

Analysis of Contingency Tables

Bayesian Analysis of Contingency Tables

MIGUEL A. GÓMEZ-VILLEGAS AND
BEATRIZ GONZÁLEZ PÉREZ

Departamento de Estadística e Investigación Operativa I, Universidad
Complutense de Madrid, Madrid, Spain

The display of the data by means of contingency tables is used in different approaches to statistical inference, for example, to broach the test of homogeneity of independent multinomial distributions. We develop a Bayesian procedure to test simple null hypotheses versus bilateral alternatives in contingency tables. Given independent samples of two binomial distributions and taking a mixed prior distribution, we calculate the posterior probability that the proportion of successes in the first population is the same as in the second. This posterior probability is compared with the p -value of the classical method, obtaining a reconciliation between both results, classical and Bayesian. The obtained results are generalized for $r \times s$ tables.

Keywords Bayesian statistics; Chi-square tests; Contingency tables; p -values.

Mathematics Subject Classification 62F15; 62H17.

1. Introduction

The $r \times s$ table is used for discussing different approaches to statistical inference. For example, suppose that independent random samples are drawn from two large populations, and their each member is classified as a “success” or a “failure”. The first sample is of size n_1 and produces a successes and b failures, the second is of size n_2 and produces c successes and d failures. The situation is displayed in the Table 1.

In this situation a quantitative measure of the strength of the evidence that the data gives support or rejection of the hypothesis that the proportion of successes in the first population, p_1 , is equal to the proportion of successes in the second population, p_2 , is required. This problem, apparently simple, has given rise to an extensive literature, since Karl Pearson introduced his already classical χ^2 test to value the goodness of the fit (see Pearson, 1900). This is one of the simplest natural problems to demonstrate clear differences between classical and Bayesian

Received April 7, 2004; Accepted February 3, 2005

Address correspondence to Miguel A. Gómez-Villegas, Departamento de Estadística e Investigación Operativa I, Facultad de CC. Matemáticas, Universidad Complutense de Madrid, Madrid 28040, Spain; E-mail: ma.gv@mat.ucm.es

Table 1
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

approaches, and also between different types of classical analysis. There are of course a number of variations on this problem. Some important bayesian references are given next.

Howard (1998) advocates for the more frequent use of unilateral tests and approaches to the problem from a Bayesian viewpoint. He considers as hypotheses of interest $H_1 : p_2 < p_1$ and $H_2 : p_1 < p_2$, and gives a quantitative measure of the strength of the evidence in support of the more likely hypothesis. He assumes 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. He has to assume independent Jeffreys priors about the two populations, that is to say,

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

in order to get this result. Besides, 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 $\Lambda_i = \log p_i(1-p_i)^{-1}$, and $\Lambda_i = \arcsen \sqrt{p_i}$, $i = 1, 2$, respectively, as measures of the degree in which two populations are homogeneous (see Lee, 1997, pp. 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, $\alpha = \frac{1}{2} \log \frac{4p_1p_3}{p_2^2}$ and $\beta = \frac{1}{2} \log \frac{p_1}{p_3}$. 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 Hardy-Weinberg equilibrium, $H_0 : p^2, 2p(1-p), (1-p)^2$, p being the proportion of A 's.

We consider testing equality of proportions of independent multinomial distributions when the common proportions are known. Our general approach to the problem of homogeneity consists in working directly with the simple null hypothesis and calculating its posterior probability. To do this, we will follow the

Table 2
Pearson's example

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

method used by Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002), based on assigning an initial probability π_0 to the null hypothesis and distributing the remaining probability in the points of the alternative with a prior density $\pi(p_1, p_2)$. Posterior probabilities of the null hypothesis are calculated with respect to a mixture of point prior on the null and an independent Dirichlet prior on the proportions. With this procedure, in the context of the punctual null hypothesis, it is possible to get a reconciliation between the classical p -value and the Bayesian posterior probability of the null hypothesis.

Section 2 formulates the problem in a precise way and calculates an exact expression of the posterior probability that the proportion of successes in the first population is the same as in the second, and equal to a known common value p_0 . Section 3 reaches a reconciliation between the classical and Bayesian results, and the Pearson (1947) data (see Table 2) is used to illustrate the procedure. Section 4 generalizes the results of Sec. 2 and 3 for a $r \times s$ table. Section 5 exposes a summary of conclusions.

