A CONDITION TO OBTAIN THE SAME DECISION IN THE HOMOGENEITY TESTING PROBLEM FROM THE FREQUENTIST AND BAYESIAN POINT OF VIEW

Miguel A. Gómez-Villegas and Beatriz González-Pérez
Departamento de Estadística e Investigación Operativa I
Universidad Complutense de Madrid
Facultad de CC. Matemáticas, Madrid, 28040, Spain
ma.gv@mat.ucm.es

Key Words: Contingency tables, p-values, posterior probabilities, reconciliation.

ABSTRACT

We develop a Bayesian procedure for the homogeneity testing problem of r populations using $r \times s$ contingency tables. The posterior probability of the homogeneity null hypothesis is calculated using a mixed prior distribution. The methodology consist of choosing an appropriate value of π_0 for the mass assigned to the null and spreading the remainder, $1 - \pi_0$, over the alternative according to a density function. With this method, a theorem which shows when the same conclusion is reached from both frequentist and bayesian points of view is obtained. A sufficient condition under which the p-value is less than a value α and the posterior probability is also less than 0.5 is provided.

1. INTRODUCTION

The display of the data by means of contingency tables is used for discussing different approaches to both frequentist and Bayesian Inference. For instance, when we want to investigate the behavior of a characteristic Y common to r large populations. In this situation, to get information about Y, independent random samples, $(Y_{i1}, \ldots, Y_{in_i})$, $i = 1, \ldots, r$, $\sum_{i=1}^{r} n_i = N$, are drawn, respectively, from each population. Our objective is to test if the data gives us enough evidence to reject the homogeneity null hypothesis, that is, we want to decide if r populations have a common distribution $F(y)$. To do this, we divide the common sample space into an arbitrary number, s, of exclusionary classes, C_j, $i = j, \ldots, s$. Now,

we denote by n_{ij} the observed frequency in C_j $(j = 1, \ldots, s)$ of the sample i $(i = 1, \ldots, r)$. Then, the data can be displayed in Table 1.

Table 1. Data in the $r \times s$ table.

	Class 1	Class 2	...	Class s	Total
Sample 1	n_{11}	n_{12}	...	n_{1s}	n_1
Sample 2	n_{21}	n_{22}	...	n_{2s}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Sample r	n_{r1}	n_{r2}	...	n_{rs}	n_r
Total	m_1	m_2	...	m_s	N

In this situation a quantitative measure of the strength of the evidence that the data gives in support or in rejection of the hypothesis that the proportion of elements belonging to C_j $(j = 1, \ldots, s)$ is the same in all the populations $(i = 1, \ldots, r)$, that is to say, $\mathbf{p}_{1j} = \ldots = \mathbf{p}_{rj}$, for each $j = 1, \ldots, s$.

There are of course a number of variations on this problem. In this contex, some important Bayesian references are given next.

Howard (1998) gives a Bayesian discussion of the homogeneity problem for 2×2 tables. He advocates for the more frequent use of unilateral tests, considering as hypotheses of interest $p_2 < p_1$ and $p_1 < p_2$, where p_1 and p_2 are the proportion of successes in the first and second population, respectively. He gives a quantitative measure of the strength of the evidence in support of the more likely hypothesis, assuming that p_1 and p_2 will not be exactly equal, and that neither will be 0 or 1. Given independent samples from two binomial distributions, he notes that the posterior probability that $p_2 < p_1$ can be estimated from the standard (uncorrected) χ^2 significance level. In order to reach this result, he has to suppose independent Jeffreys priors about the two populations, that is to say,

$$\pi\left(p_1,\ p_2\right) \propto p_1^{-1/2}\left(1-p_1\right)^{-1/2} p_2^{-1/2}\left(1-p_2\right)^{-1/2}.$$

Moreover, he introduces a conjugate family of priors which incorporate dependence between beliefs about the two populations.

In this same line of work, with unilateral hypotheses like $p_1 > p_2$, other Bayesian approaches to the problem of comparing two proportions for a 2×2 table can be mentioned; *log-odds-ratio* methods and *inverse-root-sine* methods, which calculate the posterior probability that $\Lambda_1 - \Lambda_2 > 0$ for beta priors, where $\Lambda_i = \log p_i \left(1 - p_i\right)^{-1}$, and $\Lambda_i = arcsen\sqrt{p_i}$, $i = 1, 2$, respectively, as measures of the degree in which two populations are homogeneous (see Lee, 2004, pages 152-154).

Quintana (1998) postulates a nonparametric Bayesian model for assessing homogeneity in $r \times s$ contingency tables with fixed right margin totals. The vectors of classification probabilities are assumed to be a sample from a distribution F, and the prior distribution of F is assumed to be a Dirichlet process, centered on a probability measure α and with weight c. He also assumes a prior distribution for c and proposes a *Bayes factor*.

Lindley (1988) gives a probability model for the formation of genotypes from two alleles. The alleles are A and a, and the genotypes are AA, Aa and aa (it is a standard notation). The model can be expressed in terms of two parameters, $\alpha = \left[\log(4p_1p_3/p_2^2)\right]/2$ and $\beta = \left[\log(p_1/p_3)\right]/2$. A Bayesian test of the hypothesis that $\alpha = 0$ versus $\alpha \neq 0$, based on a *Bayes factor*, is considered, where $\alpha = 0$ is the null hypothesis of Hardly-Weinberg equilibrium, $H_0 :\ \mathrm{p}^2, 2\mathrm{p}\left(1 - \mathrm{p}\right), \left(1 - \mathrm{p}\right)^2$, p being the proportion of A's.

