*Brazilian Journal of Probability and Statistics*
2012, Vol. 26, No. 4, 450–473
DOI: 10.1214/11-BJPS183


# Test procedures based on combination of Bayesian evidences for $H_0$

**Rosangela H. Loschi**[a]**, Cristiano C. Santos**[a] **and Reinaldo B. Arellano-Valle**[b]

[a]*Universidade Federal de Minas Gerais*
[b]*Pontificia Universidad Católica de Chile*

**Abstract.** We introduce two procedures for testing which are based on pooling the posterior evidence for the null hypothesis provided by the full Bayesian significance test and the posterior probability for the null hypothesis. Although the proposed procedures can be used in more general situations, we focus attention in tests for a precise null hypothesis. We prove that the proposed procedure based on the linear operator is a Bayes rule. We also verify that it does not lead to the Jeffreys–Lindley paradox. For a precise null hypothesis, we prove that the procedure based on the logarithmic operator is a generalization of Jeffreys test. We apply the results to some well-known probability families. The empirical results show that the proposed procedures present good performances. As a by-product we obtain tests for normality under the skew-normal one.

## 1 Introduction

Suppose that it is of interest to test the hypotheses:

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1, \tag{1.1}$$

where $\{\Theta_0, \Theta_1\}$ is a partition of $\Theta \subseteq \mathbf{R}$, the parametric space of $\theta$. Assume that the available data information is the observed value $\mathbf{x}$ of the random object $\mathbf{X}$.

In Bayesian inference, decisions about such hypotheses are usually made by taking into consideration the posterior probability of $H_i$, $i = 0, 1$, that is given by

$$P(H_i|\mathbf{x}) = \int_{\Theta_i} \pi(\theta|\mathbf{x})\, d\theta, \tag{1.2}$$

where $\pi(\theta|\mathbf{x})$ is the posterior of $\theta$. We decide for $H_i$, whenever $P(H_i|\mathbf{x})$ is the highest probability. This procedure is an intuitive and simple approach for testing. However, if $\Theta_0$ is a subset of $\Theta$ having null Lebesgue measure, $P(H_0|\mathbf{x})$ is not calculated straightforward. A solution for such a problem was provided by Jeffreys (1961) who introduces a procedure for testing precise null hypotheses. Jeffreys (1961) proves that the posteriors in these cases depend on the Bayes factor. Decisions are made as usual, say, considering $P(H_i|\mathbf{x})$ or, equivalently, taking

*Key words and phrases.* Asymmetric distributions, Bayes risk, Bayes tests, Jeffreys test, opinion pools.
Received September 2011; accepted December 2011.

into consideration the Bayes factor. Using the Bayes factor, the decision is for $H_0$, whenever it assumes high value. Following Madruga et al. (2001), the procedure presented by Jeffreys (1961) will be named the Jeffreys test.

More recently, Pereira and Stern (1999) introduced another measure of evidence for $H_0$. This measure of evidence is the region over the posterior obtained considering all points of the parametric space for which the posterior values are, at most, as large as the supremum over the subset $\Theta_0$ of $\Theta$. Therefore, the Pereira–Stern measure of evidence for the null hypothesis, $Ev(H_0, \mathbf{x})$, is the posterior probability related to the less probable points of $\Theta$. The decision is for $H_0$, whenever $Ev(H_0, \mathbf{x})$ is large. Originally, Pereira and Stern (1999) named such a test procedure the full Bayesian significance test (FBST, for short). More recently, it has been named the Pereira–Stern test by Madruga et al. (2001). One advantage of the Pereira–Stern procedure is that it makes the test for the precise hypothesis simple. A similar idea was considered by Box and Tiao (1973) when answering whether or not a particular parameter point lies inside of a posterior highest density region. Although similar, Box and Tiao (1973) did not formalize such an idea as a hypotheses tests procedure. For more details on Bayesian procedures for testing see Bernardo (2011), Pereira et al. (2008), Dahl and Newton (2007), Scott and Berger (2006), Moreno and Liseo (2003), Madruga et al. (2003), Berger and Delampady (1987), Berger and Pericchi (1996), Robert (1993) and Lavine and Schervish (1999), among many others.

The Jeffreys and Pereira–Stern procedures are both Bayesian tests for particular loss functions. A Bayesian test is understood as the procedure which is the consequence of the minimization, a posteriori, of the expected loss function (DeGroot, 1989; Madruga et al., 2001). That is, it is a coherent solution to the decision problem in (1.1).

The Jeffreys and Pereira–Stern measures of evidence are both useful posterior summaries. In general, they lead to the same decision. However, some previous works (Pereira and Stern, 2001; Madruga et al., 2003; Loschi et al., 2007, e.g.) have shown that decisions made using such measures can differ. Different decisions are expected whenever improper priors or conjugate priors with variance going to infinity are elicited to describe the prior uncertainty about the parameter. Under such priors, Jeffreys test can lead to the Jeffreys–Lindley paradox (Robert, 1993; Tsao, 2006) which is overcome if the FBST is assumed (Pereira and Stern, 1999).

This paper aims at introducing two measures of evidence for $H_0$ which consist of pooling $P(H_0|\mathbf{x})$ and the one obtained in the FBST. To aggregate these two measures of evidence, we consider the linear and the logarithmic operators. Such operators have been widely used in Group Decision Theory in order to obtain consensus probability measures. Thus, these proposed procedures are intermediate measures of evidence for $H_0$. We verify the existence of a loss function which renders decision theoretic aspects to the test procedure built assuming the linear operator. That is, we prove that it is also a Bayesian test. We also verify that this procedure does not lead to the Jeffreys–Lindley paradox. We prove that the procedure based on the logarithmic operator is a generalization of Jeffreys test. However, we could not prove that it is a Bayesian test. All four procedures are applied to test the precise null hypothesis in some probability families, including the skew-normal one. The simulation studies show that the proposed procedures tend to be better. It is noteworthy that in the skew-normal family, inference for the shape parameter considering the usual maximum likelihood approach has some problems. Particularly, the information matrix for the skew-normal distribution is singular whenever the skewness parameter is zero, preventing the use of likelihood-based methods for testing normality. Thus, as a by-product of the proposed methodologies, we obtain normality tests under the standard skew-normal distribution [see Liseo and Loperfido (2006) and Bayes and Branco (2007) for analysis under reference priors].

The problem of combining evidence for the null hypothesis is not new. Tippet (1931), Fisher (1932) and others introduced methods for pooling $p$-values. Recently, this topic was also considered by Loughin (2004) and Goutis et al. (1996) where a multivariate evidential measure is constructed from the independent univariate ones.

This paper is organized as follows. Section 2 presents Jeffreys and the full Bayesian significance tests. The connection between the FBST and the highest posterior density regions is also provided. Section 3 presents two usual mathematical methods for combining or aggregating probability distributions—the linear and the logarithmic operators. Some of their properties are also presented. Two test procedures are introduced in Section 4. They are constructed using the linear and the logarithmic operators. Some properties of the proposed tests are pointed out. In Section 5 the proposed procedures are applied to some distribution families. To evaluate their efficiency, a Monte Carlo study is performed. In order to fairly compare the test procedures, we introduce a criterion for a decision that is based on the prior Bayes risks. In Section 6 we test the returns of some Latin American emerging markets for asymmetry. Section 7 closes the paper with some conclusions.

## 2 Bayesian procedures for testing

In this section we briefly review two Bayesian procedures for testing. The posterior evidences for the null hypothesis provided by such procedures will be considered in Section 4 to build two other procedures for testing. For one-sided tests, $P(H_0|\mathbf{x})$ is obtained straightforward from (1.2). Thus, we focus our attention in precise null hypothesis tests.

Suppose that we are interested in the hypotheses test in (1.1), where $\Theta_0 = \{\theta_0\} \subset \Theta \subseteq \mathbf{R}$ and $\theta_0$ is a known value.

The Jeffreys test is the most used Bayesian procedure for testing a precise null hypothesis. For Jeffreys test, we elicit prior probabilities $P(H_i)$ for the hypotheses

$H_i$, $i = 0, 1$, and compute the posterior probability of $H_i$ through Bayes's theorem. Let $p$ be the prior probability for $H_0$. Assume that $\pi(\theta)$ is the prior distribution for $\theta$ restricted to $\Theta_1$. Thus, the prior for $\theta$ is

$$\pi^*(\theta) = p\mathbf{1}\{\theta = \theta_0\} + (1 - p)\pi(\theta)\mathbf{1}\{\theta \neq \theta_0\},$$

where $\mathbf{1}\{A\}$ is the indicator function of event $A$. Consequently, using Bayes' theorem, it follows that the posterior for $H_0$ is given by

$$P(H_0|\mathbf{x}) = \left[1 + \frac{1-p}{p}\mathrm{BF}(H_1, H_0)\right]^{-1}, \tag{2.1}$$

where $\mathrm{BF}(H_0, H_1) = f(\mathbf{x}|\theta_0)/f(\mathbf{x}|H_1) = \mathrm{BF}(H_1, H_0)^{-1}$ is the Bayes factor and $f(\mathbf{x}|H_1)$ is the prior predictive distribution restricted to $H_1$.

We accept $H_0$ whenever its posterior probability is larger than the posterior probability of $H_1$. In fact, the Jeffreys test [more generally, the test based on $P(H_0|\mathbf{x})$] is a Bayesian test whenever the following loss function is assumed:

$$\begin{cases} L(\text{Accept } H_0, \theta) = \omega_1\mathbf{1}\{\theta \in \Theta_1\}, \\ L(\text{Reject } H_0, \theta) = \omega_0\mathbf{1}\{\theta \in \Theta_0\}, \end{cases} \tag{2.2}$$

where $\omega_i > 0$, $i = 1, 2$. Notice that the loss function in (2.2) penalizes only the wrong decisions. However, it is reasonable to assume high values for both $\omega_0$ and $\omega_1$.