2. Formulation of the Problem and Posterior Probability

Consider $X_i, i = 1, 2$, independent random binomial variables, $B(n_i, p_i)$, and suppose that we wish to test

$$H_0 : p_1 = p_2 = p_0 \text{ versus } H_1 : p_1 \neq p_2, \tag{1}$$

where p_0 is a known value and the hypothesis $p_1 \neq p_2$ means that at least one of them is different from p_0 , that is to say, $(p_1, p_2) \neq (p_0, p_0)$. Moreover, suppose that our prior opinion about (p_1, p_2) is given by the density $\pi(p_1, p_2)$. Hence, in order to test (1), a mixed prior distribution is needed. We propose

$$\pi^*(p_1, p_2) = \pi_0 I_{H_0}(p_1, p_2) + (1 - \pi_0)\pi(p_1, p_2)I_{H_1}(p_1, p_2),$$

π_0 being the assigned prior probability to the null hypothesis.

Then, the posterior probability of the null hypothesis, when the data of Table 1 has been observed, is

$$P(H_0 | a, c) = \frac{p_0^{a+c}(1-p_0)^{b+d}\pi_0}{p_0^{a+c}(1-p_0)^{b+d}\pi_0 + (1-\pi_0)\int_0^1\int_0^1 p_1^a(1-p_1)^b p_2^c(1-p_2)^d \pi(p_1, p_2) dp_2 dp_1}.$$

A possible initial distribution consists in assigning independent uniform prior distributions, also called independent Laplace distributions, that is to say,

$$\pi(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2).$$

In this situation, we obtain

$$P(H_0 | a, c) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1}, \quad (2)$$

where $\eta = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}$.

A more general assignment consists in using independent beta prior distributions,

$$\pi(p_1, p_2) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\alpha-1} (1 - p_1)^{\beta-1} p_2^{\gamma-1} (1 - p_2)^{\delta-1},$$

where $p_1, p_2 \in (0, 1)$, $(\alpha, \beta, \gamma, \delta > 0)$.

Then, the posterior probability of the null hypothesis is obtained evaluating expression (2) in

$$\eta = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(a + \alpha)\Gamma(b + \beta)}{\Gamma(a + b + \alpha + \beta)} \frac{\Gamma(c + \gamma)\Gamma(d + \delta)}{\Gamma(c + d + \gamma + \delta)}.$$

The posterior probability that is calculated in expression (2) depends on π_0 , the initial prior probability assigned to the null hypothesis $H_0 : p_1 = p_2 = p_0$. Now, consider the more realistic precise hypothesis

$$H_{0_\varepsilon} : d((p_0, p_0), (p_1, p_2)) \leq \varepsilon \quad \text{versus} \quad H_{1_\varepsilon} : d((p_0, p_0), (p_1, p_2)) > \varepsilon, \quad (3)$$

with an appropriate metric d and a value of $\varepsilon > 0$ sufficiently small. Applying the method of Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002), we can use $\pi(p_1, p_2)$, our opinion about (p_1, p_2) , and calculate π_0 by means of averaging,

$$\pi_0 = \iint_{B((p_0, p_0), \varepsilon)} \pi(p_1, p_2) dp_2 dp_1, \quad (4)$$

where $B((p_0, p_0), \varepsilon) = \{(p_1, p_2) \in (0, 1) \times (0, 1), d((p_0, p_0), (p_1, p_2)) \leq \varepsilon\}$, the sphere of center (p_0, p_0) and radius ε .

Then, the prior probability assigned to H_0 and to H_{0_ε} by means of $\pi(p_1, p_2)$ is the same thing.

