1987 paper.

Suppose under $\bar{\mathcal{H}}$ that the user is only able to specify a vector of prior means $\mathbf{q} = (q_{11}, \ldots, q_{IJ})$ for $\mathbf{p}$. Then GC advocate the use of the Dirichlet mixture prior

$$\pi_{\mathrm{GC}}(\mathbf{p} \mid \mathbf{q}) = \int_0^\infty \pi_d(\mathbf{p};\ IJ,\ K\mathbf{q})\pi_{\mathrm{lC}}(K)\, dK, \tag{4.2}$$

where $\pi_{\mathrm{lC}}(K)$ is a second-stage prior (of the log-Cauchy form) on the Dirichlet precision hyperparameter K. Under the independence hypothesis, $\mathbf{p}_a$ and $\mathbf{p}_b$ are assumed independent and each assigned the mixture prior (4.2) with prior mean vectors $\mathbf{q}_a$ and $\mathbf{q}_b$, respectively.

The prior mean vectors are not assumed known, but are data-dependent in GC's approach. First consider the selection of prior means under $\mathcal{H}$. If $\mathbf{q}_a = (q_{1\cdot}, \ldots, q_{I\cdot})$, the $q_{i\cdot}$ is estimated by $\hat{q}_{i\cdot} = (x_{i\cdot} + k_0)/(n + Ik_0)$ (k_0 given in GC), which may be viewed as an estimate shrinking the observed proportion $x_{i\cdot}/n$ towards the value (of equiprobability) $1/I$. The prior mean vector $\mathbf{q}_b$ is estimated in a similar fashion. Under the unrestricted hypothesis $\bar{\mathcal{H}}$, the prior means are estimated by the independence formula $\hat{q}_{ij} = \hat{q}_{i\cdot}\hat{q}_{\cdot j}$.

5. COMPARISON OF THE FOUR BAYES FACTORS

5.1. General Comments.

The class of independence priors is similar to the conjugate class used by GD. The main feature of our class is the choice of the Dirichlet prior means to reflect a belief in independence and the averaging out of the unknown vectors of marginal prior means $\mathbf{q}_a$ and $\mathbf{q}_b$. Recall from Section 2.2 that this prior distribution on $\bar{\mathcal{H}}$ induces a conjugate prior on the independence hypothesis $\mathcal{H}$. This class of priors has a weak connection with the prior of GC. One similarity of the two approaches is that GC make the hypothesis $\bar{\mathcal{H}}$ close to the hypothesis $\mathcal{H}$ in their prior specification by setting q_{ij} (the prior mean of p_{ij} under $\bar{\mathcal{H}}$) equal to $q_{i\cdot}q_{\cdot j}$.

5.2. Numerical Comparison.

In this subsection, we numerically compare the behaviour of the Bayes factor (3.1) with test statistics proposed by GD, GC, and SS for a number of two-way tables. Since this statistic (3.1) depends on the assessment of the hyperparameter K, we will investigate the behaviour of the maximum Bayes factor $F_{\max}$. Note that, if we assume equal prior probabilities for the hypotheses $\mathcal{H}$ and $\bar{\mathcal{H}}$, then the posterior probability of $\mathcal{H}$ is $(1 + F)^{-1}$. Thus a lower bound (over K) on the posterior probability of independence is given by $P_{\min} = (1 + F_{\max})^{-1}$.

Aspects of the data which may influence the values of the test statistics include (1) the closeness of the counts to the independence configuration $x_{ij} = x_{i\cdot}x_{\cdot j}/n$ for all i, j, (2) the values of the marginal totals, (3) the dimensions of the table, and (4) the sample size n. We will investigate the effect of all four variables on the Bayes-factor statistics.

