

A SUITABLE BAYESIAN APPROACH IN TESTING POINT NULL HYPOTHESIS: SOME EXAMPLES REVISITED

Miguel A. Gómez–Villegas, Paloma Maín and Luis Sanz

Departamento de Estadística e Investigación Operativa

Universidad Complutense de Madrid 28040 Madrid, Spain

ma.gv@mat.ucm.es

Key Words: p–values; posterior probability; point null; prior distribution.

ABSTRACT

In the problem of testing the point null hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, with a previously given prior density for the parameter θ , we propose the following methodology: to fix an interval of radius ε around θ_0 and assign a prior mass, π_0 , to H_0 computed by the density $\pi(\theta)$ over the interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$, spreading the remainder, $1 - \pi_0$, over H_1 according to $\pi(\theta)$. It is shown that for Lindley’s paradox, the Normal model with some different priors and Darwin–Fisher’s example, this procedure makes the posterior probability of H_0 and the p–value matching better than if the prior mass assigned to H_0 is 0.5.

1. INTRODUCTION

1.1 HISTORY

In parametric testing point null hypothesis it is known that Bayesian and classical methods can give rise to different decisions, see Lindley (1), Berger and Sellke (2) and Berger and Delampady (3) among others. These papers show that there is a discrepancy between the classical approach, expressed in terms of the p–value, and the Bayesian one, expressed in terms of the posterior probability of the point null hypothesis and the Bayes factor. Specifically, in most of Bayesian approaches the infimum of the posterior probability of the null hypothesis or the Bayes factor, over a wide class of prior distributions, is taken 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 of these cases the mass assigned to the point null hypothesis

is 0.5. On the other hand, Casella and Berger (4) show that there is no discrepancy in the one-sided testing problem.

In most of the existing contributions a class of priors distributions is used, but our objective is to check what happens when a single prior distribution is used. The methodology to be proposed is the one introduced by Gómez-Villegas and Gómez Sánchez-Manzano (5) and justified by Gómez-Villegas and Sanz (6) where it is shown that the infimum of the posterior probability can be close to the p-value when the class of priors is the class of all unimodal and symmetric distributions.

Some relevant references, comparing classical and Bayesian measures, in addition to those mentioned above, are Pratt (7), Edwards, Lindman and Savage (8), DeGroot (9), Bernardo (10), Rubin (11), Mukhopadhyay and DasGupta (12), Berger, Boukai and Wang (13, 14), and Oh and DasGupta (15).

In Section 1.2 we present the problem. In Section 2, the methodology is applied to the Jeffreys-Lindley paradox. Section 3 contains an example with a normal model and normal prior. In Section 4 the general framework for a normal model is analyzed and an example with the Cauchy model is considered. In Section 5 we deal with the famous example of Darwin-Fisher studied by Dickey (16). Finally, Section 6 contains some additional comments.

1.2 THE PROBLEM

We consider the point null testing problem for a location parameter

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0, \quad (1.1)$$

based on observing a random variable, X , with density $f(x - \theta)$ continuous in $\theta = \theta_0$. We will suppose that the prior information about θ is given by a density $\pi(\theta)$ over the parameter space Θ .

Then, the prior to test (1.1) will be given by a mixed distribution, $\pi^*(\theta)$, assigning mass $\pi_0 > 0$ to $\theta = \theta_0$ and spreading the remainder, $1 - \pi_0$, over $\theta \neq \theta_0$ according to the density $\pi(\theta)$,

$$\pi^*(\theta) = \pi_0 I_{\{\theta_0\}}(\theta) + (1 - \pi_0)\pi(\theta)I_{\{\theta \neq \theta_0\}}(\theta). \quad (1.2)$$

To choose π_0 , the mass assigned to the point null hypothesis, we propose, as it is usually done, the replacement of (1.1) by the more realistic precise hypothesis

$$H_{0\varepsilon} : |\theta - \theta_0| \leq \varepsilon \quad \text{versus} \quad H_{1\varepsilon} : |\theta - \theta_0| > \varepsilon, \quad (1.3)$$

where ε is suitably “small”. Examples of this replacement can be found in Berger (17), Berger and Delampady (3) and Lee (18) among others. Lindley (19) presents an interesting discussion about the difference between (1.1) and (1.3).

Now, given the density $\pi(\theta)$, it is possible to fix the value of ε that makes equivalent both problems, (1.1) and (1.3) and compute π_0 as

$$\pi_0 = \int_{|\theta - \theta_0| \leq \varepsilon} \pi(\theta) d\theta. \quad (1.4)$$

We think that the choice of ε is more intuitive than just selecting an arbitrary value for π_0 , usually 0.5 in the literature.