Different ways of specifying $d((p_0, p_0), (p_1, p_2))$ can be considered. One of them could be considering an arbitrary value of ε and dividing it in two values ε_1 and ε_2 , may be $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$, and then we would build the distance starting from $|p_i - p_0| < \varepsilon_i$, $i = 1, 2$. Another way could be considering

$$B((p_0, p_0), \varepsilon) = \{(p_1, p_2) \in (0, 1) \times (0, 1), (p_1 - p_0)^2 + (p_2 - p_0)^2 \leq \varepsilon^2\}.$$

By means of this second procedure, the posterior probability obtained in (2) can be expressed in terms of ε . In this article the results are obtained first in function

of π_0 , and afterwards are specified in terms of ε employing the expression (4). In particular, it is possible to calculate the value of ε in (3) such that $\pi_0 = \frac{1}{2}$. It can be observed that if our prior opinion about (p_1, p_2) is the uniform distribution given by means of the density $\pi(p_1, p_2) = 1, p_1, p_2 \in (0, 1)$, then the value of π_0 that is obtained with the expression (4), for ε sufficiently small, is $\pi_0 = \pi\varepsilon^2$, the area of the sphere of radius ε .

Note that, in general, $H_0 : p_1 = p_2 = p_0$ in (1) is no natural null hypothesis. By this reason we consider first a value of p_0 and after take an sphere of radius ε about this value. Besides, in general, when we wish to test (1), the value of p_0 is unknown. In spite of this, (1) has a clear theoretical interest because it can be used as an auxiliary test to develop a Bayesian procedure, with the proposed methodology, when p_0 is unknown or with functional form known.

Suppose that we wish to test (1) with $p_0 = \frac{1}{2}$ and our prior opinion about (p_1, p_2) is given by the uniform density $\pi(p_1, p_2) = 1, p_1, p_2 \in (0, 1)$.

Thereby, the posterior probability of the null hypothesis is

$$P(H_0 | a, c) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1}, \tag{5}$$

where $\eta = 2^N \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}$.

It can be observed that values of ε which correspond with $\pi_0 > \frac{\eta}{\eta+1}$ get $P(H_0 | a, c) > \frac{1}{2}$. Moreover, $P(H_0 | a, c) = \frac{1}{\eta+1}$ for ε such that $\pi_0 = \frac{1}{2}$.

For the data of Table 2 we obtain $\eta = 6.7265$ and, if $\varepsilon = \frac{1}{\sqrt{2\pi}}$, then $\pi_0 = \frac{1}{2}$ and $P(H_0 | a, c) = 0.1294$, so that H_0 is rejected. Moreover, to accept H_0 with the data of Pearson's example, $\varepsilon > 0.53905$ or $\pi_0 > 0.8706$. Therefore, for the data of Table 2, we can observe that there is a wide range of values of $\varepsilon, \varepsilon < 0.53905$, for which H_0 is rejected.

3. Comparison with the Classical Method

From the classical viewpoint, instead of considering the observed data (a, c) as fixed values and permitting that (p_1, p_2) changes, the point (p_0, p_0) of the null hypothesis is fixed and after the probability of observing a point in some extreme region of the alternative hypothesis which includes (a, c) 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 probability very small has occurred.)

As usual, if we use, as measure of the evidence in support of H_1 , the discrepancy between the observed values and the expected values when H_0 is true, then, in the terms of Pearson's χ^2 statistic, the test statistical would be the random variable

$$\Lambda = \frac{a^2}{n_1 p_0} + \frac{b^2}{n_1(1 - p_0)} + \frac{c^2}{n_2 p_0} + \frac{d^2}{n_2(1 - p_0)} - N. \tag{6}$$

The sampling distribution of Λ when H_0 is true is χ^2_2 . Then, if the value of Λ in the data point is $\Lambda(a_0, c_0) = \kappa$, and the experiment was repeated independently, once again sampling n_1 subjects randomly from population 1 and n_2 subjects randomly from population 2 for (p_0, p_0) fixed, the probability that we would get a

new value of Λ as big as κ or larger than κ can be calculated. Therefore, $\{\Lambda \geq \kappa\}$ is a possible critical region, and

$$p = P\{\Lambda \geq \kappa \mid (p_0, p_0)\} = P(\chi_2^2 \geq \kappa) = e^{-\frac{\kappa}{2}},$$

is the p -value.

With this procedure, the decision of accepting or rejecting H_0 depends on the size of the p -value, namely, H_0 is rejected when $p < p^*$, $p^* \in (0, 1)$ being a value sufficiently small.

Now, we are going to suppose that we wish to test (1) with $p_0 = \frac{1}{2}$ by means of the previous classical method.