To measure the deviation of the table from independence we use the p-value $p_{X^2} = P[X^2(I-1)(j-1) > y]$. If this p-value is viewed as an approximate posterior probability that the hypothesis $\mathcal{H}$ is true, and equal prior probabilities on $\mathcal{H}$ and $\bar{\mathcal{H}}$ are assumed, then the corresponding Bayes factor is $F_C = p_{X^2}^{-1} - 1$. We will use, as in Good (1967), a $\log_{10}$ scale to compare Bayes factors. Approximately, $\log_{10} F_C = -\log_{10} p_{x^2}$. Using this scale, a p-value of 0.05 corresponds to $\log_{10} F_C = 1.3$.

The value of the test statistic may be influenced by the margins of the table. A set of marginal totals will be called flat if the individual totals are approximately equal (otherwise it will be called rough). Good and Crook (1987) make the general observation that the marginal totals can influence the test for independence, but this influence is only significant for tables with rough margins.

In the following examples, we compare the statistic $F_{\max}$ with the statistics of GD, GC, and SS. For the conjugate-prior approach of GD, we will use the uniform choice $\mathbf{a} = \mathbf{1}$. GC's approach requires the specification of the lower and upper quartiles of the log-Cauchy density for the hyperparameter K. We let the quartiles be 1 and 20; this choice appears to give representative values for the Bayes factor for the examples in Good and Crook (1987).

First we let $n = 200$ and observe the behaviour for thirty 2×2 tables, where we vary the roughness of the margins and the agreement of the table with independence. In this example, we will say that a particular set of margins is smooth if the smaller marginal proportion lies between 0.4 and 0.5, rough if the smaller proportion lies between 0.2 and 0.4, and very rough if the smaller proportion lies between 0 and 0.2. Table 6 gives values of $\log_{10} F_{\max}, \log_{10} F_{\mathbf{GD}}, \log_{10} F_{\mathrm{GC}}$, and $\log_{10} F_{\mathrm{SS}}$ for tables classified by p-value and flatness of the margins.

First note that, for a given value of $\log_{10} F_C$ the value of any one of the Bayes factors is generally not influenced by the flatness of the margins. The sole exception is when both margins are very rough and the evidence against independence is significantly decreased. It is simplest to compare the four Bayes factors in the summary row which takes the mean of the five values (very-rough–very-rough excluded) in a given column. Generally, the GD values give relatively strong support for the independence hypothesis for tables close to independence but relatively weak support to $\mathcal{H}$ for tables far from independence. The SS values are generally 0.1–0.2 smaller than the GD values. This behaviour is likely due to the different noninformative priors used by the two methods. The GC values, in contrast, give weak support to $\mathcal{H}$ for tables where $\log_{10} F_C \leq 1$ and stronger support for tables where $\log_{10} F_C \geq 3$. Note that the $F_{\max}$ values are bounded between the GD and GC values. When $\log_{10} F_C = 2$, the GC and $F_{\max}$ values are approximately equal for five of the six tables.

To see if the above observations carry over to other sample sizes, Table 7 summarizes calculations made on more 2×2 tables where $n = 50, 200$, and 800. For each value of $\log_{10} F_C$ and sample size, four different tables were generated and the values of the weights of evidence averaged (tables with very rough margins which gave atypical values were excluded). In this table, we note that the GD and SS values are relatively sensitive to the sample size. For a given value of $\log_{10} F_C$, the test statistic gives stronger support to $\mathcal{H}$ for larger sample sizes. In contrast, the GC and $F_{\max}$ values are relatively insensitive to the sample size. For example, if $\log_{10} F_C = 2$, then both $\log_{10} F_{\mathrm{GC}}$ and $\log_{10} F_{\max}$ are approximately 0.7 for all three sample sizes.

To investigate the effect of the size of the table, we compute values of the four test statistics for a set of 41 tables where at least one of the two dimensions exceeds two. These tables vary with respect to the sample size n and the roughness of the margins. It is of interest to see if any clear relationship is present between the p-value and the value of any of the Bayesian statistics.

