



# Bayesian and frequentist evidence in one-sided hypothesis testing

Elías Moreno<sup>1</sup> · Carmen Martínez<sup>1</sup>

Received: 26 October 2020 / Accepted: 12 May 2021 / Published online: 8 June 2021

© Sociedad de Estadística e Investigación Operativa 2021

## Abstract

In one-sided testing, Bayesians and frequentists differ on whether or not there is discrepancy between the inference based on the posterior model probability and that based on the  $p$  value. We add some arguments to this debate analyzing the discrepancy for moderate and large sample sizes. For small and moderate samples sizes, the discrepancy is measured by the probability of disagreement. Examples of the discrepancy on some basic sampling models indicate the somewhat unexpected result that the probability of disagreement is larger when sampling from models in the alternative hypothesis that are not located at the boundary of the hypotheses. For large sample sizes, we prove that the Bayesian one-sided testing is, under mild conditions, consistent, a property that is not shared by the frequentist procedure. Further, the rate of convergence is  $O(e^{nA})$ , where  $A$  is a constant that depends on the model from which we are sampling. Consistency is also proved for an extension to multiple hypotheses.

**Keywords** Bayesian one-sided test · Consistency · Decision rule · Intrinsic priors ·  $p$  Values · Reference priors

**Mathematics Subject Classification** 62F15 · 62F03

## 1 Introduction

While it is widely accepted that for two-sided testing the  $p$  value overstates the evidence against the null (Edwards et al. 1963; Dickey 1977; Berger and Sellke 1987), it has been argued that for one-sided testing the Bayesian and frequentist approaches agree

---

✉ Elías Moreno  
emoreno@ugr.es

Carmen Martínez  
malvarez@ugr.es

<sup>1</sup> Department of Statistics and O.R., University of Granada, Granada, Spain

in producing a data-based evaluation of the evidence on the null hypothesis (Casella and Berger 1987).

The main argument for the agreement was that the infimum of the posterior probability of the null as the prior ranges over a reasonable class of prior distributions is equal to or smaller than the  $p$  value. This yields the Casella and Berger's assertion. Thus, the  $p$  value may be on the boundary or within the range of Bayesian evidence measures. Micheas and Dey (2003) provide a similar argument for location models. However, in Berger and Mortera (1999) it is questioned whether the infimum is the best evidential summary to provide by the Bayesian inference. These authors analyzed the one-sided testing problem using empirical Bayes factors for exponential and normal models and concluded that our most important conclusion is that using the default tests provides less extreme and arguably better answers in one-sided testing than the  $p$  value (p. 553). Morris (1987) in his discussion of the paper by Casella and Berger (1987) argued that the lower bound is a misleading measure of the evidence of the null. He also asserted that typical prior beliefs should concentrate closer to the dividing line between the hypotheses.

On the other hand, the BIC approximation to the Bayes factor avoids the use of priors, but it does not apply to one-sided testing. The reason is that the dimension of the null parameter space and the whole space is the same, and hence, BIC does not correct the likelihood ratio statistic. Some modifications of BIC that can apply were given by Dudley and Haughton (1997), Kass and Vaidyanathan (1992) and Mulder and Raftery (2019).

In this paper, we compare for moderate and large sample sizes the Bayesian evidence based on the posterior model probability, and the frequentist based on the  $p$  value. The Bayesian test uses objective reference priors whenever they are proper or the intrinsic priors if the reference priors are improper. The intrinsic priors were introduced by Berger and Pericchi (1996) and further studied by Moreno (1997) and Moreno et al. (1998), and they are priors that concentrate probability around the boundary of the hypotheses, as it is required by Morris. Although intrinsic priors do not necessarily exist for nonnested models, there are some exceptions. For one-sided testing, they are constructed using an auxiliary model (Moreno 2005); for variable selection in regression, or clustering, they are constructed using an encompassing model that converts the nonnested model selection problem in a nested one (Casella et al. 2009, 2014; Moreno et al. 2010, 2015; Wang and Maruyama 2016, among others).

The comparison cannot be based on the (Type I, Type II) error vector, the standard tool for evaluating a test. The reason is that the posterior probability of the null model is typically an increasing function of the  $p$  value (Girón et al. 2006), and thus, the Bayesian and frequentist (Type I, Type II) vectors are not comparable as no one dominates the other. Then, we propose to base the comparison on the probability that the Bayesian and frequentist decision disagree. This probability gives us a measure of the strength of the disagreement. Illustrations of this analysis on common models indicate that the stronger disagreement appears when sampling from models located in the alternative hypothesis, which are not at the boundary of the hypotheses.

The comparison for large sample sizes is based on an asymptotic analysis. As far as we know, the asymptotic of the one-sided test is not yet known. Maybe, it is due to the fact that the frequentist test is by construction inconsistent under the null, although for

a wide class of sampling models it is consistent under the alternative. Our contribution to the asymptotic of the Bayesian one-sided test is a general proof of their consistency. The proof also provides the rate of consistency.

The rest of the paper is organized as follows. To facilitate the reading of the paper, we bring in Sect. 2 the intrinsic priors for one-sided testing in the absence and presence of nuisance parameters. In Sect. 3, the probability of disagreement is defined and illustrated on some basic families of sampling models. Section 4 contains the proof of the posterior model consistency of the one-sided test, and the multiple hypotheses  $H_0 : \theta = \theta_0, H_1 : \theta \leq \theta_0, H_2 : \theta \geq \theta_0$ . Concluding remarks are given in Sect. 5.

### 2 Priors for one-sided testing

Let  $X$  be a random variable with distribution  $f(x|\theta)$ , where  $\theta$  is a real parameter. The one-sided testing problem consists of the null hypothesis  $H_1 : \theta \leq \theta_0$  and the alternative  $H_2 : \theta \geq \theta_0$ , where  $\theta_0$  is a fixed real number. Families of distributions for which this problem is of interest include location and scale families.

In the absence of prior information on  $\theta$ , the reference prior  $\pi^N(\theta)$  is typically used, and hence, the one-sided Bayesian test is the model selection problem between the nonnested models

$$M_1 : \left\{ f(x|\theta), \frac{\pi^N(\theta)}{\int_{-\infty}^{\theta_0} \pi^N(\theta)d\theta} 1_{(-\infty, \theta_0)}(\theta) \right\},$$

and

$$M_2 : \left\{ f(x|\theta), \frac{\pi^N(\theta)}{\int_{\theta_0}^{\infty} \pi^N(\theta)d\theta} 1_{(\theta_0, \infty)}(\theta) \right\},$$

where  $1_A(\theta)$  is the indicator function of the set  $A$ .

Let  $P(M_i)$  denotes the prior probability of model  $M_i, i = 1, 2$ . Then, for a given sample  $\mathbf{x}_n = (x_1, \dots, x_n)$  of  $X$  the posterior probability of  $M_1$  is given by

$$\Pr(M_1|\mathbf{x}_n) = \left( 1 + B_{21}^N(\mathbf{x}_n) \frac{P(M_2)}{P(M_1)} \right)^{-1},$$

where  $B_{21}^N(\mathbf{x}_n)$  is the Bayes factor

$$B_{21}^N(\mathbf{x}_n) = \frac{\int_{-\infty}^{\theta_0} \pi^N(\theta)d\theta}{\int_{\theta_0}^{\infty} \pi^N(\theta)d\theta} \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi^N(\theta)d\theta}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi^N(\theta)d\theta}.$$

This posterior probability measures the posterior uncertainty on model  $M_1$ . If model  $M_1$  and  $M_2$  are a priori equally likely, that is,  $P(M_1) = P(M_2)$ , it seems reasonable to choose model  $M_1$  if  $\Pr(M_1|\mathbf{x}_n) > 1/2$ , and  $M_2$  otherwise.

When the prior  $\pi^N(\theta)$  is improper, the Bayes factor  $B_{21}^N(\mathbf{x}_n)$  is not well defined. Fortunately, for nested models the intrinsic methodology solves this difficulty replacing the reference prior for the parameter of the complex model with the intrinsic prior. It is well known that the resulting Bayes factor enjoys excellent sampling properties (Casella et al. 2009; Moreno and Gir3n 2005; Moreno et al. 2015).