In this situation, the test statistic is the random variable

$$\Lambda = 2 \left[\frac{a^2 + b^2}{n_1} + \frac{c^2 + d^2}{n_2} \right] - N,$$

and the evidence used is the p -value,

$$p = e^{\frac{N}{2} - \frac{a_0^2 + b_0^2}{n_1} - \frac{c_0^2 + d_0^2}{n_2}}.$$

For the data of Table 2 we obtain $\kappa = 8.33333$, and a p -value $p = 0.015504$. Observe that H_0 is rejected for $p^* = 0.05$, but for $p^* = 0.01$ there is not enough evidence to reject it, and in that sense H_0 is accepted.

To compare the proposed Bayesian method with Pearson's χ^2 classical method, which uses the value given in expression (6) as the test statistical, it would be interesting if there exists a functional dependence between both statistics, η and Λ , or between the posterior probability and the p -value, p . That it to say, $\eta = g(\Lambda)$ for some increasing function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$. However, for 2×2 tables, if $n_1 = 18$ and $n_2 = 12$, it can be observed that for the data of Pearson's example, $(a, c) = (3, 7)$, the value of η in the expression (5) is 6.72 and the value of Λ in the expression (6) is 8.33333, whereas if $(a, c) = (9, 1)$, then $\eta = 7.45$ and $\Lambda = 8.33333$. Hence, such functional dependence is not possible. Furthermore, it can be observed that in contradistinction to Λ , η distinguishes between the two previous situations. Notwithstanding, it can be verified that there exists a non-monotonous function, $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, for which $\Lambda = h(\eta)$ (see Fig. 1). Therefore, the classical test allows a representation in terms of η .

Now, the objective is to get some kind of reconciliation between the classical and the Bayesian approaches, that is to say, it would be convenient that a same number had both performances. To do this, we consider the following equation,

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1} = \frac{p}{2p^*},$$

from which the value of π_0 can be obtained,

$$\pi_0 = \left[1 + \frac{1}{\eta} \left(\frac{2p^*}{p} - 1 \right) \right]^{-1} = \frac{\eta p}{\eta p + 2p^* - p}, \tag{7}$$

