Rev Mat Complut (2010) 23: 19–35
DOI 10.1007/s13163-009-0005-z

$r \times s$ tables from a Bayesian viewpoint

M.A. Gómez-Villegas · B. González-Pérez

Received: 30 June 2008 / Accepted: 13 January 2009 / Published online: 6 November 2009

Abstract The display of the data by means of contingency tables is used for discussing different approaches to statistical inference. 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 consists of assigning an appropriate prior mass, π_0, to the null and spreading the remainder, $1 - \pi_0$, over the alternative according to a density function. With this method, it is possible to prove a theorem which shows when the p-value and the posterior probability can give rise to the same conclusion.

Keywords Contingency tables · p-values · Posterior probabilities · Reconciliation

Mathematics Subject Classification (2000) 62F15

1 Introduction

The $r \times s$ table is used for discussing different approaches to statistical inference. For example, suppose that we want to investigate the behavior of a characteristic Y common to r large populations. To get information about Y, independent random samples, $(Y_{i1}, \dots, Y_{in_i})$, $i = 1, \dots, r$, $\sum_{i=1}^{r} n_i = N$, are drawn, respectively, from each population. Our objective is to test if the data gives enough evidence to reject the null hypothesis of homogeneity, that is, we want to decide if r populations have a

---

M.A. Gómez-Villegas (✉) · B. González-Pérez
Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, 28040 Madrid, Spain
e-mail: ma_gv@mat.ucm.es

B. González-Pérez
e-mail: beatrizg@mat.ucm.es

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

	Class 1	Class 2	...	Class s	Class $s+1$	Total
Sample 1	n_{11}	n_{12}	...	n_{1s}	$n_1 - \sum_{j=1}^{s} n_{1j}$	n_1
Sample 2	n_{21}	n_{22}	...	n_{2s}	$n_2 - \sum_{j=1}^{s} n_{2j}$	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Sample r	n_{r1}	n_{r2}	...	n_{rs}	$n_r - \sum_{j=1}^{s} n_{rj}$	n_r
Total	m_1	m_2	...	m_s	$N - \sum_{j=1}^{s} m_j$	N

common distribution $F(y)$. To do this, we divide the common sample space into an arbitrary number, $s+1$, of exclusionary classes, C_j, $j = 1, \dots, s+1$. Now, we denote by n_{ij} the observed frequency in C_j $(j = 1, \dots, s+1)$ of the sample i $(i = 1, \dots, r)$. The data is displayed in Table 1.

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 p_{ij} of elements belonging to C_j $(j = 1, \dots, s+1)$ is the same in all the populations $(i = 1, \dots, r)$, that is to say, $H_0 : p_{1j} = \cdots = p_{rj}$, for any $j = 1, \dots, s+1$. Three cases can be considered for testing the homogeneity of r multinomial distributions. If we denote by p_{0j} the common value under the null H_0 of the proportion corresponding to the cell j $(j = 1, \dots, s+1)$, such vector of proportions can be known, unknown or with a functional form depending on $q < s+1$ parameters.

There are of course a number of variations, and many approaches to this problem. Some important Bayesian references are [26, 31], [24, pp. 152–154] and [29]. Furthermore, this is one of the more elemental problems that can be used to show clear differences between frequentist and Bayesian approaches, see [10], and also between different types of frequentist analysis.

Lower bounds on Bayes factors and posterior probabilities in favour of point null hypotheses, H_0, have been discussed in the literature. The startling feature of these results is that Bayesian and frequentist procedures can give rise to different decisions, see [4, 31] and [5], among others. In most of the Bayesian approaches the infimum of this Bayesian evidence measures of H_0, over a wide class of prior distributions, is considered and it is then established that the infimum is generally 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 $1/2$. On the other hand, [8] show that there is no discrepancy in the one-sided testing problem.

Other important references are [2, 3, 7, 9, 11–13, 15–21, 23–26, 32–34, 37–39].

Reference [10] compute lower bounds on Bayes factors in favour of the null hypothesis in multinomial tests of point null hypothesis $H_0 : \mathbf{p} = \mathbf{p}_0$ (a specified point), for two different classes of densities (conjugate priors with mean $\mathbf{p}_0$ and unimodal and symmetric priors about $\mathbf{p}_0$). Their general conclusion is that the lower bounds tend to be substantially larger than chi-squared p-values, raising serious questions concerning the routine use of moderately small p-values (e.g., 0.05) to represent significant evidence against the null hypothesis.

Reference [29] gives a Bayesian discussion of the homogeneity problem in 2×2 tables (see Table 2). He advocates for the more frequent use of unilateral tests,

Table 2 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

considering as hypotheses of interest $H_1 : p_2 < p_1$ and $H_2 : p_1 < p_2$, 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. He has to suppose independent Jeffreys priors on the two populations, that is, $\pi(p_1, p_2) \propto p_1^{-1/2}(1-p_1)^{-1/2} p_2^{-1/2}(1-p_2)^{-1/2}$, in order to get this result. Moreover, he introduces a conjugate family of priors which incorporate dependence between beliefs about the two populations.

In this paper, we consider in detail testing equality of proportions of independent multinomial distributions when the common proportions under the null are known. Next we indicate how a generalization can be obtained when such vector of proportions are unknown or with functional form known. Our approach to the problem of homogeneity consists of working directly with the simple null hypothesis and calculating its posterior probability. To do this, we follow the method used by [21, 22], based on assigning an appropriate initial mass π_0 to the null hypothesis and distributing the remaining probability in the points of the alternative with a prior density. Posterior probabilities of the null hypothesis are calculated with respect to a mixture of a point prior on the null and independent Dirichlet priors on the proportions.

As usual in the literature, the comparisons between frequentist and Bayesian methods are studied by searching prior distributions for which posterior probabilities and p-values are numerically equal. The innovation of this paper is to show when p-values for the chi-squared test of fit and posterior probabilities can give rise to the same decision in the following sense. Fixed a level of significance $\alpha \in (0, 1)$, when there exist an interval of values of π_0 (defined by $\ell_1 < \ell_2$), such that (whatever observed data), a Bayesian who rejects H_0 when $P(H_0|data) < 1/2$ (for convenience, this is not a restriction) by using $\pi_0 \in (\ell_1, \ell_2)$ (as the mass of the null for the mixed prior distribution to compute the posterior probability) reaches the same conclusion as a frequentist who uses α as level of significance?