However, when the models are nonnested, as it is the case here, the intrinsic priors are not necessarily unique. In fact, we may consider either the pair of priors

$$IP_1 = \{\pi^N(\theta)1_{(-\infty, \theta_0)}(\theta), \pi_2^I(\theta)1_{(\theta_0, \infty)}(\theta)\}$$

or

$$IP_2 = \{\pi_1^I(\theta)1_{(-\infty, \theta_0)}(\theta), \pi^N(\theta)1_{(\theta_0, \infty)}(\theta)\},$$

where  $\pi_1^I(\theta) = \pi^N(\theta)E_{y|\theta} B_{21}^N(y)$  and  $\pi_2^I(\theta) = \pi^N(\theta)E_{y|\theta} B_{12}^N(y)$  are intrinsic priors for the random training sample  $y$ .

We note that if the expectations  $E_{y|\theta} B_{12}^N(y)$  and  $E_{y|\theta} B_{21}^N(y)$  do exist, the Bayes factors for  $IP_1$  and  $IP_2$  are well defined although they might differ. Further, these expectations do not necessarily exist. Let us illustrate these assertions on exponential models.

**Example 1** Let  $X$  be a random variable with distribution  $f(x|\theta) = \frac{1}{\theta} \exp(-x/\theta)$ ,  $\theta > 0$ ,  $x > 0$ , and the improper Jeffreys' prior  $\pi^J(\theta) = k/\theta$ , where  $k$  is an arbitrary positive constant. Suppose we are interested in testing  $H_1 : 0 < \theta \leq 1$  versus  $H_2 : 1 \leq \theta < \infty$ . The default Bayesian models to be compared are

$$M_1 : \left\{ f(x|\theta), \pi^J(\theta) = \frac{k_1}{\theta} 1_{(0,1)}(\theta) \right\},$$

and

$$M_2 : \left\{ f(x|\theta), \pi^J(\theta) = \frac{k_2}{\theta} 1_{(1,\infty)}(\theta) \right\},$$

where  $k_1$  and  $k_2$  are arbitrary positive constant. We note that the intrinsic prior  $\pi_2^I(\theta)$  does not exist as

$$E_{y|\theta} B_{12}^J(y) = \frac{k_1}{k_2} \frac{1}{\theta} \int_0^\infty \frac{\exp\{-y(1 + 1/\theta)\}}{1 - \exp\{-y\}} dy = \infty.$$

On the other hand,  $\pi_1^I(\theta)$  exists and is given by

$$\pi_1^I(\theta) = \frac{k}{1 - \theta} 1_{(0,1)}(\theta),$$

but the Bayes factor  $B_{12}^{IP_2}(\mathbf{x}_n) = \infty$  for any sample  $\mathbf{x}_n$ . Thus, the original methodology for producing intrinsic priors does not necessarily work for one-sided testing.

An alternative proposal for defining intrinsic priors for one-sided testing was given in Moreno (2005), and it was formulated in two steps. In a first one, the Bayesian model selection problem between the auxiliary models  $M_0 : f(x|\theta_0)$  and  $M_3 : \{f(x|\theta), \pi^N(\theta), \theta \in \mathbb{R}\}$  is considered. The conditional intrinsic prior  $\pi^I(\theta|\theta_0)$  for a training sample  $y$  of minimal size is given by

$$\pi^I(\theta|\theta_0) = \pi^N(\theta)\mathbb{E}_{y|\theta} \frac{f(y|\theta_0)}{\int_{\mathbb{R}} f(y|\theta)\pi^N(\theta)d\theta}, \theta \in \mathbb{R}.$$

This is a proper prior on the real line  $\mathbb{R}$ . The second step is to take the restriction of  $\pi^I(\theta|\theta_0)$  to the parameter spaces  $\Theta_1 = \{\theta : \theta \leq \theta_0\}$  and  $\Theta_2 = \{\theta : \theta \geq \theta_0\}$ , and so we have the proper priors  $\pi_1^I(\theta|\theta_0)$  and  $\pi_2^I(\theta|\theta_0)$  for one-sided testing.

For a sample  $\mathbf{x}_n$  from  $f(x|\theta)$ , the Bayes factor for comparing model  $M_1 : \{f(x|\theta), \pi_1^I(\theta|\theta_0)\}$  and  $M_2 : \{f(x|\theta), \pi_2^I(\theta|\theta_0)\}$  is the well-defined Bayes factor for the intrinsic priors

$$B_{21}^I(\mathbf{x}_n) = \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi_2^I(\theta|\theta_0)d\theta}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1^I(\theta|\theta_0)d\theta}. \tag{1}$$

The fact that the conditional intrinsic priors  $\pi_1^I(\theta|\theta_0)$  and  $\pi_2^I(\theta|\theta_0)$  are centered around  $\theta_0$  is inherited from  $\pi^I(\theta|\theta_0)$ .

### 2.1 The presence of nuisance parameters

In the presence of nuisance parameters, the above procedure to produce intrinsic priors is adapted as follows. Let  $f(x|\theta, \xi)$  be the sampling family, where we want to test  $H_1 : \theta \leq \theta_0$  versus  $H_2 : \theta \geq \theta_0$ ,  $\xi \in \mathcal{E}$  being a nuisance parameter. Let  $\pi^N(\theta, \xi)$  be the starting default improper prior for the parameters.

The auxiliary models are now  $M_0 : f(x|\theta_0, \xi_0)$ , where  $\xi_0$  is an arbitrary but fixed point, and  $M_3 : \{f(x|\theta, \xi), \pi^N(\theta, \xi)\}$ . Assuming that  $\pi^N(\theta, \xi) = \pi^N(\theta)\pi^N(\xi)$ , the conditional intrinsic prior for  $\theta, \xi$  for the model comparison between  $M_0$  and  $M_3$  is given by

$$\pi^I(\theta, \xi|\theta_0, \xi_0) = \pi^N(\theta)\pi^N(\xi)\mathbb{E}_{y|\theta, \xi} \frac{f(y|\theta_0, \xi_0)}{\int_{\mathcal{E}} \pi^N(\xi) \int_{-\infty}^{\infty} f(y|\theta, \xi)\pi^N(\theta)d\theta d\xi}, \tag{2}$$

where  $y$  is the random training sample of minimal size. This conditional intrinsic prior is by construction a proper prior. Let us denote  $\pi_1^I(\theta, \xi|\theta_0, \xi_0)$  and  $\pi_2^I(\theta, \xi|\theta_0, \xi_0)$  the restriction of  $\pi^I(\theta, \xi|\theta_0, \xi_0)$  to  $H_1$  and  $H_2$ , respectively. Then, for a sample  $\mathbf{x}_n$ , the Bayes factor for the one-sided test is obtained as

$$B_{21}^I(\mathbf{x}_n) = \frac{\int_{\mathcal{E}} \pi^N(\xi_0)d\xi_0 \int_{\mathcal{E}} d\xi \int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta, \xi)\pi_2^I(\theta, \xi|\theta_0, \xi_0)d\theta}{\int_{\mathcal{E}} \pi^N(\xi_0)d\xi_0 \int_{\mathcal{E}} d\xi \int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta, \xi)\pi_1^I(\theta, \xi|\theta_0, \xi_0)d\theta}. \tag{3}$$

We note that in this expression the arbitrary constant appearing in  $\pi^N(\xi_0)$  cancels out in the ratio.

### 2.2 Multiple test

In many applications, the auxiliary model  $M_0$  is also of interest. For instance, if  $\theta$  represents the effectiveness of a new treatment we could be interested in testing whether  $\theta$  is equal to  $\theta_0$ , greater than  $\theta_0$  or smaller than  $\theta_0$ , where  $\theta_0$  is the effectiveness of the old treatment. In this case, we have the multiple comparison of the models

$$M_0 : f(x|\theta_0), M_1 : \{f(x|\theta), \pi_1^I(\theta|\theta_0)\}, M_2 : \{f(x|\theta), \pi_2^I(\theta|\theta_0)\}.$$

For the model prior  $\{P(M_i), i = 0, 1, 2\}$  the posterior probabilities of the models are given by

$$\Pr(M_i|\mathbf{x}_n) = \frac{m_i(\mathbf{x}_n)P(M_i)}{\sum_{j=0}^2 m_j(\mathbf{x}_n)P(M_j)}, i = 0, 1, 2, \tag{4}$$

where  $m_0(\mathbf{x}_n) = f(\mathbf{x}_n|\theta_0)$ ,

$$m_1(\mathbf{x}_n) = \int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1^I(\theta|\theta_0)d\theta,$$

and

$$m_2(\mathbf{x}_n) = \int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi_2^I(\theta|\theta_0)d\theta.$$

Other extensions to multiple hypotheses with equality and inequality constraints on multidimensional parameters have been studied in Mulder (2014). He utilizes an adjusted Fractional Bayes Factor for priors centered on the boundary of the constrained parameter space.