We decide for $H_0$ if the posterior Bayes risk of accepting the null hypothesis is the smallest. Consequently, we accept $H_0$ whenever

$$P(H_0|\mathbf{x}) > \omega_1[\omega_1 + \omega_0]^{-1}. \tag{2.3}$$

The cutoff point for the acceptance of $H_0$ depends on values $\omega_0$ and $\omega_1$ which are subjective choices. They are the "prices" to be paid if wrong decisions are made. For a detailed explanation of Jeffreys test see Jeffreys (1961), Bernardo and Smith (1994), Migon and Gamerman (1999) and many others. For multiple hypotheses tests, say, if $\Theta_0 \subseteq \mathbf{R}^q$, see Goutis et al. (1996) for a way of obtaining $P(H_0|\mathbf{x})$.

It is well known that the Jeffreys test can lead to the Jeffreys–Lindley paradox (Tsao, 2006; Lindley, 1957). The Jeffreys–Lindley paradox is an inconvenience for Bayesian theory, as it can be noticed in the following example shown in Robert (1993). Assume that $\Theta_0 = \{0\}$ in (1.1). If $x|\theta \sim N(\theta, 1)$ and the prior for $\theta$ is the normal distribution $N(0, \sigma^2)$, then the posterior of $H_0$ is given by

$$P(H_0|x) = \left[1 + \frac{1-p}{p}\frac{\phi(x(1+\sigma^2)^{-1/2})}{(1+\sigma^2)^{-1/2}\phi(x)}\right]^{-1},$$

where $\phi(\cdot)$ is the density function of a standard normal distribution. It can be noticed that $P(H_0|x)$ goes to 1 as the prior variance $\sigma^2$ goes to infinity, no matter what $p$ and $x$ are. Large values of $\sigma^2$ correspond to noninformative priors. Thus, the use of such priors is prohibited in the context of the hypothesis tests. The result is a paradox since the use of noninformative priors cannot be avoided in those practical situations (see also Hwang et al., 1992) where experts have no prior knowledge about the parameter. The FBST was introduced in the literature in order to overcome such a problem.

The Pereira–Stern test or the FBST does not introduce prior probabilities for the hypotheses $H_i$ and makes the test for the precise null hypothesis simple (Pereira and Stern, 1999, 2001). To perform one and two-sided tests using the Pereira–Stern approach, the only necessary information is the posterior distribution for $\theta$. In this case, $H_0$ is accepted if $\theta_0$ is in a high posterior probability region of $\Theta$.

Consider the highest relative surprise (HRS) set (Madruga et al., 2003), which contains all points of the parametric space for which posterior values are larger than the supremum over the subset $\Theta_0$, say,

$$T(\mathbf{x}) = \Big\{\theta \in \Theta : \pi(\theta|\mathbf{x}) > \sup_{\Theta_0}\{\pi(\theta|\mathbf{x})\}\Big\}. \tag{2.4}$$

The posterior evidence for the null hypothesis is given by $Ev(H_0, \mathbf{x}) = 1 - \Pr(\theta \in T(\mathbf{x})|\mathbf{x})$. The null hypothesis $H_0$ is accepted whenever $Ev(H_0, \mathbf{x})$ is large. [See Madruga et al. (2003) for the FBST in its invariant formulation.]

Madruga et al. (2001) defined the following loss function which renders decision theoretic aspects to the FBST, say, it is a Bayesian test if the following loss function is assumed:

$$\begin{cases} L(\text{Accept } H_0, \theta) = b + c\mathbf{1}\{\theta \in T(\mathbf{x})\}, \\ L(\text{Reject } H_0, \theta) = a[1 - \mathbf{1}\{\theta \in T(\mathbf{x})\}], \end{cases} \tag{2.5}$$

where $\mathbf{x}$ is the observed sample, and $a$, $b$ and $c$ are real positive numbers, that are subjectively chosen. The loss function in (2.5) depends on the observed data. Loss functions depending on data have been previously used in the literature by Bernardo and Smith (1994). According to Madruga et al. (2001), they are able "*to incorporate some psychological aspects from the individual's preference ordering.*"

In order to provide some ideas in how to chose $a$, $b$ and $c$ let us consider two extreme situations. First assume that $T(\mathbf{x})$ tends to the empty set. In this scenario, $\Theta_0$ is in a high density region of the parametric space, say, we have evidence favoring the null hypothesis. From (2.5) it follows that $L(\text{Accept } H_0, \theta) \to b$ and $L(\text{Reject } H_0, \theta) \to a$. Now, let us consider $T(\mathbf{x}) \to \Theta$ which means that $\Theta_0$ is in the region of the posterior with low density. In this case the evidence is against $H_0$ and we must reject it. It follows from (2.5) that $L(\text{Accept } H_0, \theta) \to b + c$ and $L(\text{Reject } H_0, \theta) \to 0$. Notice that such loss function penalizes the right decision of accepting the null hypothesis with a "price" $b$ whenever it is a true one. Thus, a reasonable proposal is to assume small values for $b$ and high values for both $a$ and $c$.

We decide for $H_0$ if the posterior Bayes risk of accepting the null hypothesis is the smallest. Consequently, we accept $H_0$ whenever

$$Ev(H_0, \mathbf{x}) > [b+c][c+a]^{-1}. \tag{2.6}$$

Decisions depend on the values $a$, $b$ and $c$. Notice that if $a < b$, then the decision will be always to reject the null hypothesis. In this case, $Ev(H_0, \mathbf{x})$ is always up to the cutoff point $[b+c][c+a]^{-1}$. If $a >> b$ and $c$ is small, there is not needed a large value of $Ev(H_0, \mathbf{x})$ to the acceptance of $H_0$. See more in Madruga et al. (2001).

There is also a useful relationship between the Pereira–Stern test and the decisions made using the highest posterior density region (HPD region). A $100(1-\gamma)\%$ HPD region for $\theta$ is the set $R(\mathbf{x}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) \geq c_\gamma\}$ where $c_\gamma$ is the largest constant such that $P(\theta \in R(\mathbf{x})|\mathbf{x}) \geq 1-\gamma$. It is usual to accept the null hypothesis if the value of $\theta$ under test—say, $\theta_0$—belongs to $R(\mathbf{x})$ Migon and Gamerman (1999). Consequently, decisions made considering the Pereira–Stern measure of evidence and the HPD region are the same if $\gamma = (b+c)(c+a)^{-1}$ or whenever

(i) $\gamma < (b+c)(c+a)^{-1} < Ev(H_0, \mathbf{x})$ or $(b+c)(c+a)^{-1} < \gamma < Ev(H_0, \mathbf{x})$, which leads to the acceptance of $H_0$;
(ii) $(b+c)(c+a)^{-1} > \gamma > Ev(H_0, \mathbf{x})$ or $\gamma > (b+c)(c+a)^{-1} > Ev(H_0, \mathbf{x})$, which leads to the rejection of $H_0$.

Otherwise, the Pereira–Stern procedure and the HPD region will lead to different decisions.

It is remarkable that both $P(H_0|\mathbf{x})$ and $Ev(H_0, \mathbf{x})$ are evidential statistics as defined in Goutis et al. (1996). They assume values in [0, 1] and their large values indicate that $H_0$ is true and their small values indicate that $H_1$ is true.

## 3 Pooling probabilities

Combination or aggregation of probabilities plays an important role in decision problems in which a group of experts express their opinions about events of interest. This subject has attracted attention in the literature for many years and many pooling procedures have been proposed in order to obtain the group consensus probability distribution. Two typical and well-known procedures for pooling probabilities are the linear and the logarithmic operators. More details about them and some other procedures for pooling probabilities can be found in French (1985), Genest and Zidek (1986) and Genest et al. (1986), and, more recently, a discussion is presented in the context of risk analysis by Clemen and Winkler (1999).

Denote by $p_i(\theta)$, $i = 1, \ldots, n$, the opinion of the $i$th expert about $\theta$ which can be a mass function in the discrete case or a density function for the continuous case. Let $\alpha_i$, $i = 1, \ldots, n$, be nonnegative weights such that $\sum_{i=1}^{n} \alpha_i = 1$. The weight $\alpha_i$ is subjectively chosen and must reveal the confidence in the expert $i$ opinion. The consensus probability distribution $p_L$ is obtained by the *linear probability pool* whenever it is given by

$$p_L(\theta) = \sum_{i=1}^{n} \alpha_i p_i(\theta). \tag{3.1}$$

The linear probability pool given in (3.1) preserves unanimity, that is, $p_L(\theta) = a$ if $p_i(\theta) = a$, for all $i$. Consequently, it satisfies the zero preservation property only if all experts unanimously declare $p_i(\theta) = 0$. On the other hand, it preserves independency only if the group is dictatorial, say, $\alpha_i = 1$ for some $i$, and expert $i$ announces that the events of interest are independent.

Consider the same notation but, now, assume that $\alpha_i > 0$, $i = 1, \dots, n$. We say that the consensus probability distribution $p_{\mathrm{NL}}$ is obtained through the *logarithmic probability pool* if it is of form

$$p_{\mathrm{NL}}(\theta) = \frac{\prod_{i=1}^{n} [p_i(\theta)]^{\alpha_i}}{\int_{\Theta} \prod_{i=1}^{n} [p_i(\theta)]^{\alpha_i}\, d\theta}. \tag{3.2}$$

The logarithmic probability pool in (3.2) also satisfies independency and zero preservation properties. However, it is not necessary unanimity for observing zero preservation property. In fact, such property follows whenever only one expert elicit $p_i(\theta) = 0$. If we assume $\sum_{i=1}^{n} \alpha_i = 1$, the logarithmic probability pool also follows the axiom of unanimity and, under this condition for the weights, the external Bayesianity property is also satisfied, which means that, receiving extra information relevant to $\theta$ after $p_i(\theta)$, $i = 1, \dots, n$, has been declared, the new consensus probability obtained by updating the original one is the same we obtain if we firstly update each expert opinion $p_i(\theta)$ and then combine them. For more details on such properties see Genest and Zidek (1986), French (1985) and many others.

