Asymptotic relationships between posterior probabilities and p–values using the hazard rate

Miguel A. Gómez–Villegas[1], Paloma Maín and Luis Sanz

Departamento de Estadística e Investigación Operativa,

Universidad Complutense de Madrid. 28040 Madrid, Spain.

Hilario Navarro

Departamento de Estadística, Investigación Operativa y Cálculo Numérico,

Facultad de Ciencias, Universidad Nacional de Eduación a Distancia, 28040 Madrid, Spain.

Abstract

In this paper the asymptotic relationship between the classical p-value and the infimum (over all unimodal and symmetric distributions) of the posterior probability in the point null hypothesis testing problem, is analyzed. It is shown that the ratio between the infimum and the classical p-value has an equivalent asymptotic behaviour to the hazard rate of the sample model.

AMS classification: 62A15.

Keywords: Hazard rate, P–value, Point null hypothesis, tail-orderings.

1. Introduction

In testing a point null hypothesis, it is well-known that the discrepancy between the classical p-value, from now on p–value, and the posterior probability of the null hypothesis for some kind of mixed prior distributions, see Berger and Sellke (1987) and Berger and Delampady (1987). Recently, it has been studied that a better approximation between Bayesian and

[1]Corresponding author: Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain. E mail:ma.gv@mat.ucm.es

classical approaches can be obtained, if the mass assigned to the null hypothesis is related to the prior density over the alternative hypothesis. The point is to work with wide classes of prior distributions, see Spiegelhalter and Smith (1982), Gómez-Villegas and Gómez Sánchez-Manzano (1992), McCulloch and Rossi (1992) and Gómez-Villegas and Sanz (1998, 2000). This better approximation is also present, in some particular cases, when only one prior distribution is used, see Gómez–Villegas, Maín and Sanz (2002). The discrepancy can also be avoided by using, in the Bayesian approach, a value that is sometime referred to as a Bayesian p–value. However, this procedure is not going to be considered in this paper. It can also be pointed out that such a discrepancy does not exist in the one-sided testing problem, see Casella and Berger (1987).

This paper deals with the asymptotic behavior of the ratio between the infimum of the posterior probability of the point null hypothesis and the classical p-value when the class of prior distributions is the class of all unimodal and symmetric distributions. In fact it is shown how this ratio depends on the hazard rate of the sample model. This relation is important to explain the influence of the model when the Bayesian and classical methodologies are compared in a point null hypothesis testing problem.

It may also be pointed out that the cited papers dealing with the discrepancy between the infimum of the posterior probability and the p-value, do not take into account the influence of the hazard rate function of the model. Using this fact, it will be shown that, if we have asymptotically high hazard rate values, it is possible to avoid the discrepancy for suitable small values of ε, but in the opposite case much larger values of ε will be needed, so that the infimum of the posterior probability and the p–value match.

In order to establish this comparison, the tail distribution classification introduced by Maín (1989) and studied by Gómez-Villegas and Maín (1992), so as the corresponding tail-ordering considered by Maín and Navarro (1997) are used.

Section 2 reviews the Bayesian framework of testing point null hypotheses that we have used and gives previous definitions including the asymptotic tail ordering used in this work. Section 3 presents the main result and its application to different sample models. Finally, in Section 4 some conclusions and comments are also given.

2. Preliminaries

Consider a point null testing problem

$$H_0^* : \theta = \theta_0 \quad \text{versus} \quad H_1^* : \theta \neq \theta_0, \tag{1}$$

based on observing a random variable, X, with density $f(x - \theta)$ continuous in $\theta = \theta_0$. The Bayesian approach considered in this paper supposes, as is often done, that the probability of $\theta = \theta_0$ is $\pi_0 > 0$, and such that the prior information is given by a mixed distribution assigning π_0 to the point $\theta = \theta_0$ and spreading the remainder, $1 - \pi_0$, according to a density $\pi(\theta)$ over $\theta \neq \theta_0$. In order to make comparisons with the p-value, $\pi(\theta)$ is usually chosen from a class of distributions. Furthermore, in many situations the choice of a particular prior distribution can be difficult. Thus, it will be assumed that $\pi(\theta)$ belongs to the class $G_{US} = \{$all distributions unimodal and symmetric about $\theta_0\}$, which is a reasonable class of priors since the symmetry is a natural "objective" assumption. Besides, that requirement is equivalent to the assumption that $\pi(\theta)$ is nonincreasing in $|\theta - \theta_0|$. So, when $\pi(\theta)$ is in G_{US} it seems that alternative values of θ (to θ_0) are not favored according to the point null hypothesis being studied. Some other justifications for using this class of priors can be found in Berger (1994), Casella and Berger (1987), Berger and Sellke (1987) and Gómez-Villegas and Sanz (1998).

