Sociedad de Estadística e Investigación Operativa
Test (0000) *Vol.* **00**, *No.* 0, pp. 000–000

$\varepsilon-$Contaminated Priors in Contingency Tables

Miguel A. Gómez-Villegas* **and Beatriz González-Pérez**
Departamento de Estadística e Investigación Operativa I
Universidad Complutense de Madrid, Spain

Abstract

The $r \times s$ table is used for discussing different approaches to statistical inference. We develop a Bayesian procedure to test simple null hypotheses versus bilateral alternatives in contingency tables. We consider testing equality of proportions of independent multinomial distributions when the common proportions are known. A lower bound of the posterior probabilities of the null hypothesis is calculated with respect to a mixture of point prior on the null and a $\varepsilon-$contaminated prior on the proportions under the alternative. The resulting Bayes tests are compared numerically to Pearson's χ^2 in a number of examples. For the examined examples the lower bound and the p-value can be made close. The obtained results are generalized when the common proportions vector under the null is unknown or with known functional form.

Key Words: Contingency tables, $\varepsilon-$contaminated priors, test statistic, critical region, p-values, posterior probabilities, reconciliation.

AMS subject classification: 62F15.

1 Introduction

We suppose that independent random samples are drawn from two large populations, and each of their members 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 Table 1.

It is necessary to have a quantitative measure of the strength of the evidence that the data gives in support or in rejection of the hypothesis that

*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

Received: Month 0000; Accepted: Month 0000

the proportion of successes in the first population, p_1, is equal to the proportion of successes in the second population, p_2. Therefore, the parameter of interest for the problem of homogeneity outlined is $\theta = (p_1, p_2)$.

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

To develop a Bayesian analysis concerning an unknown parameter, θ, it is necessary to indicate the prior beliefs about θ through a prior distribution of probability. However, the case in which the initial information can be expressed in terms of a concrete probability distribution is not usual, due to the fact that, frequently, this prior information is diffuse. In some situations, due to this absence of precision, the prior information is expressed in terms of a class of distributions, Γ, in which all possible prior distributions for θ are included. Moreover, to compare the posterior probability of the null hypothesis of the Bayesian method with the *p-value* of the classical method it is reasonable to consider a class of prior distributions instead of a concrete distribution, given that the *p-value* does not use prior information.

An interesting way of describing prior opinions is to consider the class of $\varepsilon-$contaminated distributions given by

$$\Gamma = \{\pi = (1-\varepsilon)\, q_0 + \varepsilon q,\ \ q \in Q\}\,. \tag{1.1}$$

q_0 is a particular prior distribution, the prior that one would use in a Bayesian analysis with only one prior distribution. Q is the class of probability distributions that represents the possible (and reasonable) deviations from q_0. A fixed ε, with $0 < \varepsilon < 1$, represents the degree of contamination that we want to introduce into q_0.

Referring to the class Q, there are several possibilities to keep in mind. We are going to use the class of all probability distributions. Gómez-Villegas and Sanz (2000) use this class of distributions to study the problem of testing a simple null hypothesis from a Bayesian perspective. Huber (1973) and Sivaganesan (1988) also use this class although in another context. Berger and Berliner (1986), Sivaganesan and Berger (1989) and Berger (1985, 1994) give relevant information about other classes of contamination.

Section 2 formulates the homogeneity problem studied in a precise way and introduces the procedure to make up the mixed prior distribution. In section 3 the notation used in this paper is detailed. Section 4 calculates the infimum of the posterior probability that the proportion of successes in the first population is the same as in the second and both equal to a known value, when our initial opinion is in the class of $\varepsilon-$contaminated distributions. Relevant examples are contained in section 5. Section 6 compares the lower bound of the posterior probability obtained in section 4 with the *p-value*. Section 7 treats the case of $r \times s$ tables. Section 8 generalizes the proposed methodology when the common proportions vector under the null is unknown or with known functional form. Finally, section 9 contains some closing comments.

2 Formulation of the Problem

Consider X_i, i =1, 2, independent random variables distributed, respectively, $B\left(n_i, p_i\right)$, with $n_i \in \mathrm{N}$, i =1, 2, fixed and known, and suppose that we wish to test

$$H_0 : p_1 = p_2 = p_0 \text{ versus } H_1 : \exists i,\ p_i \neq p_0, \tag{2.1}$$

with p_0 known.

Moreover, suppose that our prior opinion about (p_1, p_2) is given by the density $\pi\left(p_1, p_2\right) \in \Gamma$, Γ being the class of $\varepsilon-$contaminated distributions given in expression (1.1).

Then, a mixed prior distribution is needed to test (2.1). We propose

$$\pi^*\left(p_1, p_2\right) = \pi_0 I_{H_0}\left(p_1, p_2\right) + \left(1-\pi_0\right)\pi\left(p_1, p_2\right) I_{H_1}\left(p_1, p_2\right), \tag{2.2}$$

where π_0 is the assigned prior mass to the null hypothesis. However, there is no rule to fix the value of π_0, usually $\pi_0 = \frac{1}{2}$ (see Robert, 2001, chapter 5).

Now, consider the more realistic precise hypothesis

$$H_{0\delta} : d\left(\left(p_0, p_0\right), \left(p_1, p_2\right)\right) \leq \delta \text{ versus } H_{1\delta} : d\left(\left(p_0, p_0\right), \left(p_1, p_2\right)\right) > \delta, \quad (2.3)$$

with a proper metric d and a value of $\delta > 0$ sufficiently small. Therefore, any point (p_1, p_2) such that $d\left(\left(p_0, p_0\right), \left(p_1, p_2\right)\right) \leq \delta$ is considered indistinguishable from (p_0, p_0).

An interesting discussion and a justification of this construction by using the Kullback-Leibler information measure about the change of (2.1) for (2.3) can be found in Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002).

