

## Article

# An Objective and Robust Bayes Factor for the Hypothesis Test One Sample and Two Population Means

Israel A. Almodóvar-Rivera <sup>1,\*</sup>  and Luis R. Pericchi-Guerra <sup>2</sup> <sup>1</sup> Department of Mathematical Sciences, University of Puerto Rico at Mayagüez, Mayagüez, PR 00680, USA<sup>2</sup> Department of Mathematics, University of Puerto Rico at Rio Piedras, San Juan, PR 00930, USA; luis.pericchi@upr.edu

\* Correspondence: israel.almodovar@upr.edu

**Abstract:** It has been over 100 years since the discovery of one of the most fundamental statistical tests: the Student's  $t$  test. However, reliable conventional and objective Bayesian procedures are still essential for routine practice. In this work, we proposed an objective and robust Bayesian approach for hypothesis testing for one-sample and two-sample mean comparisons when the assumption of equal variances holds. The newly proposed Bayes factors are based on the intrinsic and Berger robust prior. Additionally, we introduced a corrected version of the Bayesian Information Criterion (BIC), denoted BIC-TESS, which is based on the effective sample size (TESS), for comparing two population means. We studied our developed Bayes factors in several simulation experiments for hypothesis testing. Our methodologies consistently provided strong evidence in favor of the null hypothesis in the case of equal means and variances. Finally, we applied the methodology to the original Gosset sleep data, concluding strong evidence favoring the hypothesis that the average sleep hours differed between the two treatments. These methodologies exhibit finite sample consistency and demonstrate consistent qualitative behavior, proving reasonably close to each other in practice, particularly for moderate to large sample sizes.



**Citation:** Almodóvar-Rivera, I.A.; Pericchi-Guerra, L.R. An Objective and Robust Bayes Factor for the Hypothesis Test One Sample and Two Population Means. *Entropy* **2024**, *26*, 88. <https://doi.org/10.3390/e26010088>

Academic Editors: Carlos Alberto De Bragança Pereira, Paulo Canas Rodrigues and Mark Andrew Gannon

Received: 30 November 2023

Revised: 10 January 2024

Accepted: 17 January 2024

Published: 20 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** student's  $t$  test; Bayes factors; intrinsic priors; robust prior

## 1. Introduction

One of the fundamental topics in statistics revolves around the one-sample population means and the comparison of two-sample means. The go-to method for addressing this question is typically the Student's  $t$  test [1]. Conducting a hypothesis test for the population mean holds significant importance in the scientific research community and various fields where making inferences about population parameters is pivotal. Frequentists heavily rely on  $p$ -values to determine whether to reject or not reject the null hypothesis [2]. However,  $p$ -values, along with significance testing based on fixed  $\alpha$ -levels, tend to exaggerate evidence against null hypotheses for large sample sizes and lack the operational meaning of a probability [3–5]. While the Bayesian approach has gained attention in hypothesis testing and model selection [6,7], its application in essential statistics topics remains somewhat limited [8]. This raises the question: Why is a Bayesian Student's  $t$  test necessary? We argue for two main reasons. Firstly, Bayesian tests provide evidence for a hypothesis of interest that naturally adapts to any sample size. Secondly, the Bayes factor can be easily converted to posterior model probabilities and support one of the testing frameworks. Another crucial consideration is that scientific questions often have a Bayesian nature, such as, “What is the probability that these two treatments differ?”.

Bayesian hypothesis testing and model selection have been undergoing extensive development because of recent advances in the creation of “default” Bayes factors that can be used in the absence of substantial subjective prior information [5,9–11]. The study in [12] proposed some arguments for the choice of the prior, such as (i) the fact that it

is located around zero, (ii) the scale parameter  $\sigma$ , (iii) the fact that it is symmetric, and (iv) that it should have no moments. Bayes factors are attractive in terms of interpretation as odds, and the direct probability of the posterior model is readily understandable by general users of statistics [13]. Methods based on conjugate priors for the Student's  $t$  test have a long history. Perhaps the most transparent approach for the two-sample Student's  $t$  is in [14]. However, natural conjugate priors do not lead to robust procedures; they have tails that are typically of the same form as the likelihood function and will hence remain influential when the likelihood function is concentrated in the prior tails, which can lead to inconsistency [15]. This conjugate Bayes factor for comparing two samples based on the Student's  $t$  is finite sample-inconsistent, i.e., it does not go to zero when the estimates go to infinity.

In this work, we proposed an objective and robust Bayes factor for testing the hypothesis of one-sample and two-sample means based on the  $t$ -statistic. Our Bayes factors can be easily implemented, allowing researchers to determine support for a particular hypothesis. This manuscript proceeds as follows. In Section 2, we derive these objectives and robust Bayes factors for one-sample and two-sample scenarios and demonstrate their finite sample consistency. In Section 3, we compare our Bayes factors with existing methodologies under several experimental frameworks. In Section 4, we apply our methodologies to real-life datasets such as the original Gosset sleep data and to comparisons of changes in blood pressure in mice according to their assigned diet. We conclude this work with a discussion in Section 5.

## 2. Methodology

Statistical inference for the mean (one or two samples) has an important rule in statistics and several fields. For instance, it is very common to test in terms of the average or population mean. Suppose that we are comparing two hypotheses,  $H_0 : \theta_0 \in \Theta$  vs.  $H_1 : \theta_1 \in \Theta$ . Suppose that we have available prior densities  $\pi_i, i = 1, 2$  for each hypothesis and let  $f_i(x|\theta_i)$  be the probability density function under the  $i$ th hypothesis. Define the marginal or predictive densities for each hypothesis of interest (or model),

$$m_i(x) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i, \quad (1)$$

which are sometimes called the evidence of the  $i$ th hypothesis or model. The Bayes factor for comparing  $H_0$  to  $H_1$  is then given by

$$B_{01} = \frac{m_0(x)}{m_1(x)} = \frac{\int f_0(x|\theta_0)\pi_0(\theta_0)d\theta_0}{\int f_1(x|\theta_1)\pi_1(\theta_1)d\theta_1}. \quad (2)$$

The interpretation of the Bayes factor proceeds as follows. If  $B_{01} > 1$ , then the evidence is in favor of the null hypothesis, while  $B_{01} < 1$  gives evidence in favor of the alternative hypothesis. If prior probabilities  $P(H_i), i = 0, 1$  of the hypotheses are available, then one can compute the posterior probabilities of it from the Bayes factors. The posterior probability of  $H_0$ , given the data  $x$ , is

$$\mathbb{P}(H_0|x) = \frac{m_0(x)P(H_0)}{\sum_{j=0}^1 m_j(x)P(H_j)} = \frac{1}{1 + \frac{P(H_1)}{P(H_0)} B_{10}}; \quad (3)$$

where  $B_{10} = 1/B_{01}$ .

### 2.1. One-Sample Mean Hypothesis Testing

A one-sample hypothesis test for the population mean is one of the most fundamental statistics topics, either as an introductory topic or to address research questions. Suppose

we have a random sample from a normal distribution, i.e.,  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , with an unknown standard deviation  $\sigma > 0$ . We are interested in testing for the population mean  $\mu$ .

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0. \quad (4)$$

A Bayesian approach to test this hypothesis is based on the theory of intrinsic priors [16,17]. The authors begin with the noninformative priors for the null and alternative hypotheses,  $\pi_0^N(\sigma) = 1/\sigma$  and  $\pi_1^N(\mu, \sigma) = 1/\sigma^2$ . After some calculations, the authors showed that the conditional proper intrinsic prior under the alternative Hypothesis  $H_1$  is given by

$$\pi^I(\mu|\sigma) = \frac{1}{2\sqrt{\pi}\sigma} \frac{1 - e^{-\mu^2/\sigma^2}}{(\mu^2/\sigma^2)}.$$

One can express  $\pi^I(\mu, \sigma) = \pi(\mu|\sigma)\pi(\sigma)$ . The resulting intrinsic prior under  $H_1$  is defined as

$$\pi_1^I(\mu, \sigma) = \pi_1^I(\mu|\sigma)/\sigma = \frac{1}{2\sqrt{\pi}} \frac{1 - e^{-\mu^2/\sigma^2}}{\mu^2}.$$

The approximate Bayes factors based on the intrinsic prior ( $B_{01}^{IP}$ ) for a one-sample population mean are

$$B_{01}^{IP} \approx B_{01}^N \cdot \frac{\pi_1^N(\hat{\mu}, \hat{\sigma})}{\pi_1^I(\hat{\mu}, \hat{\sigma})} (1 + o(1)).$$

Here,  $\hat{\mu}$  and  $\hat{\sigma}$  are the Maximum Likelihood Estimators (MLEs) under  $H_1$ . The resulting Bayes factor for the hypothesis in (4) is

$$B_{01}^{IP} \approx \sqrt{2n} \left( 1 + \frac{t^2}{n-1} \right)^{-n/2} \frac{t^2/(n-1)}{1 - e^{-t^2/(n-1)}}; \quad (5)$$

where  $t = (\bar{x} - \mu_0)/s/\sqrt{n}$ , where  $\bar{x}$  is the sample mean and  $s$  is the sample standard deviation. Large values of  $B_{01}$  give evidence in favor of the null hypothesis. Also, we can transform these Bayes factors using the natural logarithm scale ( $2 \log B_{01}$ ), and values above 3 give some evidence in favor of the null hypothesis, while values above 10 give stronger evidence in favor of the null hypothesis; see [13].

This Bayes factor satisfies the finite sample consistency principle. Suppose that we are comparing the alternative hypothesis with the null hypothesis,  $H_0 : \beta = 0$ . As the least squares estimate  $\hat{\beta}$  (and the noncentrality parameter) goes to infinity, so that one becomes sure that  $H_0$  is wrong, the Bayes factor of  $H_0$  to  $H_1$  goes to zero.

**Theorem 1.** For a fixed sample size  $n \geq 2$ , the Bayes factor based on the intrinsic prior ( $B_{01}^{IP}$ ) for the one-sample mean  $\mu$  is finite sample-consistent.

**Proof.** For a fixed sample,  $n \geq 2$ , and letting  $t^2 \rightarrow \infty$  or equivalently  $|t| \rightarrow \infty$ , the Bayes factor based on the intrinsic prior goes to 0, i.e.,

$$\lim_{|t| \rightarrow \infty} B_{01}^{IP} \rightarrow 0; \text{ or equivalently } \lim_{|t| \rightarrow \infty} \mathbb{P}(H_0|\mathbf{x}) \rightarrow 0.$$

The Bayes Factor based on the intrinsic prior  $B_{01}^{IP}$  is finite sample consistent.  $\square$

### Robust Bayes Factor for the One-Sample Test for the Mean

Even though the Bayes Factor constructed using the intrinsic prior is finite sample-consistent, it is only an approximation. Evidence has been found that priors with flatter tails than those of the likelihood function tend to be fairly robust, [18,19]. The robust prior proposed here is developed by [20]; we call it the Berger robust prior. This prior is hierarchical; by such a choice, we can obtain robustness while keeping the calculations

relatively simple, and the computations are exact. The definition of this robust prior, denoted  $\pi^R(\xi)$ , can be defined as follows:

1.  $\xi|\lambda \sim \mathcal{N}_p(\mu, \mathbf{B}(\lambda))$ , where  $\mathbf{B}(\lambda) = \rho\lambda^{-1}(b+d) - d$  and  $\rho = \frac{p+1}{p+3}$ .
2.  $\lambda$  has a density  $\pi(\lambda) = \frac{1}{2}\lambda^{-1/2}$  on  $(0, 1)$ .

where  $p$  is the rank of the design matrix. Recall that we are interested in testing (4); therefore, under the null hypothesis, the likelihood is in the form

$$f_0(\mathbf{x}|\mu_0, \sigma) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}\right\}.$$

The noninformative prior under the null hypothesis is  $\pi^N(\sigma) = 1/\sigma$ . The marginal density  $m(\mathbf{x})$  under the null hypothesis is given by

$$\begin{aligned} m_0(\mathbf{x}) &= \int f_0(\mathbf{x}|\sigma)\pi^N(\sigma)d\sigma = \int_0^\infty (2\pi)^{-n/2}\sigma^{-n}e^{-\frac{SS_0^2}{2\sigma^2}}\sigma^{-1}d\sigma \\ &= (2\pi)^{-n/2}(SS_0^2)^{-n/2}\Gamma\left(\frac{n}{2}\right); \end{aligned} \quad (6)$$

where  $SS_0^2 = \sum_{i=1}^n (x_i - \mu_0)^2$ , the sums of squares under  $H_0$  and  $\Gamma(\cdot)$  is the gamma function. Similarly, we can obtain an alternative likelihood under the alternative Hypothesis  $H_1$ :

$$f_1(\mathbf{x}|\mu, \sigma) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{SS_1^2}{2\sigma^2} - \frac{n(\mu - \bar{x})^2}{2\sigma^2}\right\}.$$

Here,  $\bar{x} = n^{-1}\sum_{i=1}^n x_i$  is the sample mean and  $SS_1^2 = \sum_{i=1}^n (x_i - \bar{x})^2$  is the sum of squares under the alternative. The Berger robust prior will be considered under the alternative hypothesis  $\pi_1(\mu, \sigma) = \pi^R(\mu|\sigma)/\sigma$ . The marginal density under  $H_1$  is

$$\begin{aligned} m_1(\mathbf{x}) &= \int f_1(\mathbf{x}|\mu, \sigma)\pi_1(\mu, \sigma)d\mu d\sigma = \int f_1(\mathbf{x}|\mu, \sigma)\pi^R(\mu|\sigma)\frac{1}{\sigma}d\mu d\sigma \\ &= \frac{(2\pi)^{-n/2}}{n \cdot (\bar{x} - \mu_0)^2} \sqrt{\frac{n+1}{2}} \left( \frac{\Gamma((n-2)/2)}{2(\frac{s^2}{2})^{(n-2)/2}} - \frac{\Gamma((n-2)/2)}{2(\frac{s^2}{2} + \frac{(\bar{x}-\mu_0)^2}{(n+1)/n})^{(n-2)/2}} \right). \end{aligned} \quad (7)$$

Computing the ratio of the marginals from (6) and (7), the Bayes factor based on the Berger robust prior is given by

$$B_{01}^R = \sqrt{\frac{2}{n+1}} \frac{n-2}{n-1} t^2 \left(1 + \frac{t^2}{n-2}\right)^{-n/2} \left(1 - \left(1 + \frac{2t^2}{n^2-1}\right)^{-(n-2)/2}\right)^{-1}. \quad (8)$$

Here,  $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$  is the usual  $t$ -statistic with  $n-1$  degrees of freedom, where  $\bar{x}$  and  $s$  are the sample mean and sample standard deviation, respectively.