To choose the mass assigned to the null hypothesis, π_0, we propose the following procedure: a precise hypothesis can be represented as

$$H_0 : |\theta - \theta_0| \leq \varepsilon \quad \text{versus} \quad H_1 : |\theta - \theta_0| > \varepsilon, \tag{2}$$

where ε is "small" and the point null hypothesis is replaced by this interval hypothesis. Then, given $\pi(\theta)$, a value of ε can be fixed to compute

$$\pi_0 = \int_{|\theta - \theta_0| \leq \varepsilon} \pi(\theta)\, d\theta \tag{3}$$

leading to the mixed prior distribution

$$\pi^*(\theta) = \pi_0 I_{\{\theta = \theta_0\}}(\theta) + (1 - \pi_0) I_{\{\theta \neq \theta_0\}}(\theta) \pi(\theta), \tag{4}$$

where $I_A(\theta) = 1$ if $\theta \in A$ and $I_A(\theta) = 0$ if $\theta \in A^c$. To justify this choice of the mixed distribution see Gómez–Villegas and Sanz (2000), the idea is to make compatible both problems, the point and the interval null hypotheses.

Furthermore, if $\pi_0 = 0.5$ is used, as it is usually done in the literature, the corresponding value of ε computed by (3) is very large. Then, the mixed prior distribution given by (4) for this π_0 does not seem reasonable because the point and the interval null hypotheses would not be equivalent problems.

Now, we are going to use the hazard rate function, $r_{f_X}(x) = f_X(x)/(1 - F_X(x))$ for the continuous case, to describe the influence that the tail behaviour of the sample model has on the asymptotic discrepancy between the p–value and the infimum of the posterior probability of the point null hypothesis over the class G_{US}. For some other uses and properties of the hazard rate function see Barlow and Proschan (1975).

The asymptotic hazard tail ordering to be considered, Maín and Navarro (1997), can be defined by

$$F \preceq_{t_h} G \text{ if and only if there exists a value } c < \infty \text{ such that } \lim_{x\to\infty} r_g(x)/r_f(x) = c.$$

For example, some of the usual distributions are ordered with this tail ordering as follows:

$$Normal \prec_{t_h} Gamma \sim_{t_h} Logistic \prec_{t_h} Lognormal \prec_{t_h}$$

$$Student \sim_{t_h} Pareto \sim_{t_h} Cauchy.$$

There are some other tail orderings (see Rojo 1992 and Shaked and Shanthikumar 1994) using some other features of the tail distributions, but in our problem the hazard rate function seems to be the most proper tool.

3. Main results

Let us suppose that a random variable X, having density $f(x - \theta)$, θ being an unknown parameter is observed. For the point null testing problem (1), the usual frequentist measure of evidence against H_0^* is the p-value, that is

$$p(x) = Pr_{\theta=\theta_0}(|T(X)| \geq |T(x)|), \tag{5}$$

where $T(X)$ is an appropriate statistic.

The next result justifies the different approximations between the p–value and the posterior probability, observed when distributions with different tails are used.

Theorem 3.1 *If the function f is continuous in θ_0 and symmetric and all the limits to be handled exist, then*

$$\lim_{x\to\infty} \frac{\inf_{\pi\in G_{US}} Pr(H_0^*|x)}{r_f(x)p(x)} = \varepsilon, \tag{6}$$

where ε is the half–length of the interval hypothesis in (2).

Proof: For testing (1) the posterior probability of the null H_0^* is

$$Pr(H_0^*|x) = \frac{f(x-\theta_0)\pi_0}{f(x-\theta_0)\pi_0 + (1-\pi_0)\int_{\theta\neq\theta_0} f(x-\theta)\pi(\theta)d\theta}$$

with $\pi(\theta) \in G_{US}$.

Being $\pi(\theta)$ unimodal and symmetric it can be written as a mixture of uniform distributions (see Brandwein and Strawderman 1978).

And, computing the infimum over G_{US} or over the class G_U, of all uniform distributions $U(\theta_0 - k, \theta_0 + k)$ with $k \in \Re$, is the same (see Casella and Berger 1987). Replacing π_0 by (3) for the class G_U

$$Pr(H_0^*|x) = \frac{2f(x-\theta_0)}{2f(x-\theta_0) + (\frac{1}{\varepsilon} - \frac{1}{k})\int_{|\theta-\theta_0|\leq k} f(x-\theta)d\theta},$$

which is decreasing in k, the infimum is reached when k goes to infinity. Observing that,

$$\int_{|\theta-\theta_0|\leq k} f(x-\theta)d\theta \leq \int_{\Re} f(x-\theta)d\theta = 1,$$

the infimum is

$$\inf_{\pi\in G_{US}} Pr(H_0^*|x) = \left(1 + \frac{1}{2\varepsilon}\frac{1}{f(x-\theta_0)}\right)^{-1}. \tag{7}$$