The equality of cell probabilities null hypothesis in contingency tables may be considered as a special simple hypothesis. In parametric testing of a simple null hypothesis, it is known that frequentist and Bayesian procedures can give rise to different decisions, see Lindley (1957), Berger and Sellke (1987) and Berger and Delampady (1987), among others. On the other hand, Casella and Berger (1987) show that there is no discrepancy in the one-sided testing problem.

It is needed to remark that, in the literature, the comparison between frequentist and Bayesian methods, for a particular testing problem, is usually carried out by searching for a prior distribution which does p-values approximately equal to posterior probabilities. In most of the Bayesian approaches the infimum of the posterior probability of the null hypothesis or the Bayes factor, over a wide class of prior distributions, is considered and it is then obtained

that the infimum is substantially larger than the corresponding p-value. It is necessary to point out that in all these situations the mass assigned to the simple null hypothesis is 1/2.

However, Lindley (1957) introduces this discrepancy for normal distributions with respect to the decision derived from both analysis. He produces an example to show that, if H is a simple hypothesis and x the result of an experiment, the following two phenomena can occur simultaneously: "a significance test for H reveals that x is significant, at, say, the 5% level" and "the posterior probability of H, given x, is for quite small prior probabilities of H, as high as 95%".

Motivated in Lindley's paradox our objective is to show when and how, to test (1), there is no discrepancy between the decision derived from frequentist and Bayesian approaches when a single prior distribution is used.

Recently, Gómez-Villegas and González-Pérez (2005) have developed a Bayesian procedure to test equality of proportions of independent multinomial distributions when the common proportions are known. Their approach to the homogeneity testing problem consists of working directly with the simple null hypothesis and calculating its posterior probability. To do this, they follow the methodology used by Gómez-Villegas, Maín and Sanz (2004) for the multivariate point null testing problem. This methodology is based on choosing an appropriate value of π_0 for the probability of the point null and distributing the remaining probability over the alternative with a prior density. Furthermore, Gómez-Villegas and González-Pérez (2005) calculate posterior probabilities of the null hypothesis with respect to a mixture of a point prior on the null and an independent Dirichlet prior on the proportions. They reconcile frequentist and Bayesian evidence in terms of a sufficient condition under which the same decision is reached with both methods. To do this they introduce an appropriate value of π_0 which verifies that the p-value is less (or higher) than α and the posterior probability is also less (or higher) than 0.5.

Usually, we only want to investigate the equality of cell probabilities, without knowing anything about the common value under the null. In this work, we develop three Bayesian methods to test equality of proportions of independent multinomial distributions when the

common proportions are unknown, generalizing the results obtained by Gómez-Villegas and González (2005). Three Bayesian evidence measures are calculated, using appropriate mixed prior distributions, and conditions under which the p-value is less (or higher) than α and the posterior probability is also less (or higher) than 0.5. This is a new approximation because it permits one to reach the same decision from both points of view.

Table 2. Pearson's example.

	Successes	Failures	Total
Sample 1	3	15	18
Sample 2	7	5	12
Total	10	20	30

Section 2 formulates the problem in a precise way. In section 3, three Bayesian methods to test the homogeneity null hypothesis with $r \times s$, when the common proportions vector under the null is unknown, are developed. Section 4 reconciles frequentist and Bayesian approaches in terms of a sufficient condition and Pearson's (1947) data (see Table 2) is used to illustrate the procedure. Section 5 provides a summary of conclusions.

2. THE PROBLEM

Let X_i, $i = 1, \cdots, r$, be independent multinomial random variables, $MB(n_i, \mathbf{p}_i)$, with $\mathbf{p}_i = (p_{i1}, \cdots, p_{is}) \in \Theta$, where $\Theta = \{ \mathbf{p} = (p_1, \cdots, p_s) \in (0,\ 1)^s,\ \sum_{i=1}^{s} p_j = 1\} \subset \mathrm{R}^{s-1}$.

In this situation, we wish to test

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r,\ versus\ H_1 : \exists i \neq j,\ \mathbf{p}_i \neq \mathbf{p}_j. \tag{1}$$

Therefore, a mixed prior distribution is needed to test (1).

Consider that our prior opinion about $\mathbf{P} = (\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Theta^r \subset \mathrm{R}^{r(s-1)}$ is given by means of the density $\pi(\mathbf{P}) = \prod_{i=1}^{r} \pi(\mathbf{p}_i)$.

Denote by $\mathbf{p}_0 = (p_{01}, \cdots, p_{0s}) \in \Theta$ the unknown value under the null. Therefore, if we denote by $\mathbf{P}_0 = (\mathbf{p}_0, \cdots, \mathbf{p}_0) \in \Theta^r \subset \mathrm{R}^{r(s-1)}$, then $H_0 : \mathbf{P} = \mathbf{P}_0$ is the null hypothesis in (1).

