A BAYESIAN ANALYSIS FOR THE HOMOGENEITY TESTING PROBLEM USING $\varepsilon-$CONTAMINATED PRIORS

Miguel A. Gómez-Villegas and Beatriz González-Pérez

Departamento de Estadística e Investigación Operativa I

Universidad Complutense de Madrid

Facultad de CC. Matemáticas, Madrid, 28040, Spain

ma.gv@mat.ucm.es and beatrizg@mat.ucm.es

Key Words: $\varepsilon-$contaminated priors, homogeneity testing problem, contingency tables, posterior probabilities, p-values.

ABSTRACT

In this paper the problem of testing if r populations have the same distribution from a Bayesian perspective is studied using $r \times s$ contingency tables and $\varepsilon-$contaminated priors. A procedure to build a mixed prior distribution is introduced and a justification for this construction based on a measure of discrepancy is given. A lower bound for the posterior probabilities of the homogeneity null hypothesis, when the prior is in the class of $\varepsilon-$contaminated distributions, is calculated and compared numerically with the usual p-value. Examples show that the discrepancy between both is more acute when the mass assigned to the null in the mixed prior distribution is 0.5.

1. INTRODUCTION

1.1 $\varepsilon-$CONTAMINATED CLASS

To carry out a Bayesian analysis concerning an unknown parameter, θ, it's necessary to model our prior information about θ through a prior distribution. Frequently, the prior information is diffuse and the initial opinion can not be expressed in terms of a concrete probability distribution. This absence of precision is the reason why, often, the prior information is expressed in terms of a class of distributions, Γ, where all possible prior distributions about θ are included. Furthermore, while considering a class of prior distributions instead of a concrete prior it looks reasonable to compare the posterior probability of the null hypothesis with the usual p-value since the p-value does not use prior information. Therefore, if there is statistical coherence, a

Bayesian which uses the class of all prior distributions must reach the same decision as a frequentist.

An interesting way to describe deviations of our prior opinions about the parameter of interest involves considering the $\varepsilon-$contaminated class, for instance, see Gómez-Villegas and Sanz (2000). Let $q_0 = q_0(\theta)$ be a known prior distribution that summarizes the initial opinion about the parameter θ and Q be a class of probability distributions $q = q(\theta)$ which represents the possible (and reasonable) deviations from q_0. When a fixed and known degree of contamination ε is introduced into q_0 by the statistician, with $0 < \varepsilon < 1$, the $\varepsilon-$contaminated class Γ is the class of all distributions $\pi = \pi(\theta)$ built from a convex combination of q_0 and $q \in Q$,

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

Note that if $\varepsilon = 0$, no contamination is introduced in q_0, and $\Gamma \equiv q_0$.

We can take into account several possibilities for the class Q. We are going to work with the class of all probability distributions. Huber (1973), Sivaganesan (1988), Gómez-Villegas and Sanz (2000) and Gómez-Villegas and González-Pérez (2008) use this class in other contexts. Berger and Berliner (1986), Berger (1985, 1994) and Sivaganesan and Berger (1989) give relevant information about other choices.

In subsection 2 we introduce a procedure to make up a mixed prior distribution to approach the homogeneity testing problem from a Bayesian viewpoint and in subsection 3 a justification of this construction is provided. In subsection 4 we introduce the notation and in subsection 5 we expose the focus of the paper.

1.2 THE PROBLEM

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

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

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

Then, we can model this situation as follow: let X_i, $i = 1, \cdots, r$, be independent multinomial random variables, $MB(n_i, \mathbf{p}_i)$, with $\mathbf{p}_i = (p_{i1}, \cdots, p_{is-1}) \in \Theta$, where

2

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

and $p_s = 1 - \sum_{i=1}^{s-1} p_j$.