**Theorem 2.** For a fixed sample size  $n \geq 3$ , the Bayes factor based on the Berger robust prior ( $B_{01}^R$ ) for the one-sample mean  $\mu$  is finite sample-consistent.

**Proof.** For a fixed sample,  $n \geq 3$ , and letting  $t^2 \rightarrow \infty$ , or equivalently  $|t| \rightarrow \infty$ , the Bayes factor based on the Berger robust prior goes to 0, i.e.,

$$\lim_{|t| \rightarrow \infty} B_{01}^R \rightarrow 0; \text{ or equivalently } \lim_{|t| \rightarrow \infty} \mathbb{P}(H_0|\mathbf{x}) \rightarrow 0.$$

The Bayes factor based on the Berger robust prior  $B_{01}^R$  is finite sample-consistent.  $\square$

Unlike the Bayes factor derived with the intrinsic prior, this robust Bayes factor has a closed form. We conclude the derivations for the Bayesian approach based on the intrinsic and robust prior that are finite sample-consistent. We now extend the objective and robust Bayesian approach to the two-sample scenarios.

## 2.2. Two-Sample Mean Hypothesis Test

Another fundamental research question of interest is whether or not the two groups are similar. This problem is usually addressed in the two-sample Student's  $t$  test to compare if these groups differ in means. Let  $X_1, \dots, X_{n_1} \sim N(\mu_1, \sigma^2)$  and let  $Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma^2)$  independent of  $X$  with  $\sigma > 0$  unknown. At first, we noticed that we were assuming that these two samples arise from a normal distribution with different means but equal variances. It is common interest to determine if these two samples are equal, or at least that they do not differ in location. To answer this, a hypothesis test for comparing two-sample means is performed, i.e.,  $H_0 : \mu_1 = \mu_2$  against  $H_1 : \mu_1 \neq \mu_2$ . To answer this question, Ref. [14] proposed the conjugate Bayes factor. This Bayes factor is based on the conjugate prior:  $(\mu_1 - \mu_2)/\sigma = \delta/\sigma | \sigma \sim N(\lambda, \sigma_\delta^2)$ . Centering the prior assessment on the null hypothesis, i.e., making  $\lambda = 0$ , is usually a very reasonable choice. Then, the conjugate Bayes factor is simplified as  $B_{01}^C = \left[ \frac{1+t_v^2/\nu}{1+t_v^2/(\nu(1+n_\delta\sigma_\delta^2))} \right]^{-(\nu+1)/2} \sqrt{1+n_\delta\sigma_\delta^2}$ . However, this Bayes factor is not finite sample-consistent, as  $|t| \rightarrow \infty$ , or  $t^2 \rightarrow \infty$ ; the  $B_{01}^C$  does not go to zero, or equivalently, the posterior probability of the null hypothesis  $\mathbb{P}(H_0|data)$  does not go to zero. In fact, as  $|t| \rightarrow \infty$ , then  $B_{01}^C \rightarrow (1+n_\delta\sigma_\delta^2)^{-\nu/2} > 0$ , where  $n_\delta = 1/(1/n_1 + 1/n_2)$  and  $\nu = n_1 + n_2 - 2$  are the degrees of freedom.

### 2.2.1. Intrinsic Bayes Factor for Two-Sample Means

To address the limitation of the conjugate prior, our first approach is based on the theory of intrinsic priors introduced in [16,17]. Similar to the one-sample case, the method is to dig out a prior that yields, for moderate to large sample sizes, results equivalent to an established method for scaling the intrinsic Bayes factors. The resulting set of equations typically has solutions, at least in the nested hypothesis scenario, which is our case, and has been successfully applied coupled with the intrinsic Bayes factor method. Consider the hypotheses tests for the comparison of two populations means with unknown and equal variance  $\sigma^2 > 0$ ,  $H_0 : \mu_1 = \mu_2$  vs.  $H_1 : \mu_1 \neq \mu_2$ . Let  $\delta_0 = (\mu_1 + \mu_2)/2$  and  $\delta_1 = (\mu_1 - \mu_2)/2$ , then  $\mu_1 = \delta_0 + \delta_1$  and  $\mu_2 = \delta_0 - \delta_1$ . This transformation leads us to the following design matrix  $\mathbf{X}$  based on the training samples:

$$\mathbf{X}\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{2 \times 1} & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{2 \times 1} & \mathbf{1}_{2 \times 1} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{4 \times 1} & \begin{bmatrix} \mathbf{1}_{2 \times 1} \\ -\mathbf{1}_{2 \times 1} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \delta_0 \\ \delta_1 \end{bmatrix} = [\mathbf{X}_0(l) : \mathbf{X}_1(l)] \boldsymbol{\delta};$$

where  $\mathbf{1}_{k \times 1}$  and  $\mathbf{0}_{k \times 1}$  are vectors of 1's and 0's of length  $k$ . The parameter of non-centrality can be computed as

$$\begin{aligned} \lambda(l) &= \sigma^{-2} \delta_1^t \mathbf{X}_1^t(l) \left( \mathbf{I} - \mathbf{X}_0(l) [\mathbf{X}_0^t(l) \mathbf{X}_0(l)]^{-1} \mathbf{X}_0^t(l) \right) \mathbf{X}_1(l) \delta_1 \\ &= \sigma^{-2} \left( \frac{\mu_1 - \mu_2}{2} \right)^2 \cdot 4 = \frac{(\mu_1 - \mu_2)^2}{\sigma^2}; \end{aligned} \quad (9)$$

which becomes, in the comparison of two means,  $\lambda(l) = (\mu_1 - \mu_2)^2 / \sigma^2$ . Following the general theory of the intrinsic Bayes factor for linear models [16,17], we have that an intrinsic prior is of the following form:

$$\pi^I(\lambda(l)|\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \frac{(1 - e^{-\lambda(l)/2})}{\lambda(l)}.$$

Substitution from  $\lambda(I)$  using the non-centrality parameter of (9), then the transformation to the conditional of the parameter under the simple test, is  $H_0 : \delta_1 = 0$  vs.  $H_1 : \delta_1 \neq 0$ . The conditional intrinsic prior for the hypothesis test is

$$\pi_1^I(\delta_1|\sigma) = \frac{1 - e^{-8\delta_1^2/\sigma^2}}{4\sqrt{2\pi}(\delta_1^2/\sigma)}. \quad (10)$$

This conditional prior is proper, i.e.,  $\int \pi_1^I(\delta_1|\sigma)d\delta_1 = 1$  and it satisfies the condition discussed by [12]. The intrinsic prior, under the alternative Hypothesis  $H_1$ , is of the form  $\pi_1^I(\delta_1, \delta_0, \sigma) = \pi_1^I(\delta_1|\delta_0, \sigma) \cdot \pi_1^I(\delta_0, \sigma)$ , where  $\pi_1^I(\delta_0, \sigma) = 1/\sigma$ , i.e.,

$$\pi_1^I(\delta_0, \delta_1, \sigma) = \frac{1}{\sigma} \cdot \frac{1 - e^{-8\delta_1^2/\sigma^2}}{4\sqrt{2\pi}\delta_1^2/\sigma} = \frac{1 - e^{-8\delta_1^2/\sigma^2}}{4\sqrt{2\pi}\delta_1^2}. \quad (11)$$

Setting up this framework, we can derive the intrinsic Bayes factor to compare two-sample means. We will first obtain the marginal density under the null hypothesis  $m_0(\mathbf{x}, \mathbf{y})$ . First, consider the joint likelihood function of the two samples under the null Hypothesis  $H_0$ :

$$f_0(\mathbf{x}, \mathbf{y}|\delta_0, \sigma) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{S_x^2 + S_y^2 + n_1(\bar{x} - \delta_0)^2 + n_2(\bar{y} - \delta_0)^2}{2\sigma^2}\right\}.$$

Here,  $\bar{x}$  is the sample mean of the first group,  $\bar{y}$  is the sample mean of the first group, and  $S_x^2 = \sum_{i=1}^{n_1}(x_i - \bar{x})^2$ ,  $S_y^2 = \sum_{i=1}^{n_2}(y_i - \bar{y})^2$  are the sums of squares under the null hypothesis. The marginal density using the non-informative prior  $\pi^N(\delta_0, \sigma) = 1/\sigma$  is computed as

$$\begin{aligned} m_0(\mathbf{x}, \mathbf{y}) &= \int \int f_0(\mathbf{x}, \mathbf{y}|\delta_0, \sigma) \pi(\delta_0, \sigma) d\delta_0 d\sigma \\ &= \frac{2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right) (S_x^2 + S_y^2)^{-(n-1)/2} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2}}{\sqrt{n}}, \end{aligned} \quad (12)$$

where  $t = (\bar{x} - \bar{y}) / (S_p \sqrt{n_\delta})$ , where  $n = n_1 + n_2$ ,  $S_p$  is the pooled standard deviation, i.e.,  $S_p^2 = (S_x^2 + S_y^2) / (n - 2)$  and  $n_\delta = 1/n_1 + 1/n_2$ . Similarly, the joint likelihood function of the two samples under the alternative Hypothesis  $H_1$  is given by

$$f_1(\mathbf{x}, \mathbf{y}|\delta_0, \delta_1, \sigma) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{S_x^2 + S_y^2 + n_1(\bar{x} - (\delta_0 + \delta_1))^2 + n_2(\bar{y} - (\delta_0 - \delta_1))^2}{2\sigma^2}\right\}.$$

The marginal density using the intrinsic prior  $\pi^I(\delta_0, \delta_1, \sigma)$  defined in (11) is given by

$$m_1(\mathbf{x}, \mathbf{y}) = \int \int \int f_1(\mathbf{x}, \mathbf{y}|\delta_0, \delta_1, \sigma) \pi^I(\delta_0, \delta_1, \sigma) d\delta_0 d\delta_1 d\sigma. \quad (13)$$

As in the one-sample framework, this Bayes factor can be approximated using the non-informative prior  $\pi^N(\delta_0, \delta_1, \sigma) = 1/\sigma^2$  in the asymptotic result as

$$\begin{aligned} m_1(\mathbf{x}, \mathbf{y}) &\approx m_1^N(\mathbf{x}, \mathbf{y}) \frac{\pi^I(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})}{\pi^N(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})} \\ &= \int \int \int f_1(\mathbf{x}, \mathbf{y}|\delta_0, \delta_1, \sigma) \pi^N(\delta_0, \delta_1, \sigma) d\delta_0 d\delta_1 d\sigma \frac{\pi^I(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})}{\pi^N(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})} \\ &= \frac{\sqrt{\pi} 2^{(n-3)/2-1/2} \Gamma\left(\frac{n-1}{2}\right) (S_x^2 + S_y^2)^{-(n-1)/2}}{\sqrt{n_1 n_2}}. \end{aligned} \quad (14)$$

Using (12) and (14), we can compute  $B_{01}^N$ :

$$B_{01}^N = \frac{m_0^N(\mathbf{x}, \mathbf{y})}{m_1^N(\mathbf{x}, \mathbf{y})} = \sqrt{\frac{2}{\pi n_\delta}} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2}; \quad (15)$$

where  $t^2 = (\bar{x} - \bar{y})^2 / (S_p n_\delta)$ ,  $S_p$  is the pooled estimate of the variance,  $S_p^2 = (S_x^2 + S_y^2) / (n-2)$  and  $n_\delta = (1/n_1 + 1/n_2)$ . Let  $\hat{\delta}_1$  and  $\hat{\sigma}^2$  be the corresponding maximum likelihood estimator (MLE). Let  $\hat{\delta}_1 = (\bar{x} - \bar{y})/2$  and  $\hat{\sigma}^2 = (S_x^2 + S_y^2)/n = (n-2)/nS_p^2$ ; where  $S_p^2$  is the variance pooled estimates and  $n = n_1 + n_2$ . We can express the  $\hat{\delta}_1^2/\hat{\sigma}^2 = n_\delta n t^2 / (n-2)$  in terms of the  $t$ -statistic. Then, the approximate intrinsic Bayes factor  $B_{01}^{IP}$  can be obtained by

$$\begin{aligned} B_{01}^{IP} &\approx B_{01}^N \frac{\pi^N(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})}{\pi^I(\hat{\delta}_0, \hat{\delta}_1, \hat{\sigma})} = \sqrt{\frac{2}{\pi n_\delta}} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2} \frac{\hat{\delta}_1^2/\hat{\sigma}^2}{\frac{1-e^{-8\hat{\delta}_1^2/\hat{\sigma}^2}}{4\sqrt{2\pi}}} \\ &= \frac{\sqrt{n_\delta} n t^2}{n-2} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2} \left(1 + \coth\left(\frac{n_\delta n t^2}{n-2}\right)\right). \end{aligned} \quad (16)$$

Here,  $t = (\bar{x} - \bar{y}) / (S_p \sqrt{n_\delta})$  and  $\coth(\cdot)$  is the hyperbolic cotangent function defined as  $\coth(x) = (e^{2x} + 1) / (e^{2x} - 1)$ .