Now, we are going to consider the more realistic precise hypotheses,

$$H_{0\delta} : \mathbf{P} \in C(\delta) \; versus \; H_{1\delta} : \mathbf{P} \notin \mathbf{C}(\delta),$$

with $C(\delta) = \bigcup_{\mathbf{p}_0 \in \Theta} B(\mathbf{P}_0, \delta)$,

$$B(\mathbf{P}_0, \delta) = \left\{ \mathbf{P} \in \Theta^r, \; \sum_{i=1}^{r} \sum_{j=1}^{s-1} (p_{ij} - p_{0j})^2 \leq \delta^2 \right\}$$

and a value of $\delta > 0$ sufficiently small.

We propose to assign a prior mass, π_0, to the null by means of averaging,

$$\pi_0 = \int_{C(\delta)} \pi(\mathbf{P}) d\mathbf{P}. \tag{2}$$

3. THREE BAYESIAN APPROACHES

In this section we develop three Bayesian methods to test (1).

3.1. FIRST METHOD

If the prior opinion about $\mathbf{p}_0$, the unknown value of the common proportions under the null in (1), is given by $\pi(\mathbf{p}_0)$, then, to test (1), we propose the following mixed prior distribution:

$$\pi^*(\mathbf{P}) = \pi_0 \pi(\mathbf{p}_0) I_{H_0}(\mathbf{P}) + (1 - \pi_0) \pi(\mathbf{P}) I_{H_1}(\mathbf{P}),$$

with $\pi_0 = \pi_0(\delta)$ as in (2).

We can note that the prior probability assigned to H_0 by means of $\pi^*(\mathbf{P})$ and to $H_{0\delta}$ by means of $\pi(\mathbf{P})$ are the same thing.

In this situation, the posterior probability of H_0, when the data of Table 1 has been observed, is

$$\frac{\pi_0 \int_\Theta \prod_{j=1}^{s} p_{0j}^{\sum_{i=1}^{r} n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0}{\pi_0 \int_\Theta \prod_{j=1}^{s} p_{0j}^{\sum_{i=1}^{r} n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0 + (1 - \pi_0) \prod_{i=1}^{r} \int_\Theta \prod_{j=1}^{s} p_{ij}^{n_{ij}} \pi(\mathbf{p}_i) d\mathbf{p}_i}.$$

Consider $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{is})$, with $\alpha_{ij} > 0$, $j = 1, \cdots, s$, $i = 1, \cdots, r$ and assign to each $\mathbf{p}_i$ a Dirichlet prior distribution of parameter α_i, $D(\alpha_i)$, $i = 1, \cdots, r$, (see Ghosh and Ramamoorthi, 2003, chapter 3),

$$\pi\left(\mathbf{p}_i\right)=\frac{\Gamma\left(\sum_{j=1}^{s}\alpha_{ij}\right)}{\prod_{j=1}^{s}\Gamma\left(\alpha_{ij}\right)}\prod_{j=1}^{s}p_{ij}^{\alpha_{ij}-1},\quad \mathbf{p}_i=\left(p_{i1},\cdots,p_{is}\right)\in\Theta,\quad i=1,\cdots,r.$$

In this case, the posterior probability of H_0 is

$$\int_{\Theta}\prod_{j=1}^{s}p_{0j}^{\sum_{i=1}^{r}n_{ij}}\pi\left(\mathbf{p}_0\right)d\mathbf{p}_0=\frac{\Gamma\left(\sum_{j=1}^{s}\alpha_{0j}\right)}{\prod_{j=1}^{s}\Gamma\left(\alpha_{0j}\right)}\int_{\Theta}\prod_{j=1}^{s}p_{0j}^{m_j+\alpha_{0j}-1}d\mathbf{p}_0$$

$$=\frac{\Gamma\left(\sum_{j=1}^{s}\alpha_{0j}\right)}{\prod_{j=1}^{s}\Gamma\left(\alpha_{0j}\right)}\frac{\prod_{j=1}^{s}\Gamma\left(m_j+\alpha_{0j}\right)}{\Gamma\left(N+\sum_{j=1}^{s}\alpha_{0j}\right)}.$$

Therefore, such posterior probability can be expressed as

$$\mathrm{B}_1\left(\pi_0\right)=\left[1+\frac{1-\pi_0}{\pi_0}\eta_1\right]^{-1},\tag{3}$$

where

$$\eta_1=\frac{\prod_{j=1}^{s}\Gamma\left(\alpha_{0j}\right)}{\Gamma\left(\sum_{j=1}^{s}\alpha_{0j}\right)}\frac{\Gamma\left(N+\sum_{j=1}^{s}\alpha_{0j}\right)}{\prod_{j=1}^{s}\Gamma\left(m_j+\alpha_{0j}\right)}\frac{\prod_{i=1}^{r}\Gamma\left(\sum_{j=1}^{s}\alpha_{ij}\right)}{\prod_{i=1}^{r}\prod_{j=1}^{s}\Gamma\left(\alpha_{ij}\right)}\frac{\prod_{i=1}^{r}\prod_{j=1}^{s}\Gamma\left(n_{ij}+\alpha_{ij}\right)}{\prod_{i=1}^{r}\Gamma\left(n_i+\sum_{j=1}^{s}\alpha_{ij}\right)}$$

is a statistic which quantifies the strength of the evidence against H_0.

With this procedure we reject H_0 when $\mathrm{B}_1>1/2$.