For the procedures $F_{\mathbf{GD}}$ and F_{SS} no such relationship appears to exist. These values, for large $I \times J$ tables, appear to be greatly influenced by the margins and the sample size. Clearer relationships appear to exist for F_{GC} and $F_{\max}$. The values of the two Bayesian statistics are plotted as functions of $\log_{10} F_C$ in Figure 4. The line $y = x$ is also plotted in this figure so comparisons can be made between the Bayes factors and the p-value. Note that the comments made earlier about the two statistics for 2×2 tables appear to

TABLE 6: Weights of evidence computed using four methods for thirty 2 × 2 tables, $n = 200$.

Margins of table	Method	Weight of evidence $\log_{10} F_C = 0.5$	1.0	2.0	3.0	4.0
Smooth-smooth	GD	−0.37	0.06	0.91	1.66	2.66
	GC	0.29	0.41	0.81	1.33	2.13
	SS	−0.63	−0.18	0.65	1.41	2.32
	max K	0.00	0.18	0.83	1.49	2.36
Rough-rough	GD	−0.36	0.01	0.75	1.57	2.46
	GC	0.30	0.39	0.74	1.29	2.01
	SS	−0.53	−0.18	0.62	1.41	2.43
	max K	0.00	0.16	0.73	1.44	2.24
Smooth-rough	GD	−0.44	−0.07	0.65	1.71	2.54
	GC	0.30	0.37	0.69	1.43	2.05
	SS	−0.62	−0.35	0.51	1.52	2.23
	max K	0.00	0.10	0.66	1.60	2.27
Very rough–very rough	GD	−0.69	−0.47	0.44	1.08	1.73
	GC	0.28	0.34	0.62	1.03	1.47
	SS	−0.45	−0.01	0.73	1.49	2.22
	max K	0.00	0.03	0.57	1.18	1.71
Rough–very rough	GD	−0.66	−0.24	0.50	1.52	2.45
	GC	0.30	0.36	0.67	1.29	2.06
	SS	−0.36	−0.13	0.70	1.44	2.32
	max K	0.00	0.08	0.66	1.45	2.31
Smooth–very rough	GD	−0.61	−0.18	0.46	1.64	2.53
	GC	0.30	0.38	0.72	1.55	2.11
	SS	−0.43	−0.14	0.68	1.38	2.11
	max K	0.00	0.12	0.75	1.80	2.34
Summary (very rough–very rough excluded)	GD	−0.59	−0.08	0.65	1.60	2.50
	GC	0.30	0.38	0.73	1.40	2.10
	SS	−0.51	−0.20	0.63	1.43	2.28
	max K	0.00	0.13	0.73	1.50	2.30

hold for larger tables. The F_{GC} and $F_{\max}$ statistics are roughly equal for tables where $\log_{10} F_C = 2$. For tables closer to independence, F_{GC} gives significantly less support than $F_{\max}$ to independence.

In Figure 4, note that a weak relationship appears to exist between $\log_{10} F_C$ and $\log_{10} F_{\mathrm{GC}}$. In fact, for these examples, the values for $\log_{10} F_{\mathrm{GC}}$ are nearly constant for tables where $0 \le \log_{10} F_C \le 2$. In contrast, a strong linear relationship appears to exist between $\log_{10} F_C$ and $\log_{10} F_{\max}$. For $\log_{10} F_C < 0.7$, $\log_{10} F_{\max} = 0$; for $0.7 < \log_{10} F_C < 4$, the line $\log_{10} F_{\max} = 0.74\,(\log_{10} F_C - 0.7)$ appears to give a good fit.

SUMMARY

The Bayes factor is an attractive alternative to classical test statistics such as X^2 and G^2.

TABLE 7: Summary values for weights of evidence for 2 × 2 tables for sample sizes n = 50, 200, and 800.

