

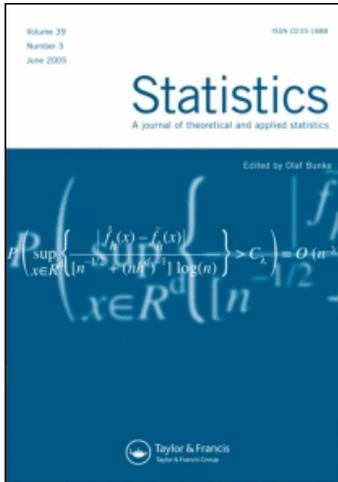
This article was downloaded by: [Complutensian University of Madrid]

On: 3 July 2009

Access details: Access Details: [subscription number 908033671]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Statistics

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713682269>

A Bayesian analysis for the multivariate point null testing problem

Miguel A. Gómez-Villegas ^a; Paloma Maín ^a; Luis Sanz ^a

^a Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, Madrid, Spain

First Published: August 2009

To cite this Article Gómez-Villegas, Miguel A., Maín, Paloma and Sanz, Luis(2009)'A Bayesian analysis for the multivariate point null testing problem', *Statistics*, 43:4, 379 — 391

To link to this Article: DOI: 10.1080/02331880802505173

URL: <http://dx.doi.org/10.1080/02331880802505173>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Bayesian analysis for the multivariate point null testing problem

Miguel A. Gómez-Villegas*, Paloma Maín and Luis Sanz

Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, Madrid, Spain

(Received 4 April 2006; final version received 21 August 2008)

A Bayesian test for the point null testing problem in the multivariate case is developed. A procedure to get the mixed distribution using the prior density is suggested. For comparisons between the Bayesian and classical approaches, lower bounds on posterior probabilities of the null hypothesis, over some reasonable classes of prior distributions, are computed and compared with the p -value of the classical test. With our procedure, a better approximation is obtained because the p -value is in the range of the Bayesian measures of evidence.

Keywords: posterior probability; multivariate point null hypothesis; p -value; mixed prior distributions

AMS 2000 Subject Classifications: 62F15; 62F03

1. Introduction

Let X be a random variable having density $f(x|\theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^p$ and suppose that we want to test

$$H_0 : \theta = \theta^0 \quad \text{versus} \quad H_1 : \theta \neq \theta^0, \quad (1)$$

where $\theta^0 = (\theta_1^0, \dots, \theta_p^0)$ is a known vector and $\theta \neq \theta^0$ means that at least one element of θ is different from the corresponding element of θ^0 . This problem does not have only a theoretical aspect because it appears in many practical situations also, for example, the classical data set in [1] gives cork deposits on trees, with the thickness of cork deposits in the four directions (North, East, South, West) measured by cork bearings on $n = 28$ trees. If the average cork deposits $(\theta_N, \theta_E, \theta_S, \theta_W)$ are not equal, this might indicate that the thickness of cork depends on ecological circumstances, such as dominant wind direction. The corresponding null hypothesis would be $\theta_N - \theta_E = \theta_E - \theta_S = \theta_S - \theta_W = 0$, that is $H_0 : (\theta_1, \theta_2, \theta_3) = (0, 0, 0)$ versus $H_1 : (\theta_1, \theta_2, \theta_3) \neq (0, 0, 0)$. Moreover, Equation (1) is widely used in social and educational sciences, for example Timm [2] reported the results of an experiment where subjects responded to ‘probe words’ at five positions in a sentence and tested a point null hypothesis and it is also used in many other

*Corresponding author. Email: ma.gv@mat.ucm.es

problems that contrast the difference of the means in paired data as, for example, an application to psychological tests in different stages of a person's life.

There are many approaches to the univariate two-sided hypothesis test, both in frequentist and Bayesian tests, but not to the multivariate two-sided one from a Bayesian perspective. Some exceptions are cited next. Berger and Delampady [3] show the conflict between Bayes factors or posterior probabilities and the frequentist p -values. Consequently, they recommended some alternatives to the p -values, using $\pi_0 = 0.5$ as prior probability to the null hypothesis, and with this choice they conclude that the posterior probabilities or the Bayes factors are larger than the p -values. Berger *et al.* [4] show that the conditional frequentist method and the Bayesian methodology for testing precise hypotheses are equivalent in the sense that both methods report the same error probabilities upon rejecting or accepting. Oh [5] deals with the multivariate normal distribution. In [6] the relevance of π_0 , the prior probability of the sharp null hypothesis, to the difference between the infimum of the posterior probability and the p -value is explored. They used a fixed value of π_0 whereas we will use the prior distribution to get π_0 . Other papers compare the p -value with the Bayes factor, for example [7–9], although the Bayes factors are not equally calibrated by different authors.

Let us suppose that our prior opinion about θ is given by the density $\pi(\theta)$. Then, to test Equation (1) we need a mixed prior distribution

$$\pi^*(\theta) = \pi_0 I_{\{\theta^0\}}(\theta) + (1 - \pi_0)\pi(\theta)I_{\{\theta \neq \theta^0\}}(\theta), \quad (2)$$

with π_0 the prior probability assigned to H_0 .

From our point of view, in a Bayesian context using a prior density $\pi(\theta)$, the problem is the 0.5 prior probability usually assigned to the null hypothesis as is the case in [3]. It gives too much weight to the null hypothesis, H_0 , making Bayesian and frequentist approaches very different. In addition, our proposal is to give the prior mass to the point null hypothesis using only our prior density, $\pi(\theta)$. Bayesian and frequentist approaches have a different meaning, but if the p -value and the infimum of the posterior probabilities are numerically nearly equal it means that both approaches are coherent.

Now, consider the more realistic precise hypothesis

$$H_{0\varepsilon} : d(\theta^0, \theta) \leq \varepsilon \quad \text{versus} \quad H_{1\varepsilon} : d(\theta^0, \theta) > \varepsilon \quad (3)$$

with a proper metric d and ε 'small' enough. What we propose is to use $\pi(\theta)$, our prior opinion about θ , and compute π_0 by means of

$$\pi_0 = \int_{E(\theta^0, \varepsilon)} \pi(\theta) d\theta, \quad (4)$$

where $E(\theta^0, \varepsilon) = \{\theta \in \mathbb{R}^p, d(\theta^0, \theta) \leq \varepsilon\}$. Thus, the prior probabilities assigned to H_0 and $H_{0\varepsilon}$, through $\pi(\theta)$, are equal.