### 3 Comparing Bayesian and frequentist one-sided test

The discrepancy between the Neyman fixed Type I error probability, the Fisher  $p$  value, and the Jeffreys posterior probability methodologies has been deeply discussed in the literature. Keys papers are Efron and Gous (2001); Berger (2003). In Berger (2003), it is asserted that For many types of testing, Fisher, Jeffreys and Neyman disagreed as to the basic numbers to be reported and could report considerably different conclusions in actual practice. Here we provide a different measure of the discrepancy between the conclusions from the Fisher  $p$  value and the Bayesian analysis. As we already mentioned, the Bayesian analysis uses the standard default priors whenever they are proper or the intrinsic priors when the default priors are improper.

The frequentist decision of rejecting the null  $M_1$ , conditional on a given sample  $\mathbf{x}_n = (x_1, \dots, x_n)$ , is made for any sample point  $\mathbf{x}_n$  in the critical region

$$W_n^F(a_n) = \{\mathbf{x}_n : p(\mathbf{x}_n) \leq a_n\}, a_n \in (0, 1),$$

where  $p(\mathbf{x}_n)$  is the  $p$  value of  $\mathbf{x}_n$  and  $a_n$  a specified value, typically  $a_n = 0.05$ . The Bayesian rule rejects  $M_1$  for any sample point  $\mathbf{x}_n$  in the critical region

$$W_n^B(b_n) = \{\mathbf{x}_n : \Pr(M_1|\mathbf{x}_n) \leq b_n\}, b_n \in (0, 1),$$

where  $b_n$  is a specified value, typically  $b_n = 0.5$ . The frequentist Type I and II errors are given by

$$\mathbf{E}^F(\theta) = (\alpha_n^F, [1 - \beta_n^F(\theta)]1_{(\theta_0, \infty)}(\theta)),$$

and the Bayesian by

$$\mathbf{E}^B(\theta) = (\alpha_n^B, [1 - \beta_n^B(\theta)]1_{(\theta_0, \infty)}(\theta)),$$

where  $\alpha_n^C = \sup_{\theta \leq \theta_0} \Pr_{\mathbf{X}_n|\theta}(W_n^C(a_n))$  and  $\beta_n^C(\theta) = \Pr_{\mathbf{X}_n|\theta}(W_n^C(b_n))$  for  $C = F, B$ . The vectors  $\mathbf{E}^F(\theta)$  and  $\mathbf{E}^B(\theta)$  are not comparable except if one of them dominates the other. The latter situation is not common as  $\Pr(M_1|\mathbf{x}_n)$  used to be an increasing function of  $p(\mathbf{x}_n)$ .

An alternative way of comparing both testing methods is that of computing the probability of the sample region on which the decision rules disagree. The disagreement region  $D_n(a_n, b_n)$  is given by

$$D_n(a_n, b_n) = [\bar{W}_n^F(a_n) \cap W_n^B(b_n)] \cup [W_n^F(a_n) \cap \bar{W}_n^B(b_n)],$$

where  $\bar{W}$  denotes the complement of  $W$ . When the true model is  $f(x|\theta)$  the probability of  $D_n(a_n, b_n)$ , conditional on  $\theta$ ,

$$d_n(\theta) = \Pr_{\mathbf{X}_n|\theta}(D_n(a_n, b_n)),$$

is our quantity of interest. The larger the probability  $d_n(\theta)$ , the larger the frequentist and Bayesian disagreement, conditional on  $\theta$ .

For small or moderate sample sizes, we illustrate how large this disagreement can be on binomial, exponential, and normal models. The first two do not contain nuisance parameter, and the last one contains one nuisance parameter.

We note that when the sample size  $n$  goes to infinity and sampling from a null model, the probability of disagreement is the frequentist size, and when sampling from an alternative model it is the limit of the frequentist Type II error. For, from the

consistency results in Section 4 it follows that

$$\lim_{n \rightarrow \infty} \Pr_{\mathbf{X}_n|\theta}(W_n^B(0.5)) = \begin{cases} 1, & \text{for } \theta > \theta_0, [P_\theta], \\ 0, & \text{for } \theta < \theta_0, [P_\theta]. \end{cases}$$

Thus, for any  $\theta \geq \theta_0$ ,

$$\lim_{n \rightarrow \infty} \Pr_{\mathbf{X}_n|\theta}(D_n(0.05, 0.5)) = \lim_{n \rightarrow \infty} \Pr_{\mathbf{X}_n|\theta}(\bar{W}_n^F(0.05)) = 1 - \lim_{n \rightarrow \infty} \beta_n^F(\theta), [P_\theta],$$

and for  $\theta \leq \theta_0$ ,

$$\lim_{n \rightarrow \infty} \Pr_{\mathbf{X}_n|\theta}(D_n(0.05, 0.5)) = \lim_{n \rightarrow \infty} \Pr_{\mathbf{X}_n|\theta}(W_n^F(0.05)), [P_\theta].$$

The symbol  $[P_\theta]$  means that the limit is in probability when sampling from  $f(x|\theta)$ .

### 3.1 Bernoulli model

Let  $X$  be a random variable with distribution  $Ber(x|\theta) = \theta^x(1 - \theta)^{1-x}$ ,  $x = 0, 1$ ,  $\theta \in (0, 1)$ . We want to test  $H_1 : \theta \leq \theta_0$  versus  $H_2 : \theta \geq \theta_0$ . Since the Jeffreys' prior for  $\theta$  is the proper Beta distribution  $Be(\theta|1/2, 1/2)$ , we compare the Bayesian models

$$M_1 : \left\{ Ber(x|\theta), \pi_1^J(\theta|\theta_0) = \frac{\theta^{-1/2}(1 - \theta)^{-1/2}}{I_{\theta_0}} 1_{(0, \theta_0)}(\theta) \right\},$$

and

$$M_2 : \left\{ Ber(x|\theta), \pi_2^J(\theta|\theta_0) = \frac{\theta^{-1/2}(1 - \theta)^{-1/2}}{\pi - I_{\theta_0}} 1_{(\theta_0, 1)}(\theta) \right\},$$

where  $I_{\theta_0} = \int_0^{\theta_0} \theta^{-1/2}(1 - \theta)^{-1/2}d\theta$  is the incomplete beta function. For a sample  $\mathbf{x}_n = (x_1, \dots, x_n)$  from  $Ber(x|\theta)$  and model prior  $P(M_1) = P(M_2) = 1/2$ , the posterior probability of  $M_1$  is given by

$$\Pr(M_1|\mathbf{x}_n) = \left( 1 + \frac{I_{\theta_0}}{\pi - I_{\theta_0}} \frac{\int_{\theta_0}^1 \theta^{t_n+1/2}(1 - \theta)^{n-t_n+1/2}d\theta}{\int_0^{\theta_0} \theta^{t_n+1/2}(1 - \theta)^{n-t_n+1/2}d\theta} \right)^{-1},$$

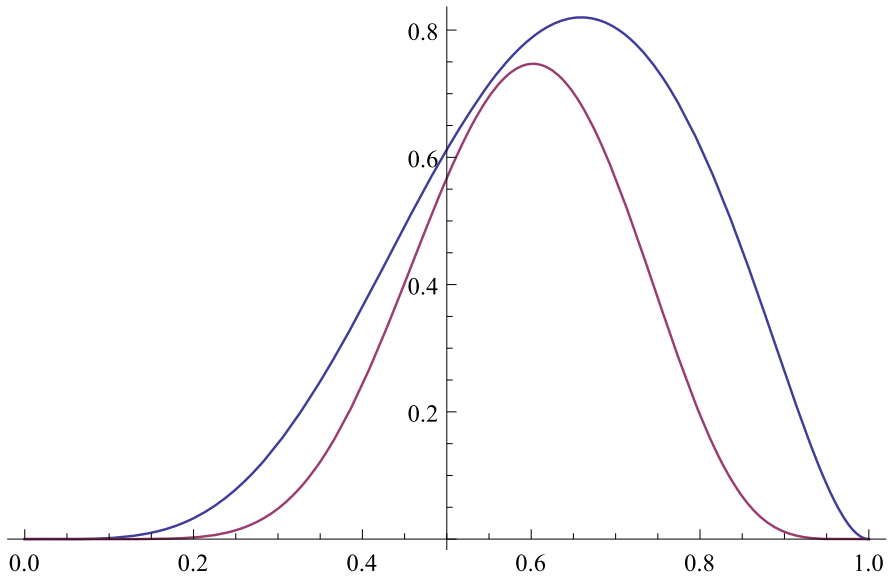
where  $t_n = \sum_{i=1}^n x_i$ . For  $\theta_0 = 1/2$ , the critical and discrepancy regions for the conventional values  $a_n = 0.05$ ,  $b_n = 0.5$ , and sample sizes  $n = 5, 10, 20, 40$  are given in Table 1.

From Table 1, we note that  $W_n^F(0.05) \subset W_n^B(0.5)$ , and hence,  $D_n$  contains critical points under the Bayesian approach that are not critical under the frequentist.



**Table 1** Critical and disagreement regions

$n$	$W_n^F(0.05)$	$W_n^B(0.5)$	$D_n(0.05, 0.5)$
5	{5}	{3, 4, 5}	{3, 4}
10	{9, 10}	{5, ..., 10}	{5, ..., 8}
20	{15, ..., 20}	{10, ..., 20}	{10, ..., 14}
40	{26, ..., 40}	{20, ..., 40}	{20, ..., 25}
80	{48, ..., 80}	{40, ..., 80}	{40, ..., 47}



**Fig. 1** Probabilities of disagreement as a function of  $\theta$  for testing the null  $H_0 : \theta \leq 0.5$  versus  $H_1 : \theta \geq 0.5$  in Bernoulli model for  $n = 10$  (higher curve) and  $n = 20$  (lower curve)

Figure 1 displays the probability  $d_n(\theta)$  for the sample sizes  $n = 10, 20$  as  $\theta$  ranges over  $(0, 1)$ . From Fig. 1, it follows that the curve of disagreement probabilities is a nonsymmetric curve around the boundary  $\theta_0 = 0.5$ . The points of large disagreement probabilities are located in the alternative hypothesis. For  $n = 10$ , the mode of the curve is as large as 0.82, and it is attained  $\theta = 0.66$ . Further, for  $\theta \in (0.45, 0.84)$  the probability of disagreement is greater than 0.5. The asymmetry arises because of in the frequentist decision rule the null hypothesis  $\theta \leq 0.5$  plays a different role than the alternative  $\theta \geq 0.5$ .

### 3.2 Exponential model

Let  $X$  be a random variable with the exponential distribution  $f(x|\theta) = 1/\theta \exp(-x/\theta)$ ,  $x > 0, \theta > 0$ , and consider the one-sided testing with null  $H_1 : 0 < \theta \leq 1$  versus the alternative  $H_2 : 1 \leq \theta < \infty$ . The Jeffreys' prior for  $\theta$  is the improper density  $\pi^J(\theta) = k/\theta$  that cannot be used for the Bayesian test. To derive the intrinsic

**Table 2** Critical and disagreement regions

$n$	$W_n^F(0.05)$	$W_n^B(0.5)$	$D_n(0.05, 0.5)$
5	(1.83, $\infty$ )	(0.940, $\infty$ )	(0.940, 1.83)
10	(1.57, $\infty$ )	(0.968, $\infty$ )	(0.968, 1.57)
20	(1.39, $\infty$ )	(0.984, $\infty$ )	(0.984, 1.39)
40	(1.27, $\infty$ )	(0.990, $\infty$ )	(0.990, 1.27)

priors, we follow the two steps presented in Sect. 2.2. First, we find the intrinsic prior  $\pi^I(\theta|\theta_0 = 1)$  arising from the model comparison  $M_0 : f(x|\theta_0)$  versus  $M_3 : \{f(x|\theta), \pi^J(\theta)\}$ , which turns out to be  $\pi^I(\theta|1) = (1 + \theta)^{-2}, \theta \geq 0$ , and the restriction of  $\pi^I(\theta|1)$  to the interval  $H_1$  and  $H_2$  yields the Bayesian models

$$M_1 : \left\{ f(x|\theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \pi_1^I(\theta|1) = \frac{2}{(1 + \theta)^2} 1_{(0,1)}(\theta) \right\},$$

and

$$M_2 : \left\{ f(x|\theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \pi_2^I(\theta|1) = \frac{2}{(1 + \theta)^2} 1_{(1,\infty)}(\theta) \right\}.$$

Hence, for a sample  $\mathbf{x}_n$  from  $f(x|\theta)$  and the model prior  $P(M_1) = P(M_2) = 1/2$ , the posterior probability of model  $M_1$  is given by

$$\Pr(M_1|n, \bar{x}_n) = \left( 1 + \frac{\int_1^\infty (1 + \theta)^{-2}\theta^{-n} \exp\{-n\bar{x}_n/\theta\}d\theta}{\int_0^1 (1 + \theta)^{-2}\theta^{-n} \exp\{-n\bar{x}_n/\theta\}d\theta} \right)^{-1},$$

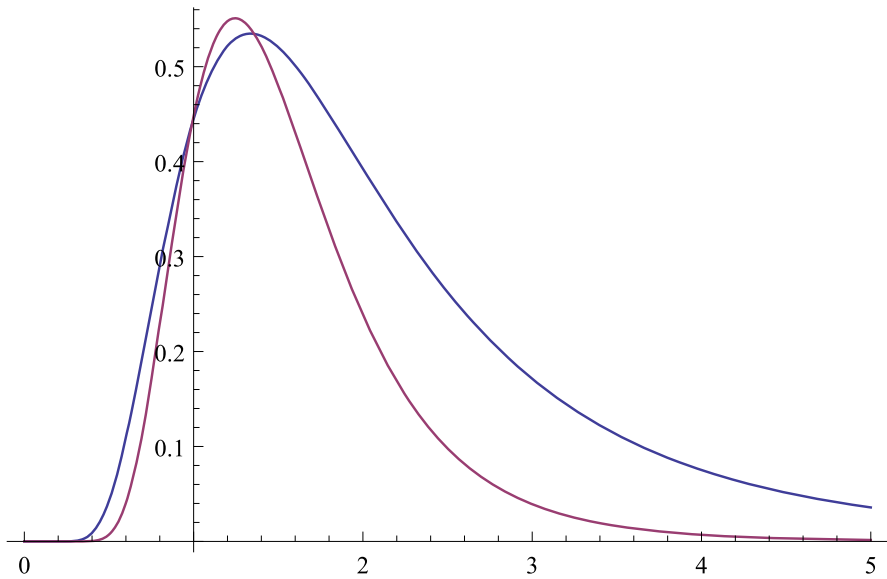
where  $\bar{x}_n = \sum_{i=1}^n x_i/n$ .

Table 2 presents the critical regions  $W_n^F(0.05)$  and  $W_n^B(0.5)$  and the disagreement region  $D_n(0.05, 0.5)$  for the sample sizes  $n = 5, 10, 20, 40$ .

Since  $W_n^F(0.05) \subset W_n^B(0.5)$  the set  $D_n$  contains points  $\bar{x}_n$  for which the null  $H_1$  is rejected under the Bayesian analysis, but it is accepted under the frequentist. Thus, the frequentist is a more conservative test than the Bayesian. Figure 2 displays the probability  $d_n(\theta)$  for the sample size  $n = 5, 10$  as  $\theta$  ranges in  $(0, 5)$ . This figure shows that the probability of disagreement is a nonsymmetric curve around the boundary  $\theta = 1$ . The probability of disagreement attains its maximum in the alternative hypothesis  $H_2$ , a similar behavior as that in the preceding Bernoulli one-sided testing.

### 3.3 Normal model

Let us consider the random variable  $X$  with normal distribution  $N(x|\mu, \sigma^2)$ , where  $\mu$  is the parameter of interest and  $\sigma$  is the nuisance parameter, and consider the one-sided testing  $H_1 : \mu \leq 0$  versus  $H_2 : \mu \geq 0$ . The Jeffreys' prior is the improper density  $\pi^J(\mu, \sigma) = k/\sigma$ . Then, we derive the intrinsic priors using the auxiliary models  $M_0 : N(x|0, \sigma_0^2)$  and  $M_3 : \{N(x|\mu, \sigma^2), \pi^J(\mu, \sigma) = k/\sigma\}$ , where  $\sigma_0$  is a unknown



**Fig. 2** Probabilities of disagreement as a function of  $\theta$  for testing the null  $H_0 : \theta \leq 1$  versus  $H_1 : \theta \geq 1$  in exponential model for  $n = 5$  (thinner curve) and  $n = 10$  (thicker)

but fixed value. The intrinsic prior of  $(\mu, \sigma)$ , conditional on  $(0, \sigma_0)$ , turns out to be the proper density

$$\pi^I(\mu, \sigma | 0, \sigma_0) = N\left(\mu | 0, \frac{\sigma_0^2 + \sigma^2}{2}\right) HC^+(\sigma | 0, \sigma_0),$$

where  $HC^+(\sigma | 0, \sigma_0)$  represents the half Cauchy distribution on the positive part of the real line. The restriction of  $\pi^I(\mu, \sigma | 0, \sigma_0)$  to the sets  $\{\mu : \mu \leq 0\}$  and  $\{\mu : \mu \geq 0\}$  yields the priors

$$\pi_i^I(\mu, \sigma | 0, \sigma_0) = 2N\left(\mu | 0, \frac{\sigma_0^2 + \sigma^2}{2}\right) HC^+(\sigma | 0, \sigma_0) 1_{H_i}(\mu), \quad i = 1, 2.$$

Integrating  $\sigma_0$  with respect to the improper prior  $\pi^N(\sigma_0) = k/\sigma_0$ , we have the unconditional intrinsic priors

$$\pi_i^I(\mu, \sigma | 0) = 2k 1_{H_i}(\mu) \int_0^\infty N\left(\mu | 0, \frac{\sigma_0^2 + \sigma^2}{2}\right) HC^+(\sigma | 0, \sigma_0) \frac{1}{\sigma_0} d\sigma_0, \quad i = 1, 2.$$

This yields the Bayesian model  $M_i : \{N(x|\mu, \sigma^2), \pi_i^I(\mu, \sigma | 0)\}$  for  $i = 1, 2$ . For a sample  $\mathbf{x}_n = (x_1, \dots, x_n)$  from the normal distribution  $N(x|\mu, \sigma^2)$  and the model

**Table 3** Critical and disagreement regions

$n$	$W_n^F(0.05)$	$W_n^B(0.5)$	$D_n(0.05, 0.5)$
5	(2.132, $\infty$ )	(0, $\infty$ )	(0, 2.132)
10	(1.833, $\infty$ )	(0, $\infty$ )	(0, 1.833)
20	(1.729, $\infty$ )	(0, $\infty$ )	(0, 1.729)
40	(1.685, $\infty$ )	(0, $\infty$ )	(0, 1.685)

prior  $P(M_1) = P(M_2) = 0.5$ , the posterior probability of  $M_1$  turns out to be

$$\Pr(M_1|n, \bar{x}_n, s_n) = \left( 1 + \frac{\int_0^\infty \int_0^{\pi/2} g(\mathbf{x}_n, \varphi, \mu) d\varphi d\mu}{\int_{-\infty}^0 \int_0^{\pi/2} g(\mathbf{x}_n, \varphi, \mu) d\varphi d\mu} \right)^{-1},$$

where

$$g(\mathbf{x}_n, \varphi, \mu) = (\sin \varphi)^{-n} \left( 2\mu^2 + \frac{ns_n^2 + n(\bar{x}_n - \mu)^2}{\sin^2 \varphi} \right)^{-(n+1)/2},$$

$\bar{x}_n = \sum_{i=1}^n x_i/n$  and  $s_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2/n$ . Thus, the posterior probability of  $M_1$  depends on the data  $(\bar{x}_n, s_n^2)$  through the statistic  $t_{n-1} = \sqrt{n} \bar{x}_n/s_n$ . Further,  $\Pr(M_1|n, t_{n-1})$  is a decreasing function of  $t_{n-1}$ . Since  $\Pr(M_1|n, t_{n-1}) = 0.5$  if and only if  $t_{n-1} = 0$ , the Bayesian critical region is given by

$$W_n^B = \{t_{n-1} : t_{n-1} \geq 0\}.$$

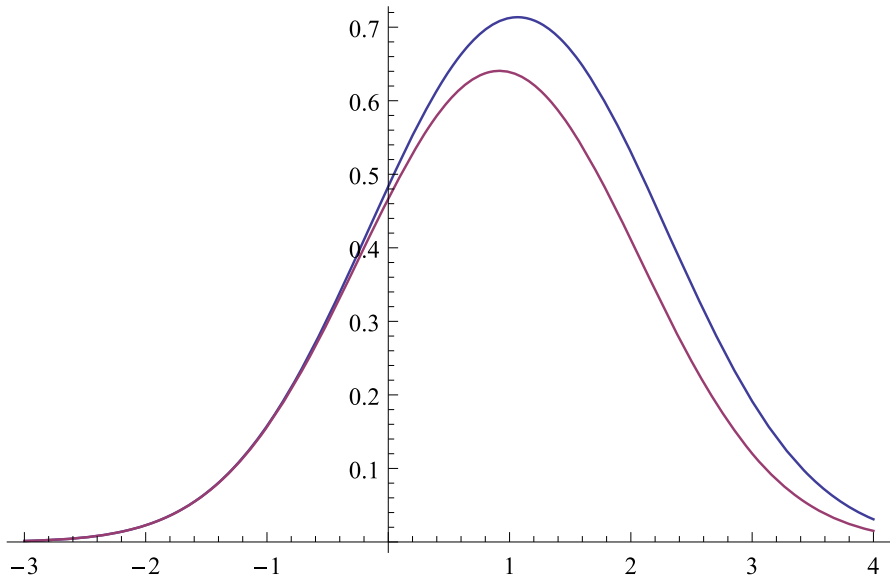
The frequentist critical region  $W_n^F$  for rejecting the null  $H_1$  is given by the quantile 0.95 of the student  $t$  distribution with  $n - 1$  degrees of freedom. Critical and disagreement regions for sample sizes  $n = 5, 10, 20, 40, 80$  are displayed in Table 3.

From Table 3, it follows that  $W_n^F(0.05) \subset W_n^B(0.5)$ , and hence,  $D_n$  contains points  $\bar{x}_n$  that yield the Bayesian rejection of the null but lead to the frequentist acceptance. Thus, the frequentist is a more conservative test than the Bayesian.

Figure 3 displays the probabilities of  $D_n(\mu, \sigma)$  for the sample sizes  $n = 5, 10$  as  $\mu$  ranges over  $(-3, 4)$  and  $\sigma = 1$ . The message of these probabilities is that the probability of discrepancy is again a nonsymmetric curve around the boundary  $\mu = 0$ . For  $n = 5$ , the maximum probability is as large as 0.712 and it is attained at  $\theta = 1.06$ , a point located in the alternative hypothesis  $H_2$ . Further, for  $\mu \in (0.08, 2.09)$  the probability of disagreement is larger than 0.5.

### 4 Asymptotic

Consistency is a key property of any statistical testing procedure as it means that the procedure provides the correct decision without uncertainty when the sample size goes to infinity. An inconsistent procedure should not be used for hypothesis testing.



**Fig. 3** Probabilities of disagreement as a function of  $\mu$  for testing the null  $H_0 : \mu \leq 0$  versus  $H_1 : \mu \geq 0$  in normal models for  $n = 5$  (higher curve) and  $n = 10$  (lower curve)

**Definition 1** Given a class of models  $\mathcal{M} = \{M_i, P(M_i), i = 1, \dots, p\}$ , where  $M_i = \{f_i(x|\theta_i), \pi_i(\theta_i), \theta_i \in \Theta_i\}$  and  $P(M_i)$  is the prior probability of model  $M_i$ , the Bayesian model selection is posterior model consistent in  $\mathcal{M}$  if the equations

$$\lim_{n \rightarrow \infty} \Pr(M_i | \mathbf{x}_n) = \begin{cases} 1, [P_{\theta_j}], \text{ for } j = i, \\ 0, [P_{\theta_j}], \text{ for } j \neq i, \end{cases} \tag{5}$$

hold for any  $i, j = 1, \dots, p$ . The symbol  $[P_{\theta_j}]$  reminds us that the limit is in probability when sampling from  $f_j(x|\theta_j)$ .

From the expression of  $\Pr(M_i | \mathbf{x}_n)$ , it follows that posterior model consistency in  $\mathcal{M}$  holds if and only if for any  $\theta_j \in \Theta_j$  we have that

$$\lim_{n \rightarrow \infty} \sum_{\substack{k=1 \\ k \neq i}}^p B_{ki}(\mathbf{x}_n) \frac{P(M_k)}{P(M_i)} = \begin{cases} 0, [P_{\theta_j}], j = i, \\ \infty, [P_{\theta_j}], j \neq i, \end{cases} \text{ for } i, j = 1, 2, \dots, p. \tag{6}$$

Theorem 1 proves that under mild conditions on the likelihood and priors, posterior model consistency holds for the Bayesian one-sided testing.

**Theorem 1** Let us consider the class of models  $\mathcal{M} = \{M_i, P(M_i), i = 1, 2\}$ , where  $M_i = \{f(x|\theta), \pi_i(\theta), \theta \in \Theta_i\}$  with  $\Theta_1 = \{\theta : \theta \leq \theta_0\}$  and  $\Theta_2 : \{\theta : \theta \geq \theta_0\}$ , and  $P(M_i) > 0, i = 1, 2$ . Then, if the likelihood  $f(\mathbf{x}_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$  and prior  $\pi_i(\theta)$

are continuous functions such that  $f(\mathbf{x}_n|\theta_0) > 0$  and  $\pi_i(\theta_0) > 0$ , the Bayes factor

$$B_{21}(\mathbf{x}_n) = \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi_2(\theta)d\theta}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta}$$

satisfies

$$\lim_{n \rightarrow \infty} B_{21}(\mathbf{x}_n) = \begin{cases} 0, & [P_\theta], \text{ if } \theta < \theta_0, \\ \infty, & [P_\theta], \text{ if } \theta > \theta_0. \end{cases}$$

Further, the speed of the limit in probability  $[P_\theta]$  is  $O(e^{-nA(\theta)})$  if  $\theta < \theta_0$  with  $A(\theta) = E_\theta \log f(x|\theta) - E_{\theta_0} \log f(x|\theta_0) > 0$ , and  $O(e^{nA(\theta)})$  if  $\theta > \theta_0$ .

**Proof** Let us write the Bayes factor  $B_{21}(\mathbf{x}_n)$  as

$$B_{21}(\mathbf{x}_n) = B_{01}(\mathbf{x}_n)B_{20}(\mathbf{x}_n),$$

where

$$B_{01}(\mathbf{x}_n) = \frac{f(\mathbf{x}_n|\theta_0)}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta} \text{ and } B_{20}(\mathbf{x}_n) = \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi_2(\theta)d\theta}{f(\mathbf{x}_n|\theta_0)}.$$

When sampling from  $f(x|\theta_1)$  for  $\theta_1 < \theta_0$ , we can write

$$-2 \log B_{01}(\mathbf{x}_n) = -2 \log \frac{f(\mathbf{x}_n|\theta_0)}{f(\mathbf{x}_n|\theta_1)} - 2 \log \frac{f(\mathbf{x}_n|\theta_1)}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta}. \tag{7}$$

Using the Laplace approximation for the integral  $\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta$ , the second term of the right side of (7) can be written as

$$\begin{aligned} & -2 \lim_{n \rightarrow \infty} \log \frac{f(\mathbf{x}_n|\theta_1)}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta} \\ &= -2 \lim_{n \rightarrow \infty} \log \left( \frac{f(\mathbf{x}_n|\theta_1)}{f(\mathbf{x}_n|\hat{\theta}_1)} \frac{1}{n^{-1/2}} \right) \\ &= \lim_{n \rightarrow \infty} \left( -2 \log \frac{f(\mathbf{x}_n|\theta_1)}{f(\mathbf{x}_n|\hat{\theta}_1)} \right) - 2 \lim_{n \rightarrow \infty} \log n^{1/2}, [P_{\theta_1}], \end{aligned}$$

where  $f(\mathbf{x}_n|\hat{\theta}_1) = \sup_{\theta < \theta_0} f(\mathbf{x}_n|\theta)$ . Further, the limit in probability  $[P_{\theta_1}]$  of  $-2 \log f(\mathbf{x}_n|\theta_1)/f(\mathbf{x}_n|\hat{\theta}_1)$  is a positive random variable that does not degenerate to a constant (in fact, it has a Chi-squared distribution; see Wilks 1963, chapter 13).

Moreover, the first term in the right-hand side of (7) can be written as

$$-2 \log \frac{f(\mathbf{x}_n|\theta_0)}{f(\mathbf{x}_n|\theta_1)} = 2n \left( \frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta_1) - \frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta_0) \right). \tag{8}$$

If  $A(\theta_1)$  is the limit in probability  $[P_{\theta_1}]$  of the expression inside the parenthesis in (8), we have by the Law of Large Numbers that

$$\begin{aligned} A(\theta_1) &= \lim_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta_1) - \frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta_0) \right) \\ &= E_{\theta_1} \log f(x|\theta_1) - E_{\theta_1} \log f(x|\theta_0) > 0, [P_{\theta_1}], \end{aligned}$$

where from the Jensen inequality  $A(\theta_1) > 0$ . Thus,

$$\lim_{n \rightarrow \infty} \left( -2 \log \frac{f(\mathbf{x}_n|\theta_0)}{f(\mathbf{x}_n|\theta_1)} \right) = \lim_{n \rightarrow \infty} 2nA(\theta_1) = \infty, [P_{\theta_1}]. \tag{9}$$

We note that this convergence is not only in probability, but it is almost surely  $[P_{\theta_1}]$ . Therefore,

$$\lim_{n \rightarrow \infty} B_{01}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} n^{1/2} e^{-nA(\theta_1)} = 0, [P_{\theta_1}]. \tag{10}$$

On the other hand, when sampling from  $f(x|\theta_1)$  the Bayes factor  $B_{20}(\mathbf{x}_n)$  is finite, in fact it is smaller than or equal to 1. For,

$$\begin{aligned} \lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n) &= \lim_{n \rightarrow \infty} \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta) \pi_2(\theta) d\theta}{f(\mathbf{x}_n|\theta_0)} \\ &\leq \lim_{n \rightarrow \infty} \frac{\sup_{\theta \geq \theta_0} f(\mathbf{x}_n|\theta)}{f(\mathbf{x}_n|\theta_0)} = 1, [P_{\theta_1}], \end{aligned} \tag{11}$$

where the last equality follows from the fact that the MLE  $\hat{\theta}_n$  of  $\theta$  in the set  $\{\theta \geq \theta_0\}$  converges in probability  $[P_{\theta_1}]$  to  $\theta_0$ . Thus, from (10) and (11) it follows that

$$\lim_{n \rightarrow \infty} B_{21}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n)B_{01}(\mathbf{x}_n) = 0, [P_{\theta_1}]. \tag{12}$$

This proves the first assertion.

Similar arguments to those used for proving (10) yield that when sampling from  $f(x|\theta_2)$  for  $\theta_2 > \theta_0$ ,

$$\lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n) = \infty, [P_{\theta_2}], \tag{13}$$

and the rate of convergence is  $O(e^{nA(\theta_2)})$ , where  $A(\theta_2) = E_{\theta_2} \log f(x|\theta_2) - E_{\theta_2} \log f(x|\theta_0) > 0$ . Further, when sampling from  $f(x|\theta_2)$ , the Bayes factor  $B_{01}(\mathbf{x}_n)$  is greater than zero. Indeed, this result follows from

$$\lim_{n \rightarrow \infty} B_{01}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\theta_0)}{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta) \pi_1(\theta) d\theta}$$

$$\geq \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\theta_0)}{\sup_{\theta \leq \theta_0} f(\mathbf{x}_n|\theta)} = 1, [P_{\theta_2}]. \tag{14}$$

Thus, from (13) and (14) we have that

$$\lim_{n \rightarrow \infty} B_{21}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n)B_{01}(\mathbf{x}_n) = \infty, [P_{\theta_2}]. \tag{15}$$

This proves the second assertion and completes the proof of the theorem. □

**Corollary 1** *Let  $f(x|\theta, \xi)$  be a sampling model, where  $\xi \in \mathcal{E}$  is a nuisance parameter, and consider the one-sided testing  $H_1 : \theta \leq \theta_0$  versus  $H_2 : \theta \geq \theta_0$ . Then, the Bayes factor  $B_{21}^I(\mathbf{x}_n)$  for the intrinsic priors  $\pi_i^I(\theta, \xi|\theta_0)$  for  $i = 1, 2$  given in (2) is consistent.*

**Proof** The Bayes factor  $B_{21}^I(\mathbf{x}_n)$  can be written as

$$B_{21}^I(\mathbf{x}_n) = \frac{\int_{\theta_0}^{\infty} h_2(\mathbf{x}_n|\theta)d\theta}{\int_{-\infty}^{\theta_0} h_1(\mathbf{x}_n|\theta)d\theta},$$

where

$$h_i(\mathbf{x}_n|\theta) = \int_{\mathcal{E}} \int_{\mathcal{E}} f(\mathbf{x}_n|\theta, \xi)\pi_i^I(\theta, \xi|\theta_0, \xi_0)\pi^N(\xi_0)d\xi d\xi_0.$$

This proves the assertion. □

Let us consider the multiple model comparison between model  $M_0 : f(x|\theta_0)$ ,  $M_1 : \{f(x|\theta_1), \pi_1(\theta_1), \theta_1 < \theta_0\}$  and  $M_2 : \{f(x|\theta_2), \pi_2(\theta_2), \theta_2 > \theta_0\}$  with model prior  $P(M_i) > 0$  for  $i = 0, 1, 2$ . Theorem 2 proves that the Bayesian model selection is posterior model consistent and the rate of convergence is exponential except when sampling from  $M_0$  in which case the rate is potential.

**Theorem 2** *Under the conditions in Theorem 1, and model prior  $P(M_i) > 0, i = 0, 1, 2$ , we have that*

(i)

$$\lim_{n \rightarrow \infty} \Pr(M_0|\mathbf{x}_n) = \begin{cases} 1, [P_{\theta}], \text{ if } \theta = \theta_0, \\ 0, [P_{\theta}], \text{ if } \theta \neq \theta_0, \end{cases}$$

and the rate of convergence for  $\theta = \theta_0$  is  $O(n^b)$  with  $b > 0$ , and for  $\theta \neq \theta_0$  is  $O(e^{-nA(\theta)})$ , where  $A(\theta) = \lim_{n \rightarrow \infty} E_{\theta} \log f(x|\theta) - E_{\theta} f(x|\theta_0) > 0$ .

(ii) For  $i \neq 0$ , we have

$$\lim_{n \rightarrow \infty} \Pr(M_i|\mathbf{x}_n) = \begin{cases} 1, [P_{\theta}], \text{ if } \theta < \theta_0, \\ 0, [P_{\theta}], \text{ if } \theta \geq \theta_0, \end{cases}, i = 1, 2$$

and the rate of convergence is again exponential.



**Proof** We prove that Eq. (6) holds for  $i, j = 0, 1, 2$ , and any  $\theta$  in the real line  $\mathbb{R}$ .

To prove i) we have to prove that when sampling from  $f(x|\theta_0)$

$$\lim_{n \rightarrow \infty} \left( B_{10}(\mathbf{x}_n) \frac{P(M_1)}{P(M_0)} + B_{20}(\mathbf{x}_n) \frac{P(M_2)}{P(M_0)} \right) = 0, [P_{\theta_0}], \tag{16}$$

and when sampling from  $\theta \neq \theta_0$ ,

$$\lim_{n \rightarrow \infty} \left( B_{10}(\mathbf{x}_n) \frac{P(M_1)}{P(M_0)} + B_{20}(\mathbf{x}_n) \frac{P(M_2)}{P(M_0)} \right) = \infty, [P_\theta], \theta \neq \theta_0. \tag{17}$$

To prove (16), we use the Laplace approximation to the integral on  $B_{10}(\mathbf{x}_n)$ , and write

$$\lim_{n \rightarrow \infty} B_{10}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} \frac{\int_{-\infty}^{\theta_0} f(\mathbf{x}_n|\theta)\pi_1(\theta)d\theta}{f(\mathbf{x}_n|\theta_0)} = \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\hat{\theta}_1)}{f(\mathbf{x}_n|\theta_0)} n^{-1/2} = 0, [P_{\theta_0}]. \tag{18}$$

The last equality follows from the fact that  $\lim_{n \rightarrow \infty} f(\mathbf{x}_n|\hat{\theta}_1)/f(\mathbf{x}_n|\theta_0)$ ,  $[P_{\theta_0}]$  is a positive random variable. The same arguments yields

$$\lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} \frac{\int_{\theta_0}^{\infty} f(\mathbf{x}_n|\theta)\pi_2(\theta)d\theta}{f(\mathbf{x}_n|\theta_0)} = \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\hat{\theta}_2)}{f(\mathbf{x}_n|\theta_0)} n^{-1/2} = 0, [P_{\theta_0}]. \tag{19}$$

Then, (18) and (19) prove (16) and the first assertion in (i) is proved. If we now sampling from  $f(x|\theta_1)$  for  $\theta_1 < \theta_0$ , we have

$$\lim_{n \rightarrow \infty} B_{10}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\hat{\theta}_1)}{f(\mathbf{x}_n|\theta_1)} \frac{f(\mathbf{x}_n|\theta_1)}{f(\mathbf{x}_n|\theta_0)} n^{-1/2} = \infty, [P_{\theta_1}]. \tag{20}$$

The last equality follows from equality (10) in Theorem 1. A similar argument proves that when sampling from  $f(x|\theta_2)$  for  $\theta_2 > \theta_0$ ,

$$\lim_{n \rightarrow \infty} B_{20}(\mathbf{x}_n) = \lim_{n \rightarrow \infty} \frac{f(\mathbf{x}_n|\hat{\theta}_2)}{f(\mathbf{x}_n|\theta_2)} \frac{f(\mathbf{x}_n|\theta_2)}{f(\mathbf{x}_n|\theta_0)} n^{-1/2} = \infty, [P_{\theta_2}]. \tag{21}$$

Expressions (20) and (21) prove the second assertion in (i). This completes the proof of part (i).

To prove (ii) for  $i = 1$ , we have to prove that

$$\lim_{n \rightarrow \infty} \left( B_{01}(\mathbf{x}_n) \frac{P(M_0)}{P(M_1)} + B_{21}(\mathbf{x}_n) \frac{P(M_2)}{P(M_1)} \right) = 0, [P_{\theta_1}],$$

and

$$\lim_{n \rightarrow \infty} \left( B_{01}(\mathbf{x}_n) \frac{P(M_0)}{P(M_1)} + B_{21}(\mathbf{x}_n) \frac{P(M_2)}{P(M_1)} \right) = \infty, [P_\theta], \theta \neq \theta_1.$$

And for  $i = 2$ , arguments are similar to that used for proving (i) show that

$$\lim_{n \rightarrow \infty} \left( B_{01}(\mathbf{x}_n) \frac{P(M_0)}{P(M_1)} + B_{12}(\mathbf{x}_n) \frac{P(M_1)}{P(M_2)} \right) = 0, [P_{\theta_2}],$$

and

$$\lim_{n \rightarrow \infty} \left( B_{01}(\mathbf{x}_n) \frac{P(M_0)}{P(M_1)} + B_{12}(\mathbf{x}_n) \frac{P(M_1)}{P(M_2)} \right) = \infty, [P_{\theta}], \theta \neq \theta_2.$$

This completes the proof of the theorem. □

We note that these theorems are valid for the intrinsic priors  $\pi_1^I(\theta|\theta_0)$  and  $\pi_2^I(\theta|\theta_0)$ .

### 5 Concluding remarks

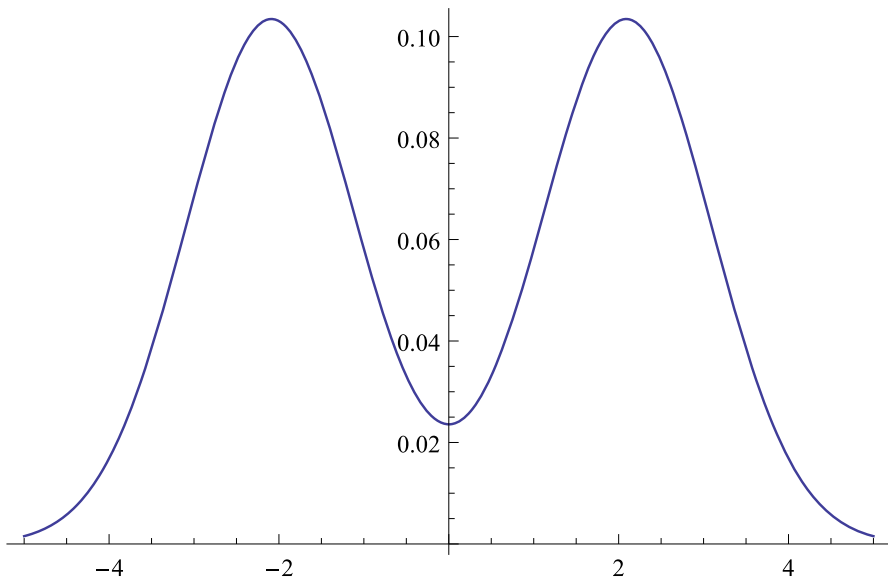
For small and moderate samples sizes, a comparison between the decisions of Bayesian and frequentist one-sided testing has been presented. The comparison is based on the probability that the procedures make opposite decisions. This comparison indicates that for usual sampling families the probability is maximum when the true model is in the alternative hypothesis and not at the boundary of the hypotheses. This asymmetric behavior is inherited from the asymmetric role the frequentist approach assigns to the null and the alternative hypothesis.

It is interesting to point out that in our one-sided testing examples the sampling region in which the decisions disagree contains points that the frequentist analysis accepts the null hypothesis and the Bayesian rejects it ( $W^F \subset W^B$ ). For other hypothesis testing, the disagreement region is not necessarily of this type, as the following simple example shows.

**Example 2** Let us consider the two-sided testing on the mean  $\theta$  of the normal distribution  $N(x|\theta, 1)$ . For the null  $H_1 : \theta = 0$  and the alternative  $H_2 : \theta \neq 0$  with the prior  $N(\theta|0, 2)$ , the frequentist rejection region contains the Bayesian rejection region ( $W^B \subset W^F$ ), which is exactly the opposite what occurs in the one-sided testing.

Figure 4 displays the probability of disagreement, conditional on  $\theta$ , as  $\theta$  ranges over  $(-5, 5)$  for  $n = 1$ . The disagreement region is given by  $D_1 = \{x : 1.96 \leq |x| \leq 2.22\}$ . The curve as a function of  $\theta$  represents the probability that the Bayesian analysis accepts the null and the frequentist rejects the null. The maximum probability of disagreement is equal to 0.103 and it is attained at  $|\theta| = 2.105$ . This probability grows as  $n$  grows. For instance, for  $n = 20$  the shape of the curve is similar to that in Fig. 4, but the mode is as large as 0.64 and it is attained at  $|\theta| = 0.34$ .

On the other hand, for large sample sizes we have proved that under mild conditions the Bayesian procedure provides consistent posterior model probabilities. This assertion is also valid in the presence of nuisance parameters. Moreover, the speed of the convergence is surprisingly fast. For the multiple test  $H_0 : \theta = \theta_0, H_1 : \theta \leq \theta_0$ ,



**Fig. 4** Disagreement probabilities for the frequentist and Bayesian test for the null  $H_0 : \theta = 0$  versus  $H_1 : \theta \neq 0$  for the normal model  $N(x|\theta, 1)$  and  $n = 1$

$H_2 : \theta \geq \theta_0$  the Bayesian testing procedure is also consistent, and the speed of convergence of the posterior probabilities is also exponential except when sampling from  $H_0$  in which case it is  $O(n^{1/2})$ . We recall that the frequentist test finds serious difficulties for dealing with multiple tests.

**Acknowledgements** The first author of this research was supported by the Junta de Andalucía Grant A-FQM-456-UGR18. We are grateful to two anonymous Referees for their comments that lead to improve the presentation of the article.

## References

- Berger JO (2003) Could Fisher, Jeffreys and Neyman have agreed on testing? *Stat Sci* 18:1–32
- Berger JO, Mortera J (1999) Default Bayes factors for one-sided hypothesis testing. *J Am Stat Assoc* 94:542–554
- Berger JO, Pericchi LR (1996) The intrinsic Bayes factor for linear models (with discussion). In: Bernardo JM et al (eds) *Bayesian statistics*, vol 5. Oxford University Press, New York, pp 25–44
- Berger JO, Sellke T (1987) Testing a point null hypothesis: the irreconcilability of p values and evidence. *J Am Stat Assoc* 82:112–122
- Casella G, Berger R (1987) Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J Am Stat Assoc* 82:106–111
- Casella G, Girón FJ, Martínez ML, Moreno E (2009) Consistency of Bayesian procedures for variable selection. *Ann Stat* 37:1207–1228
- Casella G, Moreno E, Girón FJ (2014) Cluster analysis, model selection, and prior distributions on models, Bayesian. *Analysis* 9:613–658
- Dickey JM (1977) Is the tail area useful as an approximate Bayes factor? *J Am Stat Assoc* 72:138–142
- Dudley RM, Haughton D (1997) Information criteria for multiple data sets and restricted parameters. *Stat Sin* 7:265–284

- Edwards W, Lindman H, Savage L (1963) Bayesian statistical inference for psychological research. *Psychol Rev* 70:193–242
- Efron B, Gous A (2001) Scales of evidence for model selection: Fisher versus Jeffreys. In: Lahiri P (ed) *Lecture notes-monograph series*, vol 38. Institute of Mathematical Statistics, Hayward, pp 208–246
- Girón FJ, Martínez ML, Moreno E, Torres F (2006) Objective testing procedures in linear models: calibration of the p-values. *Scand J Stat* 33:765–784
- Kass RE, Vaidyanathan SK (1992) Approximate Bayes factors and orthogonal parameters. With application to testing equality of two binomial proportions. *J R Stat Soc Ser B* 54:129–144
- Micheas AC, Dey DK (2003) Prior and posterior predictive p-values in the one-side location parameter testing problem. *Sankhya Ser A* 65:158–178
- Moreno E (1997) Bayes factors for intrinsic and fractional priors in nested models, Bayesian robustness. In: Dodge Y (ed) *L<sub>1</sub> Statistical procedures and related topics*, vol 31 of *lecture notes-monograph series*. Institute of Mathematical Statistics, Hayward, CA, pp 257–270
- Moreno E (2005) Objective Bayesian analysis for one-sided testing. *Test* 14:181–198
- Moreno E, Girón FJ (2005) Consistency of Bayes factors for linear models. *C R Acad Sci Paris Ser I* 340:911–914
- Moreno E, Bertolino F, Racugno W (1998) An intrinsic limiting procedure for model selection and hypothesis testing. *J Am Stat Assoc* 93:1451–1460
- Moreno E, Girón FJ, Casella G (2010) Consistency of objective Bayes factors as the model dimension grows. *Ann Stat* 38:1937–1952
- Moreno E, Girón FJ, Casella G (2015) Posterior model consistency in variable selection as the model dimension grows. *Stat Sci* 30:228–241
- Morris C (1987) Comment to Casella and Berger (1987). *J Am Stat Assoc* 82:112–139
- Mulder J (2014) Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses. *Comput Stat Data Anal* 71:448–463
- Mulder J, Raftery AE (2019) BIC extensions for order-constrained model selection. *Sociol Methods Res.* <https://doi.org/10.1177/0049124119882459>
- Wang M, Maruyama Y (2016) Consistency of Bayes factor for nonnested model selection when the model dimension grows. *Bernoulli* 22:2080–2100
- Wilks SS (1963) *Mathematical statistics*. Wiley, New York

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.