A Bayesian Analysis for the Multivariate Point Null Testing Problem

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

Dpto. Estadística e I. O., Universidad Complutense de Madrid. 28040 Madrid, Spain.

Abstract

A Bayesian test for simple versus bilateral hypotheses, in the multivariate case, is developed. A procedure to get the mixed distribution, for the bilateral hypothesis, 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. A better approximation is obtained because the p–value is in the range of the Bayesian measures of evidence.

MSC: 62F15; 62F03.

Keywords: Posterior probability, Multivariate point null hypothesis, P-value, Mixed prior distributions, Robust Bayesian analysis.

1. Introduction

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

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

*Corresponding author.

E–mail address: ma.gv@mat.ucm.es (M.A. Gómez–Villegas)

where $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^0, \ldots, \boldsymbol{\theta}_p^0)$ is a known vector and $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ means that at least one element of $\boldsymbol{\theta}$ is different from the corresponding element of $\boldsymbol{\theta}^0$. For example, classical data set in Rao (1948) gives cork deposits on trees. The thickness of cork deposits in four directions (North, East, South, West) was 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 thickness of cork depends on ecological circumstances, such as dominant wind direction. The corresponding null hypothesis would be $H_0 : \theta_N - \theta_E = \theta_E - \theta_S = \theta_S - \theta_W = 0$.

There are many approaches for the univariate two–sided hypothesis test, both in classical and Bayesian tests, but not for the multivariate two–sided one. Some exceptions are Oh (1998), who deals with the multivariate normal distribution, and Oh and DasGupta (1999), who explore the relevance of π_0, the prior probability of the sharp null hypothesis, in the difference between the infimum of the posterior probability and the p–value for some classes of priors on the alternative hypothesis.

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

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

with π_0 the prior probability assigned to H_0.

Now, consider the more realistic precise hypothesis

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

with a proper metric d and ε “small”. What we propose is to use $\pi(\boldsymbol{\theta})$, our prior opinion about $\boldsymbol{\theta}$, and compute π_0 by means of

$$\pi_0 = \int_{B(\boldsymbol{\theta}^0, \varepsilon)} \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{4}$$

being $B(\boldsymbol{\theta}^0, \varepsilon) = \{\boldsymbol{\theta} \in R^p, d(\boldsymbol{\theta}^0, \boldsymbol{\theta}) \leq \varepsilon\}$, a sphere of radius ε centered at $\boldsymbol{\theta}^0$. Thus, the prior probabilities assigned to H_0 and $H_{0\varepsilon}$, through $\pi(\boldsymbol{\theta})$, are equal.

There are several ways to specify $d(\boldsymbol{\theta}^0, \boldsymbol{\theta})$. One way is to take an arbitrary value of ε and divide it in values ε_i, $i = 1, \ldots, p$ –perhaps $\varepsilon_i = \varepsilon/p$, for all i– so that the uncertainty is shared among every coordinate, and then to built the distance starting from $|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i}| \leq \varepsilon_i$, $i = 1, \ldots, p$. Another way is to consider $B(\boldsymbol{\theta}^0, \varepsilon)$ as the sphere of radius ε, centered at $\boldsymbol{\theta}^0$. This last approach will be used in this paper because of its computational tractability and intuitive appeal.

Several reasons can justify the choice of π_0 as in (4), despite of the usual value taken for π_0 is 0.5. Firstly, in one dimension, when using (2) and (4) with suitable small values of ε – in case of normal likelihood $\varepsilon \in (0.1, 0.3)$ – and $\pi(\boldsymbol{\theta})$ 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 Gómez–Villegas and Gómez (1992), Gómez–Villegas and Sanz (1998, 2000) and in Gómez–Villegas, Maín and Sanz (2002).

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

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

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

for some suitable measure of discrepancy, δ.

One of the most popular measures of discrepancy is

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

see, by example, Bernardo and Smith (1994) pag. 76 and, for our problem, we have

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

We think this is a desirable property. Usually in the literature, at least in the unidimensional case, the expression (2) is used with $\pi_0 = 0.5$. However, for $\pi_0 = 0.5$, (7) gives $\delta(\pi^*|\pi) = 0.693$ that seems a high discrepancy between these two distributions, π^* and π.