In the next section, we consider these two procedures for aggregating probabilities to obtain new measures of evidence for $H_0$.

## 4 Proposed procedures for testing

Since the evidences for $H_0$, $P(H_0|\mathbf{x})$ and $Ev(H_0, \mathbf{x})$, assume values in the $[0, 1]$ interval, we consider them as the opinions of two different experts about the same event and aggregate them in order to obtain a consensus or intermediate measure of evidence for $H_0$.

Considering the linear probability pool given in (3.1), we have a new measure of evidence for the null hypothesis that is given by

$$EvL(H_0|\mathbf{x}) = \alpha Ev(H_0, \mathbf{x}) + (1 - \alpha) P(H_0|\mathbf{x}), \tag{4.1}$$

where $\alpha \in [0, 1]$ is the weight of the Pereira–Stern measure in the intermediate measure generated by pooling $Ev(H_0, \mathbf{x})$ and $P(H_0|\mathbf{x})$. We decide for the null hypothesis $H_0$ whenever $EvL(H_0|\mathbf{x})$ is large. This procedure is named throughout this paper the *linear-pool-based test*.

Considering the logarithmic probability pool in (3.2), another consensus measure of evidence for $H_0$ is obtained and assumes the following form:

$$EvNL(H_0|\mathbf{x}) = \frac{[Ev(H_0, \mathbf{x})]^{\alpha}[P(H_0|\mathbf{x})]^{1-\alpha}}{[Ev(H_0, \mathbf{x})]^{\alpha}[P(H_0|\mathbf{x})]^{1-\alpha} + [1 - Ev(H_0, \mathbf{x})]^{\alpha}[P(H_1|\mathbf{x})]^{1-\alpha}}, \tag{4.2}$$

where $\alpha \in [0, 1]$. Similarly, we decide for the null hypothesis $H_0$ whenever $EvNL(H_0|\mathbf{x})$ is large. We name this procedure the *logarithmic-pool-based test*. Although it is not necessary, in (4.2) we assume $\alpha \in [0, 1]$ because, under this condition, the logarithmic probability pool follows the unanimity property.

The idea of combining evidences for $H_0$ of different sources was previously considered to aggregate $p$ values (Fisher, 1932; Tippet, 1931; Goutis et al., 1996 and Wassmer, 2000, for instance). Goutis et al. (1996) also introduced a rule to combine the posterior of $H_0$ which generalizes the method introduced by Fisher (1932) to combine $p$ values. However, the focus of such works is to obtain $P(H_0|\mathbf{x})$ or the $p$ value for the joint null hypothesis, that is, $H_0 : (\theta_1, \dots, \theta_k) \in \boldsymbol{\Theta}_0 \subseteq \mathbf{R}^k$. Such measures of evidence for $H_0$ are obtained by combining the individual ones, that is, the $p$ values or the $P(H_0|\mathbf{x})$ for each hypothesis $H_0 : \theta_i \in \Theta_{0i}$. An interesting discussion about the use of such ideas in sequential tests can be found in Wassmer (2000). More about multiple test problems can also be found in Pigeot (2000).

If compared to such measures of evidence for $H_0$, the ones introduced in (4.1) and (4.2) are based on different foundations. The proposed measures combine evidences for the same hypothesis provided by different test procedures. Despite this, they can be extended straightforward to test a joint null hypothesis. Moreover, $EvL(H_0|\mathbf{x})$ and $EvNL(H_0|\mathbf{x})$ are evidential statistics as defined in Goutis et al. (1996). $EvL(H_0|\mathbf{x})$ also generalizes the average method by assuming different weights for "individual" measures of evidence (Goutis et al., 1996).

### 4.1 Some properties of $EvNL(H_0|\mathbf{x})$ and $EvL(H_0|\mathbf{x})$

Firstly, we should notice that, for precise null hypothesis tests, $EvNL(H_0|\mathbf{x})$ generalizes the Jeffreys's measure of evidence for $H_0$ given in (2.1). After some calculations, we have that

$$EvNL(H_0|\mathbf{x}) = \left\{1 + \left[\frac{1 - Ev(H_0, \mathbf{x})}{Ev(H_0, \mathbf{x})}\right]^{\alpha}\left[\frac{1-p}{p}\mathrm{BF}(H_1, H_0)\right]^{1-\alpha}\right\}^{-1}, \tag{4.3}$$

which turns into expression (2.1) if $\alpha = 0$.

$EvL(H_0|\mathbf{x})$ and $EvNL(H_0|\mathbf{x})$ are both nondecreasing in $Ev(H_0, \mathbf{x})$ as well as in $P(H_0|\mathbf{x})$. Such measures also tend to 1 whenever $Ev(H_0, \mathbf{x})$ and $P(H_0|\mathbf{x})$ go to 1, simultaneously. Additionally, $EvNL(H_0|\mathbf{x})$ is zero if at least one procedure, $Ev(H_0, \mathbf{x})$ or $P(H_0|\mathbf{x})$, provides strong evidence against $H_0$. Such properties correspond to some axioms established by Goutis et al. (1996) which must be followed by the evidential statistics.

Besides, for the nontrivial case where $\alpha \neq 0$, if $Ev(H_0, \mathbf{x}) \neq 0$ [even for $Ev(H_0, \mathbf{x})$ very close to 0 which is strong evidence against the null hypothesis] and $P(H_0|\mathbf{x}) \to 1$, we have that $EvNL(H_0|\mathbf{x}) \to 1$. A similar result is observed for $Ev(H_0, \mathbf{x}) \to 1$ and $P(H_0|\mathbf{x}) \neq 0$. It is well known that the Jeffreys test can lead to the Jeffreys–Lindley paradox (Lindley, 1957; Tsao, 2006; Robert, 1993). Thus, the procedure in (4.2) can also lead to the Jeffreys–Lindley paradox since it is enough having $P(H_0|\mathbf{x}) \to 1$ to observe $EvNL(H_0|\mathbf{x}) \to 1$. On the other hand, $EvL(H_0|\mathbf{x}) \to 1$ only if $P(H_0|\mathbf{x})$ and $Ev(H_0, \mathbf{x})$ tend both to 1. Since the Pereira–Stern procedure overcomes the Jeffreys–Lindley paradox, the linear-pool-based test overcomes it as well.

In the next section, we verify the existence of a loss function that confers a decision theoretic aspect to the linear-pool-based test.

### 4.2 The Bayesianity of the linear-pool-based test

Let us assume the following loss function:

$$\begin{cases} L(\text{Accept } H_0, \theta) = (1-\alpha)\gamma\mathbf{1}(\theta \in \Theta_1) + \alpha\bigl[\beta + \gamma 1\bigl(\theta \in T(x)\bigr)\bigr], \\ L(\text{Reject } H_0, \theta) = (1-\alpha)\xi\mathbf{1}(\theta \in \Theta_0) + \alpha\xi\bigl[1 - 1\bigl(\theta \in T(x)\bigr)\bigr], \end{cases} \tag{4.4}$$

where $\alpha \in [0, 1]$, $\beta \geq 0$, $\xi$ and $\gamma$ are real, positive numbers that are subjectively chosen. Notice that the loss function in (4.4) is a particular linear combination of those in (2.2) and (2.5). Let us consider two limit situations to get some guidance on how to choose the constants $\beta$, $\xi$ and $\gamma$. Assume that $H_0$ is true such that $T(\mathbf{x})$ tends to the empty set. From (4.4) it follows that $L(\text{Accept } H_0, \theta) \to \alpha\beta$ and $L(\text{Reject } H_0, \theta) \to \xi$. On the other hand, if $H_1$ is true such that $T(\mathbf{x}) \to \Theta$, it follows from (4.4) that $L(\text{Accept } H_0, \theta) \to \gamma + \alpha\beta$ and $L(\text{Reject } H_0, \theta) \to 0$. As for the FBST, such loss function penalizes the right decision of accepting the null hypothesis with a "price" $\alpha\beta$ if $H_0$ is true. In this case, $L(\text{Accept } H_0, \theta) \to 0$ if $\alpha \to 0$ or $\beta \to 0$. The constant $\alpha$ is the weight of the Pereira–Stern measure in the proposed measure. Thus, if we want to pool different evidence in favor of $H_0$, it is not reasonable to assume $\alpha$ close to zero nor close to one. Therefore, to penalize wrong decisions properly, we should assume small values for $\beta$ and high values for both $\gamma$ and $\xi$.

The combination of loss functions as (4.4) is named balanced loss function by Jozani et al. (2010). Such authors use it in the Bayesian point estimation context.

The following theorem establishes that the linear-pool-based test is a Bayesian test. It also provides the cutoff point for acceptance of $H_0$ under such procedure.