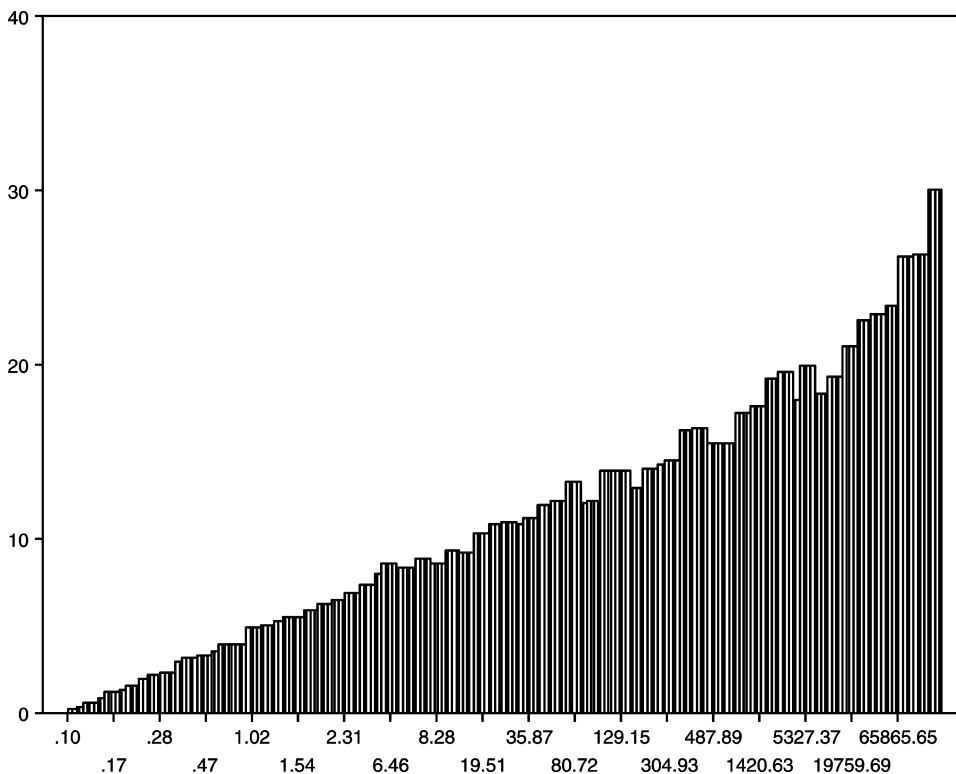


Figure 1. Bars diagram $(\eta(a, c), \Lambda(a, c))$, for 2×2 tables with $n_1 = 18$ and $n_2 = 12$. There is a non-monotonous function, $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, such that $\Lambda = h(\eta)$.

which satisfies that $P(H_0 | a, c) > \frac{1}{2}$ when $p > p^*$. Therefore, using the value of π_0 which is obtained in the expression (7), the same conclusion would be reached with both methods.

Notwithstanding, this reconciliation is too strict, since the obtained value in expression (7) depends on the data. In this sense, we do not affirm that the procedure to obtain the accord has to be by means of equaling both expressions, but that the use of a value next to the result of this equalization can furnish, when this is possible, an approximately equal numeric value from both viewpoints. The desirable reconciliation would formulate the accord so that if for example $p^* \in (0.05, 0.1)$, then $\pi_0 \in (\ell_1, \ell_2)$ for some $\ell_1, \ell_2 \in (0, 1)$, $\ell_1 < \ell_2$.

It can be noted that $0 < \pi_0 < 1$ only when $2p^* > p$. Moreover, fixing p^* , $0 < p^* < 1$, for any p -value $p, 0 < p < 2p^*$, there is an initial prior probability $\pi_0, 0 < \pi_0 < 1$, assigned to the null hypothesis of test (1) for $p_0 = \frac{1}{2}$, assuming our initial opinion about (p_1, p_2) is uniform, that allows both results, the classical and the Bayesian, to be equal. It can also be observed that if $p^* = \frac{1}{2}$, then, whatever the p -value p is, such π_0 always exists and verifies that $P(H_0 | a, c) = p$.

For the data of Table 2, if $p^* = \frac{1}{2}$, the value π_0 that reconciles the classical p -value, $p = 0.015504$, with the Bayesian posterior probability is $\pi_0 = 0.09578$. If $p^* = 0.1$ we obtain $\pi_0 = 0.36113$ and reject with a posterior probability 0.07752 .

Table 3
Summary of results for Pearson's example

κ	p	η	$(\eta + 1)^{-1}$	$\eta(\eta + 1)^{-1}$	ε
8.333333	0.015504	6.7265	0.1294	0.8706	0.53905
Table 2		$p^* = 0.5$	$p^* = 0.1$	$p^* = 0.05$	$p^* = 0.01$
$\eta p(\eta p + 2p^* - p)^{-1}$		0.09578	0.36113	0.5524	0.9587
ε		0.17461	0.33904	0.41933	0.6085
$p(2p^*)^{-1}$		0.015504	0.07752	0.15503	0.77523

If $p^* = 0.05$ we get $\pi_0 = 0.5524$ and reject with 0.15503. For $p^* = 0.01$ we get $\pi_0 = 0.9587$ and accept with 0.77533.

The obtained results are summarized in Table 3.

We can observe that the value of π_0 , and accordingly the value of ε , which obtains the agreement between the classical and the Bayesian results in the previously exposed terms, decreases when p^* increases. Besides, for the data of Pearson's example, the values of ε for which this agreement is achieved when $p^* \in (0.01, 0.5)$ are such that $\varepsilon \leq 0.6085$.

It has already been indicated that the accord between the classical and Bayesian results that is obtained by means of expression (7) is too strict. However, it gives an idea of what value of ε , when it exists, must be so that this reconciliation between both methods is possible.

To eliminate the dependence of the data, we have generated all of the possible 2×2 tables to n_1 and n_2 fixed and known. In the situation that we are studying, the entries are $n_1 = 18$ and $n_2 = 12$, and a total of 247 possible tables have been generated. Pearson's data is organized in Table 95 in the ascendant sort carried out according to the values of η (see Fig. 1). For every one of these tables, we carry out the same study that has previously been carried out for the data of Pearson's example.