**Theorem 3.** For a fixed sample size  $n \geq 4$ , the Bayes factor based on the intrinsic prior ( $B_{01}^{IP}$ ) for the comparison of two population means is finite sample-consistent.

**Proof.** For a fixed sample,  $n \geq 4$ , and letting  $t^2 \rightarrow \infty$ , or equivalently  $|t| \rightarrow \infty$ , the Bayes factor based on the intrinsic prior goes to 0, i.e.,

$$\lim_{|t| \rightarrow \infty} B_{01}^{IP} \rightarrow 0; \text{ or equivalently } \lim_{|t| \rightarrow \infty} \mathbb{P}(H_0 | \mathbf{x}) \rightarrow 0.$$

The Bayes factor based on the intrinsic prior  $B_{01}^{IP}$  is finite sample-consistent.  $\square$

### 2.2.2. Robust Bayes Factor for the Comparison of Two-Sample Means

Consider observations of a random sample from group 1 and group 2 of size  $n_1$  and  $n_2$ , respectively. We assume these groups have common variance ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ), respectively. The model of interest in this case,  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ , with  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , for  $i = 1, 2$  and  $j = 1, \dots, n_i$ . We want to compare  $H_0 : \alpha_1 = \alpha_2$  against  $H_0 : \alpha_1 \neq \alpha_2$ . Further, consider the constraint that  $\alpha_1 + \alpha_2 = 0$ ; then, the design matrix  $\mathbf{X}$  can be written as

$$\mathbf{X} = \left[ \mathbf{1}_{n \times 1} \mid \begin{bmatrix} 1/n_1 \mathbf{1}_{n_1 \times 1} \\ -1/n_2 \mathbf{1}_{n_2 \times 1} \end{bmatrix} \right];$$

This leads us to consider the following hypothesis,  $H_0 : \alpha_1 = 0$  vs.  $H_1 : \alpha_1 \neq 0$ . The reference's priors, under the null hypothesis  $H_0$  and alternative hypothesis  $H_1$ :  $\pi_0^N(\mu, \sigma) = 1/\sigma$  and  $\pi_1^N(\mu, \alpha_1, \sigma) = 1/\sigma$ . First, we proceed to find the marginal density under  $H_0$ . Consider the joint likelihood function under the null Hypothesis  $H_0$ :

$$f_0(\mathbf{y}_1, \mathbf{y}_2 | \mu, \sigma) = (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{SS_0^2 + n_1(\bar{y}_1 - \mu)^2 + n_2(\bar{y}_2 - \mu)^2}{2\sigma^2} \right\}.$$

Here,  $\mathbf{y}_1 = (y_{11}, \dots, y_{1n_1})$  and  $\mathbf{y}_2 = (y_{21}, \dots, y_{2n_2})$ ,  $\bar{y}_i$  is the sample mean of the  $i$ th group, and  $SS_0^2$  is the sum of squares under the null hypothesis. The marginal density under the null hypothesis  $m_0(\mathbf{y}_1, \mathbf{y}_2)$  is given by



$$m_0(\mathbf{y}_1, \mathbf{y}_2) = \int \int f_0(\mathbf{y}|\mu, \sigma) \sigma^{-1} d\mu d\sigma$$

$$= \frac{\Gamma((n-1)/2)}{\sqrt{n}} \frac{(\pi)^{-(n-1)/2}}{2} (SS_0^2)^{-(n-1)/2} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2} \quad (17)$$

where  $t^2 = (\bar{y}_1 - \bar{y}_2)^2 / (S_p n_\delta)$ . Here,  $S_p$  is the sample pooled estimate of the variance,  $S_p^2 = (S_1^2 + S_2^2) / (n-2)$ , and  $n_\delta = (1/n_1 + 1/n_2)$ . For the alternative Hypothesis  $H_1$ , we first consider the joint likelihood of group 1 and group 2:

$$f_1(\mathbf{y}_1, \mathbf{y}_2|\mu, \alpha, \sigma) = (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{(S_1^2 + S_2^2 + n_1(\bar{y}_1 - (\mu + \alpha))^2 + n_2(\bar{y}_2 - (\mu - \alpha))^2)}{2\sigma^2} \right\}.$$

The marginal density  $m_1(\mathbf{y}_1, \mathbf{y}_2)$  is given by

$$m_1(\mathbf{y}_1, \mathbf{y}_2) = \int \int \int f_1(\mathbf{y}|\mu, \alpha, \sigma) \times \sigma^{-1} \pi^R(\alpha) d\mu d\alpha d\sigma$$

$$= \sqrt{\frac{n_\delta(b+d)}{(2\pi)^{(n-2)}}} \frac{\Gamma((n-3)/2)}{4\sqrt{\pi n} \hat{\alpha}^2} \left[ \left( \frac{S_1^2 + S_2^2}{2} \right)^{-(n-3)/2} - \left( \frac{S_1^2 + S_2^2}{2} + \frac{\hat{\alpha}^2}{b+d} \right)^{-(n-3)/2} \right]. \quad (18)$$

Here,  $\hat{\alpha} = (\bar{y}_1 - \bar{y}_2)/2$ . The robust Bayes factor is obtained by computing the ratio of the marginal densities of (17) and (18):

$$B_{01}^R = \sqrt{\frac{8n_\delta}{b+d}} \left( \frac{t^2(n-3)}{4(n-2)} \right) \left( 1 + \frac{t^2}{n-2} \right)^{-(n-1)/2} \left( 1 - \left( 1 + \frac{t^2 n_\delta}{2(n-2)(b+d)} \right)^{-(n-3)/2} \right)^{-1}. \quad (19)$$

To finish the calculation of the robust Bayes factor  $B_{01}^R$ , the term  $b+d$  has to be defined. Therefore, we propose using the effective sample size (TESS)  $n_o^e$  of [21]. The first factor  $d = 0.25n_\delta \cdot \sigma^2$ , and the second factor  $b$  is  $b = n_o^e \cdot d = 0.25\sigma^2 \left( \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\} \right)^{-2}$ ; then,  $b+d = 0.25\sigma^2 \times \left[ n_\delta + \left( \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\} \right)^{-2} \right] = 0.25\sigma^2 \times (d^* + b^*)$ . Derivation of TESS is displayed in Appendix A.1.

**Theorem 4.** For a fixed sample size  $n \geq 4$ , the Bayes factor based on the robust prior ( $B_{01}^R$ ) for the comparison of two populations means is finite sample-consistent.

**Proof.** For a fixed sample,  $n \geq 4$ , and letting  $t^2 \rightarrow \infty$ , or equivalently  $|t| \rightarrow \infty$ , the Bayes factor based on the Berger robust prior goes to 0, i.e.,

$$\lim_{|t| \rightarrow \infty} B_{01}^R \rightarrow 0; \text{ or equivalently } \lim_{|t| \rightarrow \infty} \mathbb{P}(H_0|\mathbf{x}) \rightarrow 0.$$

The Bayes factor based on the Berger robust prior  $B_{01}^R$  is finite sample-consistent.  $\square$

The Berger robust prior yields the following (exact) expression for the correction of the main term (for group  $i$ ):

$$\pi^R(\xi_i|d_i, b_i) = \frac{1}{\sqrt{2\pi(d_i + b_i)}} \frac{[1 - \exp(-\frac{\xi_i^2}{(d_i + b_i)})]}{\sqrt{2\xi_i^2/(d_i + b_i)}}.$$

Making the change of variables  $\xi_i = \sqrt{8(d_i + b_i)} \beta^*/\sigma$  and  $\eta = \sigma$  then taking the Jacobian,

$$|J| = \left| \begin{array}{cc} \frac{\sqrt{8(d_i + b_i)}}{\sigma} & -\frac{\sqrt{8(d_i + b_i)}\beta^*}{\sigma^2} \\ 0 & 1 \end{array} \right| = \left| \frac{\sqrt{8(d_i + b_i)}}{\sigma} \right|.$$



the conditional intrinsic prior of Equation (10) is exactly recovered;  $\pi^I(\beta^*|\sigma) = \pi^R(\xi_i|d_i, b_i)|J|$ . This established a correspondence between the intrinsic and Berger's robust priors for the Student's  $t$  test.

### 2.2.3. The Effective Sample Size Bayesian Information Criterion (BIC-TESS)

Our final Bayes factor for comparing two-sample means is a variation of the Bayesian Information Criterion (BIC) of [22]. The BIC is a popular method to determine the best model in a set of competing models. However, in comparing the two-sample means, the BIC does not consider the information available in both groups but rather the entire sample. Here, we proposed replacing the sample size  $n$  with TESS. This may be used to form what may be claimed to be the corrected BIC or BIC-TESS. It can be demonstrated that BIC with TESS is:

$$B_{01}^{TESS} = \sqrt{n_o^e} \left( 1 + \frac{t^2}{n-2} \right)^{-n/2}, \quad (20)$$

where  $n_o^e$  is defined by [11]. Derivation of TESS is in Appendix A.1. If we have a balanced situation, where  $n_1 = n_2$ , the BIC-TESS is similar to the regular BIC. If the situation is unbalanced, the BIC-TESS is stabilized, since as  $n_2 \rightarrow \infty$ , the  $B_{01}^{TESS} \rightarrow \sqrt{n_1}$ .

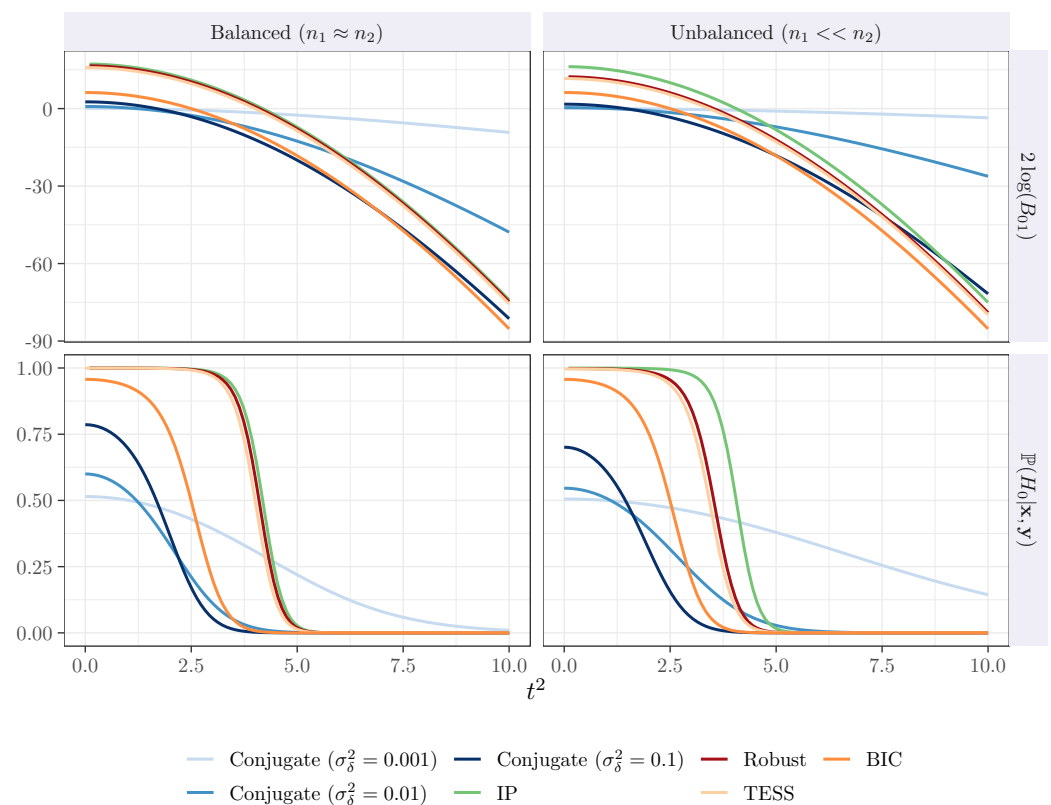
**Theorem 5.** For a fixed sample size  $n \geq 3$ , the corrected BIC ( $B_{01}^{TESS}$ ) for the two-sample mean  $\mu$  is finite sample-consistent.