**Theorem.** *Minimization of the posterior expected loss function in* (4.4) *is the linear-pool-based test.*

**Proof.** The posterior Bayes risk of accepting $H_0$ is

$$\begin{aligned}
&E_\pi(L(\text{Accept } H_0,\theta)|x)\\
&\quad = \int_\Theta \big[(1-\alpha)\gamma \mathbf{1}(\theta\in\Theta_1) + \alpha\big[\beta+\gamma 1\big(\theta\in T(x)\big)\big]\big]\pi(\theta|x)\,d\theta\\
&\quad = (1-\alpha)\gamma\int_{\Theta_1}\pi(\theta|x)\,d\theta + \alpha\beta\int_\Theta \pi(\theta|x)\,d\theta + \alpha\gamma\int_{T(x)}\pi(\theta|x)\,d\theta\\
&\quad = (1-\alpha)\gamma P(H_1|x) + \alpha\beta + \alpha\gamma\big(1-Ev(H_0,x)\big)\\
&\quad = (1-\alpha)\gamma + \alpha(\beta+\gamma) - \gamma EvL(H_0|\mathbf{x}).
\end{aligned}$$

The posterior Bayes risk of rejection is

$$\begin{aligned}
&E_\pi(L(\text{Reject } H_0,\theta)|x)\\
&\quad = \int_\Theta \big[(1-\alpha)\xi \mathbf{1}(\theta\in\Theta_0) + \alpha\xi[1-1(T(x))]\big]\pi(\theta|x)\,d\theta\\
&\quad = (1-\alpha)\xi\int_{\Theta_0}\pi(\theta|x)\,d\theta + \alpha\xi\int_\Theta \pi(\theta|x)\,d\theta - \alpha\xi\int_{T(x)}\pi(\theta|x)\,d\theta\\
&\quad = (1-\alpha)\xi P(H_0|x) + \alpha\xi Ev(H_0,x)\\
&\quad = \xi EvL(H_0|\mathbf{x}).
\end{aligned}$$

The test is to accept the null hypothesis if, and only if,

$$\begin{aligned}
E_\pi(L(\text{Accept } H_0,\theta)|\mathbf{x}) &< E_\pi(L(\text{Reject } H_0,\theta)|\mathbf{x})\\
(1-\alpha)\gamma + \alpha(\beta+\gamma) - \gamma EvL(H_0|\mathbf{x}) &< \xi EvL(H_0|\mathbf{x}) \qquad (4.5)\\
EvL(H_0|\mathbf{x}) &> [\gamma+\alpha\beta][\gamma+\xi]^{-1},
\end{aligned}$$

which concludes the proof. □

## 5 Simulation studies

The test procedures presented in the previous sections are applied to particular families of distributions. A comparison among them is done throughout a Monte Carlo study. From the Bayesian point of view, each inference problem is seen as unique, that is, it is not treated in light of the sampling replications paradigm. Consequently, a Monte Carlo study makes sense if it is understood as a way of evaluating how the decisions of different experts, who have the same prior opinion, are affected by the different sample evidences provided by the experiments they perform.

Since the usual Bayesian procedures for testing and the linear-pool-based test are all Bayesian tests, in order to fairly compare such procedures, we assume that the prior risks of accepting (rejecting) the null hypothesis are equals for all three procedures and, thus, we define the cutoff points for acceptance given by (2.3), (2.6) and (4.5). Denote respectively by $Ev(H_0)$ and $EvL(H_0)$ the prior evidences for the null hypothesis provided by the Pereira–Stern and Linear-pool-based tests. Assume that $Ev(H_0) \in (0, 1)$ and $EvL(H_0) \in (0, 1)$. By setting the prior risks of the three procedures equal, it follows that

$$\begin{aligned} c &= [\omega_1 P(H_1) - b][1 - Ev(H_0)]^{-1}, \\ \gamma &= [\omega_1 P(H_1) - \alpha\beta][1 - EvL(H_0)]^{-1}, \\ a &= [\omega_0 P(H_0)][Ev(H_0)]^{-1}, \\ \xi &= [\omega_0 P(H_0) - \alpha\beta][EvL(H_0)]^{-1}. \end{aligned}$$

The cutoff points are then defined by specifying $b$, $\beta$, $\omega_0$ and $\omega_1$. If $Ev(H_0) = 1$, the constant $c$ is arbitrarily chosen and the expressions to obtain $b$ and $a$ are simplified.

Since we have not found a loss function which renders the test constructed using the logarithmic operator (logarithmic-pool-based test), we can assume the same cutoff point as in (4.5) whenever it is possible to assume $P(H_0) = Ev(H_0) = p$. Because of the unanimity property, in this case, logarithmic and linear operators provide equal measures of evidence for $H_0$. Consequently, if $p \neq 0$ we have that $\omega_0 = a = \xi$ and the other values are obtained as before. On the other hand, if $p = 0$, it follows that $\omega_0 = a = \xi$, $b = \beta = 0$ and $c$ and $\gamma$ are arbitrarily chosen.

In this section, we apply the procedures to perform two-sided ($H_0 : \theta = \theta_0$) and one-sided ($H_0 : \theta \leq \theta_0$) tests under the normal family with a known variance. In this case, we consider a conjugate prior for the mean, with a large variance, and assume different values for $\alpha$ and $P(H_0)$. We also consider two asymmetric families of probability distributions: the exponential and the standard skew-normal distributions. In these cases, we focus the attention on tests for a precise null hypothesis. In both cases, we assume $\alpha = 0.5$. The interest is to evaluate the effect of the choice of different cutoff points (exponential case) and of informative and few informative or flat priors (skew-normal case) in the decisions. In all cases, we consider two sample sizes ($n = 10$ and 100). We generate 1000 samples of the likelihood with parameter $\theta_{\text{True}}$. Throughout this section, we assume that $b$ and $\beta$ in expressions (2.5) and (4.4) are close to zero ($b = \beta = 10^{-3}$). We also assume that $P(H_0)$ is equal or very close to $Ev(H_0)$. Thus, for all test procedures, we will accept the null hypothesis if the posterior evidence for $H_0$ is higher than the cutoff point $k = \omega_1(\omega_1 + \omega_0)^{-1}$. To establish notation, along this paper we denote by $\phi_n(\cdot; \mu, \Sigma)$ and $\Phi_n(\cdot; \mu, \Sigma)$ ($\phi_n(\cdot)$ and $\Phi_n(\cdot)$) the p.d.f. and the c.d.f., respectively, of the $n$-variate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($N_n(\mathbf{0}, I_n)$). The index $n$ is suppressed in the univariate case.

### 5.1 Tests under normal distribution

Let us consider $X_1, \ldots, X_n | \mu \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, where $\sigma^2$ is known. Assume the conjugate prior $\mu \sim N(m, v)$, $m \in \mathbf{R}$ and $v > 0$. Thus, $\mu | \mathbf{x} \sim N(m^*, \tau^2)$, where

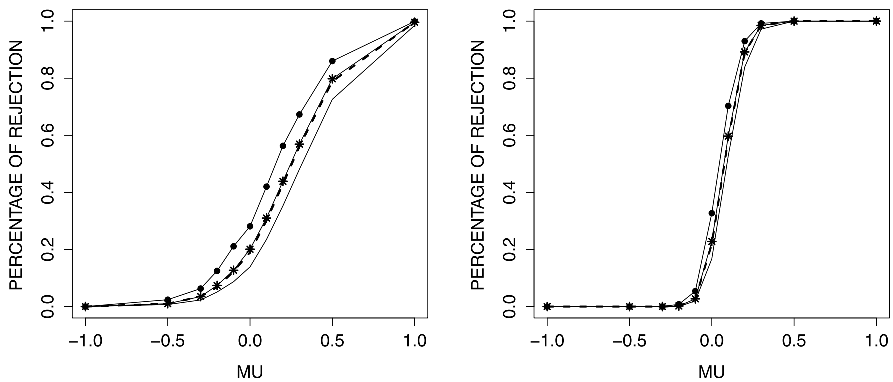


**Figure 1** *Empirical power function for Pereira and Stern (full line), usual (•), linear-pool-based (dashed line), and logarithmic-pool-based (∗) procedures, one-sided test,* $n = 10$ *(left) and* 100 *(right), normal case.*

$m^* = [v\sum_{i=1}^n x_i + m\sigma^2][nv + \sigma^2]^{-1}$ and $\tau^2 = \sigma^2 v[nv + \sigma^2]^{-1}$. The goal in this section is to evaluate the performances of the test procedures in one-sided and two-sided tests. In the later case, only tests for the precise null hypothesis are considered. We also assume different values for $\alpha$ and $P(H_0)$.

For the simulation studies (one and two-sided tests) we consider the low informative prior $\mu \sim N(0, 1000)$, and assume $\sigma^2 = 1$ and $\mu_0 = 0$. Thus, we have that $Ev(H_0) = 1.0$. We also assume the same cutoff point, $k = 0.30$, for all procedures.

5.1.1 *One-sided test.* The interest here is to test $H_0 : \mu \le \mu_0$ versus $H_1 : \mu > \mu_0$, where $\mu_0 \in \mathbf{R}$ is known. In this case, $P(H_0|\mathbf{x})$ and the Pereira–Stern measure of evidence for $H_0$ are given, respectively, by

$$P(H_0|\mathbf{x}) = \Phi(\mu_0; m^*, \tau^2),$$
$$Ev(H_0, \mathbf{x}) = 1 - \int_{T(\mathbf{x})} \phi(\mu; m^*, \tau^2)\, d\mu,$$

where $T(\mathbf{x}) = \{\mu \in \mathbf{R} : \phi(\mu; m^*, \tau^2) \ge \sup_{\Theta_0} \phi(\mu; m^*, \tau^2)\}$ and $\Theta_0 = (-\infty; \mu_0]$. If $\mu_0 \ge m^*$, then $Ev(H_0, \mathbf{x}) = 1$ since $T(\mathbf{x})$ is an empty set. If $\mu_0 < m^*$, then $T(\mathbf{x}) = \{\mu \in \mathbf{R} : |\mu - m^*| < |\mu_0 - m^*|\}$. In this case, $Ev(H_0, \mathbf{x}) = 1 - \Phi(-(\theta_0 - m^*)\tau^{-1}) + \Phi((\theta_0 - m^*)\tau^{-1})$.

Figure 1 shows the percentage of rejection of $H_0$ for different values of $\mu$ (called throughout this paper the empirical power function), all test procedures and for samples of size $n = 10$ and 100.