On the other hand, the p-value of observed data is, from (5) with $T(X) = X$

$$p(x) = 2(1 - F(x-\theta_0)),$$

then the ratio between the infimum and the p-value is given by

$$\begin{aligned} \frac{\inf_{\pi\in G_{US}} Pr(H_0^*|x)}{p(x)} &= \frac{\varepsilon f(x-\theta_0)}{(2\varepsilon f(x-\theta_0)+1)(1-F(x-\theta_0))} \\ &= \frac{\varepsilon r_f(x-\theta_0)}{2\varepsilon f(x-\theta_0)+1}, \end{aligned} \tag{8}$$

where $r_f(x-\theta_0)$ is the hazard rate of the sample model that, for large x, reflects the different tail behavior of the sample distributions.

From (8), the result (6) is immediately obtained. $\square$

Theorem 3.1 shows that for large x we have

$$\frac{\inf_{\pi\in G_{US}} Pr(H_0^*|x)}{p(x)} \approx \varepsilon r_f(x)$$

explaining how the comparison between the p-value and the infimum of the posterior probability of the point null hypothesis depends asymptotically on the hazard rate of the sample model, that is on the tail behavior of the sample model.

In fact we get that for a Normal distribution and for large x

$$\frac{\inf_{\pi\in G_{US}} Pr(H_0^*|x)}{p(x)} \approx \varepsilon x.$$

It means that the value of ε must be small to match the different measures. More concretely for a value $x = 3$ taking $\varepsilon = 1/3$, the posterior probability and the p-value are close. Otherwise, the point null hypothesis can be changed to the interval one with $\varepsilon = 1/3$ and in this case the infimum of the posterior probability is 0.99 times the p-value.

Whereas for a Cauchy distribution

$$\frac{\inf_{\pi\in G_{US}} Pr(H_0^*|x)}{p(x)} \approx \varepsilon \frac{1}{x},$$

then for a large x a large ε is required to make the Bayesian approach we have used and the classical one agree. Then for $x = 3$ a very large $\varepsilon = 3$ is necessary to make the Bayesian and classical measures of evidence equal. Alternatively the point null hypothesis might be changed by the interval one with $\varepsilon = 3$. For this heavy tailed model the needed value of ε is too large to consider both the point and the interval hypotheses equivalent, and using a proper small ε makes the infimum of the posterior probability strictly less than the p-value.

The following two examples show the effect produced by the use of increasing values of x for a couple of sample models. Thus in the first case, a heavy-tailed distribution, the convergence is slower than in the second one where a medium-tailed distribution is considered.

Example 1.(*Heavy-tailed distribution*). Let X be a Pareto random variable with density

$$f(x-\theta) = \frac{a}{2x_0}\left(\frac{x_0}{x_0+|x-\theta|}\right)^{a+1}; \quad -\infty < x < \infty, \quad a > 0$$

To test (1) , with $\theta_0 = 0$, the ratio between the infimum of the posterior probability and the p-value results

$$\frac{\inf_{\pi \in G_{US}} Pr(H_0^*)}{p(x)} = \frac{\varepsilon a (x_0 + |x|)^a}{\varepsilon a x_0^a + (x_0 + |x|)^{a+1}}.$$

Obviously, if this last expression is multiplied by $(r_f(x))^{-1}$ it gives ε. Table 1 shows, in a particular case, how this limit is attained for increasing values of x. The infimum is noted by $\underline{Pr}(H_0^*|x)$.

Table 1: Comparisons for the Pareto distribution with $a = 2$, $x_0 = 1$ and $\varepsilon = 0.2$.

x	$p(x)$	$\underline{Pr}(H_0^*\|x)$	$\underline{Pr}(H_0^*\|x)/p(x)$	$\varepsilon \times r_f(x)$
1	0.250	0.0476	0.1905	0.4
5	0.0278	0.0019	0.06654	0.08
10	0.0083	0.0003	0.03635	0.02
50	3.85×10^{-4}	3.0×10^{-6}	0.00784	0.008
100	9.8×10^{-5}	3.9×10^{-7}	0.00396	0.004
300	1.1×10^{-6}	1.47×10^{-8}	0.00133	0.00133

Example 2.(*Medium-tailed distribution*). Let X be a random variable with double-exponential density

$$f(x-\theta) = \frac{1}{2}e^{-|x-\theta|}; \quad -\infty < x < \infty.$$

To test (1), with $\theta_0 = 0$, the ratio between the infimum of the posterior probability and the p-value is

$$\frac{\inf_{\pi \in G_{US}} Pr(H_0^*|x)}{p(x)} = \frac{\varepsilon}{1+\varepsilon e^{-|x|}}.$$

Numerical results for some specific values are given in Table 2.