3.2. SECOND METHOD

Gómez-Villegas and González-Pérez (2005) calculate the posterior probability of H_0 when $\mathbf{P}_0\in\Theta$ is a known value using the mixed prior distribution

$$\pi^*\left(\mathbf{P}|\mathbf{p}_0\right)=\pi_0 I_{H_0}\left(\mathbf{P}\right)+\left(1-\pi_0\right)\pi\left(\mathbf{P}\right)I_{H_1}\left(\mathbf{P}\right).$$

In this situation, if we assign to each $\mathbf{p}_i$ a Dirichlet prior distribution of parameter α_i, $i=1,\cdots,r$, this posterior probability is

$$P\left(H_0|n_{11},\ldots,n_{rs},\mathbf{p}_0\right)=\left[1+\frac{1-\pi_0}{\pi_0}\eta_2\right]^{-1},\tag{4}$$

where

$$\eta_2 = \eta_2(\mathbf{p}_0) = \prod_{j=1}^{s} p_{0j}^{-m_j} \frac{\prod_{i=1}^{r} \Gamma\left(\sum_{j=1}^{s} \alpha_{ij}\right)}{\prod_{i=1}^{r} \prod_{j=1}^{s} \Gamma\left(\alpha_{ij}\right)} \frac{\prod_{i=1}^{r} \prod_{j=1}^{s} \Gamma\left(n_{ij} + \alpha_{ij}\right)}{\prod_{i=1}^{r} \Gamma\left(n_i + \sum_{j=1}^{s} \alpha_{ij}\right)}.$$

If the prior opinion about $\mathbf{p}_0$ is given by $\pi(\mathbf{p}_0)$, then, the following Bayesian evidence measure to test (1) can be considered:

$$\mathrm{B}_2 = \int_{\Theta} P\left(H_0 | n_{11}, \ldots, n_{rs}, \mathbf{p}_0\right) \pi\left(\mathbf{p}_0\right) d\mathbf{p}_0.$$

From this Bayesian viewpoint we reject H_0 when $\mathrm{B}_2 > 1/2$.

3.3. THIRD METHOD

In the same context of the second method, the idea is to consider the supremum value of $P\left(H_0 | n_{11}, \ldots, n_{rs}, \mathbf{p}_0\right)$ when $\mathbf{p}_0 \in \Theta$ as a Bayesian quantitative measure to test (1).

In this situation, if we assign to each $\mathbf{p}_i$ a Dirichlet prior distribution of parameter α_i, $i = 1, \cdots, r$, as the infimum of $\prod_{j=1}^{s} p_{0j}^{-m_j}$ when $\mathbf{p}_0 \in \Theta$ is reached in $\hat{p}_{0j} = m_j/N$, $j = 1, \cdots, s$, such a measure would be

$$\mathrm{B}_3 = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_3\right]^{-1}, \tag{5}$$

where

$$\eta_3 = N^N \prod_{j=1}^{s} m_j^{-m_j} \frac{\prod_{i=1}^{r} \Gamma\left(\sum_{j=1}^{s} \alpha_{ij}\right)}{\prod_{i=1}^{r} \prod_{j=1}^{s} \Gamma\left(\alpha_{ij}\right)} \frac{\prod_{i=1}^{r} \prod_{j=1}^{s} \Gamma\left(n_{ij} + \alpha_{ij}\right)}{\prod_{i=1}^{r} \Gamma\left(n_i + \sum_{j=1}^{s} \alpha_{ij}\right)}.$$

Therefore, with this method we reject H_0 when $\mathrm{B}_3 < 1/2$.

4. RECONCILIATION BETWEEN FREQUENTIST AND BAYESIAN APPROACHES

From the frequentist viewpoint, instead of considering the observed data $(n_{11}, \ldots, n_{rs})$ in Table 1 as fixed values and permitting that $\mathbf{P}$ changes, the point $\mathbf{P}_0$ of the null hypothesis is fixed and later, the probability of observing a point in some extreme region which includes $(n_{11}, \ldots, n_{rs})$ is calculated. That is to say, instead of calculating the posterior probability of the null hypothesis, the p-value is calculated. (The idea is basically that or H_0 is false, or an event with very small probability has occurred.)

As usual, we are going to use as frequentist measure of the evidence, the discrepancy between the observed and expected values under H_0, in the terms of Pearson's χ^2 statistic. Therefore, the *test statistic* is the random variable

$$\Lambda = N\left(\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{n_{ij}^2}{n_i m_j} - 1\right).$$

If λ_0 is the value of Λ at an observed point, $\Lambda\left(n_{ij0},\ i=1,\cdots,r,\ j=1,\cdots,s\right)=\lambda_0$, then $\{\Lambda \geq \lambda_0\}$ is a possible *critical region* and the corresponding p-value is

$$p\left(\lambda_0\right) = \sup_{\mathbf{p_0}\in\Theta} P\left(\Lambda \geq \lambda_0 | \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}_0\right) = P\left(\chi^2_{(r-1)(s-1)} \geq \lambda_0\right).$$

With this procedure, the decision of accepting or rejecting H_0 depends on the size of the p-value. For instance, H_0 is rejected when $p\left(\lambda_0\right) < \alpha$, where $\alpha \in (0,\ 1)$ is a sufficiently small value (the significance level of the test).

We can note that the three Bayesian evidence measures given in expressions (3), (4) and (5), respectively, depends on $\pi_0 = \pi_0(\delta)$ given in (2).