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

Anyhow, in this paper the results are obtained as a function of π_0 and then they can be specified for every π_0 as in (4). In particular, it is possible to compute the value of ε, for the precise hypothesis (3), that gives $\pi_0 = 0.5$.

In Section 2, lower bounds on posterior probabilities over some reasonable classes of prior distributions are given: unimodal and symmetric priors are analyzed 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 will take wide classes of prior distributions and then we will compute the infimum of the posterior probabilities over these classes. This is the usual procedure to compare the Bayesian and classical approaches, because a classical might behave like a Bayesian using a large class of priors.

2.1. Lower bounds for Unimodal and Symmetric priors

Because of the structure of the problem, it looks reasonable to deal first with the class

$$\Gamma_{US} = \{\text{Unimodal and symmetric distributions about } \boldsymbol{\theta}^0\}. \tag{8}$$

Furthermore, if $\pi(\boldsymbol{\theta})$ is in Γ_{US} it can be expressed as a mixture of uniform and symmetric distributions over spheres of radius k centered at $\boldsymbol{\theta}^0$. Then, to find the infimum of the posterior probability of the point null hypothesis over the class Γ_{US}, it is sufficient to find it over the much smaller class, see Casella and Berger (1987),

$$U_S = \{\text{Uniform and symmetric distributions over spheres of radius } k \text{ centered at } \boldsymbol{\theta}^0\}. \tag{9}$$

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

Theorem 1. *If $\pi^*(\boldsymbol{\theta})$ is given by (2) with π_0 as in (4), then*

$$\inf_{\pi \in \Gamma_{US}} P(H_0|\boldsymbol{x}) = \left(1 + \frac{1}{V(\boldsymbol{\theta}^0, \varepsilon)} \int_{R^p} \frac{f(\boldsymbol{x}|\boldsymbol{\theta})}{f(\boldsymbol{x}|\boldsymbol{\theta}^0)} \, d\boldsymbol{\theta}\right)^{-1}, \tag{10}$$

where $V(\boldsymbol{\theta}^0, \varepsilon)$ is the volume of $B(\boldsymbol{\theta}^0, \varepsilon)$, the sphere of radius ε centered at $\boldsymbol{\theta}^0$.

Proof: See the Appendix.

To check how this result works, we consider the multivariate normal distribution.

Example 1. Suppose $\boldsymbol{X}$ is $N_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$ distributed, σ^2 known, where $\boldsymbol{X} = (X_1, \ldots, X_p)'$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)'$. It is desired to test (1) with a sample of size n. The classical significance test statistic is

$$T(\overline{\boldsymbol{X}}, \boldsymbol{\theta}^0) = \frac{n}{\sigma^2} |\overline{\boldsymbol{X}} - \boldsymbol{\theta}^0|^2, \tag{11}$$

with $\overline{\boldsymbol{X}} = (\overline{X}_1, \ldots, \overline{X}_p)$. Under the null hypothesis, H_0, $T(\overline{\boldsymbol{X}}, \boldsymbol{\theta}^0)$ has a χ^2_p distribution. Therefore, the p–value of the observed data, $\overline{\boldsymbol{x}}$, is given by $p(\overline{\boldsymbol{x}}) = P\left\{\chi^2_p \geq T(\overline{\boldsymbol{x}}, \boldsymbol{\theta}^0)\right\}$.

Using Theorem 1, the infimum of the posterior probability of the point null over the class Γ_{US} is

$$\inf_{\pi \in \Gamma_{US}} P(H_0|\overline{\boldsymbol{x}}) = \left\{1 + \frac{1}{V(\boldsymbol{\theta}^0, \varepsilon)} \int_{R^p} \frac{\exp\left(-\frac{n}{2\sigma^2}|\overline{\boldsymbol{x}} - \boldsymbol{\theta}|^2\right)}{\exp\left(-\frac{n}{2\sigma^2}|\overline{\boldsymbol{x}} - \boldsymbol{\theta}^0|^2\right)} \, d\boldsymbol{\theta}\right\}^{-1} \tag{12}$$