Section 2 formulates the problem in a precise way and calculates an exact expression of the posterior probability that the proportion of elements belonging to C_j $(j = 1, \dots, s+1)$ is the same in the r populations, and equal to a known common value p_{0j}. In Sect. 3, Theorem 3.1 provides the required constructions of ℓ_1 and ℓ_2 and the data of [36] (see Table 3) is used to illustrate the procedure. Section 4 considers a possible change of the parameter space. Section 5 generalizes the results of Sect. 2 to different versions of the homogeneity testing problem. Section 6 gives a summary of conclusions.

Table 3 Pearson's example

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

2 A Bayesian approach

Let $X_i \sim MB(n_i, \mathbf{p}_i)$, $i = 1, \dots, r$, be independent multinomial random variables, with unknown cell probabilities $\mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta$, where

$$\Theta = \left\{ \mathbf{p} = (p_1, \dots, p_s) \in (0,1)^s, \sum_{i=1}^{s} p_j \leq 1 \right\}$$

and $p_{s+1} = 1 - \sum_{j=1}^{s} p_j$. In this situation, we wish to test

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}_0, \quad \text{versus} \quad H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j, \tag{1}$$

where $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ is a specified value and $p_{0s+1} = 1 - \sum_{j=1}^{s} p_{0j}$.

To develop a Bayesian analysis about an unknown parameter, $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_r)$, indicating the prior beliefs about $\mathbf{P}$ through a prior distribution of probability is necessary. It is clear that a mixed prior distribution is needed to test (1). Assume that our prior opinion about $\mathbf{P}$ under the alternative H_1 is given by means of the density $\pi(\mathbf{P}) = \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \prod_{i=1}^{r} \pi(\mathbf{p}_i)$. The main problem here is the selection of appropriate initial mass π_0 of the null H_0.

A precise null hypothesis is often better understood by a non statistician as an approximation to a situation where the parameter (vector) of interest resides in a very small ball in the parameter space. Let us to consider the hypotheses

$$\begin{aligned} &H_{0\delta} : \mathbf{P} \in B(\mathbf{P}_0, \delta) \quad \text{versus} \quad H_{1\delta} : \mathbf{P} \notin B(\mathbf{P}_0, \delta), \\ &B(\mathbf{P}_0, \delta) = \left\{ \mathbf{P} \in \Theta^r, \sum_{i=1}^{r} \sum_{j=1}^{s} (p_{ij} - p_{0j})^2 \leq \delta^2 \right\}, \\ &\mathbf{P_0} = (\mathbf{p}_0, \dots, \mathbf{p}_0), \quad \delta > 0. \end{aligned}$$

At this point, we want to remark that the suitable choice of δ, which depends on the problem we are dealing with, may be more intuitive for a non statistician than just selecting an arbitrary value of π_0 in the following sense. It is easier for a no statistician gives us a value of δ based on thinking about a small region $C(\delta)$ where θ_0 is included. That is to say, he tells us, for instance: "I agree $\theta = \theta_0$ when $\theta \in C(\delta)$ for $\delta = 0.3$". From a Bayesian viewpoint, we propose to use the following mixed prior distribution to test (1):

$$\begin{aligned} &\pi^*(\mathbf{P}) = \pi_0 I_{H_0}(\mathbf{P}) + (1 - \pi_0)\pi(\mathbf{P}) I_{H_1}(\mathbf{P}), \\ &\pi_0 = \pi_0(\delta) = \int_{B(\mathbf{P}_0, \delta)} \pi(\mathbf{P}) d\mathbf{P}. \end{aligned} \tag{2}$$

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. For the multivariate point null testing problem, [21, 22] give several reasons to justify the choice of π_0 as in (2), despite of the usual value taken for $\pi_0 = 1/2$. Basically the principal reason is that the Kullback-Leibler measure of discrepancy between π and π^*,

$$D(\pi^*|\pi) = \int_\Theta \pi(\mathbf{P}) \ln\left[\frac{\pi(\mathbf{P})}{\pi^*(\mathbf{P})}\right] d\mathbf{P},$$

goes to zero when δ goes to zero with this choice. This is a consequence of the following reasoning.

Remarking the differences between these two probability measures is important. For instance, if $\boldsymbol{\theta}$ is the parameter of interest, and μ and μ^* denote the probability measures defined by π and π^*, then

$$\mu(A) = \int_A d\pi(\boldsymbol{\theta}) = \int_A \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

and

$$\mu^*(A) = \int_A d\pi^*(\theta) = \begin{cases} \pi_0 + (1-\pi_0)\mu(A) & \text{if } \theta_0 \in A, \\ (1-\pi_0)\mu(A) & \text{if } \theta_0 \notin A. \end{cases}$$

It is easy to prove that μ is absolutely continuous with respect to μ^* ($\mu \ll \mu^*$), so the Radon-Nikodym derivative of μ with respect to μ^*, $\frac{d\mu}{d\mu^*}$, exists. Furthermore, it is straightforward to see that

$$\frac{d\mu}{d\mu^*}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in H_0, \\ \frac{1}{1-\pi_0} & \text{if } \boldsymbol{\theta} \notin H_0. \end{cases}$$

Thereby we can sort out the raised problem by defining the discrepancy between μ and μ^* as

$$D(\pi^*|\pi) = \int_\Theta \ln\left[\frac{d\mu}{d\mu^*}(\boldsymbol{\theta})\right] d\mu(\boldsymbol{\theta}),$$

because of $\mu \ll \mu^*$. Then $D(\pi^*|\pi) = -\ln(1-\pi_0)$.

Several comments are in order. First, when δ goes to zero then, according to (2), π_0 goes to zero too and consequently $D(\pi^*|\pi)$ goes to zero. This is a justification for choosing π_0 as in (2) and consequently the replacement of (3) by (4) is reasonable. Secondly, when we use $\pi_0 = 1/2$ instead of using the value of π_0 given in (2), the discrepancy between μ and μ^* is perhaps a high discrepancy, $D(\pi^*|\pi) = 0.693$. Finally, the suitable choice of δ, which depends on the problem we are dealing with, may be more intuitive than just selecting an arbitrary value of π_0.