The problem involves testing if r populations have the same distribution, which can be arbitrary, when we observe the random variable $(X_1, \ldots, X_r)$. That is to say, we consider the homogeneity testing problem

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r \text{ versus } H_1 : \mathbf{p}_i \neq \mathbf{p}_j, \text{ for some } i \neq j. \tag{2}$$

In the point null testing problem $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is usual to model our prior opinion about the parameter of interest θ by means of a mixed distribution $\pi^*(\theta)$, assigning mass π_0 to the null and spreading the remainder, $1 - \pi_0$, over the alternative according to a density $\pi(\theta)$. Gómez-Villegas and Sanz (2000) study the unidimensional problem of testing a point null hypothesis by choosing $\pi_0 = \int_{|\theta - \theta_0| \leq \delta} \pi(\theta) d\theta$, with an arbitrary contamination $\pi(\theta) \in \Gamma$ (see expression (1)) and $q_0(\theta)$ known. They introduce a procedure to determine the mixed prior distribution and give a justification for their construction based on the Kullback-Leibler measure of discrepancy. They calculate and compare numerically a lower bound for the posterior probability of the point null with the p-value, when the sample model is normal in two cases: when the base prior distribution $q_0(\theta)$ is normal with both mean and variance known, and when $q_0(\theta)$ is Cauchy(0, 2). In the light of their results, they propose lower bounds to use as Bayesian evidence measures for a suitable value of δ. Examples shows that the value of δ must be chosen, in general, by making the lower bound agrees numerically with the p-value when the p-value varies in (0.01, 0.1), the usual interval where the significance level α varies. Furthermore, it seems that the discrepancy observed in testing point null hypothesis between frequentist and Bayesian approaches becomes more acute by using $\pi_0 = 0.5$ in the mixed distribution.

Note that the homogeneity null hypothesis in (2) is not completely specified. This is due to the fact that if we denote the common proportions vector under the null by $\mathbf{p}_0 = (p_{01}, \cdots, p_{0s-1}) \in \Theta$, then $\mathbf{p_0}$ is an unknown vector. Gómez-Villegas and González-Pérez (2008) study this problem when $\mathbf{p_0}$ is known and

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

In this situation, they compute a lower bound of the posterior probability of the null when $\pi (\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Gamma$, with an arbitrary contamination. Examples show that the bound is close to the p-value for a suitable value of π_0, by making the lower bound agree numerically with the p-value in (0.01, 0.1) (or equivalently, of δ computed from $\pi_0 = \int_{B(\theta_0, \delta)} \pi(\theta) d\theta$, with $\theta_0 = (\mathbf{p}_0, \cdots, \mathbf{p}_0) \in \Theta^r$ and $B(\theta_0, \delta) \equiv \left\{ \theta = (\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Theta^r, \ \sum_{i=1}^{r} \sum_{j=1}^{s-1} (p_{ij} -$ Furthermore, the discrepancy between both approaches using $\pi_0 = 0.5$ is bigger. Note that $\pi_0 = \int_{B(\theta_0, \delta)} q_0 (\theta) d\theta$ when $\varepsilon = 0$ and if independent uniform prior distributions on Θ are assigned to each $\mathbf{p}_i$, $i = 1, \cdots, r$, then

3

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

the volume of the sphere of radius δ in $\Theta^{\mathrm{r}} = [0,1]^{r(s-1)}$, for δ sufficiently small.

For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ from a Bayesian perspective, when $\theta_0 = (\mathbf{p}_0, \cdots, \mathbf{p}_0) \in \Theta^r$ and $\mathbf{p}_0 \in \Theta$, the common proportions vector under the null, is unknown with a fixed prior density $\pi^0(\mathbf{p_0})$. Let us suppose that the prior distribution about $\theta = (\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Theta^r$ is

$$\pi^*(\theta) = \pi_0 \pi^0(\mathbf{p_0}) I_{H_0}(\theta) + (1 - \pi_0) \pi(\theta) I_{H_1}(\theta), \tag{3}$$

where π_0 is the prior mass assigned to H_0 and $\pi(\theta) = \pi(\mathbf{p}_1, \cdots, \mathbf{p}_r) \in \Gamma$, where Γ is the $\varepsilon-$contaminated class defined in (1).

We propose to compute the value of π_0 in (3) as

$$\pi_0 = \int_{C(\delta)} \pi(\theta) d\theta, \tag{4}$$

with $C(\delta) \equiv \bigcup_{\mathbf{p_0} \in \Theta} B(\theta_0, \delta)$ and a value of $\delta > 0$ sufficiently small.

With this choice, if $\varepsilon = 0$ and an independent uniform prior distributions on Θ is assigned to each $\mathbf{p}_i$, $i = 1, \cdots, r$, we can remark that $\pi_0 = 2\sqrt{2}\delta + 2\delta^2 - 4\sqrt{2}\delta^3$, the area of $C(\delta)$ in $[0,1]^2$ for δ sufficiently small and 2×2 tables. These are the prior specifications for comparisons in section 3.