Applying the method of Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002), introduced by Gómez-Villegas and Gómez (1992) and justified by Gómez-Villegas and Sanz (1998), we can use $\pi\left(p_1, p_2\right)$, our opinion about (p_1, p_2), and calculate π_0 by means of

$$\pi_0 = \int_{B((p_0,p_0),\delta)} \pi\left(p_1, p_2\right) dp_2 dp_1, \quad (2.4)$$

where $B\left(\left(p_0, p_0\right), \delta\right) = \left\{\left(p_1, p_2\right) \in (0,1) \times (0,1),\ \sum_{i=1}^{2} \left(p_i - p_0\right)^2 \leq \delta^2\right\}$, the sphere having center (p_0, p_0) and radius δ.

Therefore, the prior probability assigned to H_0 by means of $\pi^*\left(p_1, p_2\right)$ and assigned to $H_{0\delta}$ by means of $\pi\left(p_1, p_2\right)$ is the same, selecting an appropriate value of δ.

In the same way of Berger and Sellke (1987), we seek to minimize $P\left(H_0 | a, c\right)$ over Γ, the class of prior distributions given in expression (1.1). From (2.4) we have $\pi_0 = (1 - \varepsilon)\, \pi_{q_0}^0 + \varepsilon \pi_q^0$, where

$$\pi_{q_0}^0 = \int_{B((p_0,p_0),\delta)} q_0\left(p_1, p_2\right) dp_2 dp_1,\ \ \pi_q^0 = \int_{B((p_0,p_0),\delta)} q\left(p_1, p_2\right) dp_2 dp_1. \quad (2.5)$$

As it is indicated in Gómez-Villegas and Sanz (2000), a reason that justifies taking the infimum is that, for a small infimum, the null hypothesis must be refused according to the interpretation of the *p-value*. More reasons can be seen in Berger and Sellke (1987). Moreover, this development is similar to the one found in Casella and Berger (1987) that reconciles Bayesian and classical measures in a one-sided testing problem.

There is an extensive literature about the comparison between Bayesian and classical measures. Some important references, besides the ones already mentioned, are Edwards et al. (1963), Pratt (1965), Dickey and Lienz (1970), Cox and Hinkley (1974), DeGroot (1974), Bernardo (1980), Spiegelhalteer and Smith (1982), Rubin (1984), Ghosh and Mukerjee (1992), McCulloch and Rossi (1992), Mukhopadhyay and Dasgupta (1997), Berger et al. (1997, 1999), Oh and DasGupta (1999), Gómez-Villegas et al. (2004b).

3 Notation

We denote the likelihood function by

$$f\left(a, c|p_1, p_2\right) = \left(\begin{array}{c} n_1 \\ a \end{array}\right)\left(\begin{array}{c} n_2 \\ c \end{array}\right) p_1^a \left(1 - p_1\right)^{n_1 - a} p_2^c \left(1 - p_2\right)^{n_2 - c},$$

which is considered as a function of $\theta = (p_1, p_2)$ for the observed value of $(X_1, X_2) = (a, c)$, $a = 0, 1, \cdots, n_1$, $c = 0, 1, \cdots, n_2$. $m\left(a, c|\pi\right)$ denotes the marginal distribution of (X_1, X_2) with respect to the prior distribution $\pi \in \Gamma$ where

$$m\left(a, c|\pi\right) = \left(1 - \varepsilon\right) m\left(a, c|q_0\right) + \varepsilon m\left(a, c|q\right).$$

Assuming the existence of the posterior distributions $q_0\left(p_1, p_2|a, c\right)$ and $q\left(p_1, p_2|a, c\right)$, the posterior distribution of (p_1, p_2) given (a, c) with respect to $\pi \in \Gamma$ is

$$\pi\left(p_1, p_2|a, c\right) = \lambda\left(a, c\right) q_0\left(p_1, p_2|a, c\right) + \left(1 - \lambda\left(a, c\right)\right) q\left(p_1, p_2|a, c\right),$$

where $\lambda\left(a, c\right) = \frac{(1-\varepsilon) m(a,c|q_0)}{m(a,c|\pi)}$.

A classical measure of the evidence against the null hypothesis, which depends on the observations, is the *p-value*. If $T = T\left(X_1, X_2\right)$ is an appropriate statistic to test (2.3), for instance a sufficient statistic, then the *p-value* of the sample point (a,c) is

$$p\,(a,c) = \sup_{(p_1,p_2)\in H_{0\delta}} P\left\{\left|T\left(X_1,X_2\right)\right| > \left|T\left(a,c\right)\right|\ \left|\,(p_1,p_2)\right.\right\}.$$

In particular, to test (2.1), the *p-value* takes the form

$$p\,(a,c) = P\left\{\left|T\left(X_1,X_2\right)\right| > \left|T\left(a,c\right)\right|\ \left|\,(p_0,p_0)\right.\right\}.$$

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 $p\,(a,c) < \alpha$, $\alpha \in (0,1)$ being the *significance level* of the test.

In section 6 we consider two different *test statistics* to test (2.1).

4 Lower Bound for the Posterior Probability

In this section we obtain a lower bound for the posterior probability of the null hypothesis to test (2.1), for a prior distribution π^* given in expression (2.2) and π_0 computed according to (2.4). Theorem 1 establishes sufficient conditions in order to achieve the infimum of the posterior probability, for an arbitrary prior distribution $\pi \in \Gamma$.

