

Bayes Factors for Comparing Two Restricted Means: An Example Involving Hypertense Individuals

Viviana Giampaoli¹ and Julio M. Singer²

¹Universidade Federal de Pernambuco and ²Universidade de São Paulo

Abstract: We are interested in comparing the mean the diastolic blood pressure of individuals submitted to a stress stimulus to that of individuals under normal conditions with the prior knowledge that the subjects in both groups are hypertense. Essentially, this may be formulated as a two sample problem for Gaussian populations with bounded means. For such purposes, we consider two different approaches to obtain Bayes factors. The first is based on predictive distributions and the second is based on Markov Chain Monte Carlo methods. The sensitivity of the Bayes factors with respect to choice of prior distributions is also investigated.

Key words: Bayes factor, blood pressure, Gibbs sampler, predictive posterior distribution, restricted parameter spaces.

1. Introduction

Giampaoli and Singer (2004) consider the problem of comparing the mean diastolic blood pressure of individuals submitted to a stress stimulus (like the death of a close relative or discharge from employment) to that of individuals under normal (no stress) conditions based on blood pressure measurements obtained on two independent samples of sizes n_1 and n_2 , respectively. The data are reproduced in Table 1, the entries of which correspond to the average of series of 30 measurements taken over periods of one hour to eliminate short term fluctuations.

Assuming normality and homocedasticity, this is the typical situation for which Student's t test is an appropriate statistical tool. In fact, a simple two sample t -test with $df=20$ yields a p -value of 0.0595. For the sake of comparison, the p -value for the corresponding Wilcoxon two sample test is 0.2929. Hence the data does not provide sufficient evidence for rejection of the null hypothesis that the mean diastolic blood pressure of subjects under normal or stress conditions are equal. Here, however, we are interested in comparing the corresponding means with the additional information that the subjects in both groups are hypertense, i.e., have mean diastolic blood pressures known to be at least 90 mmHg. This

Table 1: Diastolic blood pressure (mmHg)

Stress group										
89.6	92.4	87.1	92.2	96.4	92.2	92.7	95.0	92.2	109.2	96.8
Control group										
81.7	89.9	81.5	89.4	94.6	93.5	85.5	95.5	88.9	95.4	97.0

occurs, for example, in studies where patients are being screened for the detection of possible causes of hypertension.

More specifically, let $\mathbf{y}_1 = (y_{11}, \dots, y_{1n_1})'$ and $\mathbf{y}_2 = (y_{21}, \dots, y_{2n_2})'$ respectively denote independent random samples of $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ distributions. Under this framework, we consider tests of the hypothesis

$$H_0 : \mu_1 = \mu_2 \quad (1.1)$$

versus

$$H_1 : \mu_1 > \mu_2 \quad (1.2)$$

with or without the assumption that σ^2 is known under the restriction that $\mu_1 \geq c$ and $\mu_2 \geq c$ where c is some constant. We may take $c = 0$ without loss of generality.

Classical statistical inference under constrained parametric spaces has been addressed by many authors, among which we mention Bartholomew (1959a, 1959b, 1961), Kudô (1963), Nüesch (1966), Perlman (1969) and Barlow et al. (1972). In particular, Giampaoli and Singer (2004) proposed tests for the comparison of Gaussian distributions with restricted means (assuming or not known variance), which may be considered as alternatives to the usual Z or t tests. Using simulation studies, the author has shown that the proposed tests are more powerful than the usual Z or t tests in most circumstances. However, they are only approximate level α tests and depend on estimators of the unknown parameters. We attack the problem under a Bayesian approach, since it seems that from such a perspective, the restricted parameter space can be handled with less effort than under a classical point of view.

Testing the hypothesis (1.1) is equivalent to comparing the two following models

$$M = 1 : y_{ik} = \mu + \varepsilon_{ik}, \quad \varepsilon_{ik} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (1.3)$$

$$M = 2 : y_{ik} = \mu_i + \varepsilon_{ik}, \quad \varepsilon_{ik} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (1.4)$$

with $\mu_1 > \mu_2$, $i = 1, 2, k = 1, \dots, n_i$ and σ^2 is viewed as a nuisance parameter. Here M is an integer-valued parameter indexing the models, such that $M = 1$ corresponds to H_0 and $M = 2$ corresponds to H_1 .

In Section 2 we present details for the computation of the Bayes factors under two different approaches assuming that the common variance is known. In Section 3 we illustrate the procedures with the data depicted in Table 1. Section 4 is devoted to an extension for the unknown variance case. We conclude with a discussion of the results in Section 5.

2. Bayes Factors

The ratio of the observed marginal densities for the two models, namely

$$B_{21} = \frac{p(\mathbf{y}|M=2)}{p(\mathbf{y}|M=1)} = \frac{p(M=2|\mathbf{y})\pi_1}{p(M=1|\mathbf{y})\pi_2}, \quad (2.1)$$

where $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2)'$ denotes the vector of observations and $\pi_j = p(M=j)$, $j = 1, 2$, $\sum_{i=1}^K \pi_i = 1$ represent the prior model probabilities, is known as the Bayes factor and may be used for deciding in favor or against each model. Kass and Raftery (1995) discuss the use of these factors for hypothesis testing and suggest the rule displayed in Table 2 for deciding between one model or the other.

Table 2: Kass and Raftery's decision rule based on Bayes factors.

Bayes Factor (B_{21})	Evidence against H_0
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
>150	Very Strong

Although Bayes factors constitute important tools for statistical analysis, their use is not free of controversy. First, they may be severely affected by variation of prior distributions. Also, they lack interpretation when improper prior distributions are used, as pointed by O'Hagan (1995). Variations commonly employed for model comparison include pseudo-Bayes factors as advocated by Geisser and Eddy (1979), Gelfand et al. (1992) or Gelfand and Dey (1998), posterior Bayes factors as suggested by Aitkin (1991) or criteria that minimize some posterior loss function as proposed by Gelfand and Ghosh (1998). We can also mention intrinsic Bayes factors suggested by Berger and Pericchi (1996) and fractional Bayes factors, discussed in O'Hagan (1995) or De Santis and Spezzaferri

(1997). These techniques, however, use training samples and are unstable when the size of sample is small as in the example described above.