The proposed procedures for testing are both better than the one which considers $P(H_0|\mathbf{x})$ as the measure of evidence for $H_0$ (in this subsection named "usual test" to simplify the presentation), if the null hypothesis is true. For $\mu > \mu_0$, the usual test is the best and, in these cases, the proposed procedures present better performance than the Pereira–Stern test. For the one-sided test the empirical power

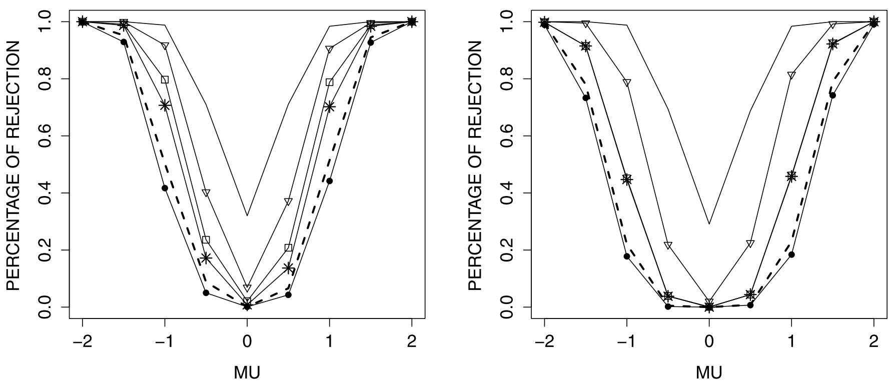


**Figure 2** *Empirical power function for Pereira and Stern (full line), Jeffreys (•), linear-pool-based with $\alpha = 1/3$ (dashed line) and $\alpha = 2/3$ (square), and logarithmic-pool-based with $\alpha = 1/3$ (∗) and $\alpha = 2/3$ (triangle) procedures, two-sided test, $n = 10$, $p = 0.50$ (left) and 0.95 (right), normal case.*

function of the linear-based and the logarithmic-based tests are very close for both sample sizes, that is, they have quite similar performance. For $n = 100$ we observe an improvement in the results for all procedures. The empirical power functions of all procedures are very close to the ideal one.

5.1.2 *Two-sided test.* Let $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$, where $\mu_0 \in \mathbf{R}$ is known. In this case, to calculate $P(H_0|\mathbf{x})$, we consider the Jeffreys strategy discussed in Section 2. The Bayes factor and the Pereira–Stern measure of evidence for $H_0$ are given, respectively, by

$$\mathrm{BF}(H_0, H_1) = \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\right\}\left(\frac{\sigma^2}{vn + \sigma^2}\right)^{-1/2} \exp\left\{\frac{n(\bar{x} - m)^2}{2(vn + \sigma^2)}\right\},$$

$$Ev(H_0, \mathbf{x}) = 1 - \int_{T(\mathbf{x})} \phi(\mu; m^*, \tau^2)\, d\mu,$$

where $T(\mathbf{x}) = \{\mu \in \mathbf{R} : |\mu - m^*| < |\mu_0 - m^*|\}$. If $\mu_0 < m^*$, it follows that $Ev(H_0, \mathbf{x}) = 1 - \Phi(-(\theta_0 - m^*)\tau^{-1}) + \Phi((\theta_0 - m^*)\tau^{-1})$. Similarly, we can obtain $Ev(H_0, \mathbf{x})$ whenever $\mu_0 > m^*$.

As it was assumed for the one-sided test, we also assume $\mu \sim N(0, 1000)$ which here can lead to the Jeffreys–Lindley paradox. The main goal is to evaluate the effect of $\alpha$, the weights of $Ev(H_0, \mathbf{x})$ in (4.1) and (4.2), and of $p = P(H_0)$ in the posterior evidences for the null hypothesis.

Figures 2 and 3 present the empirical power function of all test procedures for samples of size $n = 10$ and 100, respectively.

Opposite to what was observed for the one-sided test, the Pereira–Stern procedure is best for $\mu \neq \mu_0$ and the Jeffreys test is best for $\mu = \mu_0$. The proposed procedures are both better than the Pereira–Stern test if the null hypothesis is true.

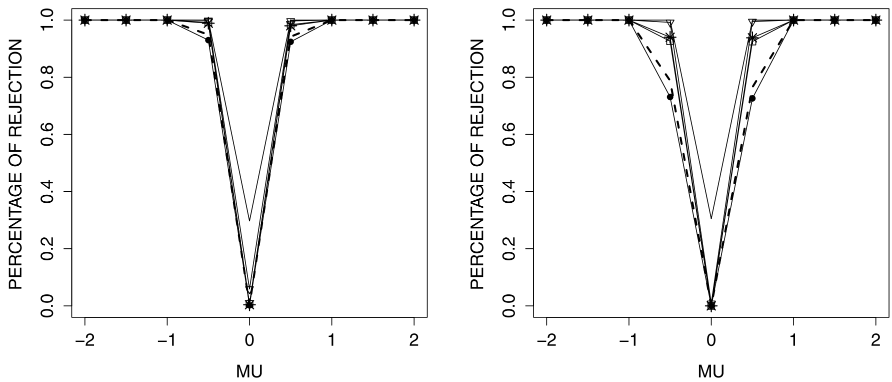


**Figure 3** *Empirical power function for Pereira and Stern* (*full line*), *Jeffreys* (•), *linear-pool-based with* $\alpha = 1/3$ (*dashed line*) *and* $\alpha = 2/3$ (*square*), *and logarithmic-pool-based with* $\alpha = 1/3$ ($*$) *and* $\alpha = 2/3$ (*triangle*) *procedures*, *two-sided test*, $n = 100$, $p = 0.50$ (*left*) *and* 0.95 (*right*), *normal case*.

They also have better performance than the Jeffreys test whenever the null hypothesis is false.

It is also noticeable from Figures 2 and 3 that the Jeffreys test presents the worst performance if the prior probability for the null hypothesis is high. The different values of $p$ affect the performance of the proposed procedures in a similar way for both values of $\alpha$. Comparing only the proposed procedures, the logarithmic-pool-based test is the best. However, the weight $\alpha$ given to $Ev(H_0, \mathbf{x})$ also influences the results. As expected, linear and logarithmic-pool-based tests present better performance whenever $\alpha = 2/3$. This behavior is expected since in the scenario considered here the Jeffreys test could wrongly lead to the acceptance of $H_0$ more frequently (Jeffreys–Lindley paradox). It is also noticeable that the empirical power functions for the linear-pool-based test with $\alpha = 1/3$ and the Jeffreys test are very close, specially whenever $p = 0.95$. Similar conclusions can be drawn for $n = 100$. In this case, the proposed procedures are much better than the Pereira–Stern test if $\mu = \mu_0$. Such procedures are comparable if $p = 0.50$ and $\mu \neq \mu_0$.

## 5.2 Tests under exponential distribution

Here we consider tests for the precise null hypothesis only, that is, $H_0 : \theta = \theta_0$. The goal is to evaluate the performances of the test procedures whenever different cutoff points, say, penalties for wrong decisions, are assumed. We consider the situation where $X_1, \ldots, X_n | \theta \overset{\text{i.i.d.}}{\sim} \exp(\theta)$, $\theta > 0$. We also assume a conjugate prior for $\theta$, say, $\theta \sim \text{Gamma}(\psi, \delta)$. Consequently, we have that $\theta | \mathbf{x} \sim \text{Gamma}(\psi + n, \delta + \sum_{i=1}^n x_i)$. Under such a model the Bayes factor and the Pereira–Stern measure of evidence for $H_0$ are given, respectively, by

$$\text{BF}(H_0, H_1) = \Gamma(\psi) \left( \delta + \sum_{i=1}^n x_i \right)^{\psi + n} [\delta^\psi \Gamma(\psi + n)]^{-1} \theta_0 e^{-\theta_0 \sum_{i=1}^n x_i},$$

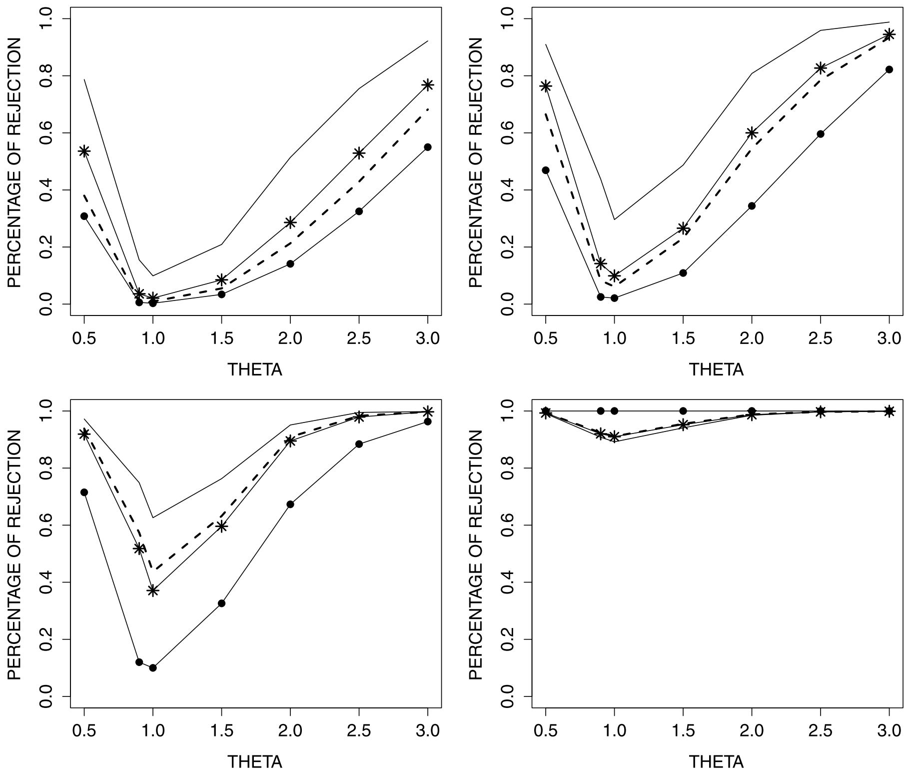


**Figure 4** *Empirical power function for Pereira and Stern* (*full line*), *Jeffreys* (•), *linear-pool-based* (*dashed line*) *and logarithmic-pool-based* (∗) *tests*, $n = 10$, *and different cutoff points*, $k = 0.10$ (*top left*), 0.33 (*top right*), 0.67 (*bottom left*) *and* 0.90 (*bottom right*), *exponential case.*