There are several comments in order to justify this approach. First, the value of ε corresponding to $\pi_0 = 0.5$ can be obtained from $\int_{|\theta - \theta_0| \leq \varepsilon} \pi(\theta) d\theta = 0.5$, but in this case the values of ε will not be suitably “small” except for excessively peaked prior densities. Secondly, if $\pi(\theta)$ is our prior information then $\pi^*(\theta)$, the mixed prior, must be near $\pi(\theta)$ in some sense and if we use the Kullback–Leibler information measure,

$$\delta(\pi^*|\pi) = \int \pi(\theta) \ln\{\pi(\theta)/\pi^*(\theta)\} d\theta, \quad (1.5)$$

as a measure of discrepancy between π and π^* , it holds that $\delta(\pi^*|\pi)$ goes to zero when ε goes to zero. However if one uses $\pi_1^*(\theta) = 0.5I_{\{\theta_0\}}(\theta) + 0.5I_{\{\theta \neq \theta_0\}}(\theta)\pi(\theta)$, then $\delta(\pi_1^*|\pi) = 0.693$ and it does not seem that H_0 can be approximated by $H_{0\varepsilon}$ in this case (see Appendix).

In any case, whichever the value of π_0 you choose, the posterior probability of the point null hypothesis is given by

$$P(H_0|x) = \left\{ 1 + \frac{1 - \pi_0}{\pi_0} \frac{m_\pi(x)}{f(x - \theta_0)} \right\}^{-1} \quad (1.6)$$

with $m_\pi(x) = \int f(x - \theta)\pi(\theta) d\theta$, the predictive distribution.

Finally, a classical measure of evidence against the null hypothesis, which depends on the observations, is the p-value. If there exists an appropriate statistic $T(X)$, for example a sufficient statistic, the p-value for testing (1.1) is given by $p(x) = P\{|T(X)| > |T(x)| | \theta_0\}$.

In this paper we wish to establish that the posterior probability of the point null hypothesis, with our methodology, is closer to the p-value than the posterior probability when the mass assigned to the point null is $\pi_0 = 0.5$, at least in the problems we have analyzed. Then, the cause of the discrepancy between the Bayesian and frequentist approximations seems to be more clear in these situations.

2. THE JEFFREYS–LINDLEY PARADOX

Lindley (1) studies the point null hypothesis (1.1) for a sample X_1, \dots, X_n when the model is $N(\theta, \sigma^2)$, with σ^2 known, and the prior distribution is the improper uniform distribution over all \Re . Then by (1.6),

$$P(H_0|\bar{x}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \sigma \left(\frac{2\pi}{n} \right)^{1/2} \exp \left\{ \frac{n}{2\sigma^2} (\bar{x} - \theta_0)^2 \right\} \right]^{-1}$$

where $\pi_0 = \int_{\theta_0 - \varepsilon}^{\theta_0 + \varepsilon} \pi(\theta) d\theta = 2\varepsilon$ in accordance with (1.4). If we take $\varepsilon = 0.1$, making H_0 equivalent to $H_{0\varepsilon}$ (see Berger and Delampady (3)), Table 1 shows that the posterior probability of the point null hypothesis, column 2, and the p-value, column 4, are close; whereas there is much more discrepancy with the posterior probability when we take $\pi_0 = 0.5$, column 5.

Table 1: Comparison between the posterior probability of the point null hypothesis and the p-value, with $X \sim N(\theta, 1)$, $\pi(\theta) = 1$, $n = 10$, $\varepsilon = 0.1$ and $t = |\bar{x} - \theta_0|n^{1/2}$.

t	$P(H_0 \bar{x})$	$P(H_{0\varepsilon} \bar{x})$	$p(\bar{x})$	$P(H_0 \bar{x}, \pi_0 = 0.5)$
1.645	0.0754	0.0670	0.1	0.2459
1.960	0.0442	0.0387	0.05	0.1563
2.576	0.0113	0.0100	0.01	0.0437
3.291	0.0014	0.0013	0.001	0.0058

If we take some other values of ε around 0.1, the results are similar. For example, if $\varepsilon = 0.15$ and $t = 1.96$, the posterior probability of H_0 is 0.0734 and the posterior probability of $H_{0\varepsilon}$ is $P(H_{0\varepsilon}|\bar{x}) = \int_{-\varepsilon}^{+\varepsilon} \pi(\theta|\bar{x}) d\theta = 0.0612$. On the other hand column 5, where $\pi_0 = 0.5$, can be obtained with our methodology just taking $\varepsilon = 0.25$.

3. AN EXAMPLE WITH NORMAL MODEL AND CONJUGATE PRIOR

Without loss of generality, consider testing the hypotheses $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ based on observing a random sample of size n from a population $N(\theta, 1)$. Now, a complete sufficient statistic for θ is the sample mean, \bar{X} , with $N(\theta, 1/n)$ distribution and if we suppose that the prior is distributed $N(0, \tau^2)$ then the predictive $m_\pi(x)$ is $N(0, \tau^2 + 1/n)$. So, using (1.4), $\pi_0 = 2\Phi(\varepsilon/\tau) - 1$, where Φ is the standard normal cumulative distribution function. We take $n = 10$, $\tau^2 = 2$ and a suitably small value of ε , $\varepsilon = 0.15$, for comparison between the p-values and the posterior probabilities of the point null hypothesis. Table 2 shows that the posterior probability of the null hypothesis and the p-value are close, whereas it is clearly shown the discrepancy between the p-value and the posterior probability for $\pi_0 = 0.5$.