Therefore, from this way the statistician is able to determine a value of δ and uses this value to compute π_0 as in (2) and develop a Bayesian test based on the precise hypothesis $H_0 : \mathbf{P} = \mathbf{P}_0$, by using a mixed prior distribution $\pi^*(\mathbf{P})$. In this sense our theorems are going to be formulated in terms of π_0 and are independent of the possible way of its selection by δ (although δ may be a useful tool in practical situations where the statistician has difficulties to determine π_0.)

Furthermore, sometimes the choice of π_0 as in (2) allows to compute a exact expression of a lower bound for the posterior probability which is easy to implement. For instance, see [21] for the multivariate point null testing problem, between others. This is not the case of this paper because of we are going to work with the posterior probability for a concrete mixed prior distribution. For lower bounds and numerical comparisons between frequentist and Bayesian approaches with $r \times s$ tables, see [18].

Consider $\alpha_i = (\alpha_{i1}, \dots, \alpha_{i(s+1)})$, with $\alpha_{ij} > 0$ for all $j = 1, \dots, s+1$ and all $i = 1, \dots, r$. If we assign to each $\mathbf{p}_i$ a Dirichlet prior distribution of parameter α_i, $D(\alpha_i)$, $i = 1, \dots, r$ (see [14, Chap. 3]), namely,

$$\pi(\mathbf{p}_i) = \frac{\Gamma(\sum_{j=1}^{s+1} \alpha_{ij})}{\prod_{j=1}^{s+1} \Gamma(\alpha_{ij})} \prod_{j=1}^{s+1} p_{ij}^{\alpha_{ij}-1},$$

$\mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta$ and $p_{i(s+1)} = 1 - \sum_{j=1}^{s} p_{ij}$, $i = 1, \dots, r$.

Then

$$P(H_0|data) = \left[1 + \frac{1-\pi_0}{\pi_0}\eta\right]^{-1}, \tag{3}$$

where

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

is the Bayes factor. References [26] and [27] also considers Bayes test for multinomial problems with Dirichlet priors in other contexts.

One possibility is to assign a uniform prior distribution on Θ to each $\mathbf{p}_i$, $i = 1, \dots, r$, then

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

the volume of the sphere of radius δ in $\mathrm{R}^{r(s-1)}$, for δ sufficiently small and the posterior probability of the null hypothesis can be obtained evaluating expression (3) in

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

With this procedure, the decision of accepting or rejecting $H_0 : \mathbf{P} = \mathbf{P}_0$ depends on the size of the posterior probability given in expression (3).

Jeffreys procedure is not the only Bayesian testing method. Treating the problem as a decision problem by introducing an appropriate loss function and minimizing its expected value, is an alternative which does not require the use of a mixed prior (see [1] and [6], between others.) Absolute error loss, which corresponds with Neyman-Pearson theory, leads to 0–1 Bayes solutions, for instance see [30]. Furthermore, a mixed prior distribution allows to connect the level of significance with π_0, as we see in theorems. For this reasons and as the aim of the research presented here is to connect frequentist and Bayesian evidence, let us assume that H_0 is rejected when $P(H_0|data) < 1/2$. However, we can consider other Bayesian evidence measures, such as the Bayes factor η or a more restrictive threshold b^* instead of $1/2$.

Note that values of (δ which correspond with) $\pi_0 > \eta(\eta+1)^{-1}$ give $P(H_0|data) > 1/2$. This fact will be used in the proof of the theorem in Sect. 3. Moreover, $P(H_0|data) = (\eta+1)^{-1}$ for (δ such that) $\pi_0 = 1/2$.

For instance, to test $H_0 : p_1 = p_2 = 1/2$ with the data of Table 3 we obtain $\eta = 6.7265$ and, if $\pi_0 = 1/2$ ($\delta = 1/\sqrt{2\pi}$), then $P(H_0|a, c) = 0.1294$, so that H_0 is rejected. Moreover, to accept H_0 with Pearson's data, $\pi_0 > 0.8706$ ($\delta > 0.53905$).

3 Comparison with the p-value

From the frequentist viewpoint, instead of calculating the posterior probability of the null hypothesis, the p-value is computed. The idea is basically that or H_0 is false, or an event with very small probability has occurred.

The most popular approach is to use as frequentist measure of the evidence the discrepancy between the observed and expected values under H_0, under the terms of Pearson's χ^2 test statistic,

$$\Lambda = \sum_{i=1}^{r}\sum_{j=1}^{s+1} \frac{n_{ij}^2}{n_i p_{0j}} - N. \tag{4}$$

If λ_0 denotes the value of Λ evaluated at the observed data point of Table 1, then $\{\Lambda \geq \lambda_0\}$ is a possible critical region and the corresponding p-value is

$$p = p(data) = P(\Lambda \geq \lambda_0|\mathbf{p}_0) = P(\chi^2_{r(s-1)} \geq \lambda_0).$$

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 $p < p^*$, where $p^* \in (0, 1)$ being is a sufficiently small value (the significance level of the test α).

For $p_0 = 1/2$ and the data of Table 3 we obtain $\lambda_0 = 8.33$ and a p-value $p(a, c) = 0.0155$. Observe that H_0 is rejected for $p^* = 0.05$, but when $p^* = 0.01$ there is not enough evidence to reject it, and in that sense H_0 is accepted.

Our goal is to show when and how there is no discrepancy between the decisions derived from this frequentist and Bayesian approaches to test (1) for a specified significance level.

We can note that the Bayesian evidence measure given in expression (3) depends on $\pi_0 = \pi_0(\delta)$ given in (2). When it may be possible to calculate the value of π_0,

$$\pi_0 = \pi_0(data) = \frac{\eta p}{\eta p + 2p^* - p}, \tag{5}$$

which verifies

$$P(H_0|data) = \left[1 + \frac{1-\pi_0}{\pi_0}\eta(data)\right]^{-1} = \frac{p(data)}{2p^*},$$

the same conclusion would be reached with both approaches (if π_0 is as in (5), then $P(H_0|data) < 1/2$ when $p(data) < p^*$).