$$Ev(H_0, \mathbf{x}) = 1 - \int_{T(\mathbf{x})} e^{-\theta(\delta+n\bar{x})}\theta^{\psi+n-1}(\delta + n\bar{x})^{\psi+n}[\Gamma(\psi + n)]^{-1}\,d\theta,$$

where $T(\mathbf{x}) = \{\theta \in \mathbf{R}_+ : (\psi + n - 1)\log(\theta/\theta_0) \geq (\delta + n\bar{x})(\theta - \theta_0)\}$. Assume that $\theta_0$ is smaller than the posterior mode. Since the Gamma distribution has a unique mode, $T(\mathbf{x}) = \{\theta : \theta_0 \leq \theta \leq a\}$, where $a$ is such that $\log(\theta_0) - \theta_0(\delta + \sum_{i=1}^n x_i)[\psi + n - 1]^{-1} = \log(a) - a(\delta + \sum_{i=1}^n x_i)[\psi + n - 1]^{-1}$. Thus, denoting by $\Gamma_a(\alpha, \delta)$ the c.d.f. of the Gamma distribution with parameters $\alpha$ and $\delta$ evaluated in $a$, it follows that $Ev(H_0, \mathbf{x}) = 1 - \Gamma_a(\psi + n, \delta + \sum_{i=1}^n x_i) + \Gamma_{\theta_0}(\psi + n, \delta + \sum_{i=1}^n x_i)$. We obtain $Ev(H_0, \mathbf{x})$ for $\theta_0$ greater than the posterior mode similarly.

Figures 4 and 5 provide the empirical power function of all tests for samples of size $n = 10$ and 100, respectively. Different cutoff points $k$ are considered. We assume that $\theta_0 = 1$ and that, a priori, $\theta \sim \text{Gamma}(0.001, 0.001)$.

We have that $Ev(H_0) = P(H_0) = 0.0063$. If, compared to Jeffreys test, the proposed test procedures are better whenever $\theta_{\text{True}} \neq \theta_0$, for $\theta_{\text{True}} = \theta_0$ they tend to have better performances than the FBST. The test based on the logarithmic operator is better than the test constructed assuming the linear operator whenever

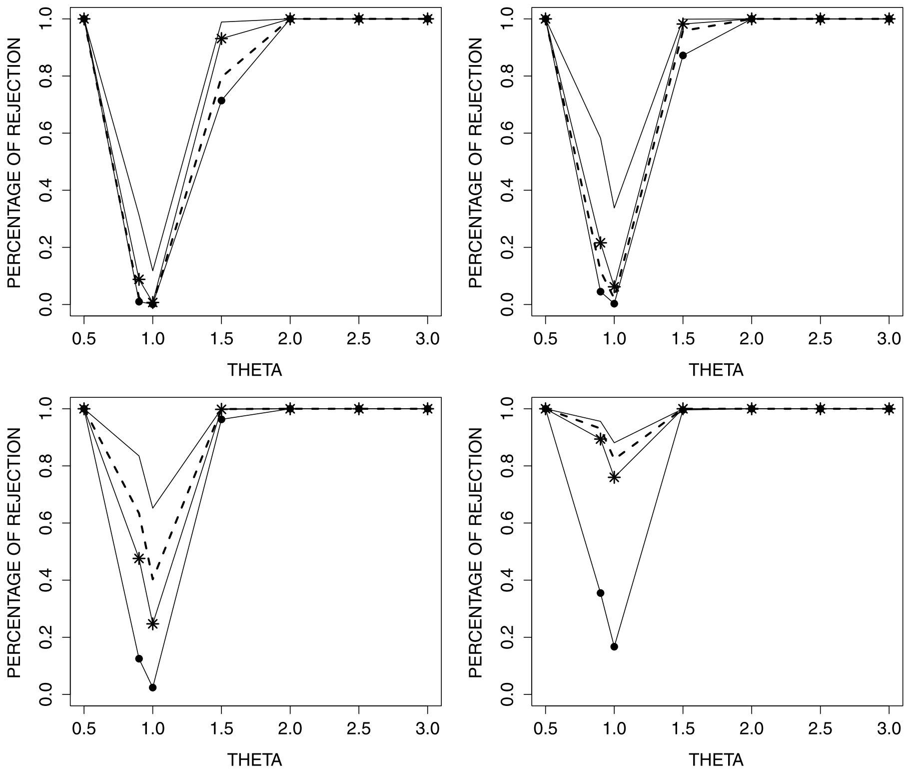


**Figure 5** *Empirical power function for Pereira and Stern (full line), Jeffreys (•), linear-pool-based (dashed line) and logarithmic-pool-based* $(*)$ *tests*, $n = 100$, *and different cutoff points*, $k = 0.10$ (*top left*), 0.33 (*top right*), 0.67 (*bottom left*) *and* 0.90 (*bottom right*), *exponential case.*

$\theta_{\text{True}} \neq \theta_0$ and for $k$ up to 0.33. For such cutoff points the test based on the linear operator is better if the null hypothesis is true. For $n = 100$ we observe an improvement in the results for all procedures mainly for small values of $k$. It is noteworthy that the FBST presents better performance for $\theta_{\text{True}}$ close to $\theta_0$.

## 5.3 Tests under skew-normal distribution

The skew-normal distribution family considered here was introduced by Azzalini (1985). Such a family includes the normal as a special case and also preserves some nice properties of the normal family. However, inference for the shape or skewness parameter $\lambda$ using the usual maximum likelihood approach has some problems, such as the existence of local maximum and also the maximum likelihood estimator for $\lambda$ can be infinite (Sartori, 2006). Moreover, if $\lambda = 0$, the Fisher information matrix is singular which prevents the use of maximum likelihood-based procedures for testing normality under the skew-normal family. Such a problem can be overcome by the use of Bayesian tests. Bayesian inference for the skewness parameter $\lambda$ has been considered, for instance, by Liseo and Loperfido (2006), Bayes and Branco (2007) and Arellano-Valle et al. (2009). The Jeffreys test for $\lambda$ was firstly considered by Bayes and Branco (2007) that assume two centered student-$t$ prior distributions for $\lambda$, with small degrees of freedom—one of which shows to be a good approximation for the reference prior introduced by Liseo and Loperfido (2006).

In this section, the goals are to consider test procedures for the skewness parameter $\lambda$ and to evaluate the effect of informative and low informative priors for $\lambda$ in the decisions. As a by-product we introduce Bayesian tests for normality under the skew-normal family. We extend previous works by assuming different test procedures for that and by assuming normal priors for the skewness parameter.

Suppose that, given $\lambda \in \mathbf{R}$, the random variables $X_1, \dots, X_n$ are i.i.d. with standard skew-normal distribution which has density $f(\mathbf{x}|\lambda) = 2^n \phi_n(\mathbf{x})\Phi_n(\lambda\mathbf{x})$. Assume that $\lambda \sim N(m, v)$. As shown in Arellano-Valle et al. (2009), under such assumptions, conjugacy is observed. Thus, the posterior is also a skewed distribution which has p.d.f. $\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v)\Phi_n(\lambda\mathbf{x})\ [\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)]^{-1}$. We have that the Bayes factor and the Pereira–Stern measure of evidence for $H_0 : \lambda = \lambda_0$ are given, respectively, by

$$\mathrm{BF}(H_0, H_1) = \Phi_n(\lambda_0\mathbf{x})[\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)]^{-1},$$

$$Ev(H_0, \mathbf{x}) = 1 - \int_{T(\mathbf{x})} \phi(\lambda; m, v)\Phi_n(\lambda\mathbf{x})[\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)]^{-1}\, d\lambda,$$

where $T(\mathbf{x}) = \{\lambda \in \mathbf{R} : \phi(\lambda; m, v)\Phi_n(\lambda\mathbf{x}) \geq \phi(\lambda_0; m, v)\Phi_n(\lambda_0\mathbf{x})\}$. Extensions for skew-normal families with unknown location and scale parameters can be done using some results presented in Arellano-Valle et al. (2009) and in Liseo and Loperfido (2006).

In the Monte Carlo study in the following, we assume $\lambda_0 = 0$, that is, we are testing for normality, and assume prior distributions centered in $m = 0$. As a consequence, the tangential set is given by $T(\mathbf{x}) = \{\lambda \in \mathbf{R} : \lambda^2 \leq 2v \sum_{i=1}^n \log[2\Phi(\lambda x_i)]\}$.

We assume two priors for $\lambda$, $\lambda \sim N(0, 1)$ and $\lambda \sim N(0, 50)$. Thus, the prior evidence for $H_0$ is $Ev(H_0) = 1.0$. We also assume $P(H_0) = 0.95$.

Figures 6 and 7 show the empirical power function for all test procedures for samples of sizes 10 and 100, respectively.

All test procedures have better performance when we assume a less informative prior for $\lambda$ whatever $k$ is. The chosen cutoff point also influences in performance of the test procedures. They tend to have better performance for $k = 0.67$. The Jeffreys test works poorly for more informative priors and $k = 0.33$. Similarly to what was observed for the exponential case, the Jeffreys test is the best if the null hypothesis is true and the Pereira–Stern test works better whenever the null hypothesis is false. Also, the proposed procedures are better than the Pereira–Stern (Jeffreys) test if $H_0$ is true (false).