Generalizing the method introduced by Gómez-Villegas and Sanz (2000), this construction is inspired on the supposition that if a density $\pi(\theta)$ represents our prior beliefs about θ, as to test (1) is not possible with $\pi(\theta)$, we approach (1) by

$$H_{0\delta} : \theta \in C(\delta) \quad \text{versus} \quad H_{1\delta} : \theta \notin C(\delta), \tag{5}$$

choosing an appropriate value of δ.

1.3 JUSTIFICATION

The choice of π_0 as in (4) is basic for the next calculations. A way of justifying this construction is by using the Kullback-Leibler information measure

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

as a discrepancy measure between π and π^*. However, $\pi(\theta)$ is a density function and the mixed distribution $\pi^*(\theta)$ given in (3) is not a density. Therefore there is a problem.

4

It is important to remark the differences between these two probability measures on $(\mathbb{R}^r, \mathcal{B}^r)$. For instance, if μ and μ^* denote the probability measures defined by π and π^* then

$$\mu(A) = \int_A d\pi(\theta) = \int_A \pi(\theta)d\theta$$

and

$$\mu^*(A) = \int_A d\pi^*(\theta) = \begin{cases} \pi_0 \int_{A \cap H_0} \pi^0(\mathbf{p}_0)d\mathbf{p}_0 + (1-\pi_0)\mu(A) & \text{if } A \cap H_0 \neq \varnothing \\ (1-\pi_0)\mu(A) & \text{if } A \cap H_0 = \varnothing \end{cases},$$

for all $A \in \mathcal{B}^r$.

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

$$\frac{d\mu}{d\mu^*}(\theta) = \begin{cases} 0 & \text{if } \theta \in H_0 \\ \frac{1}{1-\pi_0} & \text{if } \theta \notin H_0 \end{cases}. \tag{6}$$

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

$$D(\pi^*|\pi) = \int_\Theta ln\left[\frac{d\mu}{d\mu^*}(\theta)\right]d\mu(\theta), \tag{7}$$

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 (4), π_0 goes to zero too and consequently $D(\pi^*|\pi)$ goes to zero. This is a justification of the choice for π_0 as in (4) and to replace (2) by (5). Secondly, when we use $\pi_0 = 0.5$ instead of using the value of π_0 given in (4), the discrepancy between μ and μ^* is perhaps a high discrepancy, $D(\pi^*|\pi) = 0.693$.

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 non statistician to give 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 $\delta = 0.3$". Therefore, from this way he is able to determine a value of δ and the statistician uses this value to compute π_0 as in (4) and develop a Bayesian test based on the precise hypothesis $H_0 : \theta = \theta_0$, by using the mixed prior distribution $\pi^*(\theta)$ introduced in (3). Furthermore, sometimes choosing π_0 as in (4) allows the computation of an exact expression of a lower bound for the posterior probability which is easy to implement. For instance, see Gómez-Villegas, Maín and Sanz (2004) for the multivariate point null testing problem, between others.

1.4 NOTATION

We denote the likelihood function by

5

$$f\left(n_{11}, \ldots, n_{rs}|\theta\right) = \prod_{i=1}^{r} \frac{n_i}{\prod_{j=1}^{s} n_{ij}!} \prod_{j=1}^{s} p_{ij}^{n_{ij}}$$

which is considered as a function of θ for the observed value of $(X_1, \ldots, X_r) = (n_{11}, \ldots, n_{rs})$. If the marginal distribution of $(X_1, \ldots, X_r)$ with respect to the prior distribution $\pi \in \Gamma$ is denoted by $m\left(n_{11}, \ldots, n_{rs}|\pi\right)$, then

$$m\left(n_{11}, \ldots, n_{rs}|\pi\right) = (1 - \varepsilon) m\left(n_{11}, \ldots, n_{rs}|q_0\right) + \varepsilon m\left(n_{11}, \ldots, n_{rs}|q\right).$$