Sample size n	Method	Weight of evidence $\log_{10} F_C = 0.5$	1.0	2.0	3.0	4.0
50	GD	−0.13	0.25	1.05	2.0	2.7
	GC	0.15	0.29	0.70	1.4	2.0
	SS	−0.27	0.11	0.86	1.7	2.2
	max K	0.00	0.14	0.69	1.6	2.2
200	GD	−0.59	−0.08	0.65	1.6	2.5
	GC	0.30	0.38	0.73	1.4	2.1
	SS	−0.51	0.20	0.63	1.4	2.3
	max K	0.00	0.13	0.73	1.5	2.3
800	GD	−0.81	−0.41	0.47	1.4	2.3
	GC	0.51	0.54	0.73	1.3	2.1
	SS	−0.77	−0.43	0.37	1.3	2.2
	max K	0.00	0.15	0.77	1.6	2.4

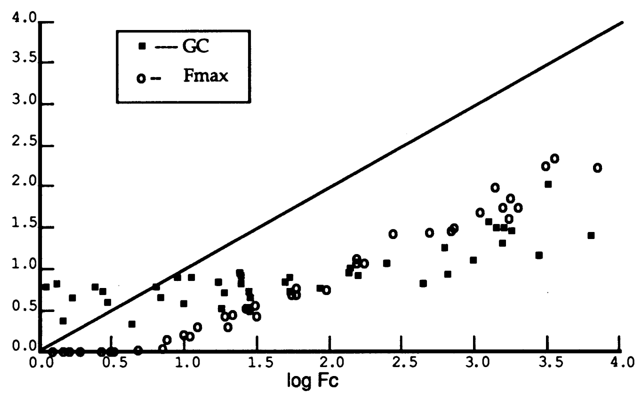


FIGURE 4: Values of $\log_{10} F_{\max}$ and $\log_{10} F_{GC}$ plotted as a function of $\log_{10} F_C$ for 41 large two-way tables.

However, from Section 5, we see that the value of a Bayes factor can depend significantly on the prior distributions placed on the hypotheses $\mathcal{H}$ and $\bar{\mathcal{H}}$. Therefore, it is important for the user to understand the implications of various prior distributions chosen.

One advantage of the independence prior class over the GC, GD, and SS priors is the simplicity of interpretation and assessment. The distribution for $\bar{\mathcal{H}}$ (2.1) is appropriate for use in the situation where the cell probabilities are believed to be located in a neighbourhood of the independence hypothesis. The strength of this belief in independence is reflected by the single hyperparameter K. If K is difficult to assess, then we recommend computing F_K for a range of plausible values of F and computing $F_{\max}$.

A second comment is that the independence class of priors is related to the random-effects model discussed in Diaconis and Efron (1985). Recall that, if $\mathbf{X}$ has the Dirichlet-multinomial density (3.3), then $\mathbf{Y} = \tau\mathbf{X}$ has the same moment structure (up to second order) as a multinomial random variable. If the mean vector $\mathbf{q}$ has the independence

TABLE 8: Values of the normal upper bound and fitted values for $\log_{10} F_{\max}$ from Figure 1.

$\log_{10} F_C$	$\log_{10} B_N$	$\max(O, 0.74(\log_{10} F_C - 0.7))$
0.5	0.00	0.00
1.0	0.15	0.22
2.0	0.18	0.96
3.0	1.62	1.70
4.0	2.48	2.44

configuration, then the statistics X^2 and G^2 evaluated at the **Y** data are both asymptotically chi-squared with $(I-1)(J-1)$ d.f. In particular, since $X^2(\mathbf{Y}) = \tau X^2$, this Bayesian model is equivalent asymptotically to the random-effects model discussed by Diaconis and Efron. Those authors used this model as an intermediate model between the hypotheses of independence and uniformity [see Dickey (1985) for a Bayesian analysis of this model].