The value of π_0 which verifies

$$B_k(\pi_0) = \frac{p}{2p^*} \quad (k=1,2,3) \tag{6}$$

satisfies $P(H_0|n_{11},\ldots,n_{rs}) > \frac{1}{2}$ when $p(n_{11},\ldots,n_{rs}) > p^*$. Therefore, using the value of π_0 which is obtained from (6), the same conclusion would be reached with both approaches. If we denote this value by π_{k0}^0, we can note that

$$\pi_{k0}^* = \frac{\eta_k p}{\eta_k p + 2p^* - p},\ k=1,3,$$

while, with the second method, π_{20}^* must be calculated numerically.

Notwithstanding, this reconciliation is too strict, since π_{k0}^* $(k=1,2,3)$ depends on the data. In this sense, we do not affirm that the procedure to obtain the agreement has to be done by means of equaling both expressions but using of a value next to the result of this equalization. Consequently, the value of π_0, and accordingly the value of δ, which obtains the agreement must decrease when p^* increases.

The desirable reconciliation is to formulate the agreement so that if for example $p^* \in (0.05,\ 0.1)$, then there exists an interval of values of $\pi_0(\delta) \in (\ell_1,\ \ell_2)$, for some $\ell_1,\ \ell_2 \in (0,\ 1)$, $\ell_1 < \ell_2$, such that the decision obtained using the p-value to test (1) is the same as the conclusion reached with some Bayesian measure.

In order to eliminate the dependence of the data, we consider the sample space formed by all of the possible $r \times s$ tables to $n_i,\ i = 1, \ldots, r$ fixed and known.

Remember that the three Bayesian evidence measures given in expressions (3), (4) and (5) depend on $\pi_0 = \pi_0(\delta)$ given in (2).

Let $\pi_0^k = \pi_0^k(\delta_k)$ the value of π_0 which verifies $\mathrm{B}_k(\pi_0) > 1/2$ when $\pi_0 > \pi_0^k$, $k = 1, 2, 3$. For example, $\pi_0^k = \eta_k(\eta_k + 1)^{-1}$, when $k = 1, 3$.

Fixed p^*, denote by means of

$$\ell_1^k = \ell_1^k(p^*, n_1, \cdots, n_r) = \max_{(n_{ij}),\ p > p^*} \pi_0^k,$$

$$\ell_2^k = \ell_2^k(p^*, n_1, \cdots, n_r) = \min_{(n_{ij})\ p \leq p^*} \pi_0^k,$$

$k = 1, 2, 3$.

The following theorem shows how and when it is possible to achieve a reconciliation in the exposed terms.

Theorem 3.1. Let $n_i,\ i = 1, \ldots r$ and p^* be fixed and known.

If p^* satisfies $\ell_1^k \leq \ell_2^k$ with the $k-$Bayesian method ($k = 1, 2, 3$), then there exists an interval of values of π_0, $I_k = I_k(p^*, n_1, \ldots n_r) = \left(\ell_1^k,\ \ell_2^k\right)$, such that one and only one of the two following postulates is verified:

$$\text{“}p(n_{110}, \cdots, n_{rs0}) > p^* \ \&\ B_k(\pi_0 | n_{110}, \cdots, n_{rs0}) > 1/2\text{”},$$

$$\text{“}p(n_{110}, \cdots, n_{rs0}) \leq p^* \ \&\ B_k(\pi_0 | n_{110}, \cdots, n_{rs0}) \leq 1/2\text{”},$$

whatever $(n_{110}, \cdots, n_{rs0})$ may be.

Proof. The three Bayesian evidence measures given in expressions (3), (4) and (5) verifies that $B_k(\pi_0)$ is an increasing function of π_0 and $B_k(\pi_0) > 1/2$, when $\pi_0 > \pi_0^k$, $k = 1, 2, 3$.

Moreover, if $\lambda_1 < \lambda_2$, then $p(\lambda_1) = P\{\Lambda \geq \lambda_1 | \theta_0\} \geq P\{\Lambda \geq \lambda_2 | \theta_0\} = p(\lambda_2)$.

Let λ^* and λ_* be

$$\lambda^* = \min_{(n_{11},\cdots,n_{rs}),\ p(\lambda)\leq p^*} \Lambda,$$

$$\lambda_* = \max_{(n_{11},\cdots,n_{rs}),\ p(\lambda)> p^*} \Lambda.$$

Thereby, $\lambda_* \leq \lambda^*$.

Furthermore, if p^* satisfies $\ell_1^k \leq \ell_2^k$, then $(\ell_1^k,\ \ell_2^k)$ is an interval of values in $(0,\ 1)$.

Fixed $\pi_0 \in \left(\ell_1^k,\ \ell_2^k\right)$ and $(n_{11}, \cdots, n_{rs})$ such that $\Lambda(n_{11}, \cdots, n_{rs}) = \lambda$, with $\lambda < \lambda_*$, then $\pi_0 > \ell_1^k \geq \pi_0^k(n_{11}, \ldots, n_{rs})$ and $P\{\Lambda \geq \Lambda(x_1, \cdots, x_n) | \theta_0\} > p^*$.

On the other hand, fixed $\pi_0 \in \left(\ell_1^k,\ \ell_2^k\right)$ and $(n_{11}, \cdots, n_{rs})$ such that $\lambda \geq \lambda^*$, then $\pi_0 < \ell_2^k \leq \pi_0^k(n_{11}, \ldots, n_{rs})$ and $P\{\Lambda \geq \Lambda(x_1, \cdots, x_n) | \theta_0\} \leq p^*$.