Let p^* be fixed and denote by means of

$$\ell_1 = \ell_1(p^*, n_1, \ldots, n_r) = \max_{tables,\, p > p^*} \eta(\eta + 1)^{-1},$$

$$\ell_2 = \ell_2(p^*, n_1, \ldots, n_r) = \min_{tables,\, p \leq p^*} \eta(\eta + 1)^{-1}.$$

Reference [35] revisit the issue of the apparent irreconcilability for testing a point null with iid observations from a multivariate normal distributions. They consider the threshold value of π_0 required for the smallest posterior probability (over appropriate families of priors on the alternative) and the p-value to coincide (although we can do it by $\pi_0^* = 1/2$ in the last equation, but in this case the goal is different on taking the same decision, typically, the p-value and the posterior probability are exactly equal). Reference [35] compute then the null distribution of the threshold (which depends on the data through the standardized $z = \sqrt{n}\bar{x}$) for shedding light on the issue of the criticality of the $\pi_0 = 1/2$ assumption made in the existing literature.

Naturally, the threshold value computed by (5) will be different from different data, since the p-value p and the Bayes factor η depends on the data. In this sense, we do not affirm that the procedure to obtain the agreement (same decision not numerical equality) has to be done by means of equaling both expressions, but the use of a value next to the result of this equalization can lead to the value of π_0 which we are looking for. For instance, if π_0 (non depending on the data) exists, we think that this value of π_0 (and accordingly δ) which achieves the agreement must decrease when p^* increases (note that this property which is verified by (5), is very intuitive). Furthermore, we ask for the following question, when there exists an interval of values (ℓ_1, ℓ_2) such that, whatever data, the decision derived from the posterior probability to text (1) (depending on $P(H_0|data) \leq$ or $\geq 1/2$, see expression (3)) computed with $\pi_0 \in (\ell_1, \ell_2)$ is the same taken by a frequentist who uses p^* as the level of significance to quantify the p-value?

To eliminate the dependence on the data, we consider the sample space formed by all of the possible $r \times s$ tables with n_i, $i = 1, \ldots, r$ fixed and known. For instance, in the context of Pearson's example (see Table 3), 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 ascending order carried out according to the values of η (see Fig. 1).

The following theorem gives a response to the question introduced previously.

Theorem 3.1 *Let n_i, $i = 1, \ldots, r$ and p^* be fixed and known. If p^* satisfies that $\ell_1 \leq \ell_2$, then there exists an interval of values of π_0, $I = I(p^*, n_1, \ldots, n_r) = (\ell_1, \ell_2)$, such that one and only one of the two following postulates is verified* (*whatever data*)

$$\text{“}p(data) > p^* \& P(H_0|data) > 1/2\text{”},$$

$$\text{“}p(data) \leq p^* \& P(H_0|data) \leq 1/2\text{”}.$$

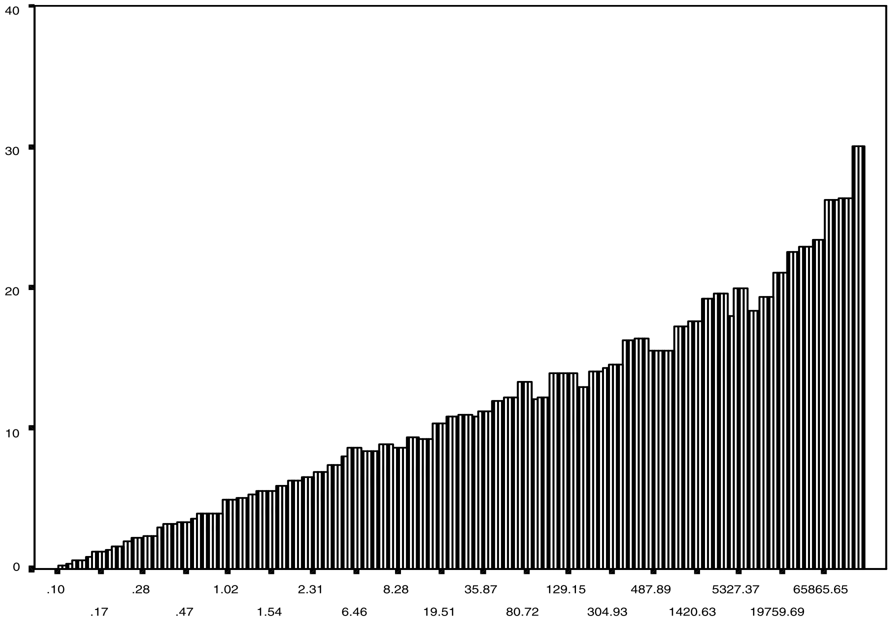

Fig. 1 Bars diagram $(\eta(a,c), \Lambda(a,c))$ to test $H_0 : p_1 = p_2 = 1/2$ using 2×2 tables with $n_1 = 18$, $n_2 = 12$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$. There is a non-monotonous function, $h : \mathrm{R}^+ \to \mathrm{R}^+$, such that $\Lambda = h(\eta)$

Proof Denote by κ the value of η in the observed data point. The posterior probability given in expression (3) verifies that

$$P(H_0|\kappa) = \left[1 + \frac{1-\pi_0}{\pi_0}\kappa\right]^{-1} > 1/2, \quad \text{when } \pi_0 > \pi^0(\kappa),$$

where $\pi^0(\kappa) = \kappa(\kappa+1)^{-1}$ is a strictly increasing function.

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 κ^*, κ_* and λ^* be

$$\kappa^* = \min_{tables, p \leq p^*} \eta, \kappa_* = \max_{tables, p > p^*} \eta, \lambda^* = \min_{tables, p \leq p^*} \Lambda.$$

Thereby, if p^* satisfies $\ell_1 \leq \ell_2$, the critical region $\{\Lambda \geq \lambda^*\} = \{\eta \geq \kappa^*\}$ and it is verified that $\kappa_* \leq \kappa^*$ and

$$\pi^0(\kappa_*) = \max_{tables, p > p^*} \pi^0(\eta) = \ell_1 \leq \ell_2 = \min_{tables, p \leq p^*} \pi^0(\eta) = \pi^0(\kappa^*).$$

Fixed $\pi_0 \in (\ell_1, \ell_2)$ and $(n_{11}, \ldots, n_{rs})$ a data point such that $\eta(data) = \kappa$, with $\kappa < \kappa_*$, then $\pi_0 > \ell_1 > \kappa(\kappa+1)^{-1}$ and $P\{\Lambda \geq \Lambda(data)|\theta_0\} > p^*$.