Consider the hypothesis introduced in (2.1), an arbitrary prior distribution $\pi \in \Gamma$ as in (1) and a mixed prior distribution as in (2.2) with assigned mass to the null hypothesis in (2.3) according to (2.4). Then,

$$P\left(H_0|a,c\right) = \left[1 + \frac{1-\pi_0}{\pi_0}\frac{m\left(a,c|\pi\right)}{f\left(a,c|p_0,p_0\right)}\right]^{-1}.$$

To calculate a lower bound of the posterior probability of H_0, it is sufficient to compute an upper bound of $\frac{1-\pi_0}{\pi_0}m\left(a,c|\pi\right)$, when $\pi \in \Gamma$. By the construction of $\pi^*\left(p_1,p_2\right)$, π_0 depends on $q \in Q$ through π_q^0 given in expression (2.5). Then, that lower bound can be calculated as the supremum in $q \in Q$ of

$$\frac{1-\pi_0}{\pi_0}m\left(a,c|\pi\right) = \left[\frac{1}{\left(1-\varepsilon\right)\pi_{q_0}^0 + \varepsilon\pi_q^0} - 1\right]\left[\left(1-\varepsilon\right)m\left(a,c|q_0\right) + \varepsilon m\left(a,c|q\right)\right]. \quad (4.1)$$

With $\pi_{q_0}^0$ and π_q^0 given in expression (2.5), as the supremum of (4.1) when $q \in Q$ is always less than or equal to the product of

$$\sup_{q\in Q}\left[\frac{1}{(1-\varepsilon)\,\pi_{q_0}^0+\varepsilon\pi_q^0}-1\right]=\frac{1}{(1-\varepsilon)\,\pi_{q_0}^0}-1$$

and $\sup_{q\in Q}\left[(1-\varepsilon)\,m\,(a,c|q_0)+\varepsilon m\,(a,c|q)\right]$, where

$$\begin{aligned} m\,(a,c|q) &\propto \int_0^1\int_0^1 p_1^a\,(1-p_1)^b\,p_2^c\,(1-p_2)^d\,q\,(p_1,p_2)\,dp_2dp_1 \\ &\leq \sup_{(p_1,p_2)\neq(p_0,p_0)} p_1^a\,(1-p_1)^b\,p_2^c\,(1-p_2)^d\,. \end{aligned}$$

Then

$$P\,(H_0|a,c)\geq\left[1+\frac{1-(1-\varepsilon)\,\pi_{q_0}^0}{(1-\varepsilon)\,\pi_{q_0}^0}\eta_\varepsilon\,(a,c)\right]^{-1}, \tag{4.2}$$

where $\eta_\varepsilon\,(a,c)=(1-\varepsilon)\,\eta(a,c)+\varepsilon\frac{\left(\frac{a}{a+b}\right)^a\left(\frac{b}{a+b}\right)^b\left(\frac{c}{c+d}\right)^c\left(\frac{d}{c+d}\right)^d}{p_0^{a+c}(1-p_0)^{b+d}}$, with

$$\eta(a,c)=\frac{m\,(a,c|q_0)}{f\,(a,c|p_0,p_0)}.$$

The previous expression furnishes a lower bound for the posterior probability of the null hypothesis to test (2.1). Therefore, the first question to propose is, when is it possible to achieve the infimum computed in expression (4.2) by a distribution of the class Γ given in (1.1). The answer is given by the following theorem.

Theorem 4.1. *Let* $(\hat{p}_1,\hat{p}_2)$ *be the maximum likelihood estimator of* (p_1,p_2) *when* $(p_1,p_2)\in H_1$.

If $(\hat{p}_1,\hat{p}_2)\notin B\left((p_0,p_0),\delta\right)$ *and* $\int_{B((\hat{p}_1,\hat{p}_2),\rho)} f\,(a,c|p_1,p_2)dp_2dp_1$*, for fixed* ρ*, is approximated by* $\pi\rho^2 f\,(a,c|\hat{p}_1,\hat{p}_2)$*, then the distribution given by*

$$\tilde{\pi}\,(p_1,p_2)=(1-\varepsilon)\,q_0\,(p_1,p_2)+\varepsilon\tilde{q}\,(p_1,p_2)\,,$$

where $\tilde{q}(p_1, p_2)$ is uniform in $B((\hat{p}_1, \hat{p}_2), \rho)$, satisfies

$$inf_{\pi\in\Gamma} P_\pi(H_0|a,c) = P_{\tilde{\pi}}(H_0|a,c) = \left[1 + \frac{1-(1-\varepsilon)\pi^0_{q_0}}{(1-\varepsilon)\pi^0_{q_0}}\eta_\varepsilon(a,c)\right]^{-1}, \tag{4.3}$$

where $\pi^0_{q_0}$ and $\eta_\varepsilon(a,c)$ are both as in expressions (2.5) and (4.2), respectively.

Proof. For (4.1), we need to compute π_0 and $m(a,c|\tilde{\pi})$. Given that for $\tilde{\pi}$,

$$\pi_0 = \int_{B((p_0,p_0),\delta)} \tilde{\pi}(p_1,p_2)dp_2dp_1$$

$$= (1-\varepsilon)\int_{B((p_0,p_0),\delta)} q_0(p_1,p_2)dp_2dp_1 + \varepsilon\int_{B((p_0,p_0),\delta)} \tilde{q}(p_1,p_2)dp_2dp_1$$

$$= (1-\varepsilon)\pi^0_{q_0}$$

and $m(a,c|\tilde{\pi}) = (1-\varepsilon)m(a,c|q_0) + \varepsilon m(a,c|\tilde{q})$, where

$$m(a,c|\tilde{q}) = \int_0^1\int_0^1 f(a,c|p_1,p_2)\,\tilde{q}(p_1,p_2)\,dp_2dp_1$$

$$= \frac{1}{\pi\rho^2}\int_{B((\hat{p}_1\hat{p}_2),\rho)} f(a,c|p_1,p_2)\,dp_2dp_1 \approx f(a,c|\hat{p}_1,\hat{p}_2)$$

$$= \binom{n_1}{a}\binom{n_2}{c}\left(\frac{a}{a+b}\right)^a\left(\frac{b}{a+b}\right)^b\left(\frac{c}{c+d}\right)^c\left(\frac{d}{c+d}\right)^d,$$

then we obtain (4.3).