Also observe that the relationship between the statistic $F_{\max}$ and the p-value is similar to the relationship found in testing for a normal mean in Edwards *et al.* (1963). These authors considered the problem of testing the sharp null hypothesis $\mathcal{H}$ that $\theta = \theta_0$, where X is distributed normally with mean θ and known variance σ^2. If one places a normal prior with mean θ_0 and variance τ^2 on the alternative space where $\theta \neq \theta_0$, then Edwards *et al.* showed that an upper bound for the Bayes factor $P(\bar{H} \mid x)/P(\mathcal{H} \mid x)$ over all $\tau > 0$ is $B_N = exp\ \{(t^2-1)/2\}/t$, where $t = (x-\theta_0)/\sigma$. Values of $\log_{10}\ B_N$ are given in Table 8 for different values of $\log_{10}\ F_C = -\log_{10}\ p$-value.

Note that the values for $\log_{10}\ B_N$ are very close to the summary values for $\log_{10}\ F_{\max}$ for the 2×2 tables given in Table 7. Also, Table 8 suggests that the normal upper bounds may also provide a rough approximation to $\log_{10}\ F_{\max}$ for larger tables.

The methodology described in this paper can be generalized to test for an arbitrary log-linear model in a multidimensional contingency table. Using the generalized linear model framework, of which log-linear modelling is a special case, Albert (1988) derives a Bayes factor for a regression null hypothesis.

ACKNOWLEDGEMENT

The author would like to express his appreciation to the associate editor and referees, whose constructive suggestions significantly improved the manuscript.

REFERENCES

Albert, J.H. (1988). Computational methods using a Bayesian hierarchical generalized linear model. *J. Amer. Statist. Assoc.*, 83, 1037–1045.

Albert, J.H., and Gupta, A.K. (1982). Mixtures of Dirichlet distributions and estimation in contingency tables. *Ann. Statist.*, 10, 1261–1268.

Altham, P.M.E. (1970). The measurement of association of rows and columns for an $r \times s$ contingency table. *J. Roy. Statist. Soc. Ser. B*, 32, 63–73.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Berger, J.O., and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *J. Amer. Statist. Assoc.*, 82, 112–122.

Brooks, R.J. (1984). Approximate likelihood ratio tests in the analysis of beta-binomial data. *Appl. Statist.*, 33, 285–289.

Crook, J.F., and Good, I.J. (1980). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, part II. *Ann. Statist.*, 8, 1198–1219.

Diaconis, P., and Efron, B. (1985). Testing for independence in a two-way table: New interpretations of the chi-square statistic. *Ann. Statist.*, 13, 845–874.

Dickey, J. (1973). Scientific reporting. *Roy. Statist. Soc. Ser. B*, 35, 285–305.

Dickey, J. (1985). Discussion to the paper by Diaconis and Efron. *Ann. Statist.*, 13, 877–881.

Edwards, W., Lindman, H., and Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.*, 70, 193–242.

Good, I.J. (1967). A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B*, 29, 399–431.

Good, I.J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.*, 4, 1159–1189.

Good, I.J., and Crook, J.F. (1987). The robustness and sensitivity of the mixed-Dirichlet Bayesian test for "independence" in contingency tables. *Ann. Statist.*, 15, 670–693.

Gunel, E., and Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, 61, 545–557.

Jeffreys, H. (1961). *Theory of Probability*. Oxford Univ. Press, Oxford.

Leonard, T. (1977). Bayesian simultaneous estimation for several multinomial distributions. *Comm. Statist. A*, 6, 619–630.

Spiegelhalter, D.J., and Smith, A.F.M. (1982). Bayes factors for linear and log-linear models with vague prior information. *J. Roy. Statist. Assoc. Ser. B*, 44, 377–387.

White, C., and Eisenberg (1959). ABO blood groups and cancer of stomach. *Yale. Biol. and Med.*, 32, 58–61.

Yule, G.U. (1900). On the association of attributes in statistics: With illustration from the material of the childhood society, etc. *Philos. Trans. Roy. Soc. London Ser. A*, 194, 257–319.

Received 7 June 1989
Revised 24 January 1990
Accepted 21 February 1990

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403-0221
U.S.A.