Table 2: Comparison between the posterior probability of the point null hypothesis and the p-value, with $X \sim N(\theta, 1)$, $\theta \sim N(0, 2)$, $n = 10$ and $\varepsilon = 0.15$, $t = |\bar{x}|n^{1/2}$.

t	$P(H_0 \bar{x})$	$P(H_{0\varepsilon} \bar{x})$	$p(\bar{x})$	$P(H_0 \bar{x}, \pi_0 = 0.5)$
1.645	0.1044	0.1133	0.1	0.5582
1.960	0.0636	0.0686	0.05	0.4238
2.576	0.0176	0.0199	0.01	0.1628
3.291	0.0024	0.0031	0.001	0.0257

The prefixed value of ε is adequate since the posterior probabilities of the point null hypothesis and interval null hypothesis are close, as it is shown in Table 2, columns 2 and 3. On the other hand, if the value to be chosen directly for π_0 is 0.5, the posterior probabilities of the point null hypothesis are much larger than the p-values. In order to make, with our methodology, $P(H_{0\varepsilon}|\bar{x})$ close to $P(H_0|\bar{x}, \pi_0 = 0.5)$, it is necessary to choose $\varepsilon = 0.95$ but in

this case ε is so great that the point null hypothesis does not seem to be equivalent to the interval hypothesis.

Now, the following question arises: is it possible to choose an interval for ε , say $(\varepsilon_1, \varepsilon_2)$, so that taking a value of ε in the interval and assigning π_0 as in (1.4), the posterior probability of the point null hypothesis, (1.1), and the p-value match?

Naturally, there is a value of ε , depending on the data, so that the p-value and the posterior probability of the point null hypothesis are equal, but this is not our objective. The analysis of the Table 3 shows that if ε is included in the interval $(1/15, 1/7)$, then the posterior probabilities of the point null hypothesis are close to the p-values, for moderate values of the observations. Furthermore, for a value of ε in $(1/15, 1/7)$ it can be observed in Table 3 that, coherently, the posterior probability of the point null hypothesis is near the posterior probability of the interval null hypothesis. Thus, in this situation, the answer to the question stated above is affirmative.

Table 3: Values of ε matching the posterior probability and the p-value, with $X \sim N(\theta, 1)$, $\theta \sim N(0, 2)$ and $n = 10$.

t	ε	$P(H_0^* \bar{x}) \simeq p(\bar{x})$	$P(H_{0\varepsilon} \bar{x})$
1.645	0.143	0.1	0.1074
1.960	0.118	0.05	0.0526
2.576	0.088	0.01	0.0104
3.291	0.065	0.001	0.0010

4. COMPARISON BETWEEN THE P-VALUE AND THE POSTERIOR PROBABILITY

In this section the different properties of the posterior probability observed in previous sections are rejoined.

Next theorem shows the behaviour of the posterior probability of the point null hypothesis considered as a function of the observations and ε , the half length of the interval null hypothesis.

The p-value is now given by

$$p(t) = 2\{1 - \Phi(|t|)\} \quad (4.1)$$

with Φ the standard normal cumulative distribution function. The mass assigned to the point null hypothesis is, by (1.1),

$$\pi_0(\varepsilon) = \int_{-\varepsilon/\tau}^{\varepsilon/\tau} \pi(\theta) d\theta \quad (4.2)$$

Theorem 4.1 *Let X_1, \dots, X_n independent random variables from a $N(\theta, \sigma^2)$ distribution, with σ^2 known. Suppose that the prior information about θ is given by $\pi_\tau(\theta) = \pi(\theta/\tau)/\tau$, with $\tau > 0$ and $\pi(\theta)$ a continuous, symmetric and unimodal density with mode at θ_0 and $\sigma/\tau \leq 1$. To test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ consider, given H_0 , the sufficient statistic $T = \bar{X}\sqrt{n}/\sigma$, then*

i) For fixed t , $P(H_0|t, \varepsilon)$ is increasing as a function of ε , with $P(H_0|t, \varepsilon = 0) = 0$ and $\lim_{\varepsilon \rightarrow \infty} P(H_0|t, \varepsilon) = 1$,

ii) For fixed ε , $P(H_0|t, \varepsilon)$ is decreasing in t , for $t > 0$, being

$$\lim_{t \rightarrow 0} P(H_0|t, \varepsilon) = \int_{-\infty}^{+\infty} \exp(-u^2/2) \pi(ku/n^{1/2}) du$$

and $\lim_{t \rightarrow +\infty} P(H_0|t, \varepsilon) = 0$.