□

It is interesting to note that the real restriction in this theorem is $(\hat{p}_1, \hat{p}_2) \notin B\left((p_0, p_0), \delta\right)$, since in this situation the approximation of the integral is always possible by choosing a sufficiently small value of ρ and with $B\left((p_0, p_0), \delta\right) \cap B\left((\hat{p}_1, \hat{p}_2), \rho\right)$ being empty. If $(\hat{p}_1, \hat{p}_2) \in B\left((p_0, p_0), \delta\right)$, (4.2) becomes a strict inequality.

5 Examples

A possible initial distribution consists of assigning independent uniform prior distributions, that is to say,

$$q_0\left(p_1,\ p_2\right) = I_{(0,\ 1)}\left(p_1\right) I_{(0,\ 1)}\left(p_2\right).$$

In this situation, the lower bound of the posterior probability of the null hypothesis to test (2.1) can be obtained evaluating $\eta_\varepsilon\left(a, c\right)$ in expression (4.2) as

$$\eta_\varepsilon\left(a, c\right) = \left(1 - \varepsilon\right) \eta\left(a, c\right) + \varepsilon \frac{\left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b \left(\frac{c}{c+d}\right)^c \left(\frac{d}{c+d}\right)^d}{p_0^{a+c}\left(1 - p_0\right)^{b+d}}, \tag{5.1}$$

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

It can be observed that for n_1, n_2 and κ fixed, there are at most four 2×2 tables in the set $A_\kappa = \{(a, c),\ \eta\left(a, c\right) = \kappa\}$. For instance, when $p_0 = 1/2$ and a and b swap places as well as c and d. Moreover, $g\left(a, c\right) = \left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b \left(\frac{c}{c+d}\right)^c \left(\frac{d}{c+d}\right)^d$ is a constant function on A_κ, and for p_0 fixed, the function $p_0^{a+c}\left(1 - p_0\right)^{b+d}$ takes at most two different values on A_κ. In this situation

$$P\left(H_0|\kappa\right) \geq \left[1 + \frac{1 - \left(1 - \varepsilon\right)\pi_{q_0}^0}{\left(1 - \varepsilon\right)\pi_{q_0}^0}\eta_\varepsilon\left(\kappa\right)\right]^{-1}, \tag{5.2}$$

where $\eta_\varepsilon\left(\kappa\right) = \left(1 - \varepsilon\right)\kappa + \varepsilon \frac{g(\kappa)}{\min_{A_\kappa} p_0^{a+c}(1-p_0)^{b+d}}$.

A more general assignment consists of using independent beta prior distributions, that is to say,

$$q_0(p_1, p_2) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\alpha-1}(1-p_1)^{\beta-1} p_2^{\gamma-1}(1-p_2)^{\delta-1},$$

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

Then, the lower bound of the posterior probability of the null hypothesis to test (2.1) is obtained evaluating $\eta(a, c)$ in expression (5.1) as

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

6 Comparison with the Classical Method

In parametric testing of a simple null hypothesis, it is known that Bayesian and classical procedures can give rise to different decisions, see Lindley (1957), Berger and Sellke (1987) and Berger and Delampady (1987), among others. 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 then it is obtained that the infimum is substantially larger than the corresponding *p-value*. It is necessary to point out that in all these situations the assigned mass to the simple null hypothesis is $\frac{1}{2}$. On the other hand, Casella and Berger (1987) show that there is no discrepancy in the one-sided testing problem.

In most of the existing contributions a class of prior distributions is used. Gómez-Villegas et al. (2004a) use the class of the unimodal and symmetric distributions to show that the *p-values* and the posterior probabilities can be close for the multivariate point null testing problem. Our objective is verify that, to test (2.1), there is no discrepancy between the classical and Bayesian approaches when the prior is in the class of $\varepsilon-$contaminated distributions.

It can be observed that the lower bound of the posterior probability given in expression (5.2) depends on the statistic $\eta(a, c)$ given in expression (5.1), which can be used as the *test statistic* to get a *critical region* and calculate the *p-value*, $p(a, c)$, for the observed point (a, c) of the sample space.

Table 2: lower bounds of the posterior probability of $H_0 : p_1 = p_2 = \frac{1}{2}$*, for tables* (a, c) *with* $P\{\eta \geq \kappa_0 | (p_0, p_0)\}$ *close to 0.1, 0.05 and 0.01,* $q_0(p_1, p_2) = I_{(0,\ 1)}(p_1) I_{(0,\ 1)}(p_2)$ *and* $\varepsilon = 0.2$.

$P\{\eta \geq \kappa_0 \| (p_0, p_0)\}$	0.1063	0.0905	0.0533	0.0442	0.0118	0.0097
$\pi^0_{q_0} = 0.5$	0.1677	0.1594	0.1013	0.0825	0.0222	0.0182
$\pi^0_{q_0} = 0.35$	0.1052	0.0996	0.0617	0.0499	0.0131	0.0107
$\pi^0_{q_0} = 0.34$	0.1014	0.0961	0.0594	0.048	0.0126	0.0103
$\pi^0_{q_0} = 0.33$	0.0978	0.0926	0.0572	0.0462	0.0121	0.0099
$\pi^0_{q_0} = 0.32$	0.0942	0.0891	0.0549	0.0444	0.0116	0.0095
$\pi^0_{q_0} = 0.31$	0.0906	0.0858	0.0528	0.0426	0.0111	0.0091
$\pi^0_{q_0} = 0.3$	0.0871	0.0824	0.0507	0.0408	0.0106	0.0087
κ_0	1.025	1.064	1.794	2.07	7.68	8.28

If the point (a_0, c_0) is observed and $\eta(a_0, c_0) = \kappa_0$, then the probability that we would get a new value of η as big as or larger than κ_0 can be computed, when the experiment is repeated independently, once again randomly sampling n_1 subjects from population 1 and n_2 subjects from population 2.