Assuming the existence of $q_0\left(\theta|n_{11}, \ldots, n_{rs}\right)$ and $q\left(\theta|n_{11}, \ldots, n_{rs}\right)$, the posterior distribution of θ with respect to $\pi \in \Gamma$ is

$$\pi\left(\theta|n_{11}, \ldots, n_{rs}\right) =$$

$$\lambda\left(n_{11}, \ldots, n_{rs}\right) q_0\left(\theta|n_{11}, \ldots, n_{rs}\right) + \left(1 - \lambda\left(n_{11}, \ldots, n_{rs}\right)\right) q\left(\theta|n_{11}, \ldots, n_{rs}\right),$$

where

$$\lambda\left(n_{11}, \ldots, n_{rs}\right) = \frac{(1 - \varepsilon) m\left(n_{11}, \ldots, n_{rs}|q_0\right)}{m\left(n_{11}, \ldots, n_{rs}|\pi\right)}.$$

From (4) we have $\pi_0 = (1 - \varepsilon)\pi_{q_0}^0 + \varepsilon\pi_q^0$, where

$$\pi_{q_0}^0 = \int_{C(\delta)} q_0(\theta)d\theta \text{ and } \pi_q^0 = \int_{C(\delta)} q(\theta)d\theta. \tag{8}$$

Then, the posterior probability of the homogeneity null hypothesis admits the following expression

$$P(H_0|n_{11}, \ldots, n_{rs}) = \left[\frac{\int_\Theta f\left(n_{11}, \ldots, n_{rs}|\theta_0\right)\pi^0(\mathbf{p}_0)d\mathbf{p}_0}{\int_\Theta f\left(n_{11}, \ldots, n_{rs}|\theta_0\right)\pi^0(\mathbf{p}_0)d\mathbf{p}_0 + \frac{1-\pi_0}{\pi_0} m\left(n_{11}, \ldots, n_{rs}|\pi\right)}\right]^{-1}$$

$$= \left[1 + \frac{1 - \pi_0}{\pi_0} \frac{\int_{\Theta^r} \prod_{i=1}^{r} \prod_{j=1}^{s} p_{ij}^{n_{ij}} \pi(\theta)d\theta}{\int_\Theta \prod_{j=1}^{s} p_{0j}^{\sum_{i=1}^{r} n_{ij}} \pi^0(\mathbf{p}_0)d\mathbf{p}_0}\right]^{-1}. \tag{9}$$

A frequentist measure of evidence against the null, which depends on our observations, is the p-value. As usual, we use the discrepancy between the observed and expected values under the null, in the terms of Pearson's χ^2 statistic, as frequentist measure of the evidence to test (1). Then, the test statistic is the random variable

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

If $\lambda_0 = \Lambda(n_{11}, \ldots, n_{rs})$ denotes the value of Λ evaluated in the data point, then $\{\Lambda \geq \lambda_0\}$ is a possible critical region and the corresponding p-value is

6

$$p(n_{11}, \ldots, n_{rs}) = P\left(\chi^2_{(r-1)(s-1)} \geq \lambda_0\right). \tag{10}$$

1.5 THE FOCUS

In the same way of Berger and Sellke (1987), we seek to minimize the posterior probability of the precise null $H_0 : \theta = \theta_0$, $P(H_0|n_{11}, \ldots, n_{rs})$, to compute a lower bound over the class Γ given in (1), as Bayesian evidence measure for the homogeneity testing problem introduced in (2).

A reason to take the infimum is that the null hypothesis must be rejected for a small infimum according to the interpretation of the p-value. Moreover, this development is similar to Casella and Berger (1987) reasoning in order to reconcile Bayesian and frequentist evidence in the one-sided testing problem.

The following works, among others, have been realized to study the discrepancy between both methods in testing problems. Delampady and Berger(1990), Ghosh and Mukerjee (1992), McCulloch and Rossi (1992), Gómez-Villegas and Gómez (1992), Mukhopadhyay and DasGupta (1997), Berger, Boukai and Wang (1997, 1999), Gómez-Villegas and Sanz (1998, 2000), Sellke, Bayarri and Berger (2001), Gómez-Villegas, Maín and Sanz (2002), Gómez-Villegas, Maín, Sanz and Navarro (2004), De la Horra (2005), Gómez-Villegas and González-Pérez (2005, 2006).

As in Gómez-Villegas and Sanz (2000) and Gómez-Villegas and González-Pérez (2008), our interest is to compare numerically the lower bound of the posterior probability of the point null with the p-value. The focus involves in finding a suitable value of π_0 (or equivalently, of δ, see subsection 1.3) which makes both values approximately equal in the range (0.01, 0.1) and showing how the discrepancy between both is more acute when $\pi_0 = 0.5$ in the mixed prior distribution.