Comparing the proposed procedures, from Figures 6 and 7, we have that the linear-pool-based test works better than the logarithmic-pool-based test whenever

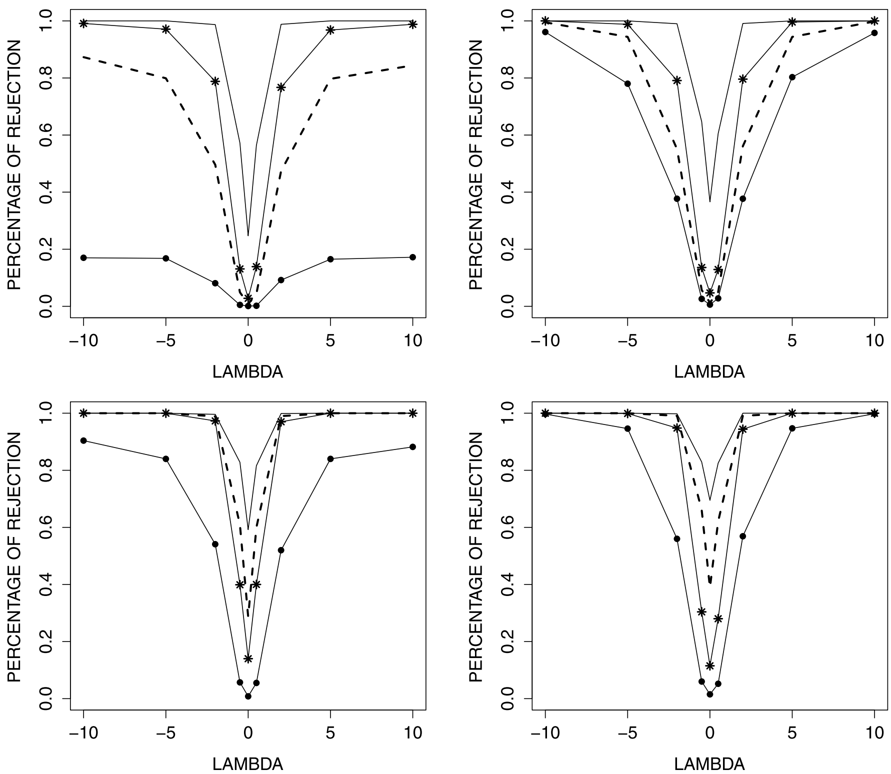


**Figure 6** *Empirical power function for Pereira and Stern (full line), Jeffreys (•), linear-pool-based (dashed line) and logarithmic-pool-based (∗) tests, $n = 10$, Skew-normal case. Cut points $k = 0.33$ (top) and 0.67 (bottom) and variance $v = 1$ (left) and 50 (right).*

we assume $k = 0.67$. The opposite is observed for $k = 0.33$. It is worthy that the proposed procedures are much better than the Jeffreys test when we elicit a more informative prior for $\lambda$ and $H_0$ is false, mainly, for $k = 0.33$. As expected, the empirical power function is closer to the ideal one if a large sample size is considered for all test procedures.

## 6 Application: Test for normality of Latin American emerging markets returns

In this section we analyze the return series of the main stock indexes of four Latin American markets, say, the MERVAL (*Índice de Mercado de Valores de Buenos Aires*) of Argentina, the IBOVESPA (*Índice da Bolsa de Valores do Estado de São Paulo*) of Brazil, the IPSA (*Índice de Precios Selectivos de Acciones*) of Chile and the IPyC (*Índice de Precios y Cotizaciones*) of Mexico. The stock returns are recorded weekly from October 31, 1995 to October 31, 2000.

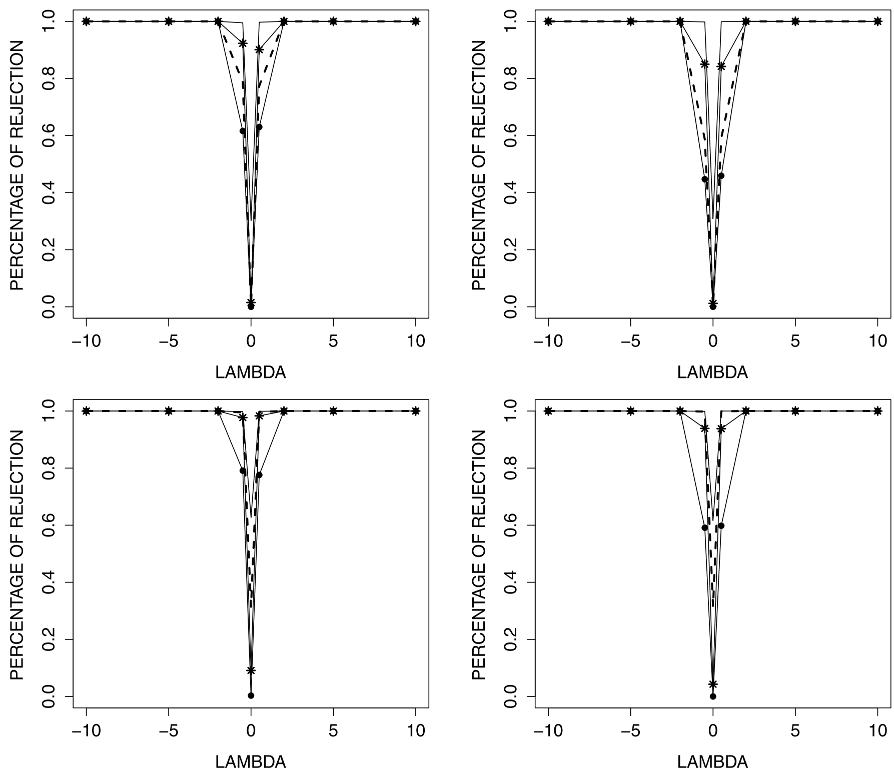


**Figure 7** *Empirical power function for Pereira and Stern (full line), Jeffreys (•), linear-pool-based (dashed line) and logarithmic-pool-based (∗) tests,* $n = 100$, *Skew-normal case. Cut points* $k = 0.33$ *(top) and* $0.67$ *(bottom) and variance* $v = 1$ *(left) and* $50$ *(right).*

It is well known that emerging markets are more susceptible to the political scenario than developed markets. Thus, their indexes tend to present more atypical observations which lead the empirical distributions of such indexes to exhibit skewness and tails that are lighter or heavier than a normal distribution. We assume the skew normal distribution to model the data behavior. In order to be better fitted by the standard skew-normal distribution, the return series $r_1, \ldots, r_n$ was transformed using the expression $y_i = r_i(\overline{r^2})^{-0.5}$, $i = 1, \ldots, n$, where $\overline{r^2} = \sum_{i=1}^{n} r_i^2/n$.

We assume that $\lambda \sim N(0, 50)$, consequently, the $Ev(H_0) = 1.0$. We also consider two prior specifications for $H_0$, a noninformative prior which establishes that $P(H_0) = 0.5$ and the other one that assumes that $P(H_0)$ is close to the prior Pereira–Stern measure of evidence for the null hypothesis, that is, we assume $P(H_0) = 0.99$. Under the last prior, we can assume the same cutoff point $k$ for accepting $H_0$ for all procedures, since in this case the linear and the logarithmic operators have similar behavior.

Figure 8 presents the posteriors for the skewness parameters for the four indexes. The area in grey represents the posterior Pereira–Stern measure of evidence

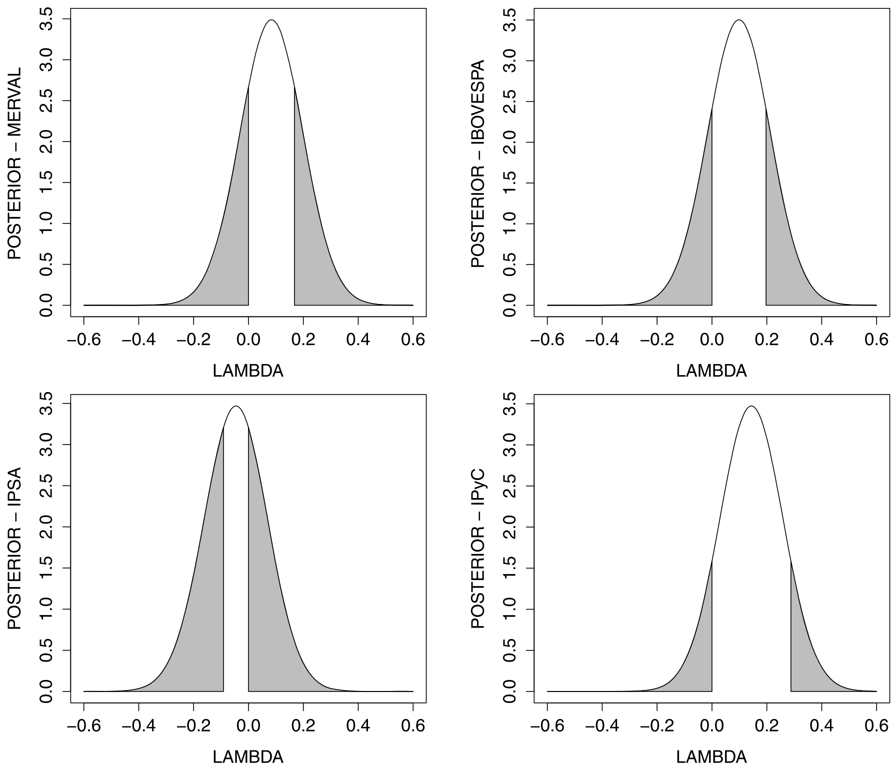


**Figure 8** *Posteriors for the skewness parameters and posterior evidences* $Ev(H_0, x)$ (*area in grey*).

**Table 1** *Posterior summaries for the skewness parameter*

| | Mean | Variance | Mode |
|---|---|---|---|
| MERVAL | 0.0838 | 0.0131 | 0.0838 |
| IBOVESPA | 0.0982 | 0.0130 | 0.0987 |
| IPSA | −0.0468 | 0.0133 | −0.0455 |
| IPyC | 0.1442 | 0.0132 | 0.1436 |

for the null hypothesis. The posteriors of $\lambda$ have unique modes and put most of their mass in small values of $\lambda$, which means that we have evidence of small asymmetry for all stock returns. The estimates for $\lambda$ are very close to zero (see Table 1). Thus, the assumption of normality can be reasonable for all stock market returns. IPyC presents the highest asymmetry. The asymmetries for all indexes are positive, except for IPSA.