Therefore, $\ell_1^k \leq \ell_2^k$ is a sufficient condition to reach the same conclusion to test (1) with the p-value, using p^*, and the $k-$Bayesian method ($k = 1, 2, 3$), using a value of δ_k with $\pi_0(\delta_k) \in (\ell_1^k, \ell_2^k)$ in the corresponding mixed prior distribution.

Table 3. Data in the 2×2 table.

	Successes	Failures	Total
Sample 1	a	b	n_1
Sample 2	c	d	n_2
Total	m_1	m_2	N

To illustrate the procedure, we are going to consider 2×2 tables (see Table 3). In this case, we want to test if the proportion of successes in the first population, p_1, is the same as in the second, p_2, that is

$$H_0 : p_1 = p_2,\ \textit{versus}\ \ H_1 : p_1 \neq p_2. \tag{7}$$

In this situation, the usual *test statistic* is the random variable

$$\Lambda = \{ad - bc\}^2 \frac{N}{n_1 n_2 m_1 m_2}.$$

and, when a data point $\lambda_0 = \Lambda(a_0, c_0)$ is observed, the evidence used is the p-value,

$$p = P\left(\Lambda \geq \lambda_0 | p_1 = p_2\right) = P\left(\chi_1^2 \geq \lambda_0\right).$$

From a Bayesian viewpoint, when p_0, p_1 and p_2 have uniform prior distributions, respectively, the Bayesian evidence measures given in expressions (3), (4) and (5) are obtained evaluating such expressions in

$$\eta_1 = \frac{\Gamma\left(N+2\right)}{\Gamma\left(m_1+1\right)\Gamma\left(m_2+1\right)} \gamma\left(a, b, c, d\right),$$

$$\eta_2 = \eta_2\left(p_0\right) = p_0^{-m_1}\left(1-p_0\right)^{-m_2} \gamma\left(a, b, c, d\right),$$

$$\eta_3 = N^N m_1^{-m_1} m_2^{-m_2} \gamma\left(a, b, c, d\right),$$

where $\gamma\left(a, b, c, d\right) = \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}$.

Moreover, this measures depends on $\pi_0 = \pi_0(\delta)$ given in (2). In this case, $\pi_0 = 2\sqrt{2}\delta + 2\delta^2 - 4\sqrt{2}\delta^3$, when δ is sufficiently small.

It is necessary to point out that none of the statistics Λ, η_1 and η_3 are *sufficient statistics*.

A summary of results to Pearson's data (see Table 2) is displayed in Table 4. We observe that the value of $\pi_{k0}^* = \pi_{k0}^*(\delta_k)$ $(k = 1, 2, 3)$ which gets strict agreement by (6) decreases when p^* increases. Furthermore, the third Bayesian method is the most conservative with respect to the H_0 in (7), whereas the second is the least. For instance, if $\pi_0 = \frac{1}{2}$, the three Bayesian methods reject H_0, but $B_3 = 0.4484$, $B_1 = 0.1463$ and $B_2 = 0.111$. Thereby, as we see in Table 4, $\pi_{30}^* < \pi_{10}^* < \pi_{20}^*$. As the values of δ_2^* are close to the values of δ_1^* and the computational cost is higher, we propose to use the first method or, with a more conservative point of view, the third.

Table 4. Summary of results for Pearson's data, with $\pi(p_0) = I_{(0,1)}(p_0)$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1)\, I_{(0,1)}(p_2)$.

Classical Method

Λ	p
5,625	0,017706

Bayesian Methods

Methods	η_k	$\mathrm{B}_k(1/2)$	π_0^k	δ_k
Method 1	5,8347	0,1463	0,8537	0,2913
Method 2		0,110919	0,995416	0,3514
Method 3	1,2301	0,4484	0,5516	0.1836

Methods	Strict Values	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
Method 1	π_{k0}^*	0,09516	0,3617	0,5566	0,9782
	δ_k^*	0,03294	0,1211	0,1852	0,3435
	$p\,(2p^*)^{-1}$	0,017706	0,08853	0,17706	0,8853
Method 2	π_{k0}^*	0,098975	0,41969	0,690015	$0,9^{11}882$
	δ_k^*	0,034244	0,14001	0,23088	0,3535
	$p\,(2p^*)^{-1}$	0,017706	0,08853	0,17706	0,8853
Method 3	π_{k0}^*	0,02169	0,1067	0,2093	0,9047
	δ_k^*	0,00763	0,03686	0,07114	0,3117
	$p\,(2p^*)^{-1}$	0,017706	0,08853	0,17706	0,8853

To eliminate the dependence on the data, we have generated a total of 247 possible 2×2 tables with $n_1 = 18$ and $n_2 = 12$. These tables are organized in ascending order carried out according to the values of η_1 and η_3 (see Figure 1 and 2), respectively. It may be checked that it is not possible to express B_1 and B_3 in terms of Λ. However, there exist non-monotonous functions $h_k : \mathrm{R}^+ \to \mathrm{R}^+$, such that $\Lambda = h_k(\eta_k)$, k =1, 3 (see Figures 1 and 2). Therefore, the *critical region* may be expressed in terms of η_1 and η_3. Also, we observe that h_3 is more irregular than h_1. Moreover, it is clear that the existence of values of p^* which satisfy

the sufficient condition that ensures the agreement between both methods depends on the *increasing tendency* which we can observe in the functional relationship that exists between both statistics, $\Lambda = h_k(\eta_k)$ $(k = 1, 3)$, although this relationship is not strictly monotonous.