But, $V(\boldsymbol{\theta}^0, \varepsilon) = \pi^{p/2} \varepsilon^p / \Gamma(p/2 + 1)$ and then

$$\inf_{\pi \in \Gamma_{US}} P(H_0|\overline{\boldsymbol{x}}) = \left\{1 + \frac{2^{p/2}\Gamma\left(\frac{p}{2} + 1\right)}{\varepsilon^{*p}} \exp\left(\frac{1}{2} T(\overline{\boldsymbol{x}}, \boldsymbol{\theta}^0)\right)\right\}^{-1}, \tag{13}$$

where $\varepsilon^* = \varepsilon\sqrt{n}/\sigma^2$. Fixed ε^*, the space dimension, p, and $\boldsymbol{\theta}^0 = \boldsymbol{0}$ for different values of $T(\overline{\boldsymbol{x}}, \boldsymbol{\theta}^0)$ 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 close, these values depend on the space dimension and they become greater as p increases.

Table 1 goes here

It can be observed, in Table 1, robustness with respect to the data for every dimension, p. For instance, when the dimension is $p = 5$ if we choose $\varepsilon^* \in (2.9, 3.5)$ the infimum of the posterior probability of H_0 is close to the different p–values. For different dimensions p, Figure 1 shows the infimum of the posterior probability with suitable values of ε^*, the p–value and the infimum of the posterior probability when $\pi_0 = 0.5$.

Figure 1 goes here

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

Table 2 goes here

It is clear, from Table 2, that 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, as the second line of Table 2 shows.

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: Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ a sample from a $N_p(\boldsymbol{\theta}, \Sigma)$, 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 $\sigma^2 known$ and being ρ the correlation coefficient. It is a permutation–symmetric multivariate normal distribution that gives exchangeable variables for $\rho \geq 0$ (Tong, 1990). Then, $|\Sigma| = \sigma^{2p}(1-\rho)^{p-1}(1+(p-1)\rho)$ and

$$\inf_{\pi \in \Gamma_{US}} P(H_0|\overline{\boldsymbol{x}}) = \left\{ 1 + \frac{\Gamma\left(\frac{p}{2}+1\right)}{\varepsilon^{*p}} (2(1-\rho)^{p-1}(1+(p-1)\rho))^{p/2} \exp\left(\frac{t}{2}\right) \right\}^{-1},$$

with $t = n\overline{\boldsymbol{x}}'\Sigma^{-1}\overline{\boldsymbol{x}}$. 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.

Figure 2 goes here

It can be also noted that if ρ increases then the posterior probability increases too, being larger than the p–value with the same value of π_0.

2.2. Lower bounds for scale mixture of Normals

In this section we assume that $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p)$ is a random sample of a $N_p(\boldsymbol{\theta}, \sigma^2\boldsymbol{I})$ distribution, σ^2 known, where $\boldsymbol{I}$ is the $p \times p$ identity matrix. Then $\overline{\boldsymbol{X}}$ is $N_p(\boldsymbol{\theta}, \sigma^2/n\boldsymbol{I})$ distributed. It is desired to test (1) with $\boldsymbol{\theta}^0 = \boldsymbol{0}$, then the appropriate statistic is $T(\overline{\boldsymbol{X}}) = n/\sigma^2\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}$. Now, the considered prior density $\pi(\boldsymbol{\theta})$ belongs to the class of scale mixture of normals

$$\Gamma_N = \{\boldsymbol{\theta}|v^2 \sim N_p(0, v^2\boldsymbol{I}),\ \pi(v^2) \text{ is a nondecreasing function on } (0,\infty)\}. \qquad (14)$$

The reason to consider this class of priors is that it assigns higher mass to the neighbors of a precise hypothesis than the class Γ_{US} can assign (see Sellke, Bayarri and Berger, 2001).

To find the lower bound on the posterior probability over the class (14) is equivalent to find it over the smaller class in which $\pi(v^2)$ is uniform on $(0, r)$, $r > 0$ (see Casella and Berger, 1987).