By means of an easy data analysis, we can check that there are values of p^* , for example $p^* = 0.5$, $p^* = 0.1$, $p^* = 0.05$, or $p^* = 0.01$, such that we can find an interval of values of π_0 , $I = I(p^*, n_1 = 18, n_2 = 12)$, which verifies that the result obtained with the proposed Bayesian method for test (1), with $p_0 = \frac{1}{2}$ and $\pi(p_1, p_2) = 1, p_1, p_2 \in (0, 1)$, using a value $\pi_0 \in I$, is the same as the result obtained with Pearson's χ^2 classical test (see the following enclosed summary of results). Hence, there exists an accord between both methods. Notwithstanding, there are also values of p^* , for example $p^* = 0.015$, such that this is not possible.

The obtained results are summarized in Table 4.

Table 4
Summary of results for 2×2 tables with $n_1 = 18$ and $n_2 = 12$

$p^* \in$	(0.46, 0.513)	(0.087, 0.143)	(0.045, 0.052)	(0.0095, 0.0138)
$\varepsilon \in$	(0.221, 0.23)	(0.353, 0.4)	(0.453, 0.462)	(0.5528, 0.5675)
$\pi_0 \in$	(0.153, 0.167)	(0.391, 0.506)	(0.643, 0.673)	(0.893, 0.914)

Moreover, it can be verified that the value of π_0 , and thereby the value of ε , such that the previous reconciliation between both methods is possible, decreases when p^* increases. Also, it can be checked that the value of π_0 computed by means of expression (7) does not always exist, and when it exists this value does not always belong to the interval of values that allow the reconciliation between both methods to be achieved.

In the general situation with fixed n_1, n_2 , and p^* , if we denote by means of

$$\ell_1 = \ell_1(p^*, n_1, n_2) = \max_{(a,c), p > p^*} \eta(\eta + 1)^{-1},$$

$$\ell_2 = \ell_2(p^*, n_1, n_2) = \min_{(a,c), p \leq p^*} \eta(\eta + 1)^{-1},$$

and p^* satisfies that $\ell_1 < \ell_2$, then there exists an interval of values of π_0 , $I = I(p^*, n_1, n_2) = (\ell_1, \ell_2)$, such that the result obtained with the developed Bayesian method to test (1), using a value of $\pi_0 \in I$, is the same conclusion obtained with the Pearson's χ^2 classical method.

It is clear that the existence of values of p^* which satisfy the sufficient condition that ensures the accord between both methods depends on the increasing tendency that we can observe (see Fig. 1) in the functional relationship that exists between both statistics, $\Lambda = h(\eta)$, although this relationship is not strictly monotonous.

Therefore, the reconciliation is possible in that sense.

4. $r \times s$ Tables

In the following, we will generalize the previously obtained results to the situation of $r \times s$ tables. To do this, we suppose that independent random samples are drawn from r sufficiently large populations, and their each member belongs to one and only one of the s classes C_1, \dots, C_s . The sample number $i, i = 1, \dots, r$, is of size n_i and yields n_{ij} individuals in the category $C_j, j = 1, \dots, s$.

The situation is displayed in Table 5.

Let $X_i, i = 1, \dots, r$, be independent multinomial random variables, $MB(n_i, \mathbf{p}_i)$, with $\mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta$, where $\Theta = \{\mathbf{p} = (p_1, \dots, p_s) \in (0, 1)^s, \sum_{i=1}^s p_j = 1\} \subset \mathbb{R}^{s-1}$. In this situation, we are going to suppose that we wish to test

$$H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0 \quad \text{versus} \quad H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j, \tag{8}$$

where $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ is an unknown value and $H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j$ means that at least one of them is different from \mathbf{p}_0 . Consider that our prior opinion about

Table 5
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

$(\mathbf{p}_1, \dots, \mathbf{p}_r)$ is given by means of the density $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \prod_{i=1}^r \pi(\mathbf{p}_i)$. Therefore, a mixed prior distribution is needed to test (8), namely

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r) = \pi_0 I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r) + (1 - \pi_0) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r),$$

π_0 being the prior probability assigned to the null hypothesis.