On the other hand, fixed $\pi_0 \in (\ell_1, \ell_2)$ and a data point such that $\kappa \geq \kappa^* \geq \kappa_*$, then $\pi_0 < \ell_2 \leq \kappa(\kappa+1)^{-1}$ and $P\{\Lambda \geq \Lambda(data)|\theta_0\} \leq p^*$. □

Table 4 Summary of results for 2 × 2 tables with $n_1 = 18$, $n_2 = 12$, $p_0 = 1/2$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$

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

Theorem 3.1 provides the required constructions of ℓ_1 and ℓ_2 (depending only on the level of significance and the sample sizes.) Therefore, $\ell_1 \leq \ell_2$ is a sufficient condition to achieve the same conclusion to test (1) with the p-value (with a fixed significance level p^*) and the posterior probability (with a value of δ such that $\pi_0(\delta) \in (\ell_1, \ell_2)$ in the corresponding mixed prior distribution). For these constructions to be useful, one needs to verify that $\ell_1 \leq \ell_2$. The existence of values p^* which satisfy this condition depends on the functional relationship that exists between the statistics Λ and η, $\Lambda = h(\eta)$, for some function $h : \mathrm{R}^+ \to \mathrm{R}^+$. It is easy to check that if h is an increasing function, $\ell_1 \leq \ell_2$ is always verified for any p^*. However, for 2 × 2 tables, if $n_1 = 18$ and $n_2 = 12$, to test (1) with $p_0 = 1/2$ and independent uniform priors, there exists a non-monotonous function $h : \mathrm{R}^+ \to \mathrm{R}^+$ for which $\Lambda = h(\eta)$ (see Fig. 1).

By means of an easy data analysis, we can check that there are values of p^*, for example $p^* = 0.5$, $p^* = 0.1$, $p^* = 0.05$, $p^* = 0.01$, 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 to test (1), with $p_0 = 1/2$ and a uniform prior distribution, using a value $\pi_0 \in I$, is the same as the result obtained with Pearson's χ^2 test of fit (see Table 4). Hence, there exists an agreement between both methods. Notwithstanding, there are also values of p^*, for example $p^* = 0.015$, such that this is not possible. Moreover, it can be verified that the value of π_0 (and thereby the value of δ), such that the previous reconciliation between both methods is possible, decreases when p^* increases. This fact is very intuitive. To illustrate this procedure, when the data point is Pearson's table (see Table 2), we can check (see Table 4) that, if $p^* = 0.01$, choosing a value of $\pi_0 = 0.9$, H_0 is accepted with both methods. However, if $p^* = 0.05$, choosing a value of $\pi_0 = 0.65$, H_0 is rejected with both approaches. Observe that $\pi_0 = 0.5$ correspond to $p^* = 0.1$.

We have mentioned that the reconciliation is possible for some p^* but not all. Observe that, for δ fixed the posterior probability in expression (3) is a decreasing function on η (the inverse of Bayes factor of H_0 relative to H_1) which can be used as a test statistic to built a critical region $\{\eta \geq \upsilon\}$, for some υ. In this case, the condition of Theorem 3.1 is verified for any p^* and the reconciliation is always possible. Furthermore, it can be checked numerically that the values of p^* and π_0 for which a reconciliation is possible when Λ is the test statistic are very similar to the corresponding calculations for η. When we detect a value of p^* for which the condition of Theorem 3.1 is not verified, is due to its associate critical region can not translate into $\{\eta \geq \upsilon\}$, for some υ (see Fig. 1). The idea is that the agreement is reached when the discrepancy between the two statistics involved in can be eliminated.

4 Change of the parameter space

As we have indicated, sometimes the statistician have difficulties for determining the prior mass π_0 in the mixed prior distribution. In this circumstances thinking about δ (the radius of a sphere centered in the null point) may be more convenient for this purpose and then computing $\pi_0 = \pi_0(\delta)$ as in (2). It has been already indicated that one possibility is to assign a uniform prior distribution on Θ to each $\mathbf{p}_i$, $i = 1, \ldots, r$. In this case, $\pi_0(\delta)$ is the volume of the sphere of radius δ in $\mathrm{R}^{r(s-1)}$, $B(\mathbf{P}_0, \delta)$, for δ sufficiently small. When the value of δ is not sufficiently small, $B(\mathbf{P}_0, \delta)$ is not contained in the parameter space $[0, 1]^{r(s-1)}$. This situation can be a nuisance in practice when we want to calculate the value of π_0 corresponding to δ. The computational difficulty is higher when the parameter dimension space increases.

In $r \times 2$ tables this computational problem can be eliminated by means of a change of the parameter space which transforms $[0, 1]^r$ in R^r. Instead of using independent uniform prior distributions, another suggestion is to consider independent Haldane's prior distributions [28], which is the following improper density:

$$\pi(p_1, \ldots, p_r) \propto \prod_{i=1}^{r} p_i^{-1}(1 - p_i)^{-1},$$

where p_i is the proportion of successes in the population i $(i = 1, \ldots, n)$. It is easy to check (by the usual change of variable argument) that Haldane's prior is equivalent to a prior uniform in the log-odds $\theta_i = \log[p_i(1 - p_i)^{-1}] \in \mathrm{R}$ $(i = 1, \ldots, r)$, $\pi(\theta_1, \ldots, \theta_r) \propto 1$.

Therefore, with this change of the parameter space, the null hypothesis in (1) is

$$H_0 : \theta_1 = \cdots = \theta_r = \theta_0,$$

where $\theta_0 = \log[p_0/(1 - p_0)^{-1}] \in \mathrm{R}$ is a known value.