Therefore, $\{\eta \geq \kappa_0\}$ is a possible *critical region*, and

$$p(a_0, c_0) = P\{\eta \geq \kappa_0 | (p_0, p_0)\} = \sum_{\eta(a,c) \geq \kappa_0} f(a, c | p_0, p_0)$$

$$= \sum_{\eta(a,c) \geq \kappa_0} \binom{n_1}{a} \binom{n_2}{c} p_0^{a+c} (1 - p_0)^{b+d} = p(\kappa_0) \tag{6.1}$$

is the *p-value*.

However, the usual Pearson's χ^2 classical method uses the random variable

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

as the *test statistic*.

In this situation, when the value of Λ at an observed point (a_0, c_0) is $\Lambda(a_0, c_0) = \lambda_0$, then the used evidence is the *p-value*

$$p(a_0, c_0) = P\{\Lambda \geq \lambda_0 | (p_0, p_0)\} = P\left(\chi_2^2 \geq \lambda_0\right) = e^{-\frac{\lambda_0}{2}}. \qquad (6.2)$$

Table 3: lower bounds of the posterior probability of $H_0 : p_1 = p_2 = \frac{1}{2}$*, for tables* (a, c) *with* $P\{\eta \geq \kappa_0 | (p_0, p_0)\}$ *close to 0.1, 0.05 and 0.01,* $q_0(p_1, p_2) = I_{(0,\ 1)}(p_1) I_{(0,\ 1)}(p_2)$ *and* $\varepsilon = 0$*.*

$P\{\eta \geq \kappa_0 \| (p_0, p_0)\}$	0.1063	0.0905	0.0533	0.0442	0.0118	0.0097
$\pi_{q_0}^0 = 0.5$	0.4938	0.4844	0.358	0.3257	0.1151	0.1077
$\pi_{q_0}^0 = 0.2$	0.1961	0.1902	0.1223	0.1077	0.0315	0.0293
$\pi_{q_0}^0 = 0.12$	0.1174	0.1135	0.0706	0.0618	0.0174	0.0162
$\pi_{q_0}^0 = 0.11$	0.107	0.104	0.0644	0.0563	0.0158	0.147
$\pi_{q_0}^0 = 0.1$	0.0978	0.094	0.0583	0.0509	0.0142	0.1324
$\pi_{q_0}^0 = 0.09$	0.0879	0.085	0.0522	0.0456	0.0127	0.0118
$\pi_{q_0}^0 = 0.08$	0.0782	0.075	0.0462	0.0403	0.0111	0.0104
$\pi_{q_0}^0 = 0.07$	0.0684	0.066	0.0403	0.0351	0.0096	0.009
κ_0	1.025	1.064	1.794	2.07	7.68	8.28

Our objective is to obtain an appropriate value of δ such that the values of the lower bound of the posterior probability given in expression (5.2) are close to the corresponding *p-values*.

Table 2 shows the values of the lower bound of the posterior probability of H_0 to test (2.1) given in expression (5.2), for some specific values of η and some $\pi_{q_0}^0$, when $p_0 = \frac{1}{2}$ and the initial opinion $q_0(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$ is contaminated with $\varepsilon = 0.2$.

We can observe that if we take an appropriate value of $\pi_{q_0}^0 = \frac{\pi_0}{1-\varepsilon} = \pi\delta^2$ (see Proof of Theorem 1), the values of the lower bounds are close to the respective *p-values* given in expression (6.1). For example, if we take $\pi_{q_0}^0 \in (0.3, 0.35)$, then for $\delta \in (0.31, 0.33)$ the lower bounds of the posterior probability are approximately equal to the *p-values*. Also, we can observe that when $\pi_{q_0}^0 = \frac{1}{2}$ there is more discrepancy between both measures.

The same study for $\varepsilon = 0$ is shown in Table 3. In this case, we can observe that if we take $\pi_{q_0}^0 \in (0.09, 0.11)$, then for $\delta \in (0.17, 0.19)$ the lower bounds of the posterior probability of H_0 are approximately equal to the *p-values*.

Table 4: lower bounds of the posterior probability of $H_0 : p_1 = p_2 = \frac{1}{2}$, for tables (a, c) with $P\{\Lambda \geq \lambda_0 | (p_0, p_0)\}$ close to 0.1, 0.05 and 0.01, $q_0(p_1,\ p_2) = I_{(0,\ 1)}(p_1) I_{(0,\ 1)}(p_2)$ and $\varepsilon = 0$.

$P\{\Lambda \geq \lambda_0 \vert (p_0, p_0)\}$	0.143	0.0868	0.052	0.0445	0.0138	0.0094
$\pi^0_{q_0} = 0.5$	0.6095	0.4938	0.358	0.3257	0.1077	0.085
$\pi^0_{q_0} = 0.2$	0.2807	0.1961	0.1223	0.1077	0.2931	0.0227
$\pi^0_{q_0} = 0.12$	0.1755	0.1174	0.0706	0.0618	0.0162	0.0125
$\pi^0_{q_0} = 0.11$	0.1617	0.107	0.0644	0.0563	0.0147	0.0113
$\pi^0_{q_0} = 0.1$	0.1478	0.0978	0.0583	0.0509	0.0132	0.0102
$\pi^0_{q_0} = 0.09$	0.1337	0.0879	0.0522	0.0456	0.0118	0.0091
$\pi^0_{q_0} = 0.08$	0.1195	0.0782	0.0462	0.0403	0.0104	0.008
$\pi^0_{q_0} = 0.07$	0.1051	0.0684	0.0403	0.0351	0.009	0.0069
λ_0	3.89	4.89	5.89	6.22	8.56	9.33

Table 4 shows the values of the lower bounds of the posterior probability of H_0 to test (2.1) given in expression (5.2), when $p_0 = \frac{1}{2}$, $q_0(p_1,\ p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$ and $\varepsilon = 0$, for some $\pi^0_{q_0}$ and tables (a, c) such that the *p-value* $P\{\Lambda \geq \lambda_0 | (p_0, p_0)\}$ given in expression (6.2) is close to the usual values 0.1,0.05 and 0.01.