A summary of results to Pearson's data (see Table 2) is displayed in Table 4. We observe that the value of $\pi^*_{k0} = \pi^*_{k0}(\delta_k)$ $(k = 1, 2, 3)$ which gets strict agreement by (6) decreases when p^* increases. Furthermore, the third Bayesian method is the most conservative with respect to the H_0 in (7), whereas the second is the least. For instance, if $\pi_0 = \frac{1}{2}$, the three Bayesian methods reject H_0, but $B_3 = 0.4484$, $B_1 = 0.1463$ and $B_2 = 0.111$. Thereby, as we see in Table 4, $\pi^*_{30} < \pi^*_{10} < \pi^*_{20}$. As the values of δ^*_2 are close to the values of δ^*_1 and the computational cost is higher, we propose to use the first method or, with a more conservative point of view, the third.

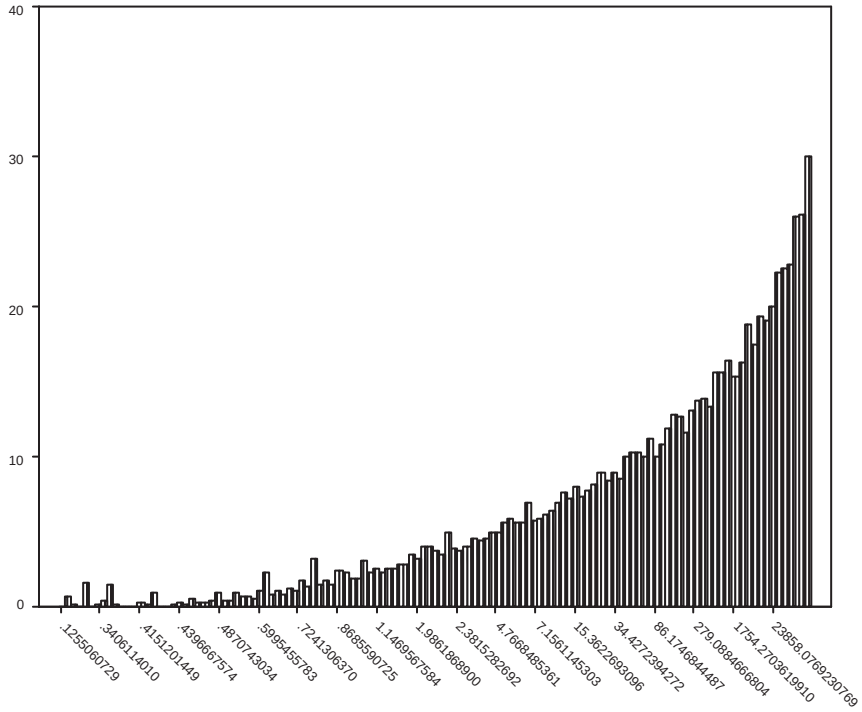

Figure 1: Bars Diagram $(\eta_1(a, c), \Lambda(a, c))$ for 2×2 tables with $n_1 = 18$, $n_2 = 12$, $\pi(p_0) = I_{(0,1)}(p_0)$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$.

By means of an easy data analysis, we can check that to test (7), with $n_1 = 18$ and $n_2 = 12$, using $\pi(p_0) = I_{(0,1)}(p_0)$ and $\pi(p_1, p_2) = I_{(0,\ 1)}(p_1)\, I_{(0,\ 1)}(p_2)$, there are values of p^* which satisfy the sufficient condition of Theorem 1. For instance, the highest value of p^* which is in agreement with the first method is $p^* = 0.0635$, while with the second this value is $p^* = 0.008$, because the third method is more conservative than the first. Moreover, when $p^* \in (0.0635,\ 0.0637)$ or $p^* \in (0.008,\ 0.0085)$, using the first method, respectively, with $\delta \in (0.2222,\ 0.223)$ (that is $\pi_0 \in (0.6651,\ 0.6675)$) or $\delta \in (0.3218,\ 0.3252)$ (that is $\pi_0 \in (0.9288,\ 0.9368)$), the obtained Bayesian decision is the same as the one obtained with the classical method. With the third method, this also happens when $p^* \in (0.08,\ 0.0085)$ using $\delta \in (0.2478,\ 0.2503)$ (that is $\pi_0 \in (0.73769,\ 0.74455)$). However, there is not agreement when $p^* = 0.5$, $p^* = 0.1$, $p^* = 0.05$ o $p^* = 0.01$ with neither of them.