Note that when an improper prior is used in testing, π_0 (the prior probability of $H_{0\delta}$) has no meaning whatsoever. Any multiple of the prior density, and hence that of π_0 serves the same purpose in deriving the posterior probabilities. However, whenever we observe at least one success and one failure in each population, the posterior probability which is obtained using independent Haldane's priors is proper. For instance, for 2×2 tables, when $n_1 = 18$ and $n_2 = 12$, there exists 187 tables which verifies the previous condition. On the other hand, thinking more carefully about the prior distribution and choosing a proper density for $\pi(\theta)$ which models the initial opinion in a more precise way are needed.

Let us to use as a Bayesian measure of evidence

$$P(H_0 | n_{11}, \ldots, n_{r2}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_H \right]^{-1},$$

where

$$\eta_H(n_{11}, \ldots, n_{r2}) = \frac{\int_{\mathrm{R}^r} f(n_{11}, \ldots, n_{r2} | \theta_1, \ldots, \theta_r) d\theta_1, \ldots, \theta_r}{f(n_{11}, \ldots, n_{r2} | \theta_0, \ldots, \theta_0)}$$

Table 5 Summary of results for 2 × 2 tables with $n_1 = 18$, $n_2 = 12$, $p_0 = 1/2$ and $\pi(p_1, p_2) \propto p_1^{-1}(1-p_1)^{-1}p_2^{-1}(1-p_2)^{-1}$

$p^* \in$	$(0.46, 0.51)$	$(0.087, 0.143)$	$(0.045, 0.052)$	$(0.006, 0.0094)$
$\delta \in$	$(0.501, 0.507)$	$(0.5473, 0.5536)$	$(0.5585, 0.5602)$	$(0.5636, 0.5639)$
$\pi_0 \in$	$(0.788, 0.807)$	$(0.941, 0.963)$	$(0.981, 0.986)$	$(0.9981, 0.999)$

$$= p_0^{-m_1}(1-p_0)^{-m_2}\prod_{i=1}^{r}\frac{\Gamma(n_{i1})\Gamma(n_{i2})}{\Gamma(n_i)},$$

$\pi_0 \in (0, 1)$ and δ such that

$$\pi_0 = \int_{B((\theta_0,\ldots,\theta_0),\delta)} d\theta_1 \ldots d\theta_r = \frac{\pi^{\frac{r}{2}}\delta^r}{\Gamma(\frac{r}{2}+1)},$$

the volume of the sphere of radius δ in R^r.

We can note that, when we use independent uniform prior distributions over the parameters $\mathbf{p}_i$, $i = 1, \ldots, r$, the posterior probability of the null depends on the statistic

$$\eta_U(n_{11}, \ldots, n_{r2}) = p_0^{-m_1}(1-p_0)^{-m_2}\prod_{i=1}^{r}\frac{\Gamma(n_{i1}+1)\Gamma(n_{i2}+1)}{\Gamma(n_i+2)} < \eta_H.$$

Consequently, $P_U(H_0|n_{11}, \ldots, n_{r2}) > P_H(H_0|n_{11}, \ldots, n_{r2})$.

Table 5 calculates some values of δ which reach agreement in the terms of the Theorem 3.1 to test (1) when $n_1 = 18$, $n_2 = 12$, $p_0 = 1/2$ and independent Haldane's prior distributions are assigned. We can observe that these values of π_0 are higher than the corresponding values when independent uniform priors are used. Moreover, by means of an easy data analysis it is easy to check that when $\pi_0 = 1/2$, $P_H(H_0|a, c) < 1/2$, whatever (a, c) may be, even more, $P_H(H_0|a, c) < p(a, c)$. Thereby, as $P_H(H_0|a, c)$ is an increasing function of π_0, to increase the value of π_0 is needed to reach agreement for p^* fixed.

5 Generalizations

The obtained results can be generalized when $\mathbf{p}_0$ is unknown or with known functional form by choosing an appropriate mixed prior distribution.

The most typical situation in homogeneity testing problem is when $\mathbf{p}_0$ is unknown. In this case, we only want to test if r populations have the same distribution which can be any one. Therefore the problem of interest is to test the hypothesis

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r, \quad \text{versus} \quad H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j. \tag{6}$$

The usual frequentist approximation uses Pearson's χ^2 test statistic

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

When the value of Λ in the observed data point is λ_0 then $\{\Lambda \geq \lambda_0\}$ is a possible critical region and

$$p = p(data) = P(\chi^2_{(r-1)(s-1)} \geq \lambda_0)$$

is the observed p-value.

Theorem 3.1 holds for this case, too, since the posterior probability of the null is monotonically decreasing in the inverse of Bayes factor using an appropriate mixed prior distribution. One option is

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

where $\pi(\mathbf{P}) = \prod_{i=1}^{r}\pi(\mathbf{p}_i)$, $\pi_0 = \int_{C(\delta)}\pi(\mathbf{P})d\mathbf{P}$, $C(\delta) = \bigcup_{\mathbf{p}_0\in\Theta} B(\mathbf{P}_0, \delta)$.

Assume that $\pi(\mathbf{p}_i)$ have a Dirichlet density with parameter $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{i(s+1)})$, $i = 0, 1, \ldots, r$ ($\alpha_{ij} > 0$, $j = 1, \ldots, s+1$). Then

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

where

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

For calculations, testing (6) using 2×2 with $n_1 = 18$, $n_2 = 12$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$, $\pi(p_0) = I_{(0,1)}(p_0)$ is considered. By means of an easy data analysis, we can check that there is not agreement when $p^* = 0.5$, $p^* = 0.1$, $p^* = 0.05$, $p^* = 0.01$. However, for $p^* \in (0.0635, 0.0637)$ o $p^* \in (0.008, 0.0085)$ we can find an interval of values of π_0, $I = I(p^*, n_1 = 18, n_2 = 12)$, where the sufficient condition of the natural extension of Theorem 3.1 is verified (see Table 6). The existence of values p^* for which the agreement is possible depends on the functional relationship, in terms of h that exists between the statistics Λ and η (see Fig. 2).