To compute Bayes factors, we need to obtain the posterior probability associated with each model, namely, $p(M = j|\mathbf{y})$, $j = 1, 2$. We consider two different techniques for such purposes. The first, due to Irony and Pereira (1995) provides an analytic expression for the Bayes factor obtained by a direct computation of the posterior distribution; it accommodates any proper prior distribution and does not require Gibbs sampling, since usual methods of numerical integration may be employed. The second, proposed by Carlin and Chib (1995), is based on the Gibbs sampler and on Markov Chain Monte Carlo methods for the computations. It requires either the use of conjugate prior distributions or the existence of conditional complete conjugation or conditional complete posterior density. Unfortunately, a direct comparison of the results obtained by the two techniques is quite complicated, since the first technique relies on a prior distribution specified for the entire parameter space while the second requires different prior distributions under the null and the alternative hypotheses.

2.1 Computation of Bayes factors via predictive distributions

First note that for $j = 1, 2$, a direct application of Bayes's Theorem yields

$$p(M = j|\mathbf{y}) = \frac{\pi_j p(\mathbf{y}|M = j)}{\pi_1 p(\mathbf{y}|M = 1) + \pi_2 p(\mathbf{y}|M = 2)}. \quad (2.2)$$

Then let

$$\begin{aligned} \Omega_1 &= \{\boldsymbol{\mu} = (\mu_1, \mu_2) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ and } \mu_1 = \mu_2\}, \\ \Omega_2 &= \{\boldsymbol{\mu} = (\mu_1, \mu_2) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ and } \mu_1 > \mu_2\} \end{aligned}$$

and note that the null ($j = 1$) and the alternative ($j = 2$) hypotheses may be written as

$$H_j : \boldsymbol{\mu} \in \Omega_j. \quad (2.3)$$

The prior density of $\boldsymbol{\mu}$ under H_j , ($j = 1, 2$) is defined as

$$g(\boldsymbol{\mu}|M = j) = \frac{g_j(\boldsymbol{\mu})}{\int_{\Omega_j} g_j(\boldsymbol{\mu}) d\boldsymbol{\mu}} \quad (2.4)$$

where g is a convenient prior density function and g_j denotes the function g restricted to the set Ω_j , i.e., the function with domain Ω_j that $g_j(\boldsymbol{\mu}) = g(\boldsymbol{\mu})$ if $\boldsymbol{\mu} \in \Omega_j$. Note that the integral $\int_{\Omega_j} g_j d\Omega_j$ represents the volume under the prior density function $g(\boldsymbol{\mu})$ in Ω_j . When the integral is null as in the case where the dimension of Ω_j is less than the dimension of the parameter space, we consider

the corresponding line integral. Naturally the prior density (2.4) is well defined if $0 < \int_{\Omega_j} g d\Omega_j < \infty$.

Under the hypothesis (2.3), $j = 1, 2$, the required predictive densities are

$$p(\mathbf{y}|M = j) = \frac{\int_{\Omega_j} g_j(\boldsymbol{\mu})l(\mathbf{y}|\boldsymbol{\mu})d\boldsymbol{\mu}}{\int_{\Omega_j} g_j(\boldsymbol{\mu})d\boldsymbol{\mu}}. \tag{2.5}$$

where $l(\mathbf{y}|\boldsymbol{\mu})$ is the likelihood function, then $p(M = j|\mathbf{y})$ may be obtained from (2.2).

2.2 Computation of Bayes factors via MCMC methods

We now describe an alternative way of computing $p(M = j|\mathbf{y})$, $j = 1, 2$ via an MCMC algorithm proposed by Carlin and Chib (1995). Essentially they consider M as an component of the random vector $(\mu, (\mu_1, \mu_2), M)$ so can it be sampled via Gibbs methods. After convergence, the estimators of $p(M = j|\mathbf{y})$, $j = 1, 2$ may be obtained as the ratios between the number of iterations for which $M = j$ and the total number of iterations. Direct sampling of the marginal distributions themselves or of the joint distribution $p(\mathbf{v}, \mathbf{y})$ is complicated, but sampling of the full conditional posterior distributions $p(\mu|(\mu_1, \mu_2), M, \mathbf{y})$, $p((\mu_1, \mu_2)|\mu, M, \mathbf{y})$ and $p(M|\mu, (\mu_1, \mu_2), \mathbf{y})$ is straightforward using the Gibbs sampler. To satisfy the MCMC conditions for convergence, we need to specify a full probability model. For such purposes we assume that μ and (μ_1, μ_2) are independent given the model indicator M , and that the prior distributions $p(\mu|M = j)$ and $p((\mu_1, \mu_2)|M = j)$, $j = 1, 2$ are proper. Thus

$$p((\mu, \mu_1, \mu_2)|M = j) = p(\mu|M = j)p(\mu_1, \mu_2|M = j). \tag{2.6}$$

We also assume that \mathbf{y} is independent of (μ_1, μ_2) given $M = 1$ and of μ given $M = 2$. We complete the Bayesian model specification by choosing proper “pseudoprior” distributions $p(\mu|M = 2)$ and $p(\mu_1, \mu_2|M = 1)$. Because of the conditional independence assumptions, these “pseudoprior” distributions do not interfere with the expression of the marginal densities and so their form is irrelevant. The joint distribution of \mathbf{y} and $(\mu, (\mu_1, \mu_2))$ given $M = j$ is

$$p(\mathbf{y}, \mu, (\mu_1, \mu_2)|M = j) = f(\mathbf{y}|\mu, (\mu_1, \mu_2), M = j)p(\mu, (\mu_1, \mu_2)|M = j)$$

where $f(\mathbf{y}|\mu, (\mu_1, \mu_2), M = j)$ is the associated density function which is equal to $f(\mathbf{y}|\mu, M = 1)\pi_1$ for $j = 1$ and $f(\mathbf{y}|\mu, (\mu_1, \mu_2), M = 2)\pi_2$ for $j = 2$. To implement the Gibbs sampler we need the full conditional distributions of μ and (μ_1, μ_2) as well as that of M . The former are given by

$$p(\mu|(\mu_1, \mu_2), M, \mathbf{y}) \propto \begin{cases} f(\mathbf{y}|\mu, M = 1)p(\mu|M = 1) & \text{if } M = 1 \\ p(\mu|M = 2) & \text{if } M = 2 \end{cases} \tag{2.7}$$

and

$$p((\mu_1, \mu_2)|\mu, M, \mathbf{y}) \propto \begin{cases} f(\mathbf{y}|\mu_1, \mu_2), M = 2)p((\mu_1, \mu_2)|M = 2) & \text{if } M = 2 \\ p((\mu_1, \mu_2)|M = 1) & \text{if } M = 1, \end{cases} \quad (2.8)$$