In section 2 a lower bound for the posterior probabilities of the homogeneity null hypothesis, when the prior is in the $\epsilon-$contaminated class, is calculated. In section 3 this lower bound is compared numerically with the usual p-value. A possible generalization of the proposed methodology for $\mathbf{p}_0$ with known functional form, $\mathbf{p}_0 = \mathbf{p}_0(\omega)$, is given in section 4. Finally, in section 5 some comments are included.

2. LOWER BOUND FOR THE POSTERIOR PROBABILITY

In this section we obtain in Theorem 1, a lower bound for the posterior probability of the null hypothesis to test (2) when the prior distribution π^* is (3) with π_0 computed according to (4). Theorem 2 establishes sufficient conditions in order to achieve the infimum of the posterior probability when $\pi \in \Gamma$.

Theorem 1. Consider the hypothesis introduced in (2), an arbitrary prior distribution $\pi \in \Gamma$ as in (1) and a mixed prior distribution as in (3) with assigned mass to the null hypothesis according to (4).

Then

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

where

$$\eta_\varepsilon\left(n_{11},\ldots,n_{rs}\right) = (1-\varepsilon)\,\eta(n_{11},\ldots,n_{rs}) + \varepsilon\frac{\sup_{\theta\notin H_0} f(n_{11},\ldots,n_{rs}|\theta)}{\int_\Theta \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}}\pi^0(\mathbf{p_0})d\mathbf{p_0}},$$

with

$$\eta(n_{11},\ldots,n_{rs}) = \frac{\int_{\Theta^r}\prod_{i=1}^r\prod_{j=1}^s p_{ij}^{n_{ij}}q_0\left(\theta\right)d\theta}{\int_\Theta \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}}\pi^0(\mathbf{p_0})d\mathbf{p_0}}.$$

Proof. Computing a lower bound of the posterior probability of H_0 is just like computing an upper bound of $\frac{1-\pi_0}{\pi_0}m\left(n_{11},\ldots,n_{rs}|\pi\right)$ when $\pi \in \Gamma$. By the construction of $\pi^*\left(\theta\right)$, π_0 depends on $q \in Q$ through π_q^0 given in expression (8). Then, that lower bound can be computed as the supremum in $q \in Q$ of

$$\frac{1-\pi_0}{\pi_0}m\left(n_{11},\ldots,n_{rs}|\pi\right) =$$

$$\left[\frac{1}{(1-\varepsilon)\,\pi_{q_0}^0 + \varepsilon\pi_q^0} - 1\right]\left[(1-\varepsilon)\,m\left(n_{11},\ldots,n_{rs}|q_0\right) + \varepsilon m\left(n_{11},\ldots,n_{rs}|q\right)\right]. \tag{12}$$

With $\pi_{q_0}^0$ and π_q^0 given in expression (8), and as the supremum of (14) 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

$$\sup_{q\in Q}\left[(1-\varepsilon)\,m\left(n_{11},\ldots,n_{rs}|q_0\right) + \varepsilon m\left(n_{11},\ldots,n_{rs}|q\right)\right],$$

where

$$m\left(n_{11},\ldots,n_{rs}|q\right) = \int_{\Theta^r} f(n_{11},\ldots,n_{rs}|\theta)q\left(\theta\right)d\theta \leq \sup_{\theta\notin H_0} f(n_{11},\ldots,n_{rs}|\theta),$$

then, we obtain (11).

Theorem 1 gives a lower bound for the posterior probability of the homogeneity null hypothesis to test (2). The first question arises when the infimum is achieved by a distribution of the class Γ defined in (1). The answer is given by the following theorem.

Theorem 2. 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 $\frac{\pi^{\frac{r(s-1)}{2}} \rho^{r(s-1)}}{\Gamma\left(\frac{r(s-1)}{2}+1\right)} f\left(n_{11}, \ldots, n_{rs}|\hat{\theta}\right)$, 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

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

$$= \left[1 + \frac{1 - (1-\varepsilon)\pi_{q_0}^0}{(1-\varepsilon)\pi_{q_0}^0} \eta_\varepsilon(n_{11}, \ldots, n_{rs})\right]^{-1}, \tag{13}$$

where $\pi_{q_0}^0$ and $\eta_\varepsilon(n_{11}, \ldots, n_{rs})$ are both as in Theorem 1.