Another important problem is when $\mathbf{p}_0 = \mathbf{p}(\boldsymbol{\omega})$, with $\mathbf{p}$: $\Omega \to \Theta$, being

$$\Omega = \left\{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q), \mathbf{p}(\boldsymbol{\omega}) = (p_1(\boldsymbol{\omega}), \ldots, p_s(\boldsymbol{\omega})) \in \Theta\right\} \subset \mathrm{R}^q$$

Table 6 Summary of results for 2 × 2 tables with $n_1 = 18$, $n_2 = 12$, p_0 unknown and $\pi(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$, $\pi(p_0) = I_{(0,1)}(p_0)$

$p^* \in$	$(0.0635, 0.0637)$	$(0.008, 0.0085)$
$\delta \in$	$(0.222, 0.223)$	$(0.322, 0.325)$
$\pi_0 \in$	$(0.665, 0.667)$	$(0.738, 0.744)$

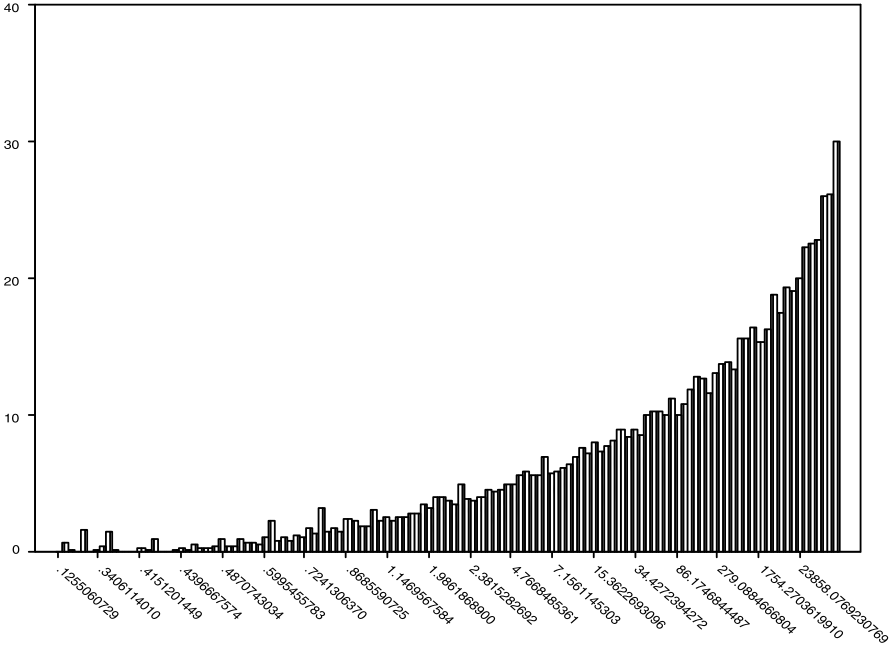

Fig. 2 Bars diagram $(\eta(a, c), \Lambda(a, c))$ to test $H_0 : p_1 = p_2 (= p_0$ unknown) using 2 × 2 tables with $n_1 = 18$, $n_2 = 12$ and $\pi(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$, $\pi(p_0) = I_{(0,1)}(p_0)$. There is a non-monotonous function, $h : \mathrm{R}^+ \to \mathrm{R}^+$, such that $\Lambda = h(\eta)$

and $q < s$ fixed. In this context, we propose

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

where π_0 is the prior probability assigned to $H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}(\boldsymbol{\omega})$ and $\pi(\mathbf{p}_1, \ldots, \mathbf{p}_r) = \prod_{i=1}^{r} \pi(\mathbf{p}_i)$.

Theorem 3.1 holds for this case, too, since the posterior probability of the null is monotonically decreasing in the inverse of Bayes factor.

6 Conclusions

The posterior probability of the null hypothesis of homogeneity of independent multinomial populations in $r \times s$ tables, when $\mathbf{p}_0$ is known, for a mixed prior distribution that assigns an initial probability π_0 to $H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}_0$ and distributes in 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 $\eta = \eta(n_{11}, \ldots, n_{rs})$ 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 $h : \mathrm{R}^+ \to \mathrm{R}^+$ is a non-monotonous function. The Kullback-Leibler discrepancy justifies the choose of $\pi_0 = \pi_0(\delta)$ as in (2). The choice of π_0 is polemic, we think that this election as a function of δ can be a good idea to help us in his practical determination and interpretation. Furthermore, literature show how this choice leads us to explicit expressions of infimum of posterior probabilities that with other selections need to be approximated by simulation (see for example, [5, 19, 20, 38]). Anyway, Theorem 3.1 does not need of this particular choice and provides the construction of an interval of values of π_0 (independently of δ) where the agreement is achieved (depending only on the level of significance and the sample sizes.)

The innovation with respect to other works about comparing frequentist and Bayesian approaches in testing problem consists of investigating when the same decision is reached with both methods instead of comparing numerically the p-value with the posterior probability or the Bayes factor. Fixing n_i, $i = 1, \ldots, r$ and $p^* \in (0, 1)$, $\ell_1 < \ell_2$, where ℓ_1 and ℓ_2 are defined similarly as in Theorem 3.1, gives a sufficient condition by which the reconciliation between both methods is possible. That is to say, if p^* satisfies that $\ell_1 \leq \ell_2$, then to any δ such that $\pi_0 = \pi_0(\delta) \in (\ell_1, \ell_2)$, whatever the observed data point may be, it is verified that we accept with the two methods or we reject with the two methods.

The existence of values p^* which satisfy the 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 $p^* \in (0.087, 0.143)$ the agreement is obtained for $\pi_0 \in (0.391, 0.506)$, that is $\delta \in (0.353, 0.4)$.