Thus, when $M = 1$, the required distribution for μ (2.7) is generated from model 1; otherwise, when $M = 2$ the distribution is generated from the corresponding “pseudoprior” distribution. Similarly, when $M = 1$, the required distribution for (μ_1, μ_2) (2.8) is generated from the corresponding “pseudoprior” distribution; otherwise, the required distribution is generated from model 2.

For M , we have

$$p(M = 1|\mu, (\mu_1, \mu_2), \mathbf{y}) = \frac{f(\mathbf{y}|\mu, M = 1)p(\mu, (\mu_1, \mu_2)|M = 1)\pi_1}{\sum_{k=1}^2 p(\mathbf{y}, \mu, (\mu_1, \mu_2), M = k)} \quad (2.9)$$

and

$$p(M = 2|\mu, (\mu_1, \mu_2), \mathbf{y}) = \frac{f(\mathbf{y}|(\mu_1, \mu_2), M = 2)p(\mu, (\mu_1, \mu_2)|M = 2)\pi_2}{\sum_{k=1}^2 p(\mathbf{y}, \mu, (\mu_1, \mu_2), M = k)} \quad (2.10)$$

Thus, samples of the posterior distribution $p(\mu, (\mu_1, \mu_2), M|\mathbf{y})$ may be obtained by samples from the full conditional distributions (2.7) - (2.8) and (2.9) - (2.10), via Gibbs algorithm. In each iteration, $l = 1, \dots, N$, a sample of size one $(\mu^{(l)}, (\mu_1, \mu_2)^{(l)}, M^{(l)})$, is obtained. The ratio

$$\hat{P}(M = j|\mathbf{y}) = \frac{\text{number of } M^{(l)} = j \text{ in the } N \text{ iterations}}{N}, \quad j = 1, 2,$$

provides a simple estimate of $p(M = j|\mathbf{y})$ that may be used to compute the Bayes factor from (2.1).

Although the form of the “pseudoprior” distributions is theoretically arbitrary, it is convenient to have them close to the conditional densities $p(\mu|(\mu_1, \mu_2), M, \mathbf{y})$ and $p((\mu_1, \mu_2)|\mu, M, \mathbf{y})$ so plausible values are generated even when the assumed model is false. Carlin and Chib (1995) recommend separate runs for each model, i.e., considering $\pi_1 = 1$ and $\pi_2 = 0$, and an approximation of the resulting posterior distribution by pseudoprior distribution for the parameter μ under model $M = 2$. Subsequently, considering $\pi_2 = 0$ and $\pi_1 = 1$, an approximation of the resulting posterior distribution may also be taken as pseudoprior distribution for the parameters (μ_1, μ_2) under model $M = 1$.

3. Analysis of the Diastolic Blood Pressure Data

In this section we consider the analysis of the data presented in Table (1) assuming that the variance is known and using either truncated Gaussian distributions or exponential distributions as prior distributions for μ_1 and μ_2 . These

distributions do not belong to conjugate families of distributions. We start with the technique proposed by Irony and Pereira (1995).

3.1 Analysis via predictive distributions

The likelihood function is

$$\begin{aligned}
 l(\mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\mu}) &= K_1 \exp \left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + n_1 (\bar{y}_1 - \mu_1)^2 \right] \right) \\
 &\quad \times \exp \left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 + n_2 (\bar{y}_2 - \mu_2)^2 \right] \right) \quad (3.1)
 \end{aligned}$$

where $\tau = \sigma^{-2}$ and $K_1 = (\tau/2\pi)^{(n_1+n_2)/2}$.

We assume that μ_1 and μ_2 have Gaussian prior distributions truncated at zero, with parameters (a_1, γ_1) and (a_2, γ_2) respectively, i.e., with density functions

$$g(\mu_j) = \frac{1}{1 - \Phi(-a_j\sqrt{\gamma_j})} \frac{\sqrt{\gamma_j}}{\sqrt{2\pi}} \exp \left[\frac{-\gamma_j}{2} (\mu_j - a_j) \right]^2 I_{[0,\infty)}, \quad (3.2)$$

with Φ denoting the standard Gaussian distribution function, $j = 1, 2$, so that

$$E(\mu_j) = a_j + \frac{1}{1 - \Phi(-a_j\sqrt{\gamma_j})} \frac{\sqrt{\gamma_j}}{\sqrt{2\pi}} \exp(-\gamma_j a_j/2), \quad (3.3)$$

$$\begin{aligned}
 Var(\mu_j) &= \gamma_j^{-1} \left\{ 1 - \frac{a_j}{1 - \Phi(-a_j\sqrt{\gamma_j})} \frac{\sqrt{\gamma_j}}{\sqrt{2\pi}} \exp(-\gamma_j a_j^2/2) \right. \\
 &\quad \left. - \left[\frac{1}{1 - \Phi(-a_j\sqrt{\gamma_j})} \frac{\sqrt{\gamma_j}}{\sqrt{2\pi}} \exp(-\gamma_j a_j^2/2) \right]^2 \right\}. \quad (3.4)
 \end{aligned}$$

Using the independence assumption on μ_1, μ_2 it follows that the corresponding joint prior density is

$$g(\mu_1, \mu_2) = K_2 \frac{\sqrt{\gamma_1\gamma_2}}{2\pi} \exp \left[-\frac{\gamma_1}{2} (\mu_1 - a_1)^2 - \frac{\gamma_2}{2} (\mu_2 - a_2)^2 \right] I_{[0,\infty) \times [0,\infty)}, \quad (3.5)$$

where $K_2 = [\Phi(a_1\sqrt{\gamma_1})\Phi(a_2\sqrt{\gamma_2})]^{-1}$.