**Proof.** For a fixed sample,  $n \geq 3$ , and letting  $t^2 \rightarrow \infty$ , or equivalently  $|t| \rightarrow \infty$ , the Bayesian Information Criterion constructed with TESS goes to 0, i.e.,

$$\lim_{|t| \rightarrow \infty} B_{01}^{TESS} \rightarrow 0; \text{ or equivalently } \lim_{|t| \rightarrow \infty} \mathbb{P}(H_0|\mathbf{x}) \rightarrow 0.$$

The corrected Bayesian Information Criterion is  $B_{01}^{TESS}$  is finite sample-consistent.  $\square$

In Figure 1, we compare the asymptotic behavior of the Bayes factors and the posterior probability of the null hypothesis when the samples are balanced ( $n_1 \approx n_2$ ) and unbalanced ( $n_1 \ll n_2$ ). The Bayes factor, based on the Berger robust prior (dark red), is very close in the range of evidence to the intrinsic Bayes factor (green) and the BIC-TESS (light orange). The robust Bayes factor, the intrinsic Bayes factor, and the BIC-TESS are relatively closed when the situation is balanced. In the unbalanced scenario, the robust Bayes factor and the BIC-TESS remain relatively close, while the intrinsic Bayes factor slightly increases. The conjugate Bayes factor (blue) is represented with different values of the prior variance  $\sigma_\delta^2$ ; darker color means higher values for the prior variance. Recall that the conjugate Bayes factor is not finite sample-consistent, and its behavior depends on the choice of  $\sigma_\delta^2$ .



**Figure 1.** Results in terms of  $2\log(B_{01})$  and posterior probability for the finite sample consistency.

### 3. Simulation Experiments

#### Experiments for the One- and Two-Sample Mean Comparisons

We generated 500 datasets from random samples taken from a normal distribution, Student's  $t$  distribution with one degree of freedom, and gamma distribution. For each of these distributions, the mean and standard deviation values were set to  $\mu_1 = 5$  and  $\sigma_1 = 3$ . The second group was created with a combination of several parameters for the location; the mean values were  $\mu_2 \in \{\mu_1, 1.5\mu_1, 2\mu_1\}$ , and for the standard deviation of the second group,  $\sigma_2 \in \{\sigma_1, 2\sigma_1, 3\sigma_1\}$ . In the case of the Student's  $t$  distribution, both groups were simulated with  $\nu = 1$  degrees of freedom. The simulated gamma samples were obtained using the method of moments for the shape parameter, with  $\alpha_i = \mu_i^2/\sigma_i^2$ , and the scale parameter, with  $\beta_i = \sigma_i^2/\mu_i$ , for  $i = 1, 2$ .

We compared our methodologies with several Bayes factors used when comparing two population means, displayed in Table 1.  $B_{01}^S$  is the classical Bayesian Information Criterion (BIC) of [22],  $(B_{01}^{ZS})$  is based on the Zellner and Siow prior [23], the two-sample Student's  $t$  Bayes factor of [14] is based on the conjugate prior with  $\sigma_\delta^2 = 1/3$ , the arithmetic Bayes factor ( $B_{10}^{EIA}$ ) of [24], and [12]'s Bayes Factor ( $B_{01}^J$ ) for the comparison of two-sample means with equal variances. One set of these Bayes factors—the BIC of Schwartz and the Zellner and Siow Bayes factors—depends only on the sample size  $n$ . The other set, based on the conjugate prior, intrinsic, Berger's (here called robust), and finally, the modified Jeffrey's prior, depends on the term  $n_\delta = 1/n_1 + 1/n_2$ . In our experiments, we do not consider the constant  $2/5$  for  $B_{01}^J$ , since we believe it satisfies the condition that the samples arise from the same distribution; for more details about the use of the constant  $2/5$ , see [12]. We also studied these Bayes factors in unbalanced situations. Heavily unbalanced samples are interesting not only from a theoretical point of view but also because they are often observed in practice in observational studies; the results of these are displayed in Figures A1–A3.

Performance was compared using the twice natural base logarithm Bayes factors ( $2\log(B_{01})$ ) for comparing the null hypothesis ( $\mu_1 = \mu_2$ ) against the alternative ( $\mu_1 \neq \mu_2$ ). This transformation allows the interpretation to be on the same scale as the deviance and likelihood ratio test statistics, as discussed in [13].

**Table 1.** Bayes factors based on the one- and two-sample means based on the Student's  $t$  test. The third column applies only to the two-sample comparison and is limiting when  $t^2 \rightarrow 0$  and  $n_2 \rightarrow \infty$ .

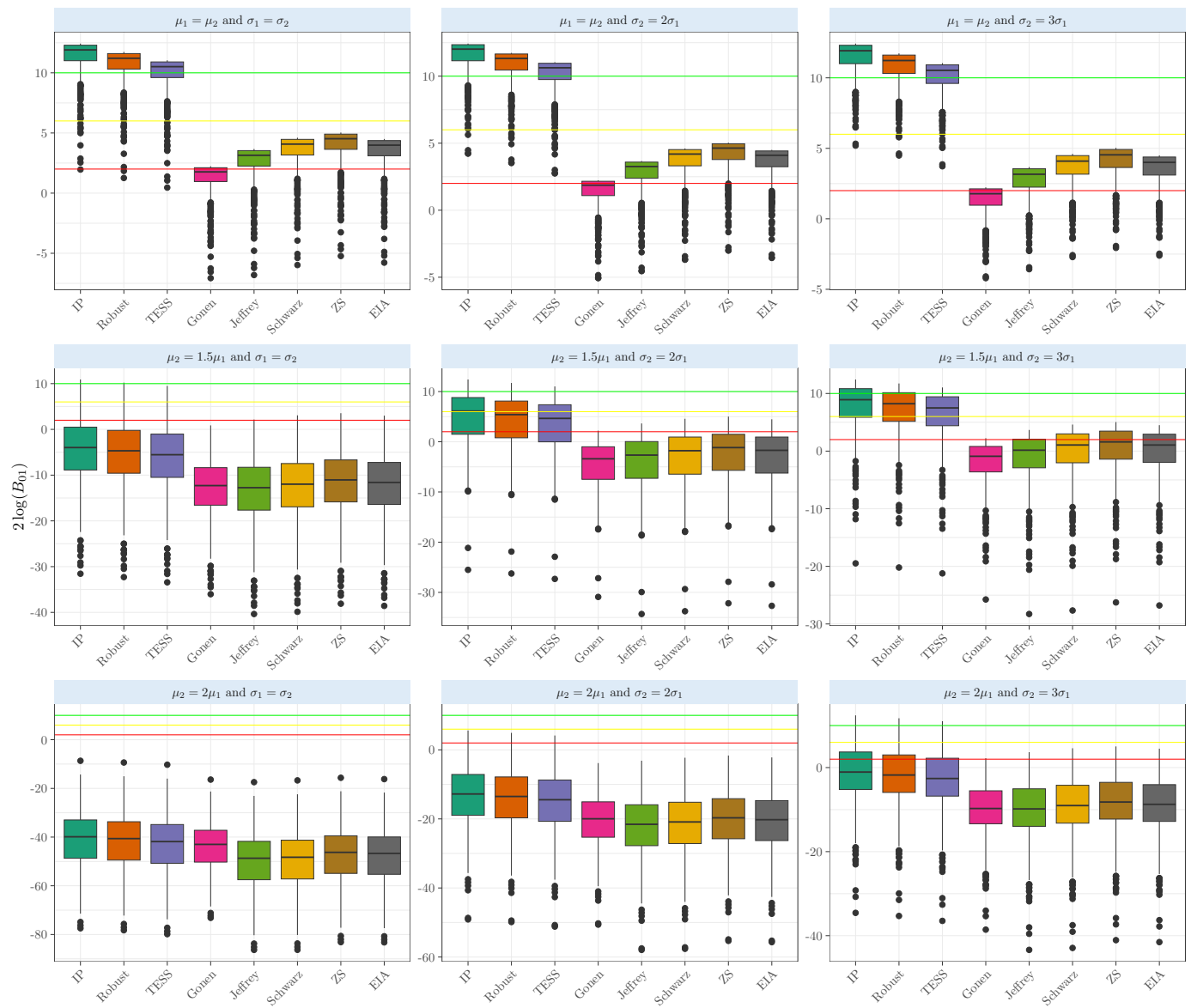
|             |                          |  |  |
|-------------|--------------------------|--|--|
| One Sample  | $B_{01}^{IP} \approx$    | $\sqrt{2n} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} \frac{t^2}{n-1} (1 - e^{-t^2/(n-1)})^{-1}$  | -                                      |
|             | $B_{01}^R =$             | $\sqrt{\frac{2}{n+1}} \frac{n-2}{n-1} t^2 \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} \left(1 - \left(1 + \frac{2t^2}{n^2-1}\right)^{-\frac{n-2}{2}}\right)^{-1}$                              | -                                      |
| Two Samples | $B_{01}^{IP} \approx$    | $\frac{\sqrt{n_\delta} n t^2}{n-2} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2} \left(1 + \coth\left(\frac{n_\delta n t^2}{n-2}\right)\right)$  | $\sqrt{n_1}$                           |
|             | $B_{01}^R =$             | $\sqrt{\frac{8n_\delta}{b+d}} \left(\frac{t^2(n-3)}{4(n-2)}\right) \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2} \left(1 - \left(1 + \frac{t^2 n_\delta}{2(n-2)(b+d)}\right)^{-(n-3)/2}\right)^{-1}$ | $\sqrt{2(1+n_1^2)}$                    |
|             | $B_{01}^{TESS} =$        | $\sqrt{n_\delta} \left(1 + \frac{t^2}{n-2}\right)^{-n/2}$  | $\sqrt{n_1}$                           |
|             | $B_{01}^J \approx$       | $\frac{2}{5} \sqrt{\frac{\pi n_\delta}{2}} \left(1 + \frac{t^2}{n-2}\right)^{-(n-1)/2}$  | $\frac{2}{5} \sqrt{\frac{\pi n_1}{2}}$ |
|             | $B_{01}^{C*} =$          | $\left[\frac{1+t^2/(n-2)}{1+t^2/((n-2)(1+\sigma_\delta^2 n_\delta))}\right]^{-(n-1)/2} \sqrt{1+\sigma_\delta^2 n_\delta}$  | $\sqrt{1+n_1 \sigma_\delta^2}$         |
|             | $B_{01}^S \approx$       | $\sqrt{n} \left(1 + \frac{t^2}{n-2}\right)^{-n/2}$   | $\infty$                               |
|             | $B_{01}^{ZS} \approx$    | $\sqrt{\frac{\pi(n-2)}{2}} \left(1 + \frac{t^2}{n-2}\right)^{-(n-3)/2}$  | $\infty$                               |
|             | $B_{01}^{EIA**} \approx$ | $\pi^{-1} \sqrt{n_\delta/(3\hat{\eta})} S^{(n-1)/2} (S_1^2 + S_2^2)^{-(n-2)/2} \frac{\Gamma((n-2)/2)}{\Gamma((n-1)/2)} \times \int_0^1 (1-x)^{-1/2} e^{-\hat{\delta}/2x} dx$                       | -                                      |

\* Ref. [14] defines  $n_\delta = (1/n_1 + 1/n_2)^{-1}$ . \*\* Ref. [24] defines  $\hat{\eta} = (S_1^2 + S_2^2)/n$  and  $\hat{\delta} = 2/3(\bar{y}_1 - \bar{y}_2)^2/\hat{\eta}$ , and  $S$  is the sum of square of the two samples combined.