Then, the posterior probability of the null hypothesis, when the data of Table 5 has been observed, is

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

Consider $\alpha_i = (\alpha_{i1}, \dots, \alpha_{is})$, with $\alpha_{ij} > 0$ for all $j = 1, \dots, s$ and all $i = 1, \dots, r$. If we assign to each \mathbf{p}_i a Dirichlet prior distribution of parameter $\alpha_i, D(\alpha_i), i = 1, \dots, r$, (see Ghosh and Ramamoorthi, 2003, Ch. 3), namely,

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

then such posterior probability is

$$\left[1 + \prod_{j=1}^s p_{0j}^{-m_j} \frac{1 - \pi_0}{\pi_0} \prod_{i=1}^r \left\{ \frac{\Gamma(\sum_{j=1}^s \alpha_{ij})}{\prod_{j=1}^s \Gamma(\alpha_{ij})} \int_{\Theta} \prod_{j=1}^s p_{ij}^{n_{ij} + \alpha_{ij} - 1} d\mathbf{p}_i \right\} \right]^{-1}.$$

Therefore, the posterior probability of the null hypothesis, when the data of Table 5 has been observed, can be expressed in the following way,

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1}, \tag{9}$$

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

We can note that if we assign a uniform prior distribution on Θ to each $\mathbf{p}_i, i = 1, \dots, r$, then the posterior probability of the null hypothesis can be obtained evaluating expression (9) in

$$\eta = \prod_{j=1}^s p_{0j}^{-m_j} \Gamma(s)^r \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + 1)}{\prod_{i=1}^r \Gamma(n_i + s)} \right\}.$$

The posterior probability calculated in expression (9) depends on π_0 , the initial prior probability that we assign to the null hypothesis, $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$.

Following, if we denote by $\mathbf{P}_0 = (\mathbf{p}_0, \dots, \mathbf{p}_0) \in \Theta^r \subset \mathbb{R}^{r(s-1)}$ and $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_r) \in \Theta^r \subset \mathbb{R}^{r(s-1)}$, then $H_0 : \mathbf{P} = \mathbf{P}_0$ is the null hypothesis of test (8). Now, we are going to consider the more realistic precise hypotheses,

$$H_{0\varepsilon} : d(\mathbf{P}_0, \mathbf{P}) \leq \varepsilon \quad \text{versus} \quad H_{1\varepsilon} : d(\mathbf{P}_0, \mathbf{P}) > \varepsilon,$$

with an appropriate metric d and a value of $\varepsilon > 0$ sufficiently small.

We propose to use $B(\mathbf{P}_0, \varepsilon) = \{\mathbf{P} \in \Theta^r, \sum_{i=1}^r \sum_{j=1}^{s-1} (p_{ij} - p_{0j})^2 \leq \varepsilon^2\}$. Then, applying the method of Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002), we can use $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \pi(\mathbf{P})$, our opinion about \mathbf{P} , to calculate π_0 by means of averaging, $\pi_0 = \int_{B(\mathbf{P}_0, \varepsilon)} \pi(\mathbf{P}) d\mathbf{P}$. We can observe that if a uniform prior distribution on Θ is assigned to each $\mathbf{p}_i, i = 1, \dots, r$, then

$$\pi_0 = \frac{\pi^{\frac{r(s-1)}{2}} \varepsilon^{r(s-1)}}{\Gamma(\frac{r(s-1)}{2} + 1)},$$

the volume of the sphere of radius ε in $\mathbf{R}^{r(s-1)}$, for ε sufficiently small.

From a classical viewpoint and considering Pearson's χ^2 test statistic,

$$\Lambda = \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i p_{0j}} - N,$$

if we denote by means of κ the value of Λ calculated in the point which the observed data of Table 5 forms, that is to say, $\Lambda(n_{ij}, i = 1, \dots, r, j = 1, \dots, s) = \kappa$, then $\{\Lambda \geq \kappa\}$ is a possible critical region and the p -value is

$$p = P(\Lambda \geq \kappa | \mathbf{p}_0) = P(\chi_{r(s-1)}^2 \geq \kappa).$$

Therefore, to search for a reconciliation between both results, the classical and the Bayesian, we can follow the same kind of reasoning developed in Sec. 3, since expression (9) has the same form as expression (2).