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

Theorem 2. *If the prior mass assigned to the null hypothesis is, from (4),*

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

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

$$\inf_{\pi \in \Gamma_N} P(H_0|t) = \left(1 + \frac{1}{\varepsilon^{*2}} \frac{\mathcal{F}_{p-2}(t)}{f_p(t)}\right)^{-1}, \qquad (15)$$

with $t = n/\sigma^2 \overline{\boldsymbol{x}}' \overline{\boldsymbol{x}}$, $\mathcal{F}_{p-2}$ *the chi–squared distribution function with* $p-2$ *degrees of freedom and* f_p *the corresponding density with* p *d.f.*

Proof: See the Appendix.

Then, fixed ε^* and the space dimension p, the infimum of the posterior probability can be obtained. Table 3 shows the values of ε^* making equal the infimum of the posterior probabilities for the class of scale mixture priors and the p–value.

Table 3 goes here

In order to compare numerically the p–value with the infimum of the posterior probability, Table 4 shows for $p = 15$ this infimum for some suitable values of ε^*, chosen from Table 3, and the infimum of the posterior probability when $\pi_0 = 0.5$.

Table 4 goes here

It can be pointed out the practical agreement between Bayesian and classical measures for the different values of ε^*. In particular, for an intermediate value of $\varepsilon^* = 1.95$, the infimun of the posterior probability and the p–value are nearly the same. However, for $\pi_0 = 0.5$ these ones are significantly different.

Figure 3 shows the graphics of the lower bounds of the posterior null probability and the p–value jointly with the posterior probability for $\pi_0 = 0.5$. We have considered just a value of ε^* for each dimension because of the robustness property.

Figure 3 goes here

3. Conclusions and comments

The most important conclusion is that the p–values and the posterior probabilities can be matched for testing multivariate precise hypothesis. The difference between this 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 paper is to give a prior probability for $\boldsymbol{\theta}^0$ equal to the probability of a sphere with a fixed radius ε and centered at $\boldsymbol{\theta}^0$, calculated from $\pi(\boldsymbol{\theta})$. This methodology shows, for the considered examples, a better approximation between the p–value and some infimum of the posterior probabilities, using further just one source of information $\pi(\boldsymbol{\theta})$.

It can be also pointed out that an apparent robustness in ε^* is observed inside every dimension p for the normal distribution and the class Γ_{US}. But for the class of scale mixtures, this robustness is observed not only inside every p but with respect to p. In both cases, for a fixed dimension p, the value of $\varepsilon^* = \varepsilon\sqrt{n}/\sigma$ does not change significantly for the different p–values.

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.

APPENDIX

Proof of Theorem 1. the posterior probability of the point null hypothesis, H_0, is given by

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

where $m_\pi(\boldsymbol{x}) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}) f(\boldsymbol{x}|\boldsymbol{\theta})\, d\boldsymbol{\theta}$. Then, computing the infimum of the posterior probability of the point null hypothesis is just like computing the supremum of $M(k) = \dfrac{1-\pi_0}{\pi_0} m_\pi(\boldsymbol{x})$ over the class (9).

Assuming $\varepsilon < k$ and denoting by $U(\boldsymbol{\theta}^0, k)$ the uniform distribution over the sphere of radius k centered at $\boldsymbol{\theta}^0$, then

$$\pi_0 = \int_{B(\theta^0,\varepsilon)} U(\boldsymbol{\theta}^0, k)\, d\boldsymbol{\theta} = \frac{V(\boldsymbol{\theta}^0, \varepsilon)}{V(\boldsymbol{\theta}^0, k)}$$

and

$$M(k) = \frac{V(\boldsymbol{\theta}^0, k) - V(\boldsymbol{\theta}^0, \varepsilon)}{V(\boldsymbol{\theta}^0, k)V(\boldsymbol{\theta}^0, \varepsilon)} \int_{B(\theta^0, k)} f(\boldsymbol{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

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,

$$\sup_k M(k) = \lim_{k\to\infty} M(k) = \frac{1}{V(\boldsymbol{\theta}, \varepsilon)} \int_{R^p} f(\boldsymbol{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta}$$

and from this we get (10). □

PROOF OF THEOREM 2. First, it can be easily checked that

$$\pi_{0r} = \frac{1}{r} \int_0^r \mathcal{F}_p\left(\frac{\varepsilon^2}{t}\right) dt.$$