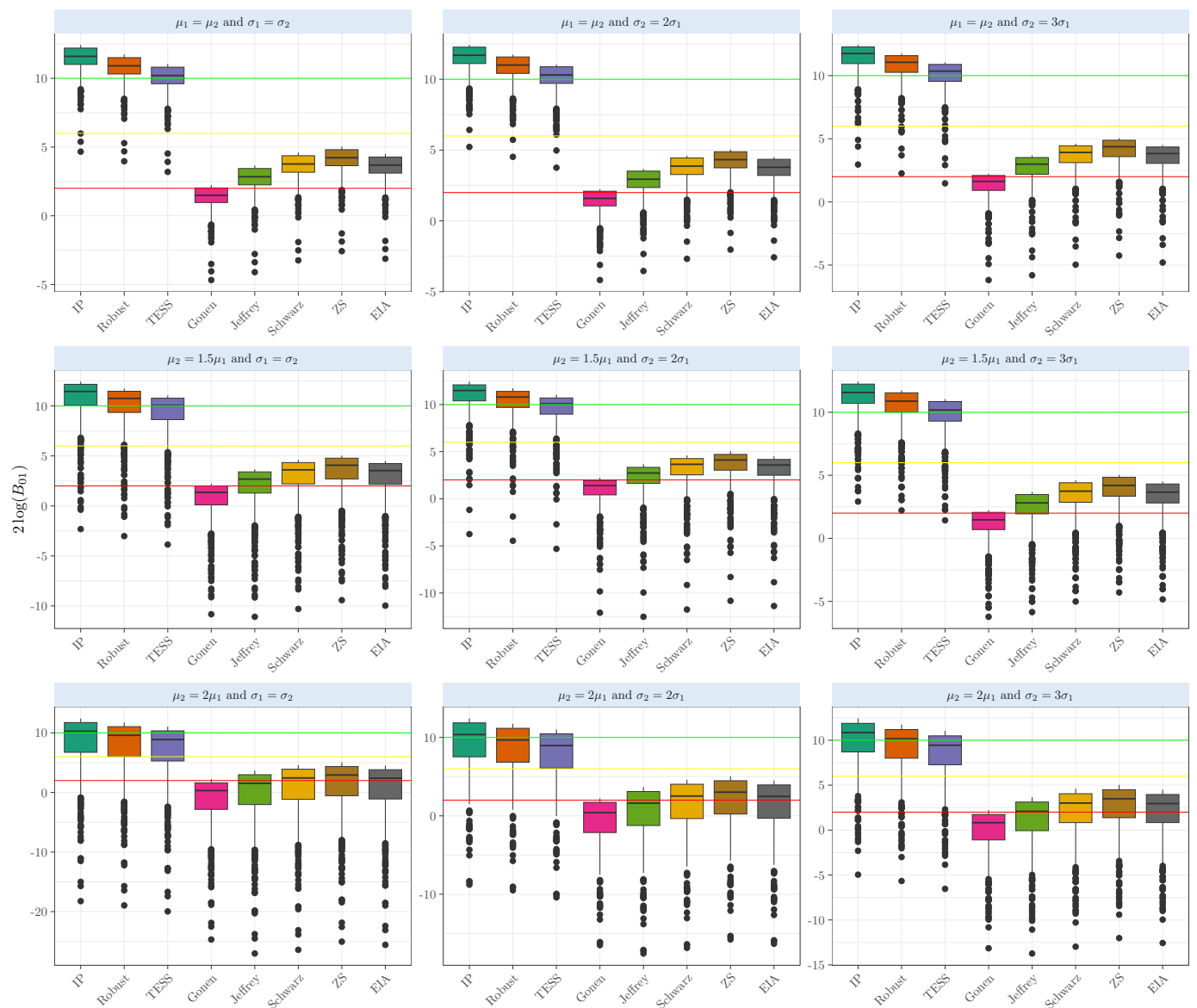
Figure 2 displays the results for the evidence based on the normal distributions when testing whether two-sample means are equal ( $\mu_1 = \mu_2$ ). The red line represents the cut-off for 10 (strong evidence), the yellow line for 6 (positive evidence), and the green for 2 (weak evidence). In the actual case when the means are equal, the Bayes factors based on the intrinsic prior and robust prior show strong evidence in favor of the null hypothesis. The average  $2\log(B_{01})$  based on the intrinsic prior shows strong evidence in favor of the null hypothesis ( $11.3 \pm 1.59$ ), while the Bayes factor based on the robust prior gives strong evidence in favor of the true case ( $10.61 \pm 1.59$ ), all above the red line. BIC-TESS also strongly supports the true case ( $9.9 \pm 1.61$ ). The other Bayes factors provide positive evidence for the true case, with averages ranging from 2.54 to 3.93. Even when the means were equal, and the samples had larger variance ( $\sigma_2 = 3\sigma_1$ ), our objectives and robust Bayes factors provided strong evidence in favor of the true case, with the average above 10. The intrinsic Bayes factor and the robust prior were above 90%, showing either strong or very strong evidence in favor of the null hypothesis when the means were equal; see Table A2 for a detailed comparison.

In the Student's  $t$  random samples, when testing whether two-sample means are equal ( $\mu_1 = \mu_2$ ), we can observe in Figure 3 the results when the means are equal. The Bayes Factors based on the intrinsic prior and robust prior show strong evidence in favor of the null hypothesis. The average  $2\log(B_{01})$  based on the intrinsic prior, robust prior, and BIC-TESS shows strong evidence in favor of the null hypothesis (averages above 10), with values of 11.41, 10.72, and 10.02; dispersion was relatively low, ranging from 1.01 to 1.03. The competing Bayes factors provide slightly positive evidence for the true case, with averages ranging from 2.65 to 4.04. It is interesting to see that in the case of  $\mu_2 = 2\mu_1$ , our Bayes factor gave positive evidence above the yellow line but was very variable; the sample standard deviation ranged from 4.98 to 5.03. Finally, in the case of gamma samples, our Bayes factors

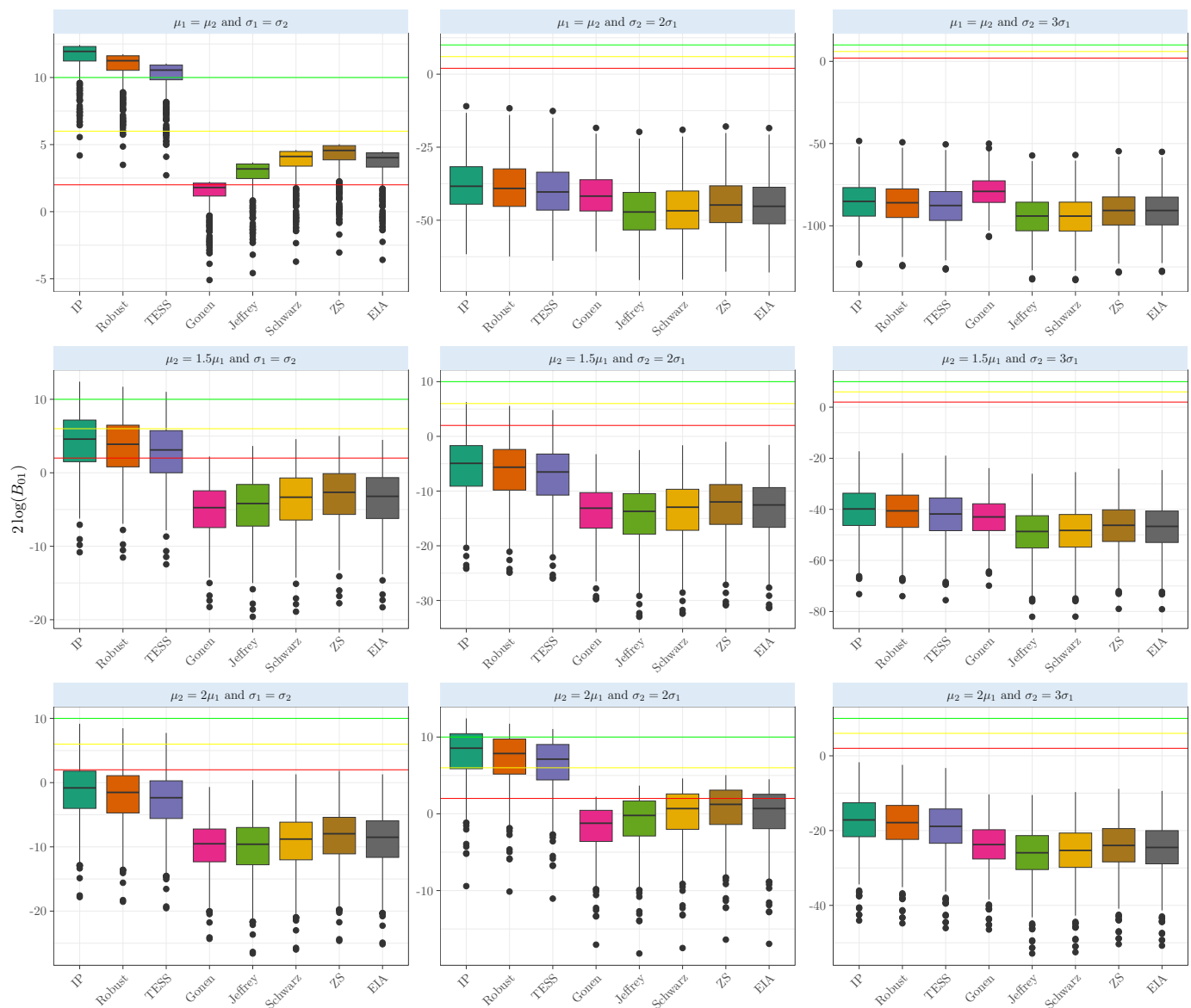
gave strong evidence only when the means and the variances were equal. Departing from any of these conditions gave strong evidence that the means were unequal; see Figure 4. For more details about the simulation results' numerical performance, see Table A1.



**Figure 2.** Evidence in the  $2 \log(B_{01})$  scale when comparing the population means of two samples that arise from a normal distribution with several means and variances with equal sizes  $n_1 = n_2 = 50$ .



**Figure 3.** Evidence in the  $2 \log(B_{01})$  scale when comparing the population means of two samples that arise from a Student's  $t$  with one degree of freedom with several means and variances with equal sizes  $n_1 = n_2 = 50$ .



**Figure 4.** Evidence in the  $2\log(B_{01})$  scale when comparing the population means of two samples that arise from a gamma distribution with several shapes and scales with equal sizes  $n_1 = n_2 = 50$ .

#### 4. Application in Real Dataset

In this section, we applied the proposed one and two Bayes factors based on the intrinsic, Berger, and robust priors, and BIC-TESS based on the Student's  $t$  statistic.

##### 4.1. Gosset Original Dataset

We first consider the century-long original Student's  $t$  sleep data from [1,25] that still raise interesting discussion; see [26,27]. In this study, the number of hours of sleep under both drugs (Dextro and Laevo) was recorded for each patient. The difference in hours was recorded to determine effectiveness, and the average number of hours of sleep gained by using each drug (Dextro and Laevo) was measured. The authors concluded that, in usual doses, Laevo was soporific, but Dextro was not. This analysis is treated as a paired sample, since it compares the sleep hours between treatments. Paired samples lead us to the one sample. The hypothesis of interest is  $H_0 : \mu_d = 0$  versus  $H_1 : \mu_d \neq 0$ ; the test statistic is  $t = -4.06$  with a  $p$ -value of 0.002. At the 5% significance level, we can conclude that there is a difference in the average sleep hours between Laevo and Dextro.

However, the Gosset original dataset has not been addressed using an objective and robust Bayesian approach. The value of the test statistics is the same as before, with  $n = 10$ . The  $(2 \log(B_{10}))$  was computed for the intrinsic and robust Bayes factor, along with the associated posterior probabilities ( $\mathbb{P}(H_1|data)$ ). The  $2 \log(B_{10}^{IP}) = 5.858$  and  $2 \log(B_{10}^R) = 5.988$  are positive, indicating strong evidence that the average sleep hours are different. Further, the posterior probability based on the intrinsic prior is 0.949, and the posterior probability based on the Berger robust prior is 0.952. Both posterior probabilities are above 90%, suggesting strong evidence favoring the average sleep difference.

This dataset is considered as a paired sample, since the recorded number of sleep hours belongs to the same participant. However, the treatments, Dextro and Laevo, might need to be considered independently. If these are considered independently, then a two-sample framework arises. We are interested in determining the sleep hours when receiving Laevo versus when receiving Dextro. Assuming equal variances between Laevo and Dextro, the hypothesis of interest is  $H_0 : \mu_L = \mu_D$  versus  $H_1 : \mu_L \neq \mu_D$ , where  $\mu_L$  is the average sleep hours when receiving Laevo and  $\mu_D$  is the average sleep hours when receiving Dextro. The two-sample test statistic is  $t = -1.86$  with a  $p$ -value of 0.079. At the 5% significance level, we can conclude that there is no difference in the average sleep hours when using Laevo versus Dextro. In our Bayesian approach,  $2 \log(B_{10}^{IP}) = -4.33$  and  $2 \log(B_{10}^R) = -3.57$ , indicating weak evidence that the average number of sleep hours differs between Laevo and Dextro. Both posterior probabilities are above 15%, suggesting weak evidence that the average number of sleep hours differs when using Laevo and Dextro.

#### 4.2. Induced Hypertension on Mice According to Diet

Our first application consists of the data from [28], but they were analyzed in a Bayesian framework using intrinsic priors by [24]. In this study, the researchers were interested in how intermittent feeding affected the blood pressure of rats. The treatment group consisted of eight rats fed intermittently for weeks, and at the final period, the rats' blood pressure measurements were taken. The blood pressure measurements of a second group of seven rats fed the usual way were defined as a control group. The hypothesis of interest is that the average blood pressure is different when the rats have intermittent fasting compared to those with their usual diet, i.e.,  $H_0 : \mu_1 = \mu_2$  versus  $H_1 : \mu_1 \neq \mu_2$ . At the 5% significance level, with a  $p$ -value = 0.044, one can conclude that there exists a difference in the mean blood pressure level according to their feeding style.