Table 2: Comparisons for the double exponential distribution with $\varepsilon = 0.2$.

x	$p(x)$	$\underline{Pr}(H_0^*\|x)$	$\underline{Pr}(H_0^*\|x)/p(x)$	$\varepsilon \times r_f(x)$
1	0.36788	0.06853	0.1863	0.2
3	0.04979	0.00986	0.1980	0.2
5	0.00674	0.00135	0.1997	0.2
10	4.54×10^{-5}	9.08×10^{-6}	0.1999	0.2
15	3.05×10^{-7}	6.12×10^{-8}	0.1999	0.2
20	2.06×10^{-9}	4.12×10^{-10}	0.2	0.2

In summary, with heavy tails and $\varepsilon = 0.2$ (Example 1), it is necessary $x > 300$ to get the ratio $\inf_{\pi\in G_{US}} Pr(H_0^*|x)/p(x)$ approximately equal to $\varepsilon \times r_f(x)$ (see Table 1). On the other hand, if the tail of the sample model is light (Example 2), $x > 10$ is enough to obtain the same result.

Until now, this kind of comparison have been done before but never taking into account the hazard rate as definitive in order to explain the different situations. For instance, Berger and Sellke(1987) in Comment 5 say that in the most statistical problems the infimum of the posterior probability is substantially larger than the p-value but this is not true when the sample model is a Cauchy distribution. This fact is also pointed out in Casella and Berger(1987).

In this paper we have shown the major influence of the hazard tail behavior on these comparisons.

4. Conclusions and comments

Summing up, in testing point null hypothesis, the asymptotic behavior of the ratio between the infimum of the posterior probability of the point null hypothesis, over a wide class of priors, and the p-value depends on the hazard rate of the sample model. This is a new argument to explain the discrepancy between $\inf_{\pi\in G_{US}} \Pr(H_0 \mid t)$, if a mixed prior is used, and $p(t)$ that has been previously observed.

So, if the sample model is a heavy-tailed distribution, for example Cauchy, t_n-Student or Pareto, the posterior probability of the point null hypothesis can be smaller, at least for a prior in the class, than the p-value for an appropriate value of ε. Whereas if the sample model is a light-tailed distribution, for example the Normal model, the posterior probability of the null is, at least for a prior in the class, equal to the p-value.

In any case, we judge that these kind of results are helpful to the better understanding of the actual peculiarity in the frame of point null hypothesis testing and they complement some other well-known ones about this particular problem.

Acknowledgements

We are very grateful to the Editor and to an anonymous referee 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.

References

Barlow, R:E: and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing.* New York, Holt, Reinhart and Winston.

Berger, J.O. (1994). An overview over robust Bayesian analysis (with discusion). *Test*, **3**, 1, 5-124.

Berger, J.O. and Delampady, M. (1987). Testing precise hypothesis. *Statist. Sci.*, **2**, 3, 217-352.

Berger, J.O. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p-values and evidence, (with discussion). *J. Amer. Statist. Assoc.*, **82**, 112-122.

Brandwein, A.C. and Strawderman, W.E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.*, **6**, 377-416.

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**, 106-111.

Gómez-Villegas, M.A. and Maín, P. (1992). The influence of prior and likelihood tail behavior on the posterior distribution. *Bayesian Statistics 4*, (J.M. Bernardo, M.H. DeGroot, D.V.

Lindley and A.F.M. Smith, Eds.). Oxford: University Press, 661-667.

Gómez-Villegas, M.A. and Gómez Sánchez-Manzano, E. (1992). Bayes Factor in Testing Precise Hypothesis. *Comm. Statist. Theory Methods,* **21**, 6, 1707-1715.

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

Gómez-Villegas, M.A. and Sanz, L. (2000). ε-contaminated prior distributions in testing point null hypothesis: a procedure to determine the prior probability. *Statist. Probab. Lett.*, **47**, 53-60.

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

Maín, P. (1989). Asymptotic behavior of reliability functions. *Statist. Probab. Lett.*, **7**, 259-263.

Maín, P. and Navarro, H. (1997). On tail behavior in Bayesian location inference. *Statist. Probab. Lett.*, **35**, 373-370.

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

Rojo, J. (1992). A pure tail ordering based on the ratio of the quantile functions. *Ann. Statist.,* ***20***, 570-579.

Shaked, M. and Shanthikumar, G. (1994). *Stochastic orders and their applications.* Academic Press, New York.

Spiegelhalter, D.J. and Smith, A.F.M. (1982). Bayes factors for linear and log-linear models with vague prior information. *J. Roy. Statist. Soc. Ser. B*, **44**, 377-387.