We can observe that if we take $\pi^0_{q_0} \in (0.09, 0.11)$, then for $\delta \in (0.17, 0.19)$ the lower bounds of the posterior probability of H_0 are approximately equal to the *p-values*.

We can remark that the interval of values of $\pi^0_{q_0}$ that stabilizes the lower bound of the posterior probability of $H_0 : p_1 = p_2 = \frac{1}{2}$ around the *p-value* of the classical method is the same for both statistics, η and Λ. This shows that the discrepancy between the respective *p-values* to both statistics is not very large.

Note that, in general, $H_0 : p_1 = p_2 = p_0$ in (1.1) is no natural null hypothesis. By this reason we consider first a value of p_0 and after take an sphere of radius δ about this value. Moreover, in general, when we wish to test (1.1), the value of p_0 is unknown. In spite of this, (1.1) has an additional 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 known functional form. This possibility is studied in section 8. The results of previous sections are generalized for $r \times s$ tables in next section.

7 $r \times s$ Tables

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 exclusionary classes $C_1, \cdots, C_s$. The sample number i, $i = 1, \cdots, r$, is of size n_i and yields n_{ij} individuals in the category C_j, $j = 1, \cdots, s$. The data are displayed in Table 5.

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

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

$$\Theta = \left\{ \mathbf{p} = (p_1, \cdots, p_s) \in (0,1)^s,\ \sum\nolimits_{i=1}^{s} p_j = 1 \right\} \subset \mathrm{R}^{s-1}.$$

In this situation, we are going to suppose that we wish to test

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}_0 \text{ versus } H_1 : \exists i,\ \mathbf{p}_i \neq \mathbf{p}_0, \tag{7.1}$$

where $\mathbf{p}_0 = (p_{01}, \cdots, p_{0s}) \in \Theta$ is a known value.

Consider that our prior opinion about the parameter of interest $\theta = (\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Theta^r$ is given by means of the density $\pi(\theta) = \pi\,(\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Gamma$, where Γ is the class of $\varepsilon-$contaminated distributions given in expression (1.1). Therefore, a mixed prior distribution is needed to test (7.1).

We propose

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

π_0 being the prior probability assigned to the null hypothesis.

Following, if we denote by $\theta_0=(\mathbf{p}_0,\cdots,\mathbf{p}_0)\in\Theta^r\subset \mathrm{R}^{r(s-1)}$, then $H_0:\theta=\theta_0$ is the null hypothesis in (7.1).

Now, we can consider that it is more realistic to test

$$H_{0\delta}:d(\theta_0,\theta)\leq\delta \text{ versus } H_{1\delta}:d(\theta_0,\theta)>\delta,$$

with an appropriate metric d and a value of $\delta>0$ sufficiently small. We propose to use $B(\theta_0,\delta)=\left\{\theta\in\Theta^r,\ \sum_{i=1}^{r}\sum_{j=1}^{s-1}(p_{ij}-p_{0j})^2\leq\delta^2\ \right\}$.

Then, applying the method of Gómez-Villegas et al. (2004a), we can use $\pi(\theta)$, our opinion about θ, to calculate π_0 by means of averaging $\pi_0=\int_{B(\theta_0,\delta)}\pi(\theta)\,d\theta$. With this choice, if $\varepsilon=0$ and independent uniform prior distributions on Θ are assigned to each $\mathbf{p}_i$, $i=1,\cdots,r$, then

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

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

In this context, the posterior probability of the null hypothesis in (2.1), when data of Table 5 has been observed, is

$$P(H_0|n_{11},\ldots,n_{rs})=\left[1+\frac{1-\pi_0}{\pi_0}\frac{\int_{\Theta^r}\prod_{i=1}^{r}\prod_{j=1}^{s}p_{ij}^{n_{ij}}\pi(\theta)\,d\theta}{\prod_{j=1}^{s}p_{0j}^{\sum_{i=1}^{r}n_{ij}}}\right]^{-1},$$

where $\pi_0=(1-\varepsilon)\,\pi_{q_0}^0+\varepsilon\pi_q^0$, with

$$\pi_{q_0}^0=\int_{B(\theta_0,\delta)}q_0(\theta)d\theta,\ \ \pi_q^0=\int_{B(\theta_0,\delta)}q(\theta)d\theta.$$

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 λ_0 the value of Λ at an observed point of Table 5, then $\{\Lambda \geq \lambda_0\}$ is a possible *critical region* and the *p-value* is

$$p\left(n_{11},\ldots,n_{rs}\right) = P\left(\Lambda \geq \lambda_0|\theta_0\right) = P\left(\chi^2_{r(s-1)} \geq \lambda_0\right)$$

With this considerations, it is possible to extend easily the previously obtained results using a similar reasoning to the developed in section 4 for $r \times s$ tables.

8 Generalizations

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.