From (3.1) and (3.5), we may write the joint density of the data and the

parameter $\boldsymbol{\mu} = (\mu_1, \mu_2)$ as

$$\begin{aligned} g(\mu_1, \mu_2)l(\mathbf{y}|\mu_1, \mu_2) &= K_1 K_2 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \exp \left\{ -\frac{\gamma_1}{2}(\mu_1 - a_1)^2 - \frac{\gamma_2}{2}(\mu_2 - a_2)^2 \right\} \\ &\quad \times \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + n_1(\bar{y}_1 - \mu_1)^2 \right] \right\} \\ &\quad \times \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 + n_2(\bar{y}_2 - \mu_2)^2 \right] \right\}. \end{aligned} \quad (3.6)$$

Let $g_1(\mu) = g(\mu, \mu)$ and $\Omega_1 = \{\mu : \mu \geq 0\}$. Then the denominator of (2.5) is

$$\begin{aligned} \int_{\Omega_1} g_1(\mu) d\mu &= K_2 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \exp \left[-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2 \right] d\mu \\ &= K_2 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \exp \left[-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)} \right] \\ &\quad \times \Phi \left[\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2} \right]. \end{aligned} \quad (3.7)$$

The corresponding numerator is

$$\begin{aligned} &\int_{\Omega_1} g_1(\mu)l(\mathbf{y}|\mu, \mu) d\mu \\ &= K_1 K_2 K_3 \exp \left[-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)} \right] \exp \left[-\frac{\tau n_1 n_2 (\bar{y}_1 - \bar{y}_2)^2}{2(n_1 + n_2)} \right] \\ &\quad \times \exp \left\{ -K_4 \left[\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} - \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2}{(n_1 + n_2)} \right]^2 \right\} \\ &\quad \times \sqrt{\frac{2\pi}{K_6}} \Phi(K_5 \sqrt{K_6}), \end{aligned} \quad (3.8)$$

with

$$K_3 = \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 \right] \right\},$$

$$K_4 = \frac{(\gamma_1 + \gamma_2)\tau(n_1 + n_2)}{2[(\gamma_1 + \gamma_2) + \tau(n_1 + n_2)]},$$

$$K_5 = \frac{(\gamma_1 a_1 + \gamma_2 a_2) + \tau(n_1 \bar{y}_1 + n_2 \bar{y}_2)}{\gamma_1 + \gamma_2 + \tau(n_1 + n_2)}.$$

and

$$K_6 = [\gamma_1 + \gamma_2 + \tau(n_1 + n_2)].$$

From (3.7) and (3.8) it follows that the predictive density (2.5) is

$$\begin{aligned}
 p(\mathbf{y}|M = 1) &= K_1 K_3 \sqrt{\frac{2\pi}{K_6}} \left[\frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \right]^{-1} \Phi(K_5 \sqrt{K_6}) \\
 &\times \exp \left[-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)} \right] \exp \left[-\frac{\tau n_1 n_2 (\bar{y}_1 - \bar{y}_2)^2}{2(n_1 + n_2)} \right] \\
 &\times \exp \left[-K_4 \left(\frac{\gamma_1 a_1 + \gamma_2 a_2}{\gamma_1 + \gamma_2} - \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2}{n_1 + n_2} \right)^2 \right] \\
 &\times \left\{ \Phi \left[\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2} \right] \right\}^{-1}. \tag{3.9}
 \end{aligned}$$

Similarly, we may compute

$$\begin{aligned}
 &\int_{\Omega_2} g(\boldsymbol{\mu}) d\boldsymbol{\mu} \\
 &= K_2 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \int_0^{\mu_1} \exp \left[-\frac{\gamma_1}{2} (\mu_1 - a_1)^2 - \frac{\gamma_2}{2} (\mu_2 - a_2)^2 \right] d\mu_2 d\mu_1 \\
 &= K_2 \left[\sqrt{\frac{\gamma_1}{2\pi}} h(\gamma_1, a_1, \gamma_2, a_2) - \Phi(-a_2 \sqrt{\gamma_2}) \Phi(a_1 \sqrt{\gamma_1}) \right], \tag{3.10}
 \end{aligned}$$

with $h(b, a, m, n) = \int_0^\infty \exp[-b(u - a)^2/2] \Phi[\sqrt{m}(u - n)] du$ as well as

$$\begin{aligned}
 &\int_{\Omega_2} g(\boldsymbol{\mu}) l(\mathbf{y}|\boldsymbol{\mu}) d\boldsymbol{\mu} \\
 &= K_1 K_2 K_3 \int_0^\infty \int_0^{\mu_1} g(\mu_1, \mu_2) l(\mu_1, \mu_2) d\mu_2 d\mu_1 \\
 &= K_1 K_2 K_3 \exp \left[-v_1 (\bar{y}_1 - a_1)^2 - v_2 (\bar{y}_2 - a_2)^2 \right] \\
 &\times \left\{ \sqrt{\frac{\pi}{u_2}} h(2u_1, \bar{\mu}_1, 2u_2, \bar{\mu}_2) - \frac{\pi}{\sqrt{u_2 u_1}} \Phi(-\bar{\mu}_2 \sqrt{2u_2}) \Phi(\bar{\mu}_1 \sqrt{2u_1}) \right\} \tag{3.11}
 \end{aligned}$$

with

$$\bar{\mu}_j = \frac{\gamma_j a_j + \tau n_j \bar{y}_j}{\gamma_j + \tau n_j}, \quad v_j = \frac{\gamma_j \tau n_j}{2(\gamma_j + \tau n_j)}, \quad u_j = \frac{\gamma_j + \tau n_j}{2},$$

for $j = 1, 2$. Using (3.10) and (3.11), we conclude that the predictive density (2.5) under the alternative hypothesis is

$$\begin{aligned}
 p(\mathbf{y}|M = 2) &= K_1 K_3 \exp[-v_1 (\bar{y}_1 - a_1)^2 - v_2 (\bar{y}_2 - a_2)^2] \\
 &\times \left[\sqrt{\frac{\pi}{u_2}} h(2u_1, \bar{\mu}_1, 2u_2, \bar{\mu}_2) - \frac{\pi}{\sqrt{u_2 u_1}} \Phi(-\bar{\mu}_2 \sqrt{2u_2}) \Phi(\bar{\mu}_1 \sqrt{2u_1}) \right] \\
 &\times \left[\sqrt{\frac{\gamma_1}{2\pi}} h(\gamma_1, a_1, \gamma_2, a_2) - \Phi(-a_2 \sqrt{\gamma_2}) \Phi(a_1 \sqrt{\gamma_1}) \right]^{-1}. \tag{3.12}
 \end{aligned}$$