Proof: *i)* Using expression (1.6), we have

$$P(H_0|t, \varepsilon) = \left\{ 1 + \frac{1 - \pi_0(\varepsilon)}{\pi_0(\varepsilon)} \frac{k}{n^{1/2}} \int_{-\infty}^{+\infty} \frac{\varphi(t-u)}{\varphi(t)} \pi\left(\frac{ku}{n^{1/2}}\right) du \right\}^{-1}, \quad (4.3)$$

where φ is the standard normal density. Since $\pi_0(\varepsilon)$, given by (4.2), is increasing, then $\{1 - \pi_0(\varepsilon)\}/\pi_0(\varepsilon)$ is decreasing and, immediately, we obtain that $P(H_0|t, \varepsilon)$ is increasing in ε .

Moreover, $\pi_0(0) = 0$ and it is easy to see that, by (4.3), $P(H_0|t, \varepsilon = 0) = 0$.

Besides, when ε tends to ∞ , then $\pi_0(\varepsilon)$ tends to 1 and it is straightforward to check that $\lim_{\varepsilon \rightarrow +\infty} P(H_0|t, \varepsilon) = 1$.

ii) In the expression (4.3) the factor depending on t is

$$M(t) = \int_{-\infty}^{+\infty} \frac{\varphi(t-u)}{\varphi(t)} \pi\left(\frac{ku}{n^{1/2}}\right) du,$$

and its derivative with respect to t is

$$M'(t) = e^{t^2/2} \int_0^\infty \left(e^{-(u-t)^2/2} - e^{-(u+t)^2/2} \right) u \pi \left(\frac{ku}{n^{1/2}} \right) du.$$

We observe that $M'(t) > 0$, therefore $M(t)$ is increasing and then $P(H_0|t, \varepsilon)$ is decreasing in t .

The $\lim_{t \rightarrow 0} P(H_0|t, \varepsilon)$ depends on $\lim_{t \rightarrow 0} M(t)$, but

$$\lim_{t \rightarrow 0} M(t) = \lim_{t \rightarrow 0} \frac{\int_{-\infty}^{+\infty} \varphi(t-u) \pi(ku/n^{1/2}) du}{\varphi(t)} = \int_{-\infty}^{+\infty} \frac{\varphi(u)}{\varphi(0)} \pi \left(\frac{ku}{n^{1/2}} \right) du$$

as asserted.

On the other hand,

$$M(t) = \int_{-\infty}^0 \frac{\varphi(t-u)}{\varphi(t)} \pi \left(\frac{ku}{\sqrt{n}} \right) du + \int_0^\infty \frac{\varphi(t-u)}{\varphi(t)} \pi \left(\frac{ku}{\sqrt{n}} \right) du$$

and while the first integral is positive, the second diverges when t tends to $+\infty$. Then, $\lim_{t \rightarrow +\infty} M(t) = +\infty$, and $\lim_{t \rightarrow +\infty} P(H_0|t, \varepsilon) = 0$. \diamond

These results give the precise behaviour of $P(H_0|t, \varepsilon)$ as a function of t and ε . Besides the intuitive meaning of this theorem is that, observing the behaviour of the posterior probability of the point null hypothesis, an interval of values for ε can be determined such that the posterior probabilities and the p-values are close, see figure 1.

Figures 1 and 2 go here

Now, to apply the previous results, let us suppose a prior distribution, $\pi(\theta)$, $N(0, \tau^2)$ with $\sigma/\tau = k \leq 1$. In this case, the posterior probability is

$$P(H_0^*|t, \varepsilon) = \left[1 + \frac{1 - \pi_0(\varepsilon)}{\pi_0(\varepsilon)} \left\{ (2\pi)^{1/2} r \varphi \left(\frac{t}{r} \right) \right\}^{-1} \right]^{-1}, \quad (4.4)$$

with $r = (1 + k^2/n)^{1/2}$.

The p-value, given by (4.1), can be numerically equal to the posterior probability of H_0 by choosing a suitable value of ε which, naturally, will depend on the observed value t . Really, we just make (4.4) and (4.1) equal, and then the value $\varepsilon(t)$ can be obtained from

$$\pi_0\{\varepsilon(t)\} = \left\{ 1 + \frac{(2\pi n)^{1/2}}{\alpha k} \varphi(\alpha t) \left(\frac{1}{p(t)} - 1 \right) \right\}^{-1} \quad (4.5)$$

where $\alpha = (1 + k^2/n)^{-1/2}$ is, clearly, $\alpha \leq 1$. Now, we can prove that the solution of (4.5) satisfies the following proposition