As usual, we can consider, as classical measure of the evidence, the discrepancy between the observed and expected values under the homogeneity null hypothesis, in the terms of Pearson's χ^2 statistic. Then, the *test statistic* is the random variable

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

If λ_0 denotes the value of Λ evaluated in the observed data point of Table 5, then $\{\Lambda \geq \lambda_0\}$ is a possible *critical region* and the corresponding *p-value* is

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

It is possible to generalize the previous results using an appropriate mixed prior distribution. One option is

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

where π_0 is the mass assigned to $H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r$, $\mathbf{p_0}$ is the unknown vector of common proportions under the null, $\pi(\theta)$ is our initial opinion in the class of $\varepsilon-$contaminated prior distributions, Γ, given in expression (1.1) and

$$\pi_0 = \int_{C(\delta)} \pi(\theta) d\theta,$$

with $C(\delta) = \bigcup_{\mathbf{p}_0 \in \Theta} B(\theta_0, \delta)$ and $\theta_0 = (\mathbf{p}_0, \ldots, \mathbf{p}_0) \in \Theta^r$. With this choice, if $\varepsilon = 0$ and independent uniform prior distributions on Θ are assigned to each $\mathbf{p}_i$, $i = 1, \cdots, r$, we can remark that $\pi_0 = 2\sqrt{2}\delta + 2\delta^2 - 4\sqrt{2}\delta^3$, for sufficiently small δ and 2×2 tables.

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

$$P(H_0|n_{11}, \ldots, n_{rs}) = \left[1 + \frac{1-\pi_0}{\pi_0} \frac{\int_{\Theta^r} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\theta) \, d\theta}{\int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0}\right]^{-1}.$$

In this case, a lower bound of the posterior probability of the homogeneity null hypothesis can be calculated by means of the same reasoning developed in section 4. For instance,

$$P(H_0|n_{11}, \ldots, n_{rs}) \geq \left[1 + \frac{1-(1-\varepsilon)\pi_{q_0}^0}{(1-\varepsilon)\pi_{q_0}^0} \eta_\varepsilon(n_{11}, \ldots, n_{rs})\right]^{-1},$$

where

$$\eta_\varepsilon(n_{11}, \ldots, n_{rs}) = (1-\varepsilon)\eta(n_{11}, \ldots, n_{rs}) + \varepsilon \frac{\prod_{i=1}^r \prod_{j=1}^s \left(\frac{n_{ij}}{n_i}\right)^{n_{ij}}}{\int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p_0}},$$

with

$$\eta(n_{11}, \ldots, n_{rs}) = \frac{\int_{\Theta^r} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\theta) \, d\theta}{\int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0}.$$

Then the following Theorem can be formulated

Theorem 8.1. *Let $\hat{\theta}$ be the maximum likelihood estimator of θ when $\theta \in H_1$.*

If $\hat{\theta} \notin C(\delta)$ and, for fixed ρ, $\int_{B(\hat{\theta},\ \rho)} f(n_{11}, \ldots, n_{rs}|\theta)d\theta$ is approximated by means of $\left[\pi^{\frac{r(s-1)}{2}} \rho^{r(s-1)} / \Gamma\left(\frac{r(s-1)}{2}+1\right)\right] f\left(n_{11}, \ldots, n_{rs}|\hat{\theta}\right)$, then the distribution given by $\tilde{\pi}(\theta) = (1-\varepsilon) q_0(\theta) + \varepsilon \tilde{q}(\theta)$, where $\tilde{q}(\theta)$ is uniform in $B\left(\hat{\theta},\ \rho\right)$, satisfies

$$inf_{\pi \in \Gamma} P_{\pi}(H_0|n_{11}, \ldots, n_{rs}) = P_{\tilde{\pi}}(H_0|n_{11}, \ldots, n_{rs})$$

$$= \left[1 + \frac{1-(1-\varepsilon)\pi^0_{q_0}}{(1-\varepsilon)\pi^0_{q_0}} \eta_{\varepsilon}(n_{11}, \ldots, n_{rs})\right]^{-1}.$$

Note that the only modification respect to Theorem 4.1 is the use of the set $C(\delta)$, the union of all the spheres centered in points of the null hypothesis and radius δ.

Another important problem is when $\mathbf{p}_0 = \mathbf{p}(\omega)$ (Lindley, 1988), with $\mathbf{p}$: $\Omega \to \Theta$, being

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

and $q < s$ fixed.

As usual, from a classical viewpoint we can use Pearson's χ^2 statistic as *test statistic*,

$$\Lambda = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n_{ij}^2}{n_i p_j(\hat{\omega})} - N,$$

where $\hat{\omega}$ is the *maximum likelihood estimator* of ω. If λ_0 is the value of Λ in the data point of Table 5, then $\{\Lambda \geq \lambda_0\}$ is a possible *critical region* and the used evidence is the *p-value*,

$$p = P\left(\chi^2_{rs-1-q} \geq \lambda_0\right).$$

In this context, for comparisons between classical and Bayesian evidence measures, we propose to use the following appropriate prior distribution

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

where π_0 is the prior probability assigned to $H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}(\omega)$, $\pi(\theta)$ is our initial opinion in the class of $\varepsilon-$contaminated prior distributions, Γ, given in expression (1.1) and

$$\pi_0 = \int_{C(\delta)} \pi(\theta) d\theta,$$

with $C(\delta) = \bigcup_{\omega\in\Omega} B(\theta_0, \delta)$ and $\theta_0 = (\mathbf{p}(\omega), \ldots, \mathbf{p}(\omega)) \in \mathbf{\Theta^r}$.

In this case, the posterior probability of the null hypothesis, when the data of Table 5 has been observed, is

$$P(H_0|n_{11},\ldots,n_{rs}) = \left[1 + \frac{1-\pi_0}{\pi_0} \frac{\int_{\Theta^r} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\theta)\, d\theta}{\int_\Omega \prod_{j=1}^s p_j(\omega)^{\sum_{i=1}^r n_{ij}} \pi(\omega) d\omega}\right]^{-1}.$$

Finally, we can note that the extension of the previous results to this situation is easy using a similar methodology.