Then, choosing, for example, equal prior probabilities for both models, i.e., ($\pi_1 = \pi_2 = 1/2$), the Bayes factor (2.1) is the ratio between (3.12) and (3.9).

Using these results, we computed Bayes factors for the data in Table 1 under different choices for the hyperparameters. We assumed σ^2 to be known with a value equal to its estimate based on residual sum of least squares; thus $\tau = \sigma^{-2} = 0.03$. To obtain the hyperparameters for the distributions of μ_1 and μ_2 , we fixed $E(\mu_j)$ and $\tau_j = [Var(\mu_j)]^{-1}$, $j = 1, 2$ and solved the system of equations defined by (3.3) and (3.4) for a_1 , γ_1 , a_2 and γ_2 . The corresponding Bayes factors obtained by fixing $E(\mu_1) = \bar{y}_1 - 90 = 4.1636$, and $E(\mu_2) = \bar{y}_2 - 90 = 0.2636$ for different values of $\tau_1 = \tau_2$ are presented in the Table 3.

Table 3: Bayes factors for different values of the hyperparameters

$\tau_1 = \tau_2$	Hyperparameters				Bayes factor (B_{21})
	a_1	γ_1	a_2	γ_2	
0.35000	4.15	0.33	-0.22	0.30	2.81
0.20000	4.12	0.16	-0.09	0.18	2.45
0.04000	4.09	0.03	0.10	0.04	1.62
0.03500	4.09	0.02	0.12	0.03	1.60
0.00350	4.13	0.00	0.22	0.00	1.60
0.00035	4.15	0.00	0.25	0.00	1.63

The Bayes factor decreases as the precision $\tau_1 = \tau_2$ decreases, and it is more stable in the case of less informative prior distributions. Also, fixing $\gamma_1 = \gamma_2 = 0.349$, we varied a_1 and a_2 in the interval $[0, 5]$ and obtained Bayes factors in the interval $[1.65, 2.97]$. These results suggest that the Bayes factor is sensitive to the choice of the hyperparameters. However, within the limits considered here, there is some weak evidence in favour of the alternative hypothesis in all cases. This evidence decreases as the prior distributions become less informative.

For the sake of comparison, we conducted a similar analysis, computing Bayes factors under unrestricted parameter spaces, i.e.,

$$\bar{\Omega}_1 = \{\boldsymbol{\mu} = (\mu_1, \mu_2) : \mu_1 = \mu_2\}, \quad (3.13)$$

$$\bar{\Omega}_2 = \{\boldsymbol{\mu} = (\mu_1, \mu_2) : \mu_1 > \mu_2\}. \quad (3.14)$$

We assumed Gaussian prior distributions with parameters $(\bar{a}_1, \bar{\gamma}_1)$ and $(\bar{a}_2, \bar{\gamma}_2)$, for μ_1 and μ_2 , respectively. The predictive densities may be computed as in the previous case, from (3.7), (3.8), (3.10) and (3.11) with the obvious modifications. Here we obtained the values of the hyperparameters by setting $\bar{a}_j = E(\mu_j)$, and $\bar{\gamma}_j = \bar{\tau}_j = \bar{\tau}$, $j = 1, 2$. In Table 4 we present the corresponding Bayes factors

for different values of $\bar{\tau}$. The results also suggest less evidence against the null hypothesis as the prior distributions become less informative.

Table 4: Bayes factors for unrestricted parameter spaces

Hyperparameter $\bar{\tau}$	Bayes factor (B_{21})
0.35000	2.16
0.20000	1.97
0.04000	0.46
0.03500	0.45
0.00350	0.40
0.00035	0.05

Comparing the results in Tables (3) and (4) for equal values of the dispersion parameters, $\tau_1 = \tau_2 = \bar{\tau} = 0.35$, we conclude that the Bayes factor obtained under the restricted parameter space ($B_{21} = 2.81$) is greater than the unrestricted parameter space counterpart ($B_{21} = 2.16$), which suggests that the restriction increases the evidence against the null hypothesis. For less informative prior distributions, however, the Bayes factors associated to the unrestricted parameter space are smaller than those obtained under the restricted parameter space.

To explore the flexibility of the methodology proposed by Irony e Pereira (1995), we maintained the same normal density for the data \mathbf{y} and considered an exponential prior distribution with parameter λ_j , for $\mu_j, j = 1, 2$, i.e.,

$$\tilde{g}(\mu_j) = \lambda_j \exp(-\lambda_j \mu_j) I_{[0, \infty)}, \tag{3.15}$$

so that $E(\mu_j) = \lambda_j^{-1}$ and $Var(\mu_j) = \lambda_j^{-2}, j = 1, 2$. The joint prior density of (μ_1, μ_2) is

$$\tilde{g}(\mu_1, \mu_2) = \lambda_1 \lambda_2 \exp(-\lambda_1 \mu_1 - \lambda_2 \mu_2) I_{[0, \infty) \times [0, \infty)},$$

and the joint likelihood of data \mathbf{y} and the parameter $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is

$$\begin{aligned} \tilde{g}(\mu_1, \mu_2) l(\mathbf{y} | \mu_1, \mu_2) &= K_1 \lambda_1 \lambda_2 \exp(-\lambda_1 \mu_1 - \lambda_2 \mu_2) \\ &\times \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + n_1 (\bar{y}_1 - \mu_1)^2 \right] \right\} \\ &\times \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 + n_2 (\bar{y}_2 - \mu_2)^2 \right] \right\} \\ &\times I_{[0, \infty) \times [0, \infty)}. \end{aligned}$$