Proof. For (10), we need to compute π_0 and $m(n_{11}, \ldots, n_{rs}|\tilde{\pi})$. Given that for $\tilde{\pi}$,

$$\pi_0 = \int_{C(\delta)} \tilde{\pi}(\theta)d\theta = (1-\varepsilon) \int_{C(\delta)} q_0(\theta)d\theta + \varepsilon \int_{C(\delta)} \tilde{q}(\theta)d\theta = (1-\varepsilon)\pi_{q_0}^0$$

and $m(n_{11}, \ldots, n_{rs}|\tilde{\pi}) = (1-\varepsilon) m(n_{11}, \ldots, n_{rs}|q_0) + \varepsilon m(n_{11}, \ldots, n_{rs}|\tilde{q})$, where

$$m(n_{11}, \ldots, n_{rs}|\tilde{q}) = \int_{\Theta^r} f(n_{11}, \ldots, n_{rs}|\theta) \tilde{q}(\theta) d\theta$$

$$= \frac{1}{\frac{\pi^{\frac{r(s-1)}{2}} \rho^{r(s-1)}}{\Gamma\left(\frac{r(s-1)}{2}+1\right)}} \int_{B(\hat{\theta}, \rho)} f(n_{11}, \ldots, n_{rs}|\theta) d\theta \approx f\left(n_{11}, \ldots, n_{rs}|\hat{\theta}\right),$$

then we obtain (13).

Observe that the maximum likelihood estimator for $\theta = (\mathbf{p_1}, \ldots, \mathbf{p_r})$ is $\hat{\theta} = (\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_r)$, where $\hat{\mathbf{p}}_i = (\hat{p}_{i1}, \ldots, \hat{p}_{is})$, with $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, r$.

To remark that the real restriction in this theorem is $\hat{\theta} \notin C(\delta)$ is of interest, since $C(\delta) \cap B(\tilde{\theta}, \rho) = \emptyset$ and the approximation of the integral is always possible by choosing a sufficiently small value of δ. In other words, (14) is a strict inequality.

We can note that the Bayes factor is a procedure which could be considered as an alternative to the posterior probability. Perhaps, the Bayes factor has the problem especially concerning what gauge gives enough evidence to reject H_0. This is the reason why we prefer to use posterior probabilities, which do not have this inconvenience, instead of Bayes factors. Definitely, the Bayes factor depends on the posterior probability.

3. COMPARISONS

In parametric testing point null hypothesis, it is known that Bayesian and frequentist procedures can give rise to different decisions, see Lindley (1957), Berger and Delampady (1987) and Berger and Sellke

(1987), amongst others. In multinomial tests of point null hypothesis $H_0 : \mathbf{p} = \mathbf{p}_0$ (a specified point), Delampady and Berger (1990) compute lower bounds on Bayes factors in favour of H_0, for two different classes of densities (conjugate priors with mean $\mathbf{p}_0$ and unimodal and symmetric priors about $\mathbf{p}_0$). 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 usually this infimum is substantially larger than the corresponding p-value. It is necessary to point out that in all these situations the mass assigned to the simple null hypothesis is $1/2$. On the other hand, Casella and Berger (1987) show that there is no discrepancy in the one-sided testing problem.

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

For numerical comparisons between the lower bound for the posterior probability computed in Theorem 1 and the usual p-value given in expression (10), we are going to consider 2×2 tables (see Table 2). In this situation, we want to test if the proportion of successes in the first population, p_1, is the same as in the second, p_2, that is

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2. \tag{14}$$

Then the usual test statistic is the random variable

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

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

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

For simplifying we consider $\varepsilon = 0$. Furthermore, we suppose for convenience that p_1 and p_2 have independent uniform prior distributions and the unknown common value under the null p_0 has a uniform distribution too. Then, from our Bayesian viewpoint, the lower bound for the posterior probability is obtained evaluating expression (11) in

$$\eta(a, c) = \frac{\Gamma(N+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)} \, \frac{\Gamma(a+1)\,\Gamma(b+1)}{\Gamma(a+b+2)} \, \frac{\Gamma(c+1)\,\Gamma(d+1)}{\Gamma(c+d+2)}.$$

Table 3 shows the values of this lower bound for some specific values of Λ and some π_0. We can observe that the values of the lower bound are close to the respective p-values by choosing an adequate value of π_0.