Proposition 4.1 *The function $g(t) = \pi_0\{\varepsilon(t)\}$ is a continuous and decreasing function and $\lim_{t \rightarrow \infty} g(t) = 0$ and $\lim_{t \rightarrow 0} g(t) = 1$.*

Proof: In the expression (4.5) the part that depends on t is

$$\varphi(\alpha t) \left(\frac{1}{p(t)} - 1 \right) = \frac{\varphi(\alpha t)}{\varphi(t)} \frac{\varphi(t)}{2\{1 - \Phi(t)\}} - \varphi(\alpha t). \quad (4.6)$$

It is easy to see that $\varphi(\alpha t)/\varphi(t)$ and $-\varphi(\alpha t)$ are increasing functions for $t \geq 0$. Furthermore, $\varphi(t)/\{1 - \Phi(t)\}$ is increasing too, since the normal distribution has increasing failure rate.

For the limits it may be pointed out, given $\alpha \leq 1$, that

$$\lim_{t \rightarrow +\infty} \frac{\varphi(\alpha t)}{p(t)} = \lim_{t \rightarrow +\infty} \frac{\varphi(\alpha t)}{2\{1 - \Phi(t)\}} \geq \lim_{t \rightarrow +\infty} \frac{\varphi(t)}{2\{1 - \Phi(t)\}}$$

and this last limit is infinite by the Mills' ratio. Then, the proposition holds. \diamond

As a consequence $\varepsilon(t)$, the value that equals the posterior probability of H_0 and the p-value, is decreasing since $g(t)$ is decreasing. Then, if we have $t_1 < t < t_2$ we can get values of ε such that $\varepsilon(t_2) < \varepsilon < \varepsilon(t_1)$ for which the posterior probability and the p-value are alike.

For example, with $n = 10$ and $k = (1/2)^{1/2}$, if $t_1 = 1.645$ then $p(t_1) = 0.1$ and, using (4.5), $g(t_1) = 0.08$ and similarly when $t_2 = 3.291$ is $p(t_2) = 0.001$ and $g(t_2) = 0.036$. Then, numerical calculus show that if we take a value of ε in the interval $(0.065, 0.143)$, the posterior probability is similar to the p-value. It may be noted that these are the same values obtained in the previous section.

Therefore it is clear that the answer raised above is affirmative and it is possible, in this setup, to choose ε between two values so that the posterior probability is near to the p-value when a prior distribution as in (1.4) is used.

On the other hand, if we use a mixed prior distribution as (1.3) with a fixed value for π_0 , but not the value given by (1.4), then the p-value is smaller than the posterior probability of the null hypothesis. The following theorem gives an upper bound for the posterior probability that makes clear what elements have an influence on the approximation between this probability and the p-value.

Theorem 4.2 For fixed $t \geq 0$, under the same conditions of theorem 4.1, we have

$$P(H_0|\varepsilon, \tau) \leq \{1 + C(\pi, \tau, \varepsilon)\}^{-1} \quad (4.7)$$

where

$$C(\pi, \tau, \varepsilon) = \left(\frac{\tau}{2\varepsilon\pi(0)} - 1 \right) \frac{k}{n^{1/2}} \int_0^{2t} \frac{\varphi(t-u)}{\varphi(t)} \pi\left(\frac{ku}{n^{1/2}}\right) du \quad (4.8)$$

and φ is the standard normal density.

Proof: The prior distribution is given by $\pi_\tau(\theta) = \pi(\theta/\tau)/\tau$, with $\pi(\theta)$ symmetric about zero, then it follows

$$\pi_0(\varepsilon) = \int_{-\varepsilon/\tau}^{\varepsilon/\tau} \pi(\theta) d\theta \leq \pi(0) \frac{2\varepsilon}{\tau}.$$

Furthermore,

$$\int_{-\infty}^{+\infty} \frac{\varphi(t-u)}{\varphi(t)} \pi\left(\frac{ku}{n^{1/2}}\right) du = \int_{-\infty}^{+\infty} e^{-u(u-2t)/2} \pi\left(\frac{ku}{n^{1/2}}\right) du$$

but $h(u) = e^{-u(u-2t)/2}$ has a maximum in $u = t$, $h(t) = e^{t^2/2}$, and $h(u) < 1$ for $u < 0$ and $u > 2t$, so that the more significant values of this function are in the interval $(0, 2t)$. Then, the last integral can be approached by

$$\int_0^{2t} e^{-u(u-2t)/2} \pi\left(\frac{ku}{n^{1/2}}\right) du$$

and then we obtain expression (4.7) with $C(\pi, \tau, \varepsilon)$ as in (4.8). \diamond

It may be pointed out that: firstly, for π and τ fixed, the bound $\{1 + C(\pi, \tau, \varepsilon)\}^{-1}$ decreases when ε decreases. Secondly, for fixed values of ε and τ the posterior probability decreases when $\pi(2kt/\sqrt{n})$ increases and this happens if we use densities with heavier tails. Thirdly, for ε and π fixed when the variance of the prior distribution increases, the bound decreases. Therefore, in all of these cases, $C(\pi, \tau, \varepsilon)$ increases and then the upper bound of the posterior probability decreases and comes closer to the p-value.