In conclusion, with fixed $n_i, i = 1, \dots, r$ and p^* , if we denote by means of

$$\ell_1 = \ell_1(p^*, n_1, \dots, n_r) = \max_{(n_{ij}), p > p^*} \eta(\eta + 1)^{-1}, \tag{10}$$

$$\ell_2 = \ell_2(p^*, n_1, \dots, n_r) = \min_{(n_{ij}), p \leq p^*} \eta(\eta + 1)^{-1}, \tag{11}$$

and p^* satisfies that $\ell_1 < \ell_2$, then there is an interval of values of $\pi_0, I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$, such that the result obtained with the proposed Bayesian method to test (8), using a value of $\pi_0 \in I$, is the same conclusion obtained when we use Pearson's χ^2 classical method.

Therefore, the accord is possible in this sense.

5. Conclusions

The posterior probability of the null hypothesis of homogeneity of independent multinomial populations in tables $r \times s$, when \mathbf{p}_0 is known for a mixed prior distribution that assigns an initial probability π_0 to $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$ and distributes of a continuous way the remaining probability in the points of the alternative hypothesis by means of a Dirichlet prior density, can be expressed as

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1},$$

where η is a statistic that measures the strength of the evidence in support of the more likely hypothesis, $\Lambda = h(\eta)$ is the test statistic for Pearson's χ^2 Classical method, and $h: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a nonmonotonous function of increasing tendency.

Fixing $n_i \in \mathbb{N}$, $i = 1, \dots, r$ and $p^* \in (0, 1)$, $\ell_1 < \ell_2$, where ℓ_1 and ℓ_2 are defined in expressions (10) and (11), respectively, gives a sufficient condition by which the reconciliation between both methods is possible. That is to say, if p^* satisfies that $\ell_1 < \ell_2$, then for some value of ε such that $\pi_0 = \pi_0(\varepsilon) \in (\ell_1, \ell_2)$, the p -value, p , verifies that $p > p^*$ and $[1 + \frac{1-\pi_0}{\pi_0}\eta]^{-1} > \frac{1}{2}$, or that $p \leq p^*$ and $[1 + \frac{1-\pi_0}{\pi_0}\eta]^{-1} \leq \frac{1}{2}$, whatever $(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s)$, the point that the observed data of Table 5 forms, is.

The existence of values p^* that satisfy such sufficient condition depends on the functional relationship, in terms of h , that exists between the statistics Λ and η . Thereby, the reconciliation between both methods is possible in that sense.

For example, for 2×2 tables with $n_1 = 18$ and $n_2 = 12$, when $p^* = 0.1$ the accord is obtained for $\varepsilon \in (0.353, 0.4)$.

The generalization of the previous results for the problem to test the homogeneity of independent multinomial populations when \mathbf{p}_0 is unknown, or with functional form known, $\mathbf{p}_0 = \mathbf{p}(\omega)$, is possible following a similar reasoning.

We are studying some robustness properties of the Bayes procedure for the ε -contaminated class of priors and we have partial results.

References

- Ghosh, J. K., Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Barcelona: Springer.
- Gómez-Villegas, M. A., Maín, P., Sanz, L. (2002). A suitable bayesian approach in testing point null hypothesis: some examples revisited. *Commun. Statist. Theor. Meth.* 31(2):201–217.
- Gómez-Villegas, M. A., Sanz, L. (2000). ε -contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statist. Probab. Lett.* 47:53–60.
- Howard, J. V. (1998). The 2×2 Table: A Discussion from a Bayesian Viewpoint. *Statist. Sci.* 13(4):351–367.
- Lee, P. M. (1997). *Bayesian Statistics: An Introduction*. London: Arnold.
- Lindley, D. V. (1988). Statistical inference concerning Hardy-Weinberg equilibrium. *Bayesian Statist.* 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* 4:139–167.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.* 5(50):157–175.
- Quintana, F. A. (1998). Nonparametric Bayesian analysis for assessing homogeneity in $k \times l$ contingency tables with fixed right margin totals. *J. Amer. Statist. Assoc. Theor. Meth.* 93(443):1140–1149.