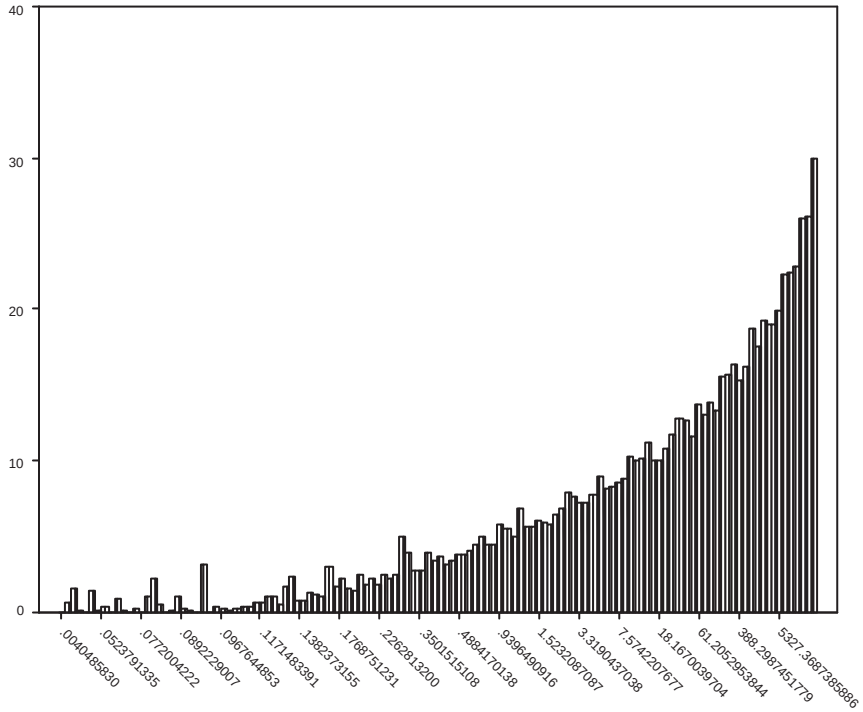

Figure 2: Bars Diagram $(\eta_3(a,c), \Lambda(a,c))$ for 2×2 tables with $n_1 = 18$, $n_2 = 12$, $\pi(p_0) = I_{(0,1)}(p_0)$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1)\, I_{(0,1)}(p_2)$.

5. CONCLUSIONS AND COMMENTS

Using $r \times s$ tables and appropriate mixed prior distributions, when independent samples are drawn from r multinomial populations, three Bayesian measures, B_k, $(k = 1, 2, 3)$ of the strength of the evidence given by the data against the homogeneity null hypothesis to test (1) can be calculated (see expressions (3), (4) and (5)).

Choosing appropriate values of π_0, the prior mass assigned to H_0 given in expression (2), it is possible to reach the same decision with frequentist and Bayesian methods. Indeed, fixing n_i, $i = 1, \cdots, r$ and $p^* \in (0, 1)$ (the value used by a frequentist statistician to quantify the usual p-value), Theorem 3.1 gives a sufficient condition by which a reconciliation between both measures is possible. That is, when $\ell_1^k \leq \ell_2^k$ (ℓ_1^k and ℓ_2^k as in Theorem 3.1) is satisfied to any of the proposed Bayesian approaches in section 3 ($k = 1, 2, 3$), a Bayesian statistician choosing $\pi_0 \in (\ell_1^k, \ell_2^k)$ in the corresponding mixed prior distribution and quantifying B_k with $1/2$, takes the same decision to test (1) as a frequentist statistician who uses p^* to quantify the usual p-value, whatever the data point may be.

The generalization of the previous results to the homogeneity testing problem of independent multinomial populations, when the common proportions under the null have a known functional form, $\mathbf{p}_0 = \mathbf{p}(\omega)$, is possible following a similar reasoning.

ACKNOWLEDGMENTS

This research was supported by Universidad Complutense-Comunidad de Madrid and Ministerio de Educación y Ciencia from Spain Grant MTM2005-05462.

We thank an Associate Editor and a referee for their valuable comments.

BIBLIOGRAPHY

Berger, J. O., Boukai, B. and Wang, Y. (1999). Simultaneous Bayesian-Frequentist Sequential Testing of Nested Hypothesis. *Biometrika*, **86**, 79-92.

Berger, J. O. and Delampady, M. (1987). Testing Precise Hyphotheses, (with Discussion). *Statistical Science*, **2**(3), 317-352.

Berger, J. O. and Sellke, T. (1987). Testing a Point Null Hyphotesis: The Irreconciliability of p-Values and Evidence, (with Discussion). *J. Amer. Statist. Assoc.*, **82**, 112-139.

Casella, G. and Berger, R. L. (1987). Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem, (with Discussion). *J. Amer. Statist. Assoc.*, **82**, 106-135.

Ghosh, J. K. and Ramamoorthi R. V. (2003). *Bayesian Nonparametrics*. Springer.

Gómez-Villegas, M. A. and González-Pérez, B. (2005). Bayesian Analysis of Contingency Tables. *Commun. Statist-Theory Meth.*, **34**, 1743-1754.

Gómez-Villegas, M. A., Maín, P. and Sanz, L. (2004). A Bayesian Analysis for the Multivariate Point Null Testing Problem. *Technical Report Dpto. EIO-I. Universidad Complutense de Madrid. 04-01.*

Howard, J. V. (1998). The 2×2 Table: A Discussion from a Bayesian Viewpoint. *Statistical Science*, **13**(4), 351-367.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction*. London: Arnold.

Lindley, D. V. (1957). A Statistical Paradox. *Biometrika*, **44**, 187-192.

Lindley, D. V. (1988). Statistical Inference Concerning Hardy-Weinberg Equilibrium. *Bayesian Statistics*, **3**, 307-326.

Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139-167.

Quintana F. A. (1998). Nonparametric Bayesian Analysis for Assessing Homogeneity in $k \times l$ Contingency Tables With Fixed Right Margin Totals. *Journal of the American Statistical Association*, **93**(443), 1140-1149.