Table 2 presents the posterior evidences for $H_0$ for all procedures. For $P(H_0) = 0.99$ and assuming that $\omega_0 = \omega_1$, which means that $k = 0.50$, the Pereira–Stern test leads to the conclusion that the returns of IPSA are symmetric and that, for

**Table 2** *Tests for the skewness parameter*

| | $Ev(Ho, \mathbf{x})$ | $P(H_0\|\mathbf{x})$ | $EvL(H_0\|\mathbf{x})$ | $EvNL(H_0\|\mathbf{x})$ |
|---|---|---|---|---|
| | | $P(H_0) = 0.99$ | | |
| MERVAL | 0.4638 | 0.9893 | 0.7265 | 0.8994 |
| IBOVESPA | 0.3873 | 0.9998 | 0.6935 | 0.9809 |
| IPSA | 0.6919 | 0.9885 | 0.8401 | 0.9330 |
| IPyC | 0.2105 | 0.9996 | 0.6051 | 0.9650 |
| | | $P(H_0) = 0.50$ | | |
| MERVAL | | 0.4802 | 0.4720 | 0.4720 |
| IBOVESPA | | 0.9772 | 0.6822 | 0.8387 |
| IPSA | | 0.4785 | 0.5851 | 0.5894 |
| IPyC | | 0.9631 | 0.5868 | 0.7253 |

the other markets, they are asymmetric. All the other test procedures lead to the conclusion that the returns for all indexes are distributed according to the standard normal distribution. If we assume that $2\omega_0 = \omega_1$, all test procedures indicate that the returns of the indexes have symmetric behavior, except for IPyC whenever the Pereira–Stern test is considered.

Assuming a strong prior evidence for $H_0$, Jeffreys and the logarithmic-pool-based tests are not able to update properly the information about the null hypothesis. Notice, for instance, that the posterior of $\lambda$ indicates weak evidence for symmetry of IPyC returns. It is worthy that such procedures put strong evidences for $H_0$ for IBOVESPA and IPyC indexes. The opposite is observed if we consider Pereira–Stern and linear-pool-based tests. A similar behavior is observed for $P(H_0) = 0.50$.

## 7 Final remarks

In this paper we introduced two Bayesian procedures for hypotheses testing which are based on aggregating the posterior of the null hypothesis and the measure of evidence for the null hypothesis provided by the FBST. These procedures were constructed considering the linear and the logarithmic operators which are typical procedures to obtain a consensus probability in Group Decision Theory. We performed a Monte Carlo study in order to compare all the four procedures assuming three distribution families, including the skew-normal. As a by-product, we obtained test procedures for normality under skew-normal distribution for which the usual likelihood procedures can not be used directly. We applied the procedures to test the returns of some Latin American emerging stock markets for asymmetry.

From the simulation study we concluded that, in general, the proposed test procedures tend to be better than the Jeffreys test whenever the null hypothesis is false, and they tend to have better performance than the Pereira–Stern test (FBST) whenever the null hypothesis is true. The logarithmic-pool-based (linear-pool-based) test tends to be better than the linear-pool-based (logarithmic-pool-based) one whenever the null hypothesis is false and small (large) cutoff points are considered. In general, compared to the Pereira–Stern test (which tends to be the most powerful test), the logarithmic-pool-based test greatly improves the power if the null hypothesis is true without losing power whenever it is false.

The use of the proposed procedures is attractive for their simplicity. Overall, the proposed procedures, mainly the logarithmic-pool-based test, bring some improvement and show themselves to be reasonable approaches for testing. They also present nice properties. In the case of testing a precise null hypothesis, the logarithmic-pool-based test generalizes the Jeffreys test and the linear-pool-based test avoids the Jeffreys–Lindley paradox.

## Acknowledgments

The authors thank one referee whose comments and suggestions have contributed to the improvement of the paper. The authors also express their gratitude to Ricardo C. Takahashi and Gustavo M. A. Rocha (UFMG-Brazil) for comments and suggestions in earlier versions of this paper. R. H. Loschi and C. C. Santos acknowledge CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, Grants 473163/2010-1, 306085/2009-7, 304505/2006-4 (RHL) and 502419/2007-5 (CCS), for a partial allowance to their research. The research of R. B. Arellano-Valle was partially sponsored by FONDECYT (Chile), Grant 1085241.

## References

Arellano-Valle, R. B, Genton, M. G. and Loschi, R. H. (2009). Shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* **100**, 91–101. MR2460479

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistic* **12**, 171–178. MR0808153

Bayes, C. L. and Branco, M. B. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* **21**(2), 141–163. MR2397043

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science* **2**, 317–352. MR0920141

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* **91**, 109–122. MR1394065

Bernardo, J. M. (2011). Integrated objective Bayesian estimation and hypothesis testing (with discussion). In *Bayesian Statistics 9* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid and D. Heckerman, eds.) 1–68. Oxford University Press.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, 1st ed. Chichester: Wiley. MR1274699

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley. MR0418321

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* **19**(2), 187–203.

Dahl, D. B. and Newton, M. A. (2007). Multiple Hypothesis testing by clustering treatment effects. *Journal of the American Statistical Association* **102**, 517–525. MR2325114

DeGroot, M. H. (1998). *Probability and Statistics*, 2nd ed. Boston, MA: Addison Wesley.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*, 4th ed. Edinburg: Oliver and Boyd.

French, S. (1985). Group consensus probability distributions: A critical survey. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 183–202. Amsterdam: North-Holland. MR0862490

Goutis, C., Casella, G. and Wells, M. T. (1996). Assessing evidence in multiple hypotheses. *Journal of the American Statistical Association* **91**(435), 1268–1277. MR1424623

Genest, C., McConway, K. J. and Schervish, M. J. (1986). Characterization of externally Bayesian pooling operators. *The Annals of Statistics* **14**(2), 487–501. MR0840510

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: A critique and an annoted bibliography. *Statistical Science* **1**(1), 114–148. MR0833278

Hwang, J. T., Casella, G., Roberts, C. P., Wells, M. and Farrell, R. (1992). Estimation of accuracy in testing. *The Annals of Statistics* **20**, 490–509. MR1150356

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Claredon Press. MR0187257

Jozani, M. J., Marchand, E. and Parsian, A. (2010). Bayesian and robust Bayesian analysis under a general class of balanced loss functions. *Statistical Papers* **53**, 51–60.

Lavine, M. and Schervish, M. J. (1999). Bayes Factors: What they are and what they are not. *The American Statistician* **53**(2), 119–122. MR1707756

Lindley, D. V. (1957). A statistical paradox. *Biometrika* **44**(1/2), 187–192.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* **136**, 373–389. MR2211345

Loschi, R. H., Monteiro, J. V. D., Rocha, G. H. M. A. and Mayrink, V. D. (2007). Testing and estimating the non-disjunction fraction in meiosis I using reference priors. *Biometrical Journal* **49**(6), 824–839. MR2416445

Loughin, T. M. (2004). A systematic comparison of methods for combining $p$-values from independent tests. *Computational Statistics and Data Analysis* **47**, 467–485. MR2086483

Madruga, M. R., Esteves, L. G. and Wechsler, S. (2001). On the Bayesianity of Pereira–Stern tests. *Test* **10**(1), 291–299. MR1881141

Madruga, M. R., Pereira, C. A. B. and Stern, J. M. (2003). Bayesian evidency test for precise hypotheses. *Journal of Statistical Planning and Inference* **117**, 185–198. MR2004654

Migon, H. S. and Gamerman, D. (1999). *Statistical Inference: An Integrated Approach*. New York: Arnold. MR1712097

Moreno, E. and Liseo, B. (2003). A default Bayesian test for the number of components in a mixture. *Journal of Statistical Planning and Inference* **111**, 129–142. MR1955877

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. *Envirometrics* **12**, 559–568.

Pereira, C. A. B., Stern, J. M. and Wechsler, S. (2008). Can a significance test be genuinely Bayesian? *Bayesian Analysis* **3**(1), 79–100. MR2383252

Pigeot, I. (2000). Basic concepts of multiple tests—a survey. *Statistical Papers* **41**, 3–36. MR1746085

Robert, C. P. (1993). A note on Jeffreys–Lindley paradox. *Statistica Sinica* **3**, 601–608. MR1243404

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew $t$ distributions. *Journal Statistical Planning and Inference* **136**, 4259–4275. MR2323415

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* **136**, 2144–2162. MR2235051

Tippet, L. H. C. (1931). *The Methods of Statistics*. London: Williams and Norgate.

Tsao, C. A. (2006). A note on Lindley's paradox. *Test* **15**(1), 125–139. MR2278951

Wassmer, G. (2000). Basic concepts of group sequential and adaptive group sequential test procedures. *Statistical Papers* **41**, 253–279. MR1788241

R. H. Loschi
C. C. Santos
Departamento de Estatística
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais
Belo Horizonte
Brazil
E-mail: loschi@est.ufmg.br
cristcarvalhosan@yahoo.com.br
URL: http://www.est.ufmg.br/~loschi/

R. B. Arellano-Valle
Departamento de Estadística
Facultad de Matemáticas
Pontificia Universidad Católica
Santiago
Chile
E-mail: reivalle@mat.puc.cl
URL: http://www.mat.puc.cl/~reivalle/