Although theoretical proofs involve awkward integrals, we introduce a case where the underlying distribution is non-normal using simulation tools. Consider independent observations from a Cauchy distribution, $C(\theta, 1)$, and suppose that the prior over θ is $C(\theta_0, 2)$. Then, Table 4 shows the values of ε for which the p-value and the posterior probability of the point null hypothesis are keep equal.

Table 4: Values of ε matching the posterior probability and the p-value, with $X \sim C(\theta, 1)$, $\theta \sim C(0, 2)$.

x	ε	$P(H_0 x) \simeq p(x)$	$P(H_{0\varepsilon} \bar{x})$
1.0	4.84	0.500	0.988
1.5	3.67	0.374	0.966
3.0	2.76	0.205	0.638
6.3	1.64	0.100	0.185
12.7	0.87	0.050	0.092

5. DARWIN'S EXAMPLE

This Darwin's example is studied by Fisher (21) in his classical book "The Design of Experiments" and has also been studied by Dickey (16). This is a typical case in which the point null hypothesis could be replaced by an interval one. The experiment tries to determine whether cross-fertilized plants have a greater growing rate than self-fertilized plants. Measurements of the differences in height of pairs of similarly grown plants were taken on a certain date.

Then, if X_1, \dots, X_n are the differences in height of n pairs of plants, it can be supposed that they come from a population $N(\theta, \sigma^2)$, with σ^2 unknown, so we want to test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. Fisher uses the statistic $T = n^{1/2}(\bar{X} - \theta_0)/s$, where s^2 is the sample variance, and the critical region for a test with $\alpha = 0.05$ is $RC = \{|T| > 2.145\}$ with $T = n^{1/2}\bar{X}/s$ given H_0 . The $n = 15$ observed differences showed $\bar{x} = 2.6166$ and $s = 4.7188$ in inches. Hence the statistic is $t_{14} = 2.1476$ and the test was barely significant.

As in Dickey (16), because σ^2 is a nuisance parameter, we consider independent prior distributions for the mean, $\pi(\theta|\nu_0, \theta_0 = 0, s_0^2) \sim t_{\nu_0}$, the Student-t density and for the variance, σ^2 , in the family $s_1^2/(\chi_{\nu_1}^2/\nu_1)$. In this case

$$\pi(\theta|\nu_0, \theta_0 = 0, s_0^2) = \left(1 + \frac{\theta^2}{\nu_0 s_0^2}\right)^{-(\nu_0+1)/2} \frac{1}{s_0} \frac{\Gamma\{(\nu_0+1)/2\}}{(\nu_0\pi)^{1/2}\Gamma(\nu_0/2)}, \quad \theta \in \mathfrak{R}$$

and

$$\pi(\sigma^2|\nu_1, s_1^2) = \frac{1}{2^{\nu_1/2}} \frac{1}{\Gamma(\nu_1/2)} e^{-\nu_1 s_1^2/(2\sigma^2)} (\nu_1 s_1^2)^{\nu_1/2} \frac{1}{(\sigma^2)^{-(\nu_1/2+1)}}, \quad \sigma^2 > 0.$$

A marginal, or integrated, likelihood function of θ proportional to a Student-t density can be obtained

$$f(\bar{x}|\theta) = \int_0^\infty \varphi(\bar{x}|\theta, \sigma^2)\pi(\sigma^2) d\sigma^2$$

and it results

$$f(\bar{x}|\theta) = \frac{\Gamma\{(m+1)/2\}}{(m\pi)^{1/2}\Gamma(m/2)} \left\{ 1 + \frac{n}{m} \frac{(\theta - \bar{x})^2}{\tau^2} \right\}^{-(m+1)/2} \frac{n^{1/2}}{\tau}$$

the Student-t distribution with $m = n - 1 + \nu_1$ degrees of freedom and $\tau^2 = \{(n - 1)s^2 + \nu_1 s_1^2\}/m$. Hence, the marginal

$$m(\bar{x}) = \int_{-\infty}^{+\infty} f(\bar{x}|\theta)\pi(\theta) d\theta.$$

Then, the posterior probability of the point null hypothesis is computed as in (1.6) with $\theta_0 = 0$ and π_0 , the mass assigned to the point null hypothesis, as in (1.4).

In this case, the prior scale for σ^2 , s_1^2 , is taken equal to the sample variance, $s_1^2 = 4.7188^2$. Then, the parameters that we need to fix are ν_0 , ν_1 , s_0^2 and ε . Table 4 shows the results when $\pi_0 = 0.5$ and different values of the parameters as in Dickey (16).