The denominator of the ratio that defines $p(\mathbf{y}|M = 1)$ is

$$\int_{\Omega_1} \tilde{g}_1(\mu) d\mu = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)},$$

where $\tilde{g}_1(\mu) = \tilde{g}(\mu, \mu)$. The corresponding numerator is

$$\begin{aligned} & \int_{\Omega_1} \tilde{g}_1(\mu) l(\mathbf{y}|\mu, \mu) d\mu \\ &= K_1 \lambda_1 \lambda_2 \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 \right] \right\} \\ & \times \int_0^\infty \exp \left\{ -(\lambda_1 + \lambda_2)\mu - \frac{\tau}{2} [n_1(\bar{y}_1 - \mu)^2 + n_2(\bar{y}_2 - \mu)^2] \right\} d\mu, \end{aligned} \quad (3.16)$$

Similarly, the denominator of the ratio that defines $p(\mathbf{y}|M = 2)$ is

$$\int_{\Omega_2} \tilde{g}(\mu_1, \mu_2) d\mu_2 d\mu_1 = 1 - \frac{\lambda_1}{(\lambda_1 + \lambda_2)}$$

and the corresponding numerator is

$$\begin{aligned} & \int_{\Omega_2} \tilde{g}(\mu_1, \mu_2) l(\mathbf{y}|\mu_1, \mu_2) d\mu_2 d\mu_1 \\ &= K_1 \lambda_1 \lambda_2 \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 \right] \right\} \\ & \times \int_0^\infty \int_0^{\mu_1} \exp \left\{ -\lambda_1 \mu_1 - \lambda_2 \mu_2 - \frac{\tau}{2} [n_1(\bar{y}_1 - \mu_1)^2 + n_2(\bar{y}_2 - \mu_2)^2] \right\} d\mu_2 d\mu_1. \end{aligned}$$

Taking the hyperparameters as the inverse of the sample means, i.e. $\lambda_1 = 0.24 \simeq 1/4.1636$, and $\lambda_2 = 3.79 \simeq 1/0.2636$, it follows that the Bayes factor (2.1) is equal to 5.22, leading to the rejection of the null hypothesis.

3.2 Analysis via MCMC methods

Now we analyse the same data via the proposal of Carlin and Chib (1995). We considered Gaussian distributions for the data and truncated prior Gaussian distributions for the parameters μ , μ_1 and μ_2 . We did not consider a pseudo-prior distribution for σ^2 , since the data are informative about such parameter, regardless of the value of M , as suggested by Carlin and Chib (1995). In all prior distributions the hyperparameters (mean and standard deviation) were set equal to the least squares estimator of the corresponding parameters in each model.

These values are presented in Table 5. Although both the likelihood and the chosen prior distributions are standard in Bayesian analyses, it is not simple to obtain expressions for the posterior distributions required to compute the Bayes factor. This difficulty is related to the restriction on the parameter space so we must rely on numerical procedures for the computations. We used the BUGS (version 0.6) software (*Bayesian Inference using Gibbs Sampling*) developed by Thomas et al. (1992) for such purposes.

Running each model separately, we obtained estimates of the posterior means and standard deviations, which we used as values for the hyperparameters of the pseudoprior distributions.

Table 5: Least squares estimates of the parameters in models (1.3) and (1.4)

Model	Parameter	Mean	Standard Deviation
1	μ	92.2	1.2
2	μ_1	94.2	1.7
	μ_2	90.3	1.7

We used the methods proposed by Geweke (1992) as convergence diagnostics. They are available in CODA (*Convergence Diagnosis and Output Analysis Software for Gibbs sampling*), a menu-driven set of S-Plus functions, originally developed by Cowles (1994), that serves as an output processor for BUGS. To compute the Bayes factors, we considered a BUGS run of 500 burn-in iterations and 10000 updating iterations. The Bayes factors for comparison of models 1 and 2 under different a priori probabilities π_j , $j = 1, 2$ are presented in Table 6.

As with the previous approach, the Bayes factors provide evidence (although weak) against the null hypothesis, (see Table 2), suggesting that the data are more compatible with model 2. To evaluate the effect of the choice of the hyperparameters of the prior distributions, we assumed $\pi_1 = \pi_2 = 0.5$ and varied the corresponding precision (inverse of the variance) parameters, τ_μ , τ_{μ_1} and τ_{μ_2} , obtaining the results displayed in Table 6.

Along the lines of the previous analyses, we observe that the Bayes factor increases as the precision increases, indicating that it is sensitive to the choice of the hyperparameters of the prior distributions. If we consider, for example, a non-informative prior distribution with precision hyperparameter equal to 0.00001, the Bayes factor is $B_{21} = 0.02$, suggesting that the null hypothesis should be accepted, i.e. that model 1 is more compatible with the data.

We also computed Bayes factors for models with unrestricted means considering $\pi_1 = \pi_2 = 0.5$, obtaining $B_{21} = 2.52$, which is greater than its counterpart

Table 6: Bayes factors computed under the Carlin and Chib proposal

A priori Probabilities		Precision Parameter			Bayes Factor
π_1	π_2	τ_μ	τ_{μ_1}	τ_{μ_2}	
0.05	0.95	0.65000	0.35000	0.35000	1.49
0.25	0.75	0.65000	0.35000	0.35000	1.64
0.50	0.50	0.65000	0.35000	0.35000	1.64
0.75	0.25	0.65000	0.35000	0.35000	1.69
0.95	0.05	0.65000	0.35000	0.35000	1.70
0.50	0.50	0.06500	0.03500	0.03500	0.69
0.50	0.50	0.00650	0.00350	0.00350	0.63
0.50	0.50	0.00065	0.00035	0.00035	0.07
0.50	0.50	0.00001	0.00001	0.00001	0.02

under the restricted parameter space ($B_{21} = 1.65$). This suggests that the inclusion of the restriction in the parameter space decreases the evidence against the null hypothesis null, an unexpected conclusion, for which we have no explanation. Finally, still working under an unrestricted parameter space, we considered a non-informative prior distribution with precision parameter equal to 0.00001, obtaining $B_{21} = 0.03$, which is in line with the value obtained under the restricted parameter space ($B_{21} = 0.02$).