10

For instance, if we take $\pi_{q_0} \in (0.10, 0.14)$, we can check that this lower bound is approximately equal to the p-values. We can also observe that the discrepancy is more acute by using $\pi_{q_0} = 0.5$ in the mixed prior distribution.

A way to choose π_{q_0} is to make the lower bound in (11) equal to the p-value in (10) and obtain $\pi_{q_0}^0$ from the expression

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

However, this implies that

$$\pi_{q_0}^0 = \frac{1}{1 - \varepsilon}\left[1 + \frac{1 - p(n_{11}, \dots, n_{rs})}{p(n_{11}, \dots, n_{rs})\eta_\varepsilon(n_{11}, \dots, n_{rs})}\right]^{-1} \tag{15}$$

and the prior probability depends on the data.

A possibility to avoid this data dependence is to use the significance level of the test α instead of $p(n_{11}, \dots, n_{rs})$. Moreover, if the value selected for $\pi_{q_0}^0$ is close to the one obtained by (15), then we get that the p-value and the infimum of the posterior probability of the null hypothesis to test (2) are close, since the infimum is a continuous function of $\pi_{q_0}^0$ (see expression (13)). In this sense, we propose to use a value of $\pi_{q_0}^0$ close to the result of this equalization which provides, by a simple data analysis as Table 3 shows, a value numerically equal from both points of view.

Table 3. lower bounds of the posterior probability of $H_0 : p_1 = p_2$ for tables (a, c) with $P\{\Lambda \geq \lambda_0 | (p_0, p_0)\}$ close to 0.1, 0.05 and 0.01, $q_0\,(p_1, p_2) = I_{(0,1)}\,(p_1)\,I_{(0,1)}\,(p_2)$, $\pi^0(p_0) = I_{(0,1)}\,(p_0)$ and $\varepsilon = 0$.

$P\{\Lambda \geq \lambda_0 \vert (p_0, p_0)\}$	0.1127	0.0980	0.0510	0.0455	0.0111	0.0086
$\pi^0_{q_0} = 0.5$	0.4658	0.3904	0.2957	0.3325	0.0936	0.0711
$\pi^0_{q_0} = 0.2$	0.1790	0.1380	0.0950	0.1107	0.0252	0.0188
$\pi^0_{q_0} = 0.15$	0.1333	0.1015	0.0690	0.0808	0.0179	0.0133
$\pi^0_{q_0} = 0.14$	0.1243	0.0944	0.0640	0.0750	0.0165	0.0123
$\pi^0_{q_0} = 0.13$	0.1153	0.0873	0.0590	0.0693	0.0152	0.0113
$\pi^0_{q_0} = 0.12$	0.1063	0.0803	0.0542	0.0636	0.0139	0.0103
$\pi^0_{q_0} = 0.11$	0.0973	0.0734	0.0493	0.0580	0.0126	0.0094
$\pi^0_{q_0} = 0.10$	0.0873	0.0664	0.0446	0.0524	0.0113	0.0084
$\pi^0_{q_0} = 0.09$	0.0794	0.0596	0.0399	0.0470	0.0101	0.0075
$\pi^0_{q_0} = 0.08$	0.0705	0.0528	0.0352	0.0415	0.0089	0.0066
$\pi^0_{q_0} = 0.07$	0.0616	0.0460	0.0306	0.0361	0.0077	0.0057
λ_0	2.5155	2.7375	3.8095	4.0000	6.4509	6.9139

4. GENERALIZATIONS

Another important problem (see Lindley, 1988) is when $\mathbf{p}_0 = \mathbf{p}(\omega)$, with $\mathbf{p}\colon \Omega \to \Theta$, 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 frequentist viewpoint we can use Pearson's χ^2 statistic as a 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 1, 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 frequentist and Bayesian evidence measures, we propose to use the following appropriate prior distribution

$$\pi^*(\theta) = \pi_0 \pi^0(\omega) I_{H_0}(\theta) + (1 - \pi_0) \pi(\theta) I_{H_1}(\theta),$$

where π_0 is the prior mass assigned to $H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}(\omega)$, $\pi^0(\omega)$ is a fixed prior, $\pi(\theta)$ is chosen in the class of $\varepsilon-$contaminated prior distributions, Γ, given in (1) and