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

References

1. Berger, J.O., Bernardo, J.M., Sun, D.: The formal definition of reference priors. Ann. Stat. **37**, 905–938 (2009)
2. Berger, J.O., Boukai, B., Wang, Y.: Unified Frequentist and Bayesian testing of a precise hypothesis. Stat. Sci. **12**(3), 133–160 (1997)
3. Berger, J.O., Boukai, B., Wang, Y.: Simultaneous Bayesian-Frequentist sequential testing of nested hypothesis. Biometrika **86**, 79–92 (1999)
4. Berger, J.O., Delampady, M.: Testing precise hypotheses (with discussion). Stat. Sci. **2**(3), 317–352 (1987)

5. Berger, J.O., Sellke, T.: Testing a point null hypothesis: the irreconcilability of p-values and evidence (with discussion). J. Am. Stat. Assoc. **82**, 112–139 (1987)
6. Bernardo, J.M.: Reference posterior distributions for Bayesian inference (with discussion). J. R. Stat. Soc. B **41**, 113–147 (1979)
7. Bernardo, J.M.: A Bayesian analysis of classical hypothesis testing (with discussion). In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.) Bayesian Statistics, pp. 605–647. University Press, Valencia (1980)
8. Casella, G., Berger, R.L.: Reconciling Bayesian and Frequentist evidence in the one-sided testing problem (with discussion). J. Am. Stat. Assoc. **82**, 1062–135 (1987)
9. Cox, D.R., Hinkcley, D.V.: Theoretical Statistics. Chapman & Hall, London (1974)
10. Delampady, M., Berger, J.O.: Lower bounds on Bayes factors for multinomial distributions, with applications to chi-squared tests of fit. Ann. Stat. **18**, 1195–1316 (1990)
11. Dickey, J.M., Lienz, B.P.: The weighted likelihood ratio, sharp hypothesis about chances, the order of a Markov chain. Ann. Stat. **41**, 214–226 (1970)
12. Edwards, W.L., Lindman, H., Savage, L.J.: Bayesian statistical inference for psychological research. Psychol. Rev. **70**, 193–248 (1963)
13. Ghosh, J.K., Mukerjee, R.: Non-informative priors (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.) Bayesian Statistics, vol. 4, pp. 195–210. University Press, Oxford (1992)
14. Ghosh, J.K., Ramamoorthi, R.V.: Bayesian Nonparametrics, pp. 193–248. Springer, Berlin (2003),
15. Gómez-Villegas, M.A., Gómez, E.: Bayes factor in testing precise hyphotheses. Commun. Stat., Theory Methods **21**, 1707–1715 (1992)
16. Gómez-Villegas, M.A., González-Pérez, B.: Bayesian analysis of contingency tables. Commun. Stat., Theory Methods **34**(8), 1743–1754 (2005)
17. Gómez-Villegas, M.A., González-Pérez, B.: A condition to obtain the same decision in the homogeneity testing problem from the Frequentist and Bayesian point of view. Commun. Stat., Theory Methods **35**, 2211–2222 (2006)
18. Gómez-Villegas, M.A., González-Pérez, B.: ε-contaminated priors in contingency tables. Test **17**, 163–178 (2008)
19. Gómez-Villegas, M.A., González-Pérez, B.: The multivariate point null testing problem: a Bayesian discussion. Stat. Probab. Lett. **78**, 3070–3074 (2008)
20. Gómez-Villegas, M.A., Maín, P., Sanz, L.: A suitable Bayesian approach in testing point null hypothesis: some examples revisited. Commun. Stat., Theory Methods **31**(2), 201–217 (2002)
21. Gómez-Villegas, M.A., Maín, P., Sanz, L.: A Bayesian analysis for the multivariate point null testing problem. Technical Report Dpto. EIO-I, Universidad Complutense de Madrid 04-01
22. Gómez-Villegas, M.A., Maín, P., Sanz, L.: Contraste de hipótesis nula puntual multivariante para variables normales correladas. Actas del XXX Congreso Nacional de Estadística e Investigación Operativa y IV Jornadas de Estadística Pública (2007)
23. Gómez-Villegas, M.A., Maín, P., Sanz, L., Navarro, H.: Asymptotic relationships between posterior probabilities and p-values using the hazard rate. Stat. Probab. Lett. **66**, 59–66 (2004)
24. Gómez-Villegas, M.A., Sanz, L.: Reconciling Bayesian and Frequentist evidence in the point null testing problem. Test **7**(1), 207–216 (1998)
25. Gómez-Villegas, M.A., Sanz, L.: ε-contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. Stat. Probab. Lett. **47**, 53–60 (2000)
26. Good, I.J.: A Bayesian significance test for multinomial distributions. J. R. Stat. Soc. A **30**, 399–431 (1967)
27. Gunel, E., Dickey, J.: Bayes factors for independence in contingency tables. Biometrika **61**(3), 545–557 (1974)
28. Haldane, J.B.S.: A note on inverse probability. Proc. Camb. Philos. Soc. **28**, 55–61 (1931)
29. Howard, J.V.: The 2×2 table: a discussion from a Bayesian viewpoint. Stat. Sci. **13**(4), 351–367 (1998)
30. Hwang, J.T., Casella, G., Robert, C., Wells, M., Farrel, R.: Estimation of accuracy in testing (with discussion). Ann. Stat. **20**, 490–509 (1992)
31. Lindley, D.V.: A statistical paradox. Biometrika **44**, 187–192 (1957)
32. McCulloch, R.E., Rossi, P.E.: Bayes factors for non-linear hypothesis and likelihood distributions. Biometrika **79**, 663–676 (1992)
33. Micheas, A.C., Dey, D.K.: Prior and posterior predictive p-values in the one-sided location parameter testing problem. Sankhya, Ser. A **65**, 158–178 (2003)

34. Mukhopadhyay, S., DasgGupta, A.: A uniform approximation of Bayes solutions and posteriors: Frequentistly valid Bayes inference. Stat. Decis. **15**, 51–73 (1997)
35. Oh, H.S., DasGupta, A.: Comparison of the p-value and posterior probability. J. Stat. Plan. Inference **76**, 93–107 (1999)
36. Pearson, E.S.: The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. Biometrika **34**, 139–167 (1947)
37. Rubin, D.B.: Bayesianly justifiable and relevant frequency calculations for the applied statistician. Ann. Stat. **12**, 1151–1172 (1984)
38. Sellke, T., Bayarri, M.J., Berger, O.: Calibration of p values for testing precise null hypotheses. Ann. Stat. **55**(1), 62–71 (2001)
39. Spiegelhalter, D.J., Smith, A.F.M.: Bayes factors for linear and log-linear models with vague prior information. J. R. Stat. Soc., Ser. B **44**, 377–387 (1982)