There are several ways to specify $d(\theta^0, \theta)$. One way is to take an arbitrary value of ε and divide it in values $\varepsilon_i, i = 1, \dots, p$ – perhaps $\varepsilon_i = \varepsilon/p$, for all i – so that the uncertainty is shared among every coordinate, and then to build the distance starting from $|\theta_i - \theta_i^0| \leq \varepsilon_i, i = 1, \dots, p$. Another way is to consider $E(\theta^0, \varepsilon)$ as an ellipsoid centred at θ^0 . This last approach will be used in this article because of its computational tractability and intuitive appeal.

Several reasons can justify the choice of π_0 as in Equation (4), even though the usual value taken for π_0 is 0.5. First, in one dimension, when using Equations (2) and (4) with suitable small values of ε – in the case of normal likelihood $\varepsilon \in (0.1, 0.3)$ – and $\pi(\theta)$ in the class of all unimodal and symmetric distributions or in the class of ε -contaminated distributions, a better approximation between the posterior probability and the p -value is obtained. These results can be seen in [10–15].

The second reason to use π_0 as in Equation (4) is that if $\pi(\theta)$ reflects our prior opinion about θ then the prior probability of θ^0 is zero, but if we use Equation (2) the prior mass assigned to θ^0 is π_0 and this probability emerges from $\pi(\theta)$.

The third reason arises because H_0 is the limit hypothesis of $H_{0\varepsilon}$ as ε goes to zero. Then, if $\pi(\theta)$ is our prior opinion to test Equation (3) and $\pi^*(\theta)$, given by Equation (2), is our prior opinion to test Equation (1), it seems natural that both $\pi(\theta)$ and $\pi^*(\theta)$ must satisfy

$$\lim_{\varepsilon \rightarrow 0} \delta(\pi^*|\pi) = 0, \tag{5}$$

for some suitable measure of discrepancy, δ . One of the most popular measures of discrepancy is the Kullback–Leibler divergence

$$\delta(\pi^*|\pi) = \int_{\Theta} \pi(\theta) \ln \frac{\pi(\theta)}{\pi^*(\theta)} d\theta \tag{6}$$

(see, [16, p. 76]). For our problem we have

$$\begin{aligned} \delta(\pi^*|\pi) &= \int_{\Theta} \pi(\theta) \ln \left\{ \frac{\pi_0}{\pi(\theta)} I_{\{\theta^0\}}(\theta) + (1 - \pi_0) I_{\{\theta \neq \theta^0\}}(\theta) \right\}^{-1} d\theta \\ &= - \int_{\Theta} \pi(\theta) \ln \left\{ \frac{\pi_0}{\pi(\theta)} I_{\{\theta^0\}}(\theta) + (1 - \pi_0) I_{\{\theta \neq \theta^0\}}(\theta) \right\} d\theta \\ &= - \int_{\{\theta \neq \theta^0\}} \pi(\theta) \ln(1 - \pi_0) d\theta \\ &= - \ln(1 - \pi_0). \end{aligned} \tag{7}$$

We think Equation (5) is a desirable property. Usually in the literature, the expression (2) is used with $\pi_0 = 0.5$. However, for $\pi_0 = 0.5$, Equation (7) gives $\delta(\pi^*|\pi) = 0.693$, which represents a high discrepancy between these two distributions, π^* and π , whereas with our approximation the result (5) is verified.

The three reasons above are enough, in our opinion, given a prior density $\pi(\theta)$, to justify the construction of $\pi^*(\theta)$ as in Equation (2) with π_0 as in Equation (4) for the problem of testing a multivariate point null hypothesis.

Finally, it must be pointed out that we do not propose to change $H_{0\varepsilon} : d(\theta^0, \theta) \leq \varepsilon$ by $H_0 : \theta = \theta^0$, since in this case it should be possible to work with $\pi(\theta)$. What we claim is that $H_{0\varepsilon}$ is, with a proper ε , a good approximation to H_0 and then we can use π_0 , calculated as in Equation (4) to test Equation (1) with $\pi^*(\theta)$ as in Equation (2). Therefore, the hypothesis (3) supports the choice of π_0 in Equation (4) as the prior probability for H_0 in Equation (1). Finally, it is easier to calibrate distances through ε than to calibrate probabilities.

Anyhow, in this article the results are obtained as a function of π_0 and then they can be specified for every π_0 as in Equation (4). In particular, it is possible to compute the corresponding value of ε for the hypothesis (1) that gives $\pi_0 = 0.5$ in Equation (4), to reflect what is happening.

In Section 2 lower bounds on posterior probabilities over some reasonable classes of prior distributions are given: elliptical priors are analysed in Subsection 2.1 and scale mixture of normal priors in Subsection 2.2. Finally, in Section 3 some comments and concluding remarks are included.

2. Lower bounds on posterior probabilities

In order to make comparisons between the p -value and the posterior probabilities we take wide classes of prior distributions and then compute the infimum of the posterior probabilities over these

classes. This is the usual procedure used in the literature to compare the Bayesian and frequentist approaches. A Bayesian with a large class of prior distributions might behave as a frequentist in the sense of reaching similar conclusions. In the limit case, a Bayesian that dealt with the class of all probability distributions as priors is a frequentist. It also allows us to interpret the p -value as a lower bound of the posterior probabilities of the null hypothesis over some classes of prior distributions.

2.1. Lower bounds for elliptical priors

Because of the structure of the problem, it looks reasonable to deal first with the class $\Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$, the class of distributions on \mathbb{R}^p having probability density functions of the type

$$\pi(\boldsymbol{\theta}) = \psi((\boldsymbol{\theta} - \boldsymbol{\theta}^0)'(\Sigma^0)^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^0))$$

with $\psi(\cdot)$ a decreasing function on $[0, \infty)$, $\boldsymbol{\theta}^0 \in \mathbb{R}^p$ and $\Sigma^0(p \times p)$ a positive definite matrix. These distributions are called elliptical and are unimodal in the sense of [17]. They have ellipsoidal contours centred at $\boldsymbol{\theta}^0$ with scale parameters Σ^0 . In particular, $\Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$ contains the spherical distributions on \mathbb{R}^p .

Furthermore, if the following additional regularity conditions are imposed:

- $\psi(r^2) \rightarrow 0$ as $r \rightarrow \infty$, and
- $\psi(r^2)$ is of bounded variation in every finite interval away from the origin,

it can be shown, see [18], that $\pi(\boldsymbol{\theta}) \in \Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$ if and only if $\pi(\boldsymbol{\theta})$ is a mixture of uniform densities on ellipsoids centred at $\boldsymbol{\theta}^0$, $E(\boldsymbol{\theta}^0, k) = \{\boldsymbol{\theta} | (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'(\Sigma^0)^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \leq k^2\}$.

Then, to find the infimum of the posterior probability of the point null hypothesis over the class $\Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$, it is sufficient to find it over the smaller class of the uniforms, see [19].

The following theorem gives the infimum of the posterior probability of the point null hypothesis when the prior density $\pi(\boldsymbol{\theta})$ is in $\Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$, using the previous result.

THEOREM 2.1 *If $\Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)$ is the class of priors and $\pi^*(\boldsymbol{\theta})$ is given by Equation (2) with π_0 as in Equation (4), then*

$$\inf_{\pi \in \Gamma_{EU}(\boldsymbol{\theta}^0, \Sigma^0)} P(H_0 | \mathbf{x}) = \left(1 + \frac{1}{V(E(\boldsymbol{\theta}^0, \varepsilon))} \int_{\mathbb{R}^p} \frac{f(\mathbf{x} | \boldsymbol{\theta})}{f(\mathbf{x} | \boldsymbol{\theta}^0)} d\boldsymbol{\theta} \right)^{-1}, \tag{8}$$

where $V(E(\boldsymbol{\theta}^0, \varepsilon))$ is the volume of $E(\boldsymbol{\theta}^0, \varepsilon) = \{\boldsymbol{\theta} | (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'(\Sigma^0)^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \leq \varepsilon^2\}$, an ellipsoid centred at $\boldsymbol{\theta}^0$.

Proof See Appendix 1 ■

Now, we consider the multivariate normal distribution to see the consequences of the previous result.

Example 2.1 Suppose \mathbf{X} is $N_p(\boldsymbol{\theta}, \sigma^2 \mathbf{I})$ distributed, σ^2 known, where $\mathbf{X} = (X_1, \dots, X_p)'$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$. It is desired to test Equation (1) with a sample of size n . The frequentist significance test statistic is

$$T(\bar{\mathbf{X}}, \boldsymbol{\theta}^0) = \frac{n}{\sigma^2} |\bar{\mathbf{X}} - \boldsymbol{\theta}^0|^2,$$

with $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_p)$. Under the null hypothesis, H_0 , $T(\bar{\mathbf{X}}, \boldsymbol{\theta}^0)$ has a χ_p^2 distribution. Therefore, the p -value of the observed data, $\bar{\mathbf{x}}$, is given by $p(\bar{\mathbf{x}}) = P\{\chi_p^2 \geq T(\bar{\mathbf{x}}, \boldsymbol{\theta}^0)\}$.

Using Theorem 2.1, the infimum of the posterior probability of the point null over the class $\Gamma_{\text{EU}}(\boldsymbol{\theta}^0, \Sigma^0)$ is

$$\inf_{\pi \in \Gamma_{\text{EU}}} P(H_0|\bar{\mathbf{x}}) = \left\{ 1 + \frac{1}{V(E(\boldsymbol{\theta}^0, \varepsilon))} \int_{\mathbb{R}^p} \frac{\exp(-n/2\sigma^2|\bar{\mathbf{x}} - \boldsymbol{\theta}^0|^2)}{\exp(-n/2\sigma^2|\bar{\mathbf{x}} - \boldsymbol{\theta}^0|^2)} d\boldsymbol{\theta} \right\}^{-1}$$

But, $V(E(\boldsymbol{\theta}^0, \varepsilon)) = \pi^{p/2} \varepsilon^p |\Sigma^0| / \Gamma(p/2 + 1)$ and taking $\Sigma^0 = I$, the identity matrix, then

$$\inf_{\pi \in \Gamma_{\text{EU}}} P(H_0|\bar{\mathbf{x}}) = \left\{ 1 + \frac{2^{p/2} \Gamma(p/2 + 1)}{\varepsilon^{*p}} \exp\left(\frac{1}{2} T(\bar{\mathbf{x}}, \boldsymbol{\theta}^0)\right) \right\}^{-1}, \tag{9}$$

where $\varepsilon^* = \varepsilon \sqrt{n}/\sigma$. By fixing ε^* , the space dimension, p , and $\boldsymbol{\theta}^0 = \mathbf{0}$ for different values of $T(\bar{\mathbf{x}}, \boldsymbol{\theta}^0)$, the infimum of the posterior probability can be obtained. Table 1 shows the values of ε^* so that the p -value and the infimum of the posterior probability are equal. These values depend on the space dimension and they become larger as p increases.

In Table 1 we observe robustness with respect to the data for every dimension, p . For instance, when the dimension is $p = 3$ and if we choose $\varepsilon^* \in (2.1, 2.3)$, then the infimum of the posterior probability of H_0 will be close to the different p -values. For different dimensions p , Figure 1 shows the infimum of the posterior probability with suitable values of ε^* , chosen from Table 1. The p -value and the infimum of the posterior probability when $\pi_0 = 0.5$ are also included. It is observed that as ε^* increases, the infimum of the posterior probability of H_0 increases too, and we can get it very close to the p -value. However, for $\pi_0 = 0.5$, the infimum of the posterior probability is always greater than the p -value, and there is a large discrepancy between these two measures of evidence. For example, the corresponding value of ε^* for $\pi_0 = 0.5$ and $p = 5$ goes from 3.96 to 5.56 for p -values from 0.1 to 0.001, which is extremely high.

Nevertheless, for high dimensions, greater or equal to 10, the value of ε^* is extreme. This fact invalidates the agreement of the p -value and the posterior probability of H_0 . The value $\varepsilon^* = 5.023$ needed to match both quantities when $p = 10$ makes the hypothesis H_0 and $H_{0,5.023}$ non-exchangeable.

Table 2 shows the infimum of the posterior probability for $p = 5$ and some suitable values of ε^* chosen from Table 1. Moreover, Table 2 includes the values of the infimum when $\pi_0 = 0.5$.

It is clear from Table 2 that, when the agreement is possible, our procedure permits a better approximation between the infimum of the posterior probability and the p -value for a proper value of ε . This is not the case if $\pi_0 = 0.5$ is chosen. The next example shows how this procedure works in a real context. Again, the difference with [3] is the choice of the prior probability of the null hypothesis.

Example 2.2 Let us consider the classical data in [20] giving skull length and breadth measured on siblings in 25 families (X_1, X_2, X_3, X_4) with X_1, X_2 of the first son and X_3, X_4 of the second

Table 1. Values of ε^* so that the p -values and the infimum of the posterior probability over Γ_{EU} are equal.

inf $P\{H_0 \bar{\mathbf{x}}\} = p$ -value	Dimension p				
	2	3	5	10	20
0.1	1.491	2.120	2.918	4.075	5.486
0.05	1.451	2.145	3.019	4.247	5.696
0.01	1.421	2.227	3.243	4.601	6.118
0.001	1.415	2.341	3.515	5.023	6.615

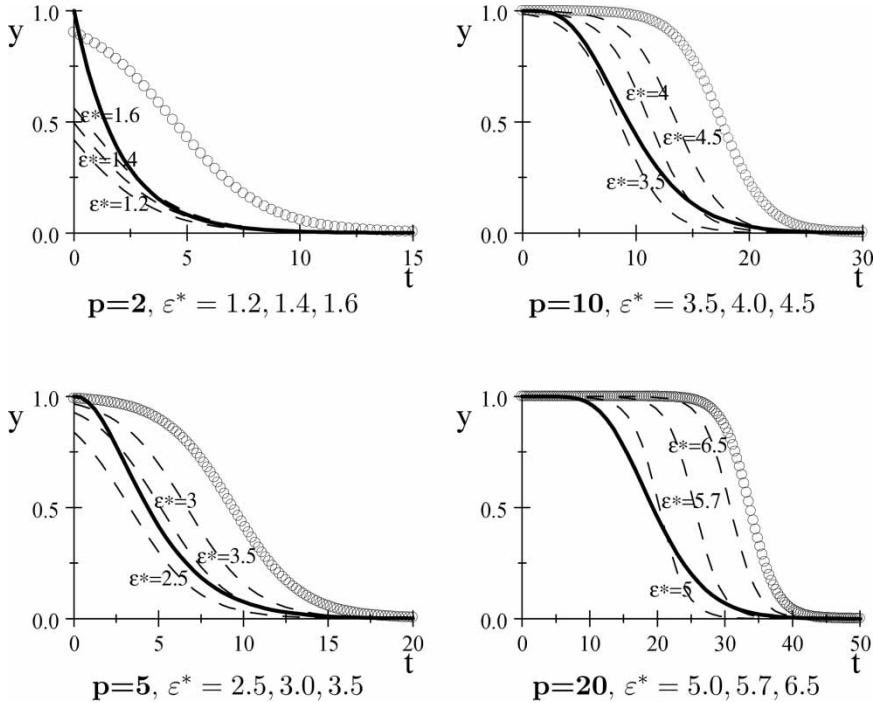


Figure 1. Thin lines: infimum of the posterior probability for some ϵ^* ; circles: infimum of the posterior probability when $\pi_0 = 1/2$; thick line: p -value.

Table 2. P -values and infimum of the posterior probabilities for elliptical priors with $p = 5$.

$p(t)$	0.1	0.05	0.01	0.001
$\Pr(H_0 t, \pi_0 = 0.5)$	0.3391	0.2350	0.0761	0.0098
$\Pr(H_0 t, \epsilon^* = 2.9)$	0.09724	0.04128	0.00578	0.00038
$\Pr(H_0 t, \epsilon^* = 3.2)$	0.14981	0.06580	0.00937	0.00063
$\Pr(H_0 t, \epsilon^* = 3.5)$	0.21619	0.09930	0.01459	0.00979

son. The point null hypothesis asserts there is no difference between the first and the second son with regard to the head size.

Dealing with the data $Y_1 = X_1 - X_3$ and $Y_2 = X_2 - X_4$ and accepting

$$\Sigma = \begin{pmatrix} 60 & 14 \\ 14 & 17 \end{pmatrix}$$

as the model covariance matrix, the mean vector is

$$\bar{x} = \begin{pmatrix} 4.08 \\ 1.76 \end{pmatrix}$$

and the classical test statistic $t = n\bar{x}'\Sigma^{-1}\bar{x} = 8.1245$ with a p -value = 0.01721.

With our methodology the corresponding $\epsilon^* = 1.4$, as shown in Figure 2, yields $\Pr(H_0 | \bar{x}) = 0.0166$. However for $\pi_0 = 0.5$ the infimum of the posterior probability is $\Pr(H_0 | \bar{x}) = 0.1373$.

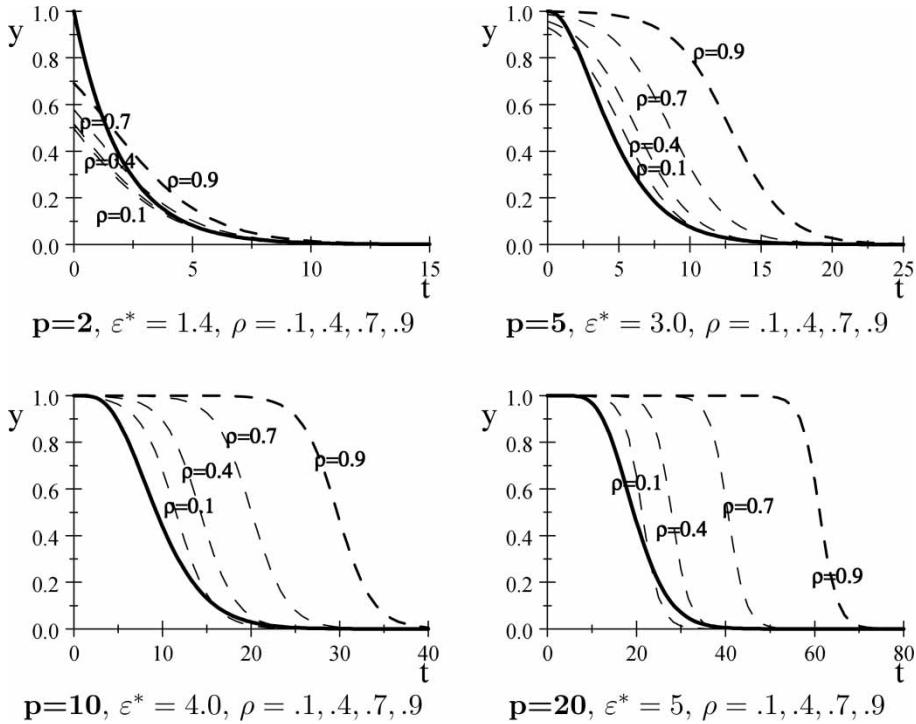


Figure 2. Thin lines: infimum of the posterior probability for some ρ ; thick line: p -value.

The next example explores the influence of the correlation between the variables in the posterior probability. We consider that the variables have the same variance, σ^2 , and a common correlation coefficient ρ .

Example 2.3 Suppose X_1, \dots, X_n is a sample from a $N_p(\theta, \Sigma)$, a multivariate normal distribution with a special correlation structure,

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix},$$

with σ^2 known and ρ being the correlation coefficient. It is a permutation-symmetric multivariate normal distribution that gives exchangeable variables for $\rho \geq 0$ (see [21]). The graphics in Figure 2 show the p -values and the infimum of the posterior probabilities for different values of p , $\varepsilon^* = \varepsilon\sqrt{n}/\sigma$ and the correlation coefficient, ρ .

It can also be noted that, in every case, it is possible to match the infimum of the posterior probability and the p -value, but the approximation is worse for high dimensions of the parameter space, especially for large values of ρ .

This fact can be explained because the covariance matrix of permutation-symmetric normal variables has certain interesting properties.

- (i) $|\Sigma| = \sigma^{2p}(1 - \rho)^{p-1}(1 + (p - 1)\rho)$, therefore Σ is a positive definite matrix if and only if $\rho \in (-1/p - 1, 1)$.

(ii) With the previous condition $\Sigma^{-1} = (\tau_{ij})$

$$\tau_{ij} = \begin{cases} \frac{1 + (p - 2)}{\sigma^2 \Delta_p(\rho)} & \text{for } i = j \\ \frac{-\rho}{\sigma^2 \Delta_p(\rho)} & \text{for } i \neq j, \end{cases}$$

where $\Delta_p(\rho) = (1 - \rho)(1 + (p - 1)\rho) > 0$ so that $\Delta_p(\rho) \sim (1 - \rho)$ as ρ goes to one.

Thus the frequentist significance test statistic $t = n\bar{\mathbf{x}}^t \Sigma^{-1} \bar{\mathbf{x}}$ converges to infinity as ρ goes to one with $t \sim 1/(1 - \rho)$.

On the other hand, using the asymptotic behaviour of the gamma tail, we have

$$P\{\chi_p^2 \geq t\} \simeq t^{(p/2)-1} e^{-t/2}$$

and taking $\Sigma^0 = I$, the infimum of the posterior probability yields

$$\underline{P}(H_0|\bar{\mathbf{x}}) = \left\{ 1 + \frac{\Gamma(p/2 + 1)}{\varepsilon^{*p}} 2^{p/2} ((1 - \rho)^{p-1} (1 + (p - 1)\rho))^{1/2} \exp\left(\frac{t}{2}\right) \right\}^{-1}.$$

Then to make comparisons for large values of ρ we have

$$\frac{P\{\chi_p^2 \geq t\}}{\underline{P}(H_0|\bar{\mathbf{x}})} \sim (1 - \rho)^{1/2}$$

as ρ goes to one.

This behaviour explains that the p -value decreases quickly with regard to the infimum of the posterior probabilities for a fixed value of ε when ρ goes to one as shown in Figure 2. Thus, for highly correlated variables, the p -value is extremely unfavourable to the point null hypothesis. On the other hand, the prior probability of the null hypothesis calculated by Equation (4) is larger, for the same value of ε , than in the uncorrelated case.

Then, if we want to match the Bayesian and frequentist perspectives, a lower ε^* must be chosen as ρ goes to one and p increases, as can be seen in Table 3. Some particular cases are visualized in Figure 3. Once again the use of $\pi_0 = 0.5$ keeps both approaches well separated.

2.2. Lower bounds for scale mixture of priors

In this section we assume that (X_1, \dots, X_p) is a random sample of a $N_p(\boldsymbol{\theta}, \sigma^2 \mathbf{I})$ distribution, and σ^2 known, where \mathbf{I} is the $p \times p$ identity matrix. Then $\bar{\mathbf{X}}$ is $N_p(\boldsymbol{\theta}, \sigma^2/n\mathbf{I})$ distributed. It is

Table 3. Values of ε^* to match the p -value and the infimum of the posterior probability for some values of ρ .

ρ	Dimension p		
	$p = 5$	$p = 10$	$p = 20$
0.1	3	3.8	3.65
0.4	1.8	0.95	0.135
0.7	0.55	5.25×10^{-2}	2.4×10^{-4}
0.9	6.65×10^{-2}	4.2×10^{-4}	8×10^{-9}

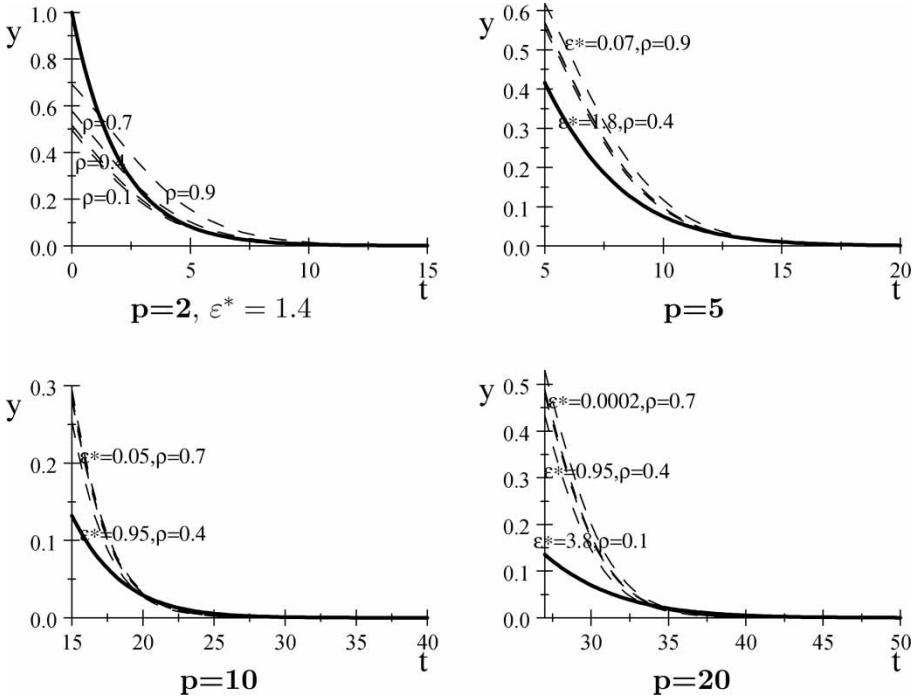


Figure 3. Thin lines: infimum of the posterior probability for some values of ε^* and ρ ; thick line: p -value.

desired to test Equation (1) with $\theta^0 = \mathbf{0}$, then the appropriate statistic is $T(\bar{\mathbf{X}}) = n/\sigma^2 \bar{\mathbf{X}}'/\bar{\mathbf{X}}$. Now, the considered prior density $\pi(\theta)$ belongs to the class of scale mixture of normals

$$\Gamma_N = \left\{ \int N_p(0, v^2 I) \pi(v^2) dv^2, \pi(v^2) \text{ a non-decreasing function on } (0, \infty) \right\}. \quad (10)$$

The reason to consider this class of priors is that it assigns higher mass to the neighbours of a precise hypothesis than the class $\Gamma_{EU}(\theta^0, \Sigma^0)$, see [7], and then we can have some large values of π_0 with moderate values of ε .

To find the lower bound on the posterior probability over the class (10) is equivalent to finding it over the smaller class in which $\pi(v^2)$ is uniform on $(0, r)$, $r > 0$, see [7].

The following theorem shows the infimum of the posterior probability over this class for $p > 2$. For notation, $\pi_r(\theta) = \int_0^r (1/r) \varphi(\theta, v^2) dv^2$, φ being the $N_p(0, v^2 I)$ density.

THEOREM 2.2 *If the prior mass assigned to the null hypothesis is, from Equation (4),*

$$\pi_{0r} = \int_{B(\mathbf{0}, \varepsilon)} \pi_r(\theta) d\theta,$$

where $B(\mathbf{0}, \varepsilon) = \{\theta, \|\theta\|^2 \leq \varepsilon^2\}$, then

$$\inf_{\pi \in \Gamma_N} P(H_0|t) = \left(1 + \frac{1}{\varepsilon^{*2}} \frac{\mathcal{F}_{p-2}(t)}{f_p(t)} \right)^{-1}, \quad (11)$$

with $t = n/\sigma^2 \bar{\mathbf{X}}'/\bar{\mathbf{X}}$, \mathcal{F}_{p-2} being the chi-squared distribution function with $p - 2$ degrees of freedom and f_p being the corresponding density with p.d.f.

Table 4. Values of ε^* such that the p -values and the infimum of the posterior probability over Γ_N are equal.

inf $P\{H_0 \bar{x}\} = p$ -value	Dimension p			
	5	10	15	20
0.1	1.71	1.93	2.07	2.18
0.05	1.64	1.83	1.95	2.06
0.01	1.56	1.71	1.81	1.90
0.001	1.52	1.63	1.74	1.78

Table 5. P -values and infimum of the posterior probabilities for scale mixture priors for $p = 5$.

$p(t)$	0.1	0.05	0.01	0.001
$\Pr(H_0 t, \pi_0 = 0.5)$	0.38037	0.28843	0.11329	0.01903
$\Pr(H_0 t, \varepsilon^* = 1.5)$	0.07846	0.04231	0.00922	0.00097
$\Pr(H_0 t, \varepsilon^* = 1.6)$	0.08832	0.04766	0.01048	0.00109
$\Pr(H_0 t, \varepsilon^* = 1.8)$	0.10921	0.05957	0.01322	0.00140

Proof See Appendix 2 ■

Then, by fixing ε^* and the space dimension p , the infimum of the posterior probability can be obtained. Table 4 shows the values of ε^* making the infimum of the posterior probabilities for the class of scale mixture priors and the p -value equal.

In fact, as asserted above, we can get moderate values of ε^* to match Bayesian and frequentist approaches.

In order to compare numerically the p -value with the infimum of the posterior probability, Table 5 shows for $p = 5$ this infimum for some suitable values of ε^* , chosen from Table 4, and the infimum of the posterior probability when $\pi_0 = 0.5$.

It can be pointed out that practical agreement exists between Bayesian and frequentist measures for the different values of ε^* . In particular, for an intermediate value of $\varepsilon^* = 1.6$, the infimums of the posterior probability and the p -value are nearly the same. However, for $\pi_0 = 0.5$, these values are significantly different because the value of ε^* goes from 4.03 to 6.69 for p -values from 0.1 to 0.001, respectively, which are extremely high.

Figure 4 shows the graphics of the lower bounds of the posterior null probability, for some values of ε^* , and the p -value jointly with the posterior probability for $\pi_0 = 0.5$.

3. Conclusions and comments

The most important conclusion is that the p -values and the posterior probabilities can be matched for testing multivariate point null hypotheses. The difference between these two measures increases when the prior mass assigned to the point null hypothesis for dimensions $p \geq 2$ is $\pi_0 = 0.5$.

The proposal in this article is to give a prior probability for θ^0 equal to the probability of a sphere or an ellipsoid with a fixed radius ε and centred at θ^0 , calculated from $\pi(\theta)$. This methodology shows, for the examples considered, a better approximation between the p -value and the infimum of the posterior probabilities using just one source of prior information $\pi(\theta)$. As pointed out in Example 2.2, where the p -value is 0.0172 and, with our procedure, $\Pr(H_0|t) = 0.0166$ when

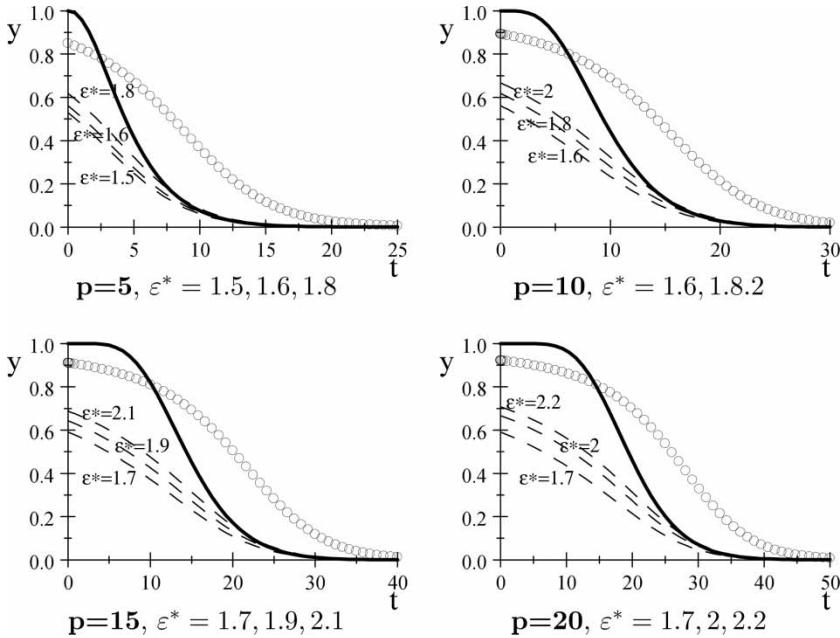


Figure 4. Thin lines: infimum of the posterior probability for some values of ϵ^* ; circles: infimum of the posterior probability when $\pi_0 = 1/2$; thick line: p -value.

$\epsilon^* = 1.4$, whereas if $\pi_0 = 0.5$, this infimum becomes 0.1373 greater than the p -value. Besides, we think that the choice of ϵ or ϵ^* is easier than the choice of π_0 in practical situations.

It can also be pointed out that an apparent robustness in ϵ^* is observed in every fixed dimension p for the normal distribution and the class Γ_{EU} , although varying with p . Furthermore, for the class of scale mixtures, this robustness is observed not only for every fixed p but for varying p as well.

Moreover, for a normal family with a common correlation coefficient, it is observed that the p -value and the infimum of the posterior probability are closer as ρ , the correlation coefficient, decreases.

Finally, if the mixed prior distribution structure is to be used for the point null testing problem with π_0 being the prior mass assigned to the null hypothesis, then a value smaller than 0.5 should be employed for π_0 to ensure a numerical agreement between the Bayesian and frequentist approaches.

Acknowledgements

This research was supported by MEC from Spain, Grant number MTM2005–5462 and CM–UCM Grant Number GR69/06. The authors are grateful to reviewers for very helpful comments.

References

- [1] C.R. Rao, *Test of significance in multivariate analysis*, Biometrika 35 (1948), pp. 58–79.
- [2] N.H. Timm, *Multivariate Analysis with Applications in Education And Psychology*, Brooks/Cole, Belmont, California, 1975.
- [3] J.O. Berger and M. Delampady, *Testing precise hypotheses*, Statist. Sci. 3 (1987), pp. 317–352.
- [4] J.O. Berger, B. Boukai, and Y. Wang, *Unified frequentist and bayesian testing of a precise hypothesis*, Statist. Sci. 3 (1997), pp. 133–160.

- [5] M.S. Oh, *A Bayes test for simple versus one-sided hypotheses on the mean vector of a multivariate normal distribution*, *Commun. Statist. Theory Meth.* 27(10) (1998), pp. 2371–2389.
- [6] H.S. Oh and A. DasGupta, *Comparison of the p-value and posterior probability*, *J. Statist. Plann. Inference* 76 (1999), pp. 93–107.
- [7] T. Sellke, M.J. Bayarri, and J. Berger, *Calibration of p-values for testing precise null hypotheses*, *Amer. Statist.* 55 (2001), pp. 62–71.
- [8] A. O'Hagan, *Properties of intrinsic and fractional Bayes factors*, *Test* 6(1) (1997), pp. 101–118.
- [9] J.B. Berger and L.R. Pericchi, *Training samples in objective Bayesian model selection*, *Ann. Statist.* 32(3) (2004), pp. 841–869.
- [10] M.A. Gómez-Villegas and E. Gómez Sánchez-Manzano, *Bayes factor in testing precise hypotheses*, *Commun. Statist. Theory Meth.* 21(6) (1992), pp. 1707–1715.
- [11] M.A. Gómez-Villegas and L. Sanz, *Reconciling Bayesian and frequentist evidence in the point null testing problem*, *Test* 7(1) (1998), pp. 207–216.
- [12] M.A. Gómez-Villegas and L. Sanz, *ϵ -contaminated priors in testing point null hypotheses: a procedure to determine the prior probability*, *Statist. Probab. Lett.* 47(1) (2000), pp. 53–60.
- [13] M.A. Gómez-Villegas, P. Maín, and L. Sanz, *A suitable Bayesian approach in testing point null hypothesis: some examples revisited*, *Commun. Statist. Theory Meth.* 31(2) (2002), pp. 201–217.
- [14] M.A. Gómez-Villegas, P. Maín, L. Sanz, and H. Navarro, *Asymptotic relationships between posterior probabilities and p-values using the hazard rate*, *Statist. Probab. Lett.* 66(1) (2004), pp. 59–66.
- [15] M.A. Gómez-Villegas, and B. González-Pérez, *Bayesian analysis of contingency tables*, *Commun. Statist. Theory Meth.* 34 (2005), pp. 1743–1754.
- [16] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley and Son, New York, 1994.
- [17] T.W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, *Proc. Amer. Math. Soc.* 6 (1955), pp. 170–176.
- [18] D.R. Jensen and I.J. Good, *A representation for ellipsoidal distributions*, *Siam J. Appl. Math.* 43(5) (1983), pp. 1194–1200.
- [19] G. Casella and R.L. Berger, *Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with discussion)*, *J. Amer. Statist. Assoc.* 82 (1987), pp. 106–135.
- [20] G.P. Frets, *Heredity of head form in man*, *Genetica* 3 (1921), pp. 193–200.
- [21] Y.L. Tong, *The Multivariate Normal Distribution*, Springer Verlag, New York, 1990.

Appendix 1.

Proof of Theorem 2.1 The posterior probability of the point null hypothesis, H_0 , is given by

$$P(H_0|\mathbf{x}) = \left(1 + \frac{1 - \pi_0}{\pi_0} \frac{m_\pi(\mathbf{x})}{f(\mathbf{x}|\theta^0)} \right)^{-1},$$

where $m_\pi(\mathbf{x}) = \int_{\Theta} \pi(\theta) f(\mathbf{x}|\theta) d\theta$. Then, computing the infimum of the posterior probability over the class $\Gamma_{EU}(\theta^0, \Sigma^0)$ of the point null hypothesis is just like computing the supremum of $M(k) = (1 - \pi_0)/\pi_0 m_\pi(\mathbf{x})$ over the class of density functions constant on ellipsoids centered at θ^0 (see [5]).

Assuming $\epsilon < k$ and denoting $\pi_k(\theta^0, k, \Sigma^0)$ by the uniform distribution over the ellipsoid centred at θ^0 , then

$$\pi_0 = \int_{E(\theta^0, \epsilon)} \pi_k(\theta^0, k, \Sigma^0) d\theta = \frac{V(E(\theta^0, \epsilon))}{V(E(\theta^0, k))}$$

and

$$M(k) = \frac{V(E(\theta^0, k)) - V(E(\theta^0, \epsilon))}{V(E(\theta^0, k))V(E(\theta^0, \epsilon))} \int_{E(\theta^0, k)} f(\mathbf{x}|\theta) d\theta.$$

It is straightforward to check that $M(k)$ is increasing in k , because $M'(k) > 0$, and then the supremum is attained as k goes to infinity,

$$\sup_k M(k) = \lim_{k \rightarrow \infty} M(k) = \frac{1}{V(E(\theta, \epsilon))} \int_{\mathbb{R}^p} f(\mathbf{x}|\theta) d\theta$$

and from this we get (8). ■

Appendix 2.

Proof of Theorem 2.2 First, it can easily be checked that

$$\pi_{0r} = \frac{1}{r} \int_0^r \mathcal{F}_p \left(\frac{\epsilon^2}{u} \right) du.$$

Then, the corresponding posterior probability, for a fixed dimension p and uniform $U(0, r)$, is

$$P_r(H_0|t) = \left(1 + \frac{1 - \pi_{0r}}{\pi_{0r}} \frac{1}{B_r}\right)^{-1}$$

where the Bayes factor, B_r , is

$$B_r = \frac{f(t|\theta^0 = 0)}{\int_{R^p} f(t|\theta)\pi_r(\theta) d\theta}$$

and

$$\begin{aligned} \frac{1}{B_r} &= \frac{1}{r} \int_{R^p} \int_0^r \frac{1}{(2\pi)^{p/2} u^{p/2}} \exp\left\{\frac{n}{\sigma^2} \bar{x}'\theta - \frac{n}{2\sigma^2} \theta'\theta - \frac{1}{2u} \theta'\theta\right\} d\theta du \\ &= \frac{1}{r} \int_0^r \left(\frac{nu}{\sigma^2} + 1\right)^{-p/2} \exp\left\{\frac{1}{2} \frac{(n/\sigma^2)^2}{n/\sigma^2 + 1/u} \bar{x}'\bar{x}\right\} du \\ &= \frac{1}{r} \exp\left\{\frac{n}{2\sigma^2} \bar{x}'\bar{x}\right\} \frac{\sigma^2}{2} \int_{\sigma^2/nr + \sigma^2}^1 z^{p/2-2} \exp\left\{-\frac{n\bar{x}'\bar{x}}{2\sigma^2} z\right\} dz \\ &= \frac{\sigma^2}{rn(p-2)} \left\{\frac{\mathcal{F}_{p-2}(t) - \mathcal{F}_{p-2}(\sigma^2 t/\sigma^2 + rn)}{f_p(t)}\right\}. \end{aligned}$$

Then, the posterior probability of the point null hypothesis, H_0 , for $p > 2$ is given by

$$P_r(H_0|t) = \left\{1 + \frac{1 - \pi_{0r}}{r\pi_{0r}} \frac{\sigma^2}{n(p-2)} \frac{\mathcal{F}_{p-2}(t) - \mathcal{F}_{p-2}(\sigma^2 t/(\sigma^2 + rn))}{f_p(t)}\right\}^{-1} \tag{B1}$$

where \mathcal{F}_{p-2} is the chi-square distribution function with $p - 2$ degrees of freedom and f_p chi-squared density function with p degrees of freedom.

Now, we look for the infimum in r but $P_r(H_0|t)$, from (B1), is decreasing in r and then

$$\inf_{\pi \in \Gamma_N} Pr(H_0|t) = \lim_{r \rightarrow \infty} P_r(H_0|t).$$

But it can be shown that

$$\lim_{r \rightarrow \infty} r\pi_{0r} = \int_0^\infty \mathcal{F}_p\left(\frac{\varepsilon^2}{t}\right) dt = \frac{\varepsilon^2}{p-2},$$

then, Equation (11) is obtained. ■