4. Extension to the Unknown Variance Case

In this section we first consider the unknown variance case under the Irony and Pereira (1995) approach. In addition to the truncated normal distributions (3.2) adopted for μ_1 and μ_2 , we chose a prior gamma distribution with parameters ν_1 and ν_2 for $\tau = \sigma^{-2}$ so that the joint prior density is

$$g^*(\mu_1, \mu_2, \tau) = K_2 K_7 \exp \left[-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2 \right] \tau^{\nu_1 - 1} \exp(-\tau \nu_2),$$

with $\tau > 0$ and $K_7 = \nu_2^{\nu_1} \sqrt{\gamma_1 \gamma_2} [\Gamma(\nu_1) 2\pi]^{-1}$. Here, the relevant (restricted) parameter space sets are

$$\begin{aligned} \Omega_1^* &= \{ \boldsymbol{\theta} = (\mu_1, \mu_2, \tau) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ and } \mu_1 = \mu_2, \tau > 0 \}, \\ \Omega_2^* &= \{ \boldsymbol{\theta} = (\mu_1, \mu_2, \tau) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ and } \mu_1 > \mu_2, \tau > 0 \}. \end{aligned}$$

The expressions (3.7) and (3.10) reduce to

$$\int_{\Omega_1^*} g^*(\mu, \mu, \tau) d\theta = K_2 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \exp \left[-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)} \right] \times \Phi \left[\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2} \right], \tag{4.1}$$

and

$$\int_{\Omega_2^*} g^*(\mu_1, \mu_2, \tau) d\theta = K_2 \left[\sqrt{\frac{\gamma_1}{2\pi}} h(\gamma_1, a_1, \gamma_2, a_2) - \Phi(-a_2 \sqrt{\gamma_2}) \Phi(a_1 \sqrt{\gamma_1}) \right], \tag{4.2}$$

Similarly, one may show that (3.8) and (3.11) reduce to

$$\int_{\Omega_1^*} g^*(\mu, \mu, \tau) l(\mu, \mu, \tau) d\theta = K_7 \left(\frac{2}{\pi} \right)^{\frac{n_1+n_2}{2}} \Gamma \left(\frac{n_1 + n_2 + 2\nu_1}{2} \right) \times \int_0^\infty \exp \left[-\frac{\gamma_1}{2} (\mu - a_1)^2 - \frac{\gamma_2}{2} (\mu - a_2)^2 \right] \times \left[\frac{1}{2} \left\{ \sum_{i=1}^{n_1} (y_{1i} - \mu)^2 + \sum_{i=1}^{n_2} (y_{2i} - \mu)^2 \right\} + \nu_2 \right]^{-\left(\frac{n_1+n_2+2\nu_1}{2}\right)} d\mu, \tag{4.3}$$

and

$$\int_{\Omega_2^*} g^*(\mu_1, \mu_2, \tau) l(\mu_1, \mu_2, \tau) d\theta = K_7 \left(\frac{2}{\pi} \right)^{\frac{n_1+n_2}{2}} \Gamma \left(\frac{n_1 + n_2 + 2\nu_1}{2} \right) \times \int_0^\infty \exp \left[-\frac{\gamma_1}{2} (\mu - a_1)^2 - \frac{\gamma_2}{2} (\mu - a_2)^2 \right] \times \left\{ \frac{1}{2} \left[\sum_{i=1}^{n_1} (y_{1i} - \mu_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \mu_2)^2 \right] + \nu_2 \right\}^{-\left(\frac{n_1+n_2+2\nu_1}{2}\right)} d\mu_1 d\mu_2. \tag{4.4}$$

The required predictive densities may then be obtained from (2.5) using either (4.1) and (4.3) or (4.2) and (4.4). Specifying the prior probabilities π_1 and π_2 we may compute the Bayes factor from (2.1). Actual computations require numerical integration methods such as the generalizations of Romberg integration, for example, routine MIDEXP (see, Press et al., 1994).

For numerical illustration we first considered an informative gamma distribution for τ by taking $\nu_1 = \nu_2 = 0.03$ (the estimator of τ based on residual sum of squares). We also considered a non-informative gamma distribution for τ by choosing $\nu_1 = 1$ and $\nu_2 = 33.3 = 1/0.03$. The results are displayed in Table 7.

Table 7: Bayes factors computed under the Irony and Pereira proposal – unknown variance case

$\tau_1 = \tau_2$	a_1	γ_1	a_2	γ_2	Bayes Factor	
					$\nu_1 = \nu_2 = 0.03$	$\nu_1 = 1, \nu_2 = 33.3$
0.35000	4.15	0.33	-0.22	0.30	0.49 (0.35)	0.49 (0.35)
0.20000	4.12	0.16	-0.09	0.18	0.78 (0.67)	0.77 (0.67)
0.04000	4.09	0.03	0.10	0.04	1.21 (1.30)	1.20 (1.26)
0.03500	4.09	0.02	0.12	0.03	1.22 (1.31)	1.21 (1.31)
0.00350	4.13	0.00	0.22	0.00	1.61 (1.63)	1.60 (1.62)
0.00035	4.15	0.00	0.25	0.00	1.71 (1.72)	1.69 (1.71)

Here too, the Bayes factor increases unexpectedly, when the precision $\tau_1 = \tau_2$ decreases as opposed to the known variance case (see Table 3). The behaviour of the Bayes factor is similar for informative or non-informative prior distributions for the precision parameter. Also, the Bayes factors computed under unknown variance situations are smaller than those obtained under the known variance case (see Tables 3 and 7).

We also repeated the computations under an unrestricted parameter space with the same choice for the hyperparameter values. The corresponding results are indicated within parentheses in Table 7. It is clear that the Bayes factors increase when $\tau_1 = \tau_2$ decreases, suggesting that the more informative the prior distribution, the smaller is the evidence against the null hypothesis, again in opposition to the results obtained under known variance (see Table 4). The corresponding Bayes factors are smaller than those obtained under the restricted parameter space for informative prior distributions ($\tau_1 = \tau_2 = 0.35$ or $\tau_1 = \tau_2 = 0.20$), but they are comparable for the non-informative prior distributions.