9 Comments

The obtained results are the consequence of the methodology based on the relation between the null punctual hypothesis in (2.1) and the more realistic hypothesis given in (2.3). In terms of the Kullback-Leibler information measure, the discrepancy between $\pi(p_1, p_2) \in \Gamma$, Γ being the class of $\varepsilon-$contaminated distributions given in expression (1.1), and the the mixed prior distribution $\pi^*(p_1, p_2)$ in (2.2) justifies the choice of π_0 as in (2.4) with an appropriate value of δ. According to this procedure, $\pi^*(p_1, p_2)$, used to test (2.1), is close to the continuous prior $\pi(p_1, p_2)$, used to test (2.3), as can be seen in Gómez-Villegas and Sanz (2000) and Gómez-Villegas et al. (2002).

When $\pi(p_1, p_2)$ is in the class of $\varepsilon-$contaminated distributions, the lower bound of the posterior probability of the point null hypothesis to test

(2.1) can be close to the *p-value*, as seen in section 6. Gómez-Villegas and Sanz (2000) obtain similar results in a different context.

The results that we have obtained indicate that, to test (2.1), the observed discrepancy between the classical and Bayesian approaches using $\pi_0 = \frac{1}{2}$ in the mixed distribution is bigger.

Finally, the methodology proposed can be used to approach the problem of testing the homogeneity of independent multinomial distributions and compare classical and Bayesian evidence measures, with $r \times s$ tables, when $\mathbf{p}_0$ is known, unknown or with its functional form known, $\mathbf{p}_0 = \mathbf{p}_0(\omega)$.

Acknowlegments

The authors are grateful to two referees, the associate editor and the editors for their remarks and criticisms that have led to improve the article.

References

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

Berger, J. O. (1994). An overview over robust bayesian analysis (with discussion). *Test*, 3(1):5–124.

Berger, J. O. and Berliner, L. M. (1986). Robust bayes and empirical bayes analysis with $\varepsilon-$contaminated priors. *Communications in Statistics, Theory and Methods*, 13:395–400.

Berger, J. O., Boukai, B., and Wang, Y. (1997). Unified frequentist and bayesian testing of a precise hypothesis. *Statistical Science*, 12(3):133–160.

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). *Journal of the American Statistical Association*, 82:112–139.

Bernardo, J. M. (1980). A bayesian analysis of classical hypothesis testing. in: Bernardo, j.m., degroot, m.h., lindley, d.v., smith, a.f.m., (eds.). *Bayesian Statistics, University Press, Valencia*, pp. 605–647. (with discussion).

Casella, G. and Berger, R. L. (1987). Reconciling bayesian and frequentist evidence in the one-sided testing problem, (with discussion. *Journal of the American Statististical Association*, 82. 106-135.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 68:966–969.

Dickey, J. M. and Lienz, B. P. (1970). The weighted likelihood ratio, sharp hypothesis about chances, the order of a markov chain. *Annals of the Mathematical Statistics*, 41:214–226.

Edwards, W. L., Lindman, H., and J., S. L. (1963). Bayesian statistical inference for psychologycal research. 70:193–248.

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors. in: Bernardo, j.m., berger, j.o., dawid, a.p., smith, a.f.m. (eds). *Bayesian Statistics, University Press, Oxford*, 4:195–210. (with discussion).

Gómez-Villegas, M. A. and Gómez, E. (1992). Bayes factor in testing precise hyphoteses. *Communications in Statistics, Theory and Methods*, 21:1707–1715.

Gómez-Villegas, M. A., Maín, P., and Sanz, L. (2002). A suitable bayesian approach in testing point null hypothesis: some examples revisited. *Communications in Statistics, Theory and Methods*, 31(2):201–217.

Gómez-Villegas, M. A., Maín, P., and Sanz, L. (2004a). A bayesian analysis for the multivariate point null testing problem. Tech. rep., Dpto. EIO-I, Universidad Complutense de Madrid. 04-01.

Gómez-Villegas, M. A., Maín, P., Sanz, L., and Navarro, H. (2004b). Asymptotic relationships between posterior probabilities and p-values using the hazard rate. *Statistics & Probability Letters*, 66:59–66.

Gómez-Villegas, M. A. and Sanz, L. (1998). Reconciling bayesian and frequentist evidence in the point null testing problem. *Test*, 7(1):207–216.

Gómez-Villegas, M. A. and Sanz, L. (2000). ε-contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statistics & Probability Letters*, 47:53–60.

Huber, P. J. (1973). The use of choquet capacities in statistics. *Bulletin of the International Statistical Institute*, 45:181–191.

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.

McCulloch, R. E. and Rossi, P. E. (1992). Bayes factors for non-linear hypothesis and likelihood distributions. *Biometrika*, 79:663–676.

Mukhopadhyay, S. and Dasgupta, A. (1997). A uniform approximation of bayes solutions and posteriors: Frequentistly valid bayes inference. *Statistics and Decisions*, 15:51–73.

Oh, H. S. and DasGupta, A. (1999). Comparison of the p-value and posterior probability. *Journal of the Statististical Planning and Inference*, 76:93–107.

Pratt, J. V. (1965). Bayesian interpretation of standard inference statements. 27:169–203.

Robert, C. P. (2001). *The Bayesian Choice*. Springer, New York.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statiscian. *Annals of Statistics*, 12:1151–1172.

Sivaganesan, S. (1988). Range of the posterior measures for priors with arbitrary contaminations. *Communications in Statistics, Theory and Methods*, 17:1591–1612.

Sivaganesan, S. and Berger, J. O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Annals of Statistics*, 17:869–889.

Spiegelhalteer, D. J. and Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. 44:377–387.