Table 5: $P(H_0|\bar{x}, \pi_0 = 0.5)$ for the Darwin-Fisher-Dickey example

s_0	ν_0	ν_1				
		0	4	8	20	100
1	7	0.3549	0.3497	0.3461	0.3399	0.3308
	100	0.3652	0.3601	0.3566	0.3505	0.3416
2.5	7	0.3031	0.2972	0.2931	0.2861	0.2761
	100	0.2934	0.2876	0.2836	0.2767	0.2668
10	7	0.5087	0.5018	0.4969	0.4885	0.4762
	100	0.4992	0.4922	0.4874	0.4790	0.4667
125	7	0.9251	0.9231	0.9218	0.9193	0.9156
	100	0.9227	0.9208	0.9193	0.9168	0.9130

For a fixed s_0 , the posterior probabilities are robust with respect to the shape ν_0 of the prior density of θ and to the degree ν_1 of the prior distribution of σ^2 . Moreover, it can be noted that the posterior probability of the point null hypothesis tends to one when the conditional prior dispersion, s_0^2 , of θ increases. That is, we can get a posterior probability close to one just increasing s_0^2 , but it means that if the prior distribution tends to give less knowledge about θ then the posterior probability of the point null hypothesis becomes greater. It does not look reasonable.

This behavior does not happen with our methodology. For example, if we take $\varepsilon = 0.2$ and compute π_0 as in (1.4), general robustness is apparent too, but now when s_0^2 increases the posterior probability decreases as it is presented in Table 5, obviously this behaviour is more intuitive.

Table 6: $P(H_0|\bar{x}, \varepsilon = 0.2)$ for Darwin's example

s_0	ν_0	ν_1				
		0	4	8	20	100
1	7	0.0903	0.0884	0.0872	0.0850	0.0819
	100	0.0975	0.0956	0.0943	0.0920	0.0889
2.5	7	0.0277	0.0270	0.0265	0.0256	0.0244
	100	0.0274	0.0267	0.0262	0.0253	0.0241
10	7	0.0159	0.0155	0.0152	0.0147	0.0140
	100	0.0159	0.0154	0.0151	0.0147	0.0140
125	7	0.0149	0.0146	0.0143	0.0139	0.0132
	100	0.0149	0.0146	0.0143	0.0139	0.0132

Although it is not included here, it can be checked that the posterior probability of the interval hypothesis, computed as $P(H_{0,0.2}|\bar{x}) = \int_{-0.2}^{0.2} \pi(\theta|\bar{x}) d\theta$ where $\pi(\theta|\bar{x})$ is the posterior distribution of θ given \bar{x} , is close to $P(H_0|\bar{x})$ the posterior probability of the point null hypothesis.

As in previous sections and with our methodology, it is possible to get values of ε so that the posterior probability and the p-value match. In Table 6 we can see that for $s_0 = 2.5$, the choice of $\varepsilon = 0.379$ leads to a posterior probability of 0.05.

Table 7: Values of ε so that the posterior probability and the p-value match in Darwin's example, $s_1 = 4.7188$.

s_0	ν_0	ν_1	ε	$p(t) \simeq P(H_0 \varepsilon, \bar{x} = 2.6166)$	$P(H_{0\varepsilon} t)$	$P(H_0 \pi_0 = 0.5)$
1	7	20	0.121	0.05	0.0481	0.3399
2.5	7	20	0.379	0.05	0.0489	0.2869
10	7	20	0.680	0.05	0.0579	0.4889
125	7	20	0.747	0.05	0.0625	0.9193

6. CONCLUDING REMARKS

In some situations, using our methodology, it is possible to get a better agreement between the posterior probability of the point null hypothesis, as a measure of Bayesian evidence, and the classical p-value, as it is shown in the examples we have studied.

Furthermore, if suitable values of ε are chosen, the differences between the posterior probability of the point null hypothesis and the p-value are not so large as if the value for π_0 is taken 0.5 directly. Really, if in testing a point null hypothesis, a mixed prior distribution with $\pi_0 = 0.5$ is used, there will be a remarkable discrepancy between Bayesian and classical evidence.

The ε chosen must be such that the posterior probabilities of the point and interval null hypotheses are similar, to be coherent with the substitution of H_0 by $H_{0\varepsilon}$. For the cases we have handled, these values of ε are within a limited range, see Tables 3 and 7, where the corresponding p-values and posterior probabilities are also very similar. Also the comment in Section 1 may be considered: the value of ε for $\pi_0 = 0.5$ can be used to get the upper bound of ε — however smaller values are recommended.

Using the Bayesian approach, this procedure gives a result close to the classical approach in testing point null hypothesis as the situation observed in the one-sided testing problem.