We now perform a similar analysis under the methodology proposed by Carlin and Chib (1995). The prior distributions for μ , μ_1 and μ_2 are the same considered in the known variance case. We also considered the same informative and non-informative prior distributions for τ used in the analysis under the Irony and Pereira (1995) methodology. We performed a BUGS run of 500 (1000) burn-in iterations and 10000 updating iterations for the informative (non-informative) prior distribution. The Bayes factors associated to the different prior distributions and different choices of π_j , $j = 1, 2$ are presented in Table 8. In all cases the Bayes factors provide weak evidence against the null hypothesis.

Finally, we computed the corresponding Bayes factors under an unrestricted parameter space, obtaining $B_{21} = 2.72$ and $B_{21} = 2.65$, for the informative and

Table 8: Bayes factors computed under the methodology of Carlin and Chib - unknown variance case

Prior Probabilities		Prior Distribution	
π_1	π_2	$\nu_1 = \nu_2 = 0.03$	$\nu_1 = 1, \nu_2 = 33.3$
0.05	0.95	1.70	1.26
0.25	0.75	1.75	1.76
0.50	0.50	1.72	1.69
0.75	0.25	1.67	1.66
0.95	0.05	1.79	1.74

non-informative prior distributions respectively; these values are greater than those computed under the restricted parameter space, namely $B_{21} = 1.72$ and $B_{21} = 1.69$. As in the known variance case the incorporation of restrictions on the parameters decreases the evidence against the null hypothesis.

5. Discussion

In this work we study two different Bayesian approaches for comparing to bounded means of Gaussian distributions, a problem of interest in situations where the phenomenon under investigation induces restrictions, as in the example described in the Introduction. The incorporation of prior knowledge about the restriction generates some technical difficulties under a frequentist approach as seen in Giampaoli and Singer (2004).

Although a direct comparison of the methodologies proposed by Irony and Pereira (1995) and Carlin and Chib (1995) is complicated in view of the differences in formulation, the Bayes factors computed under the latter are smaller, which might be an indication that the former is more efficient to detect small differences and do not present contradictory results.

The two approaches differ in the computation of the posterior probabilities $p(M = j|\mathbf{y})$, $j = 1, 2$. Under the first approach, they are directly computed from the predictive density (2.5) while under the second approach they are obtained by a Gibbs sampling of the full conditional distribution. The prior opinion on the full parameter space is expressed in a certain way via the construction of pseudoprior distributions when working under the Carlin and Chib (1995) approach. It may not be implemented with exponential prior distributions, for example, because the full conditional distribution in this case is not acceptable. The methodology proposed by Irony and Pereira (1995), on the other hand, could be inappropriate

for the comparison of non-nested models. However, not only it requires a smaller number of hyperparameters but it is also less sensitive to their choice.

We finally conclude by recognizing that the use of Bayes factors for hypothesis testing is a controversial subject and other alternatives must be considered for such purposes. In particular we mention the Bayesian reference criterion (BRC) as suggested in Bernardo and Rueda (2002) or the recent proposal of Bayesian p-values by Pereira and Stern (1999, 2001). In a future paper we will treat this problem considering the intrinsic statistic and the corresponding Bayes decision rule, the Bayesian reference criterion (BRC). This approach may be necessary the study and consequently generation of probability densities.

Acknowledgements

This work received partial financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil. The authors would like to thank Professors Heleno Bolfarine, Carlos Alberto de Bragança Pereira, Francisco Cribari-Neto and the referee for their enlightening suggestions.

References

- Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society B* **53**, 111-142.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. N. and Brunk, H. H. (1972). *Statistical Inference under Order Restrictions*. John Wiley.
- Bartholomew, D. J. (1959a). A test of homogeneity for rrdred alternatives, I. *Biometrika* **46**, 36-48.
- Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives, II. *Biometrika* **46**, 328-335.
- Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society B* **23**, 239-281.
- Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review* **70**, 351-372.
- Berger, J. O. and Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* **91**, 109-122.
- Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* **57**, 473-484.
- Cowles, M. K. (1994). *Practical Issues in Gibbs Sampler Implementation with Application to Bayesian Hierarchical Modeling of Clinical Trial Data*. PhD thesis, Division of Biostatistics, University of Minnesota.

- De Santis, F. and Spezzaferri, F. (1997). Alternative Bayes factors for model selection. *The Canadian Journal of Statistics* **4**, 503-515.
- Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153-160.
- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4* (Edited by J. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith) 147-168. Oxford University Press.
- Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society B* **56**, 501-514.
- Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1-11.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Clarendon Press.
- Giampaoli, V. and Singer, J. M. (2004). Comparison of two Normal populations with restricted means. *Computational Statistics and Data Analysis*. **46**, 511-529.
- Irony, T. Z. and Pereira, C. A. B. (1995). Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas IME-USP*, 27-46.
- Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 777-795.
- Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403-418.
- Nüesch, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *The Annals of Mathematical Statistics* **37**, 113-119.
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society B* **57**, 99-138.
- Pereira, C. A. B. and Stern, J. M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy* **1**, 69-80.
- Pereira, C. A. B. and Stern, J. M. (2001). Model selection: Full Bayesian approach. *Environmetrics* **12**, 559-568.
- Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics* **40**, 549-567.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1994). *Numerical Recipes in C, The Art of Scientific Computing*. Second Edition. Cambridge University Press.
- Thomas, A., Spielglharter, D. and Gilks, W. (1992). Bugs: a program to perform Bayesian inference using Gibbs sampling. *Bayesian Statistics 4*, 837-842.

Received April 12, 2003; accepted August 6, 2003.

Viviana Giampaoli
Departamento de Estatística
Universidade Federal de Pernambuco
CEP 50740-540
Recife, Brasil
vivig@de.ufpe.br

Julio da Motta Singer
Departamento de Estatística
Universidade de São Paulo
Caixa Postal 66281, CEP 05311-970
São Paulo, Brasil