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

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

In this case, the posterior probability of the null hypothesis, when the data of Table 1 has been observed, is

$$P(H_0 | n_{11}, \dots, n_{rs}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \frac{\int_{\Theta^r} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{n_{ij}} q_0(\theta) \, d\theta}{\int_{\Omega} \prod_{j=1}^s p_j(\omega)^{\sum_{i=1}^r n_{ij}} \pi^0(\omega) d\omega} \right]^{-1}.$$

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

5. COMMENTS

The obtained results are the consequence of the methodology based on the relation between the point null in (2) and the interval null hypothesis in (5). The discrepancy measure between $\pi(\theta)$ and $\pi^*(\theta)$ defined in (6) justifies the choice of π_0 as in (4) with an appropriate value of δ. According to this procedure the mixed prior distribution $\pi^*(\theta)$, used in the homogeneity null testing problem, is close to the density $\pi(\theta)$ used in the interval null testing problem

When $\pi(\theta)$ is in the class of $\varepsilon-$contaminated distributions, the lower bound of the posterior probability of the point null hypothesis computed in (11) can be close to the usual p-value. Gómez-Villegas and Sanz (2000) obtain similar results in a different context.

Furthermore, in the light of our results, it seems that the discrepancy observed in homogeneity testing problems between Bayesian and frequentist approaches becomes more acute by using $\pi_0 = 0.5$ in the mixed prior distribution.

Finally, the methodology proposed can be used to approach from a Bayesian viewpoint and compare with frequentist methods other point null hypothesis testing problems. For instance, to test the homogeneity of independent multinomial distributions with $r \times s$ tables when the functional form of $\mathbf{p}_0$ is known, $\mathbf{p}_0 = \mathbf{p}_0(\omega)$.

BIBLIOGRAPHY

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

Berger, J. O. (1994). An overview over robust Bayesian analysis (with discussion). *Test*, **3**, 5-124.

Berger, J. O. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with $\varepsilon-$contaminated priors. *Ann. Statist.*, **14**(2), 461-486.

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 Dedlampady, M. (1987). Testing Precise Hyphotheses, (with Discussion). *Statistical Science*, **2**(3), 317-352.

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

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

Delampady, M. and Berger, J. O. (1990). Lower bounds on Bayes factors for multinomial distributions, with application to chi-squared tests of fit. *The Annals of Statistics*, **18**(3), 1295-1316.

De la Horra, J. (2005). Reconciling classical and prior predictive p-values in the two sided location parameter testing problem. *Commun. Statist-Theory Meth.*, **34**, 575-583.

Gómez-Villegas, M. A. and Gómez, E. (1992). Bayes Factor in Testing Precise Hyphoteses. *Commun. Statist-Theory Meth.*, **21**, 1707-1715.

Gómez-Villegas, M. A. and González-Pérez, B. (2005). Bayesian Analysis of Contingency Tables. *Commun. Statist-Theory Meth.*, **34**(8), 1743-1754.

Gómez-Villegas, M. A. and González-Pérez, B. (2006). A Condition to Obtain the Same Decision in the Homogeneity Testing Problem from the Frequentist and Bayesian Point of View. *Commun. Statist-Theory Meth.*, **35**, 2211-2222.

Gómez-Villegas, M. A. and González-Pérez, B. (2008). $\varepsilon-$Contaminated Priors in Contingency Tables. *Test*, **17** (1), 163-178.

Gómez Villegas, M. A., Maín, P. and Sanz, L. (2002). A suitable Bayesian approach in testing point null hypothesis: some examples revisited. *Commun. Statist.-Theory Meth.*, **31**(2), 201-217.

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

14

Gómez-Villegas, M. A., Maín, P., Sanz, L. and Navarro, H. (2004). 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.

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*, **4**. University Press, Oxford, pp. 195-210 (with discussion).

Huber, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Int. Statist. Inst.*, **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.

Sellke, T., Bayarri, M. J. and Berger, O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, **55**(1), 62-71.

Sivaganesan, S. (1988). Range of posterior measures for priors with arbitrary contaminations. *Comm. Statist. A - Theory and Methods*, **17**, 1591-1612.

Sivaganesan, S. and Berger, J. (1989). Ranges of posterior measures for priors with unimodal contaminations. *The Annals of Statistics*, **17**(2), 868-889.