The study in Ref. [24] computed the expected arithmetic Bayes factor that favors the alternative hypothesis  $B_{10}^{EAI} = 2.035$  with  $\mathbb{P}(H_1|\mathbf{x}, \mathbf{y}) = 0.671$ , providing support that the average blood pressure measurements differ based on diet. Notably, the Bayes factors based on the intrinsic priors and robust priors yield negative values,  $2 \log(B_{10}^{IP}) = -2.412$  and  $2 \log(B_{10}^R) = -1.414$ , respectively, indicating evidence against  $H_1$ . However, the corresponding posterior probabilities ( $\mathbb{P}(H_1|\mathbf{x}, \mathbf{y})$ ) are 0.23 and 0.33, suggesting weak evidence for the alternative hypothesis that the means are different. The corrected BICTESS suggests weaker evidence against  $H_1$  with a  $2 \log(B_{10}^{TESS}) = -0.3634$  and a posterior probability of  $\mathbb{P}(H_1|\mathbf{x}, \mathbf{y}) = 0.455$ . In contrast, the conjugate  $2 \log(B_{10}^C) = 1.517$  and  $2 \log(B_{10}^{EAI}) = 1.421$ , indicating very weak evidence in favor of  $H_1$ . The associated posterior probability is 0.681.

The extreme observation in the intermittent group (115) was removed. The  $2 \log(B_{10}^{IP*}) = 2.347$ , suggesting evidence in favor of  $H_1$ , while the posterior probability of  $\mathbb{P}(H_1|\mathbf{x}^*, \mathbf{y}^*) = 0.764$  indicates a moderate level of confidence in this conclusion. The Bayes factor constructed with the Berger robust prior exhibits a higher  $2 \log(B_{10}^{R*}) = 3.535$ , along with the posterior probability of  $\mathbb{P}(H_1|\mathbf{x}^*, \mathbf{y}^*) = 0.854$ , indicating stronger support that the average of blood pressure differs by type of fasting. TESS models present even higher  $2 \log(B_{10}^{TESS*}) = 5.041$ , respectively, with corresponding posterior probabilities of 0.926, indicating substantial evidence for  $H_1$ .



## 5. Discussion

In this work, we proposed the objective and robust Bayes factors for the one-sample and two-sample comparisons. These newly proposed Bayes factors are finite sample-consistent. Both the exact and approximate forms of the Bayes factors can be easily implemented using any open-source or commercial software. Another advantage of using Bayes factors is that the posterior probabilities of the hypothesis test are easily interpretable. We reanalyzed the original study by [1] and the comparison of blood pressure in rats according to different feeding types. Our objective and robust Bayes factors showed strong evidence that the average number of hours differed between Laevo and Dextro in the mouse application. When removing potential extreme values, we concluded that there is strong evidence that the means differed. However, we reported weak evidence with the complete dataset that these averages differed according to their diet. This might occur, since the assumption of equal variances might not hold. Even though the samples might have equal means, departing from the assumption of equal variances can lead in favor of the wrong hypothesis. Although we have made a significant contribution, an aspect that might alleviate this issue is deriving an objective and robust Bayes factor for the Behrens–Fisher problem, i.e., unequal variances for both groups. Also, the Bayes factor based on the intrinsic prior depends on the maximum likelihood estimate (MLE); perhaps robust estimates can be considered, although a modified test statistic might arise. Another possible extension is to develop an objective Bayes factor for the hypothesis of several equal means; this will be an analysis of variance (ANOVA) approach in the frequentist approach.

**Author Contributions:** Conceptualization, I.A.A.-R. and L.R.P.-G.; methodology, I.A.A.-R. and L.R.P.-G.; software, I.A.A.-R.; validation, I.A.A.-R. and L.R.P.-G.; formal analysis, I.A.A.-R. and L.R.P.-G.; writing—original draft preparation, I.A.A.-R. and L.R.P.-G. All authors have read and agreed with the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available in [1,24].

**Acknowledgments:** The second author thanks Jim Berger for the fundamental discussions about the subject of this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|                 |  |
|-----------------|--|
| TESS            | The effective sample size  |
| BIC             | Bayesian Information Criterion   |
| $B_{01}^{IP}$   | Bayes Factor based on the intrinsic prior for measuring $H_0$ vs. $H_1$        |
| $B_{01}^R$      | Bayes Factor based on the Berger’s Robust prior for measuring $H_0$ vs. $H_1$  |
| $B_{01}^{TESS}$ | Corrected BIC using the effective sample size for measuring $H_0$ vs. $H_1$    |
| $B_{01}^J$      | Bayes Factor based on Jeffreys’s prior for measuring $H_0$ vs. $H_1$           |
| $B_{01}^C$      | Bayes Factor based on the conjugate prior for measuring $H_0$ vs. $H_1$        |
| $B_{01}^{ZS}$   | Bayes Factor based on the Zellner and Siow prior for measuring $H_0$ vs. $H_1$ |
| $B_{01}^{EIA}$  | Bayes Factor expected arithmetic intrinsic prior for measuring $H_0$ vs. $H_1$ |

## Appendix A

### Appendix A.1. Calculation of the Effective Sample Size

We suggest using the effective sample size of [21] to define the robust Bayes factor and BIC-Tess. The effective sample size for the parameter  $\alpha$ ,  $n_o^e = C^{-1}(X^t \Gamma^{-1} X)C^{-1}$ . Let  $X$  be the design matrix:

$$X = \left[ \mathbf{1}_{n \times 1} \mid \begin{bmatrix} 1/n_1 \mathbf{1}_{n_1 \times 1} \\ -1/n_2 \mathbf{1}_{n_2 \times 1} \end{bmatrix} \right];$$

where  $n = n_1 + n_2$ . Let  $\mathbf{X}_1$  be the second column of the design matrix  $\mathbf{X}_1 = \begin{bmatrix} 1/n_1 \mathbf{1}_{n_1 \times 1} \\ -1/n_2 \mathbf{1}_{n_2 \times 1} \end{bmatrix}$ , and let  $C_1 = \max_j \left\{ \left| \frac{1}{n_1 \sigma} \right|, \left| -\frac{1}{n_2 \sigma} \right| \right\}$ . Further, let  $\Gamma^{-1} = \sigma^{-2} \mathbf{I}_{n \times n}$ , where  $\mathbf{I}_{n \times n}$  is an identity matrix of size  $n$ . It follows from the definition of the effective sample size for the original  $\alpha$  that

$$\begin{aligned} n_o^e &= C_1^{-1} (\mathbf{X}_1^t \Gamma^{-1} \mathbf{X}_1) C_1^{-1} \\ &= C_1^{-2} \left[ \frac{\sigma^{-2}}{n_1}, \dots, \frac{\sigma^{-2}}{n_1}, -\frac{\sigma^{-2}}{n_2}, \dots, -\frac{\sigma^{-2}}{n_2} \right] \mathbf{X} \\ &= \frac{1}{C_1^2} \left[ \frac{\sigma^{-2}}{n_1^2} n_1 + \frac{\sigma^{-2}}{n_2^2} n_2 \right] \\ &= \sigma^{-2} \cdot \left( \max \left\{ \left| \frac{1}{n_1 \sigma} \right|, \left| -\frac{1}{n_2 \sigma} \right| \right\} \right)^{-2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \\ &= \sigma^{-2} \cdot \left( \max \left\{ \left| \frac{1}{n_1 \sigma} \right|, \left| -\frac{1}{n_2 \sigma} \right| \right\} \right)^{-2} n_\delta \\ &= \left( \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\} \right)^{-2} n_\delta. \end{aligned} \quad (A1)$$

Since  $\sigma > 0$ ,  $\max \left\{ \left| \frac{1}{n_1 \sigma} \right|, \left| -\frac{1}{n_2 \sigma} \right| \right\} = \frac{1}{\sigma} \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\}$ . Then, the final expression of the effective sample size is obtained. The definition of the unit information is  $d = 0.25 n_\delta \cdot \sigma^2$  and for  $b$ ,

$$b = n_o^e \cdot d = 0.25 \sigma^2 \cdot \left( \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\} \right)^{-2}. \quad (A2)$$

The factors  $b$  and  $d$  will be defined as

$$b + d = 0.25 \sigma^2 \left[ n_\delta + \left( \max \left\{ \left| \frac{1}{n_1} \right|, \left| -\frac{1}{n_2} \right| \right\} \right)^{-2} \right] = 0.25 \sigma^2 (b^* + d^*). \quad (A3)$$

This last calculation defines the factors  $b$  and  $d$  for the Robust Bayesian Student's  $t$  test,  $B_{01}^R$ .

#### Appendix A.2. Simulation Experiments

In the two-sample framework, we generated 500 datasets from a random sample from a normal distribution with parameters  $\mu_1 = 5, \sigma_1 = 3$ ; the second group was created with a combination of several parameters for the location. The mean values were  $\mu_2 \in \{\mu_1, 1.5\mu_1, 2\mu_1\}$  for the standard deviation  $\sigma_2 \in \{\sigma_1, 2\sigma_1, 3\sigma_1\}$ . In the case of the Student's  $t$  with  $\nu = 1$  degrees of freedom, the gamma distribution was simulated using the methods of moments for the shape parameter  $\alpha_i = \mu_i^2 / \sigma_i^2$  and scale parameter  $\beta_i = \sigma_i^2 / \mu_i$  for  $i = 1, 2$ . We reported the twice natural base logarithm Bayes factors ( $2 \log(B_{01})$ ) for comparing the null hypothesis ( $\mu_1 = \mu_2$ ) against the alternative ( $\mu_1 \neq \mu_2$ ). This transformation allows the interpretation to be on the same scale as deviance and likelihood ratio test statistics; see Ref. [13] for a deeper discussion.

**Table A1.** Summary statistics (means  $\pm$  standard deviation) for the balanced sample ( $n_1 = n_2 = 50$ ) of the  $2\log(B_{01})$  simulation experiments for the intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].