APPENDIX

There is a problem in using (1.5) as a measure of discrepancy between π and π^* because $\pi(\theta)$ is a density but $\pi^*(\theta)$ is not a density. We can sort out the problem considering two measures on $(\mathcal{R}, (\mathcal{B})_{\mathcal{R}})$. For all A in the Borel σ -field we define:

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

The measure $\mu(A)$ is originated by the density $\pi(\theta)$ and $\mu^*(A)$ by $\pi^*(\theta)$, with π_0 given by (1.4) and λ the Lebesgue measure.

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

$$\frac{d\mu}{d\mu^*}(\theta) = \begin{cases} 0 & \text{if } \theta = \theta_0, \\ \frac{1}{1 - \pi_0} & \text{if } \theta \neq \theta_0 \end{cases} \quad (\text{A.1})$$

Now, using that $\mu \ll \mu^*$, we can define the discrepancy between μ and μ^* as $\delta(\mu^*|\mu) = \int_{\Theta} (\ln(d\mu/d\mu^*)) d\mu$. Then, by (A.1), we have $\delta(\mu^*|\mu) = -\ln(1 - \pi_0)$.

ACKNOWLEDGEMENTS

We are very grateful to the Editor and two anonymous referees for their helpful comments and valuable suggestions on a previous version of the paper. This research has been sponsored by DGES (Spain) under grant PB-98-0797.

BIBLIOGRAPHY

- (1) Lindley, D.V. A statistical paradox. *Biometrika*, **1957**, *44*, 187–192.
- (2) Berger, J.O.; Sellke, T. Testing a point null hypotheses: The irreconcilability of p-values and evidence, (with discussion). *J. Amer. Statist. Assoc.*, **1987**, *82*, 112–139.
- (3) Berger, J.O.; Delampady, M. Testing precise hypotheses, (with discussion). *Statistical Science*, **1987**, *2*(3), 317–352.

- (4) Casella, G.; Berger, R.L. Reconciling Bayesian and frequentist evidence in the one-sided testing problem, (with discussion). *J. Amer. Statist. Assoc.* **1987**, *82*, 106–135.
- (5) Gómez-Villegas, M.A.; Gómez Sánchez-Manzano, E. Bayes factor in testing precise hypotheses. *Commun. Statist. - Theory Meth.*, **1992**, *21*, 1707-1715.
- (6) Gómez-Villegas, M.A.; Sanz, L. Reconciling Bayesian and frequentist evidence in the point null testing problem. *Test*, **1998**, *7*(1), 207–216.
- (7) Pratt, J.W. Bayesian interpretation of standard inference statements. *J. Roy. Statist. Soc. B*, **1965**, *27*, 169–203
- (8) Edwards, W.L.; Lindman, H.; Savage, L.J. Bayesian Statistical Inference for Psychological Research. *Psychol. Rev.*, **1963**, *70*, 193–248. Reprinted in *Robustness of Bayesian Analysis* (J.B. Kadane, ed.) Amsterdam: North-Holland, **1984**, 1–62.
- (9) DeGroot, M.H. Doing what comes naturally: Interpreting a tail area as a posterior probability or as a likelihood ratio. *J. Amer. Statist. Assoc.*, **1973**, *68*, 966–969.
- (10) Bernardo, J.M. A Bayesian analysis of classical hypothesis testing in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* (University Press, Valencia), **1980**, pp. 605–647 (with discussion).
- (11) Rubin, D.B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **1984**, *12*, 1151–1172
- (12) Mukhopadhyay, S.; DasGupta, A. Uniform approximation of Bayes solutions and posteriors: Frequentistly valid Bayes inference. *Statistics and Decisions*, **1997**, *15*, 51–73.
- (13) Berger, J.O.; Boukai, B.; Wang, Y. Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science*, **1997**, *12*(3), 133–160.
- (14) Berger, J.O.; Boukai, B.; Wang, Y. Simultaneous Bayesian–Frequentist sequential testing of nested hypothesis. *Biometrika*, **1999**, *86*, 79–92.

- (15) Oh, H.S.; DasGupta, A. Comparison of the p-value and posterior probability. *J. of Statist. Planning and Inference*, **1999**, *76*, 93–107
- (16) Dickey, J.M. Is the tail area useful as an approximate Bayes factor? *J. Amer. Statist. Assoc.*, **1977**, *72*, 138–142.
- (17) Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, New York, **1985**.
- (18) Lee, P.M. *Bayesian Statistics: An Introduction*, Charles Griffin, London, **1994**.
- (19) Lindley, D.V. Statistical inference concerning Hardy-Weinberg equilibrium. *Bayesian Statistics 3*, J.M. Bernardo et al. Oxford University Press. **1988**, 307–326.
- (20) Gómez-Villegas, M.A.; Sanz, L. ϵ -contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statist. and Probab. Lett.*, **2000**, *47*(1) 53–60.
- (21) Fisher, R.A. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd, **1956** / **1973**. Reprinted in **1990** within *Statistical Methods, Experimental Design and Scientific Inference* (J.M. Bennet, ed.) Oxford: University Press.