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

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

and

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

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

$$P_r(H_0|t) = \left\{1 + \frac{1-\pi_{0r}}{r\pi_{0r}} \frac{\sigma^2}{n(p-2)} \frac{\mathcal{F}_{p-2}(t) - \mathcal{F}_{p-2}\left(\frac{\sigma^2 t}{\sigma^2+rn}\right)}{f_p(t)}\right\}^{-1} \quad (16)$$

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

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

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

As we get

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

finally (15) is obtained.

References

Berger, J.O., Delampady, M., 1987. Testing precise hyphoteses. Statist. Sci. 2 (3), 317–352.

Berger, J.O., Sellke, T., 1987. Testing a point null hyphoteses: the irreconciliability of p values and evidence (with discussions). J. Amer. Statist. Assoc. 82, 112–139.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. John Wiley and Sons, New York.

Casella, G., Berger, R.L., 1987. Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with discussions). J. Amer. Statist. Assoc. 82, 106–111.

Gómez–Villegas, M.A., Gómez, E., 1992. Bayes factor in testing precise hyphoteses. Commun. Statist. - Theory Meth. 21(6), 1707-1715.

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

Gómez–Villegas, M.A., Sanz, L., 2000. ε–contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. Statistics and Probability Letters 47 (1), 53–60.

Oh, M.S., 1998. A Bayes test for simple versus one–sided hypothesis on the mean vector of a multivariate normal distribution. Commun. Statist.–Theory Meth. 27 (10), 2371–2389.

Oh, H.S., DasGupta, A., 1999. Comparison of the p–value and posterior probability. Journal of Statistical Planning and Inference 76, 93–107.

Rao, C.R., 1948. Test of significance in multivariate analysis. Biometrika 35, 58–79.

Sellke, T., Bayarri, M.J., Berger, J., 2001. Calibration of P–values for Testing Precise Null Hypotheses. Amer. Statist. 55, 62–71.

Tong, 1990. The Multivariate Normal Distribution. Springer Verlag, New York.

Table 1: Values of ε^* so that the p–values and the infimum of the posterior probability over Γ_{US} are close

	Dimension p			
p–value	2	5	10	20
0.1	1.493	2.919	4.078	5.485
0.05	1.450	3.038	4.246	5.695
0.01	1.421	3.247	4.599	6.123
0.001	1.417	3.516	5.024	6.616

Table 2: P–values and infimum of the posterior probabilities for uniform and symmetric priors

p(t)	0.1	0.05	0.01	0.001
$\underline{\Pr}\left(H_0\vert t, \pi_0 = 0.5\right)$	0.5401	0.3707	0.1053	0.0116
$\underline{\Pr}\left(H_0\vert t, \varepsilon^* = 1.4\right)$	0.0891	0.0467	0.0097	0.0010
$\underline{\Pr}\left(H_0\vert t, \varepsilon^* = 1.45\right)$	0.0949	0.0500	0.0104	0.0011
$\underline{\Pr}\left(H_0\vert t, \varepsilon^* = 1.5\right)$	0.1009	0.0533	0.0111	0.0011

Table 3: Values of ε^* such that the p–values and the infimum of the posterior probability over Γ_N are close

	Dimension p			
p–value	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 4: P–values and infimum of the posterior probabilities for scale mixture priors

p(t)	0.1	0.05	0.01	0.001
$\underline{\mathrm{Pr}}\,(H_0 \vert t, \pi_0 = 0.5)$	0.365	0.27	0.1	0.017
$\underline{\mathrm{Pr}}\,(H_0 \vert t, \varepsilon^* = 2.07)$	0.100	0.056	0.013	0.0014
$\underline{\mathrm{Pr}}\,(H_0 \vert t, \varepsilon^* = 1.95)$	0.090	0.050	0.011	0.0013
$\underline{\mathrm{Pr}}\,(H_0 \vert t, \varepsilon^* = 1.81)$	0.078	0.043	0.010	0.0011