|                 |           | $\mu_1 = \mu_2$       |                        |                        | $\mu_2 = 1.5\mu_1$    |                        |                        | $\mu_2 = 2\mu_1$      |                        |                        |
|-----------------|-----------|-----------------------|------------------------|------------------------|-----------------------|------------------------|------------------------|-----------------------|------------------------|------------------------|
|                 |           | $\sigma_1 = \sigma_2$ | $\sigma_2 = 2\sigma_1$ | $\sigma_2 = 3\sigma_1$ | $\sigma_1 = \sigma_2$ | $\sigma_2 = 2\sigma_1$ | $\sigma_2 = 3\sigma_1$ | $\sigma_1 = \sigma_2$ | $\sigma_2 = 2\sigma_1$ | $\sigma_2 = 3\sigma_1$ |
| Normal          | IP        | 11.3 $\pm$ 1.59       | 11.44 $\pm$ 1.37       | 11.42 $\pm$ 1.24       | −4.85 $\pm$ 7.44      | 4.85 $\pm$ 5.29        | 7.85 $\pm$ 4.18        | −41.19 $\pm$ 11.66    | −13.18 $\pm$ 8.77      | −1.53 $\pm$ 7.19       |
|                 | Robust    | 10.61 $\pm$ 1.59      | 10.75 $\pm$ 1.38       | 10.73 $\pm$ 1.25       | −5.55 $\pm$ 7.45      | 4.15 $\pm$ 5.29        | 7.15 $\pm$ 4.18        | −41.94 $\pm$ 11.68    | −13.9 $\pm$ 8.78       | −2.23 $\pm$ 7.19       |
|                 | TESS      | 9.9 $\pm$ 1.61        | 10.04 $\pm$ 1.39       | 10.03 $\pm$ 1.26       | −6.42 $\pm$ 7.53      | 3.38 $\pm$ 5.35        | 6.41 $\pm$ 4.22        | −43.18 $\pm$ 11.8     | −14.85 $\pm$ 8.87      | −3.07 $\pm$ 7.27       |
|                 | Conjugate | 1.23 $\pm$ 1.42       | 1.35 $\pm$ 1.23        | 1.34 $\pm$ 1.11        | −13.02 $\pm$ 6.5      | −4.5 $\pm$ 4.67        | −1.84 $\pm$ 3.7        | −43.99 $\pm$ 9.64     | −20.27 $\pm$ 7.57      | −10.11 $\pm$ 6.29      |
|                 | Jeffrey's | 2.54 $\pm$ 1.59       | 2.68 $\pm$ 1.38        | 2.66 $\pm$ 1.25        | −13.62 $\pm$ 7.45     | −3.92 $\pm$ 5.29       | −0.91 $\pm$ 4.18       | −50.01 $\pm$ 11.68    | −21.97 $\pm$ 8.78      | −10.3 $\pm$ 7.19       |
|                 | Schwarz's | 3.47 $\pm$ 1.61       | 3.61 $\pm$ 1.39        | 3.59 $\pm$ 1.26        | −12.86 $\pm$ 7.53     | −3.06 $\pm$ 5.35       | −0.02 $\pm$ 4.22       | −49.62 $\pm$ 11.8     | −21.29 $\pm$ 8.87      | −9.5 $\pm$ 7.27        |
|                 | ZS        | 3.93 $\pm$ 1.56       | 4.07 $\pm$ 1.35        | 4.05 $\pm$ 1.22        | −11.9 $\pm$ 7.3       | −2.4 $\pm$ 5.19        | 0.55 $\pm$ 4.1         | −47.56 $\pm$ 11.44    | −20.08 $\pm$ 8.61      | −8.65 $\pm$ 7.05       |
|                 | EIA       | 3.4 $\pm$ 1.57        | 3.53 $\pm$ 1.35        | 3.51 $\pm$ 1.22        | −12.45 $\pm$ 7.3      | −2.94 $\pm$ 5.19       | 0 $\pm$ 4.1            | −47.97 $\pm$ 11.36    | −20.62 $\pm$ 8.59      | −9.2 $\pm$ 7.04        |
| Student- $t(1)$ | IP        | 11.41 $\pm$ 1.01      | 11.46 $\pm$ 1.04       | 11.4 $\pm$ 1.22        | 10.55 $\pm$ 2.42      | 10.83 $\pm$ 2.01       | 11.09 $\pm$ 1.61       | 8.32 $\pm$ 4.98       | 9.05 $\pm$ 3.76        | 9.79 $\pm$ 2.97        |
|                 | Robust    | 10.72 $\pm$ 1.01      | 10.77 $\pm$ 1.04       | 10.7 $\pm$ 1.22        | 9.85 $\pm$ 2.42       | 10.13 $\pm$ 2.02       | 10.39 $\pm$ 1.62       | 7.62 $\pm$ 4.98       | 8.36 $\pm$ 3.77        | 9.09 $\pm$ 2.97        |
|                 | TESS      | 10.02 $\pm$ 1.03      | 10.06 $\pm$ 1.05       | 10 $\pm$ 1.23          | 9.14 $\pm$ 2.44       | 9.42 $\pm$ 2.04        | 9.69 $\pm$ 1.63        | 6.89 $\pm$ 5.03       | 7.63 $\pm$ 3.81        | 8.37 $\pm$ 3           |
|                 | Conjugate | 1.33 $\pm$ 0.9        | 1.37 $\pm$ 0.93        | 1.31 $\pm$ 1.08        | 0.56 $\pm$ 2.15       | 0.81 $\pm$ 1.79        | 1.04 $\pm$ 1.44        | −1.42 $\pm$ 4.4       | −0.77 $\pm$ 3.34       | −0.12 $\pm$ 2.63       |
|                 | Jeffrey's | 2.65 $\pm$ 1.01       | 2.7 $\pm$ 1.04         | 2.64 $\pm$ 1.22        | 1.79 $\pm$ 2.42       | 2.07 $\pm$ 2.02        | 2.33 $\pm$ 1.62        | −0.44 $\pm$ 4.98      | 0.29 $\pm$ 3.77        | 1.03 $\pm$ 2.97        |
|                 | Schwarz's | 3.58 $\pm$ 1.03       | 3.63 $\pm$ 1.05        | 3.56 $\pm$ 1.23        | 2.7 $\pm$ 2.44        | 2.99 $\pm$ 2.04        | 3.25 $\pm$ 1.63        | 0.45 $\pm$ 5.03       | 1.19 $\pm$ 3.81        | 1.94 $\pm$ 3           |
|                 | ZS        | 4.04 $\pm$ 0.99       | 4.09 $\pm$ 1.02        | 4.03 $\pm$ 1.19        | 3.19 $\pm$ 2.37       | 3.47 $\pm$ 1.98        | 3.72 $\pm$ 1.58        | 1.01 $\pm$ 4.88       | 1.73 $\pm$ 3.69        | 2.45 $\pm$ 2.91        |
|                 | EIA       | 3.5 $\pm$ 1           | 3.55 $\pm$ 1.02        | 3.49 $\pm$ 1.19        | 2.65 $\pm$ 2.37       | 2.93 $\pm$ 1.98        | 3.18 $\pm$ 1.59        | 0.47 $\pm$ 4.88       | 1.19 $\pm$ 3.7         | 1.91 $\pm$ 2.91        |
| Gamma           | IP        | 11.48 $\pm$ 1.26      | −38.13 $\pm$ 9.19      | −85.57 $\pm$ 13.22     | 4.16 $\pm$ 4.16       | −5.59 $\pm$ 5.51       | −40.51 $\pm$ 9.69      | −1.33 $\pm$ 4.72      | 7.9 $\pm$ 3.41         | −17.76 $\pm$ 7.26      |
|                 | Robust    | 10.79 $\pm$ 1.26      | −38.87 $\pm$ 9.21      | −86.4 $\pm$ 13.25      | 3.46 $\pm$ 4.16       | −6.3 $\pm$ 5.52        | −41.26 $\pm$ 9.7       | −2.04 $\pm$ 4.73      | 7.2 $\pm$ 3.41         | −18.48 $\pm$ 7.27      |
|                 | TESS      | 10.08 $\pm$ 1.27      | −40.08 $\pm$ 9.3       | −88.08 $\pm$ 13.38     | 2.68 $\pm$ 4.21       | −7.18 $\pm$ 5.57       | −42.49 $\pm$ 9.8       | −2.87 $\pm$ 4.77      | 6.46 $\pm$ 3.44        | −19.48 $\pm$ 7.34      |
|                 | Conjugate | 1.39 $\pm$ 1.12       | −41.48 $\pm$ 7.68      | −79.22 $\pm$ 9.99      | −5.11 $\pm$ 3.68      | −13.69 $\pm$ 4.81      | −43.46 $\pm$ 8.02      | −9.96 $\pm$ 4.15      | −1.8 $\pm$ 3.02        | −24.24 $\pm$ 6.23      |
|                 | Jeffrey's | 2.72 $\pm$ 1.26       | −46.94 $\pm$ 9.21      | −94.46 $\pm$ 13.25     | −4.6 $\pm$ 4.16       | −14.37 $\pm$ 5.52      | −49.33 $\pm$ 9.7       | −10.1 $\pm$ 4.73      | −0.86 $\pm$ 3.41       | −26.55 $\pm$ 7.27      |
|                 | Schwarz   | 3.64 $\pm$ 1.27       | −46.52 $\pm$ 9.3       | −94.52 $\pm$ 13.38     | −3.75 $\pm$ 4.21      | −13.61 $\pm$ 5.57      | −48.93 $\pm$ 9.8       | −9.31 $\pm$ 4.77      | 0.02 $\pm$ 3.44        | −25.92 $\pm$ 7.34      |
|                 | ZS        | 4.1 $\pm$ 1.23        | −44.55 $\pm$ 9.02      | −91.12 $\pm$ 12.98     | −3.07 $\pm$ 4.08      | −12.64 $\pm$ 5.4       | −46.89 $\pm$ 9.51      | −8.46 $\pm$ 4.63      | 0.59 $\pm$ 3.34        | −24.57 $\pm$ 7.12      |
|                 | EIA       | 3.57 $\pm$ 1.24       | −44.99 $\pm$ 8.97      | −91.11 $\pm$ 12.83     | −3.62 $\pm$ 4.08      | −13.18 $\pm$ 5.4       | −47.32 $\pm$ 9.44      | −9.01 $\pm$ 4.63      | 0.05 $\pm$ 3.35        | −25.1 $\pm$ 7.1        |

**Table A2.** Frequency distribution for the balanced framework ( $n_1 = n_2 = 50$ ) in the normal random variables of the  $2 \log(B_{01})$  based on the scale proposed by [13]. Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].

| $\mu_1 = \mu_2$       |             |             |             |             |             |                        |             |             |             |             |                        |             |             |             |             |
|-----------------------|-------------|-------------|-------------|-------------|-------------|------------------------|-------------|-------------|-------------|-------------|------------------------|-------------|-------------|-------------|-------------|
| $\sigma_1 = \sigma_2$ |             |             |             |             |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |             |             |             |             |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    |
| IP                    | 434 (86.8%) | 56 (11.2%)  | 9 (1.8%)    | 1 (0.2%)    | 0 (0%)      | 439 (87.8%)            | 57 (11.4%)  | 4 (0.8%)    | 0 (0%)      | 0 (0%)      | 438 (87.6%)            | 60 (12%)    | 2 (0.4%)    | 0 (0%)      | 0 (0%)      |
| Robust                | 400 (80%)   | 85 (17%)    | 13 (2.6%)   | 2 (0.4%)    | 0 (0%)      | 407 (81.4%)            | 85 (17%)    | 8 (1.6%)    | 0 (0%)      | 0 (0%)      | 402 (80.4%)            | 93 (18.6%)  | 5 (1%)      | 0 (0%)      | 0 (0%)      |
| TESS                  | 339 (67.8%) | 141 (28.2%) | 17 (3.4%)   | 3 (0.6%)    | 0 (0%)      | 344 (68.8%)            | 144 (28.8%) | 12 (2.4%)   | 0 (0%)      | 0 (0%)      | 329 (65.8%)            | 164 (32.8%) | 7 (1.4%)    | 0 (0%)      | 0 (0%)      |
| Conjugate             | 0 (0%)      | 0 (0%)      | 171 (34.2%) | 264 (52.8%) | 65 (13%)    | 0 (0%)                 | 0 (0%)      | 200 (40%)   | 241 (48.2%) | 59 (11.8%)  | 0 (0%)                 | 0 (0%)      | 182 (36.4%) | 261 (52.2%) | 57 (11.4%)  |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 392 (78.4%) | 73 (14.6%)  | 35 (7%)     | 0 (0%)                 | 0 (0%)      | 402 (80.4%) | 68 (13.6%)  | 30 (6%)     | 0 (0%)                 | 0 (0%)      | 396 (79.2%) | 78 (15.6%)  | 26 (5.2%)   |
| Schwarz               | 0 (0%)      | 0 (0%)      | 437 (87.4%) | 40 (8%)     | 23 (4.6%)   | 0 (0%)                 | 0 (0%)      | 444 (88.8%) | 39 (7.8%)   | 17 (3.4%)   | 0 (0%)                 | 0 (0%)      | 445 (89%)   | 45 (9%)     | 10 (2%)     |
| ZS                    | 0 (0%)      | 0 (0%)      | 452 (90.4%) | 30 (6%)     | 18 (3.6%)   | 0 (0%)                 | 0 (0%)      | 458 (91.6%) | 31 (6.2%)   | 11 (2.2%)   | 0 (0%)                 | 0 (0%)      | 459 (91.8%) | 34 (6.8%)   | 7 (1.4%)    |
| EIA                   | 0 (0%)      | 0 (0%)      | 435 (87%)   | 42 (8.4%)   | 23 (4.6%)   | 0 (0%)                 | 0 (0%)      | 444 (88.8%) | 39 (7.8%)   | 17 (3.4%)   | 0 (0%)                 | 0 (0%)      | 444 (88.8%) | 46 (9.2%)   | 10 (2%)     |
| $\mu_2 = 1.5\mu_1$    |             |             |             |             |             |                        |             |             |             |             |                        |             |             |             |             |
| $\sigma_1 = \sigma_2$ |             |             |             |             |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |             |             |             |             |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    |
| IP                    | 2 (0.4%)    | 19 (3.8%)   | 67 (13.4%)  | 48 (9.6%)   | 364 (72.8%) | 69 (13.8%)             | 184 (36.8%) | 110 (22%)   | 54 (10.8%)  | 83 (16.6%)  | 176 (35.2%)            | 195 (39%)   | 87 (17.4%)  | 21 (4.2%)   | 21 (4.2%)   |
| Robust                | 1 (0.2%)    | 12 (2.4%)   | 60 (12%)    | 45 (9%)     | 382 (76.4%) | 50 (10%)               | 176 (35.2%) | 122 (24.4%) | 52 (10.4%)  | 100 (20%)   | 137 (27.4%)            | 214 (42.8%) | 97 (19.4%)  | 23 (4.6%)   | 29 (5.8%)   |
| TESS                  | 0 (0%)      | 11 (2.2%)   | 45 (9%)     | 45 (9%)     | 399 (79.8%) | 24 (4.8%)              | 167 (33.4%) | 137 (27.4%) | 46 (9.2%)   | 126 (25.2%) | 76 (15.2%)             | 248 (49.6%) | 116 (23.2%) | 26 (5.2%)   | 34 (6.8%)   |
| Conjugate             | 0 (0%)      | 0 (0%)      | 0 (0%)      | 2 (0.4%)    | 498 (99.6%) | 0 (0%)                 | 0 (0%)      | 8 (1.6%)    | 63 (12.6%)  | 429 (85.8%) | 0 (0%)                 | 0 (0%)      | 36 (7.2%)   | 147 (29.4%) | 317 (63.4%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 1 (0.2%)    | 5 (1%)      | 494 (98.8%) | 0 (0%)                 | 0 (0%)      | 50 (10%)    | 77 (15.4%)  | 373 (74.6%) | 0 (0%)                 | 0 (0%)      | 132 (26.4%) | 128 (25.6%) | 240 (48%)   |
| Schwarz               | 0 (0%)      | 0 (0%)      | 2 (0.4%)    | 9 (1.8%)    | 489 (97.8%) | 0 (0%)                 | 0 (0%)      | 77 (15.4%)  | 97 (19.4%)  | 326 (65.2%) | 0 (0%)                 | 0 (0%)      | 186 (37.2%) | 117 (23.4%) | 197 (39.4%) |
| ZS                    | 0 (0%)      | 0 (0%)      | 3 (0.6%)    | 9 (1.8%)    | 488 (97.6%) | 0 (0%)                 | 0 (0%)      | 101 (20.2%) | 96 (19.2%)  | 303 (60.6%) | 0 (0%)                 | 0 (0%)      | 228 (45.6%) | 99 (19.8%)  | 173 (34.6%) |
| EIA                   | 0 (0%)      | 0 (0%)      | 2 (0.4%)    | 9 (1.8%)    | 489 (97.8%) | 0 (0%)                 | 0 (0%)      | 74 (14.8%)  | 103 (20.6%) | 323 (64.6%) | 0 (0%)                 | 0 (0%)      | 185 (37%)   | 119 (23.8%) | 196 (39.2%) |
| $\mu_2 = 2\mu_1$      |             |             |             |             |             |                        |             |             |             |             |                        |             |             |             |             |
| $\sigma_1 = \sigma_2$ |             |             |             |             |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |             |             |             |             |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    |
| IP                    | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 11 (2.2%)   | 13 (2.6%)   | 476 (95.2%) | 12 (2.4%)              | 52 (10.4%)  | 109 (21.8%) | 53 (10.6%)  | 274 (54.8%) |
| Robust                | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 10 (2%)     | 9 (1.8%)    | 481 (96.2%) | 5 (1%)                 | 46 (9.2%)   | 102 (20.4%) | 54 (10.8%)  | 293 (58.6%) |
| TESS                  | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 5 (1%)      | 12 (2.4%)   | 483 (96.6%) | 5 (1%)                 | 38 (7.6%)   | 90 (18%)    | 53 (10.6%)  | 314 (62.8%) |
| Conjugate             | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 1 (0.2%)    | 11 (2.2%)   | 488 (97.6%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 5 (1%)      | 19 (3.8%)   | 476 (95.2%) |
| Schwarz               | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 13 (2.6%)   | 21 (4.2%)   | 466 (93.2%) |
| ZS                    | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 18 (3.6%)   | 25 (5%)     | 457 (91.4%) |
| EIA                   | 0 (0%)      | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 0 (0%)      | 0 (0%)      | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 13 (2.6%)   | 21 (4.2%)   | 466 (93.2%) |

**Table A3.** Frequency distribution for the balanced framework ( $n_1 = n_2 = 50$ ) Student's  $t$  random variables of the  $2 \log(B_{01})$  based on the scale proposed by [13]. Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].

| $\mu_1 = \mu_2$       |             |             |             |             |           |                        |             |             |             |           |                        |             |             |             |           |
|-----------------------|-------------|-------------|-------------|-------------|-----------|------------------------|-------------|-------------|-------------|-----------|------------------------|-------------|-------------|-------------|-----------|
| $\sigma_1 = \sigma_2$ |             |             |             |             |           | $\sigma_2 = 2\sigma_1$ |             |             |             |           | $\sigma_2 = 3\sigma_1$ |             |             |             |           |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative  | Very Strong            | Strong      | Positive    | Weak        | Negative  | Very Strong            | Strong      | Positive    | Weak        | Negative  |
| IP                    | 456 (91.2%) | 41 (8.2%)   | 3 (0.6%)    | 0 (0%)      | 0 (0%)    | 455 (91%)              | 44 (8.8%)   | 1 (0.2%)    | 0 (0%)      | 0 (0%)    | 448 (89.6%)            | 49 (9.8%)   | 3 (0.6%)    | 0 (0%)      | 0 (0%)    |
| Robust                | 410 (82%)   | 87 (17.4%)  | 3 (0.6%)    | 0 (0%)      | 0 (0%)    | 416 (83.2%)            | 82 (16.4%)  | 2 (0.4%)    | 0 (0%)      | 0 (0%)    | 410 (82%)              | 83 (16.6%)  | 7 (1.4%)    | 0 (0%)      | 0 (0%)    |
| TESS                  | 310 (62%)   | 187 (37.4%) | 3 (0.6%)    | 0 (0%)      | 0 (0%)    | 320 (64%)              | 178 (35.6%) | 2 (0.4%)    | 0 (0%)      | 0 (0%)    | 310 (62%)              | 181 (36.2%) | 8 (1.6%)    | 1 (0.2%)    | 0 (0%)    |
| Conjugate             | 0 (0%)      | 0 (0%)      | 131 (26.2%) | 326 (65.2%) | 43 (8.6%) | 0 (0%)                 | 0 (0%)      | 156 (31.2%) | 301 (60.2%) | 43 (8.6%) | 0 (0%)                 | 0 (0%)      | 155 (31%)   | 297 (59.4%) | 48 (9.6%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 404 (80.8%) | 85 (17%)    | 11 (2.2%) | 0 (0%)                 | 0 (0%)      | 412 (82.4%) | 74 (14.8%)  | 14 (2.8%) | 0 (0%)                 | 0 (0%)      | 396 (79.2%) | 86 (17.2%)  | 18 (3.6%) |
| Schwarz               | 0 (0%)      | 0 (0%)      | 459 (91.8%) | 37 (7.4%)   | 4 (0.8%)  | 0 (0%)                 | 0 (0%)      | 457 (91.4%) | 39 (7.8%)   | 4 (0.8%)  | 0 (0%)                 | 0 (0%)      | 454 (90.8%) | 36 (7.2%)   | 10 (2%)   |
| ZS                    | 0 (0%)      | 0 (0%)      | 479 (95.8%) | 18 (3.6%)   | 3 (0.6%)  | 0 (0%)                 | 0 (0%)      | 474 (94.8%) | 24 (4.8%)   | 2 (0.4%)  | 0 (0%)                 | 0 (0%)      | 474 (94.8%) | 17 (3.4%)   | 9 (1.8%)  |
| EIA                   | 0 (0%)      | 0 (0%)      | 457 (91.4%) | 39 (7.8%)   | 4 (0.8%)  | 0 (0%)                 | 0 (0%)      | 457 (91.4%) | 39 (7.8%)   | 4 (0.8%)  | 0 (0%)                 | 0 (0%)      | 454 (90.8%) | 36 (7.2%)   | 10 (2%)   |

| $\mu_2 = 1.5\mu_1$    |             |             |             |             |             |                        |             |             |             |             |                        |             |             |             |            |
|-----------------------|-------------|-------------|-------------|-------------|-------------|------------------------|-------------|-------------|-------------|-------------|------------------------|-------------|-------------|-------------|------------|
| $\sigma_1 = \sigma_2$ |             |             |             |             |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |             |             |             |            |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative   |
| IP                    | 378 (75.6%) | 92 (18.4%)  | 22 (4.4%)   | 5 (1%)      | 3 (0.6%)    | 395 (79%)              | 88 (17.6%)  | 14 (2.8%)   | 1 (0.2%)    | 2 (0.4%)    | 421 (84.2%)            | 69 (13.8%)  | 10 (2%)     | 0 (0%)      | 0 (0%)     |
| Robust                | 327 (65.4%) | 132 (26.4%) | 31 (6.2%)   | 5 (1%)      | 5 (1%)      | 347 (69.4%)            | 129 (25.8%) | 19 (3.8%)   | 3 (0.6%)    | 2 (0.4%)    | 375 (75%)              | 111 (22.2%) | 14 (2.8%)   | 0 (0%)      | 0 (0%)     |
| TESS                  | 259 (51.8%) | 187 (37.4%) | 39 (7.8%)   | 9 (1.8%)    | 6 (1.2%)    | 271 (54.2%)            | 199 (39.8%) | 23 (4.6%)   | 4 (0.8%)    | 3 (0.6%)    | 284 (56.8%)            | 191 (38.2%) | 24 (4.8%)   | 1 (0.2%)    | 0 (0%)     |
| Conjugate             | 0 (0%)      | 0 (0%)      | 125 (25%)   | 259 (51.8%) | 116 (23.2%) | 0 (0%)                 | 0 (0%)      | 109 (21.8%) | 289 (57.8%) | 102 (20.4%) | 0 (0%)                 | 0 (0%)      | 139 (27.8%) | 283 (56.6%) | 78 (15.6%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 321 (64.2%) | 102 (20.4%) | 77 (15.4%)  | 0 (0%)                 | 0 (0%)      | 342 (68.4%) | 100 (20%)   | 58 (11.6%)  | 0 (0%)                 | 0 (0%)      | 366 (73.2%) | 90 (18%)    | 44 (8.8%)  |
| Schwarz               | 0 (0%)      | 0 (0%)      | 387 (77.4%) | 53 (10.6%)  | 60 (12%)    | 0 (0%)                 | 0 (0%)      | 403 (80.6%) | 60 (12%)    | 37 (7.4%)   | 0 (0%)                 | 0 (0%)      | 424 (84.8%) | 47 (9.4%)   | 29 (5.8%)  |
| ZS                    | 0 (0%)      | 0 (0%)      | 403 (80.6%) | 46 (9.2%)   | 51 (10.2%)  | 0 (0%)                 | 0 (0%)      | 426 (85.2%) | 44 (8.8%)   | 30 (6%)     | 0 (0%)                 | 0 (0%)      | 443 (88.6%) | 32 (6.4%)   | 25 (5%)    |
| EIA                   | 0 (0%)      | 0 (0%)      | 387 (77.4%) | 53 (10.6%)  | 60 (12%)    | 0 (0%)                 | 0 (0%)      | 401 (80.2%) | 63 (12.6%)  | 36 (7.2%)   | 0 (0%)                 | 0 (0%)      | 423 (84.6%) | 49 (9.8%)   | 28 (5.6%)  |

| $\mu_2 = 2\mu_1$      |             |           |             |             |             |                        |             |             |             |             |                        |             |             |            |             |
|-----------------------|-------------|-----------|-------------|-------------|-------------|------------------------|-------------|-------------|-------------|-------------|------------------------|-------------|-------------|------------|-------------|
| $\sigma_1 = \sigma_2$ |             |           |             |             |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |             |             |            |             |
|                       | Very Strong | Strong    | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong      | Positive    | Weak       | Negative    |
| IP                    | 264 (52.8%) | 125 (25%) | 60 (12%)    | 12 (2.4%)   | 39 (7.8%)   | 263 (52.6%)            | 144 (28.8%) | 64 (12.8%)  | 11 (2.2%)   | 18 (3.6%)   | 324 (64.8%)            | 114 (22.8%) | 48 (9.6%)   | 7 (1.4%)   | 7 (1.4%)    |
| Robust                | 225 (45%)   | 150 (30%) | 65 (13%)    | 18 (3.6%)   | 42 (8.4%)   | 231 (46.2%)            | 166 (33.2%) | 72 (14.4%)  | 9 (1.8%)    | 22 (4.4%)   | 270 (54%)              | 157 (31.4%) | 55 (11%)    | 8 (1.6%)   | 10 (2%)     |
| TESS                  | 162 (32.4%) | 190 (38%) | 79 (15.8%)  | 22 (4.4%)   | 47 (9.4%)   | 183 (36.6%)            | 196 (39.2%) | 81 (16.2%)  | 16 (3.2%)   | 24 (4.8%)   | 183 (36.6%)            | 229 (45.8%) | 64 (12.8%)  | 10 (2%)    | 14 (2.8%)   |
| Conjugate             | 0 (0%)      | 0 (0%)    | 73 (14.6%)  | 194 (38.8%) | 233 (46.6%) | 0 (0%)                 | 0 (0%)      | 73 (14.6%)  | 193 (38.6%) | 234 (46.8%) | 0 (0%)                 | 0 (0%)      | 82 (16.4%)  | 245 (49%)  | 173 (34.6%) |
| Jeffrey               | 0 (0%)      | 0 (0%)    | 224 (44.8%) | 87 (17.4%)  | 189 (37.8%) | 0 (0%)                 | 0 (0%)      | 229 (45.8%) | 102 (20.4%) | 169 (33.8%) | 0 (0%)                 | 0 (0%)      | 261 (52.2%) | 110 (22%)  | 129 (25.8%) |
| Schwarz               | 0 (0%)      | 0 (0%)    | 268 (53.6%) | 77 (15.4%)  | 155 (31%)   | 0 (0%)                 | 0 (0%)      | 269 (53.8%) | 98 (19.6%)  | 133 (26.6%) | 0 (0%)                 | 0 (0%)      | 330 (66%)   | 75 (15%)   | 95 (19%)    |
| ZS                    | 0 (0%)      | 0 (0%)    | 287 (57.4%) | 72 (14.4%)  | 141 (28.2%) | 0 (0%)                 | 0 (0%)      | 298 (59.6%) | 82 (16.4%)  | 120 (24%)   | 0 (0%)                 | 0 (0%)      | 349 (69.8%) | 64 (12.8%) | 87 (17.4%)  |
| EIA                   | 0 (0%)      | 0 (0%)    | 268 (53.6%) | 78 (15.6%)  | 154 (30.8%) | 0 (0%)                 | 0 (0%)      | 268 (53.6%) | 100 (20%)   | 132 (26.4%) | 0 (0%)                 | 0 (0%)      | 330 (66%)   | 77 (15.4%) | 93 (18.6%)  |

**Table A4.** Frequency distribution for the balanced framework ( $n_1 = n_2 = 50$ ) from the gamma random variables of the  $2 \log(B_{01})$  based on the scale proposed by [13]. Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].

| $\mu_1 = \mu_2$       |             |             |             |             |           |                        |        |          |        |            |                        |        |          |        |            |
|-----------------------|-------------|-------------|-------------|-------------|-----------|------------------------|--------|----------|--------|------------|------------------------|--------|----------|--------|------------|
| $\sigma_1 = \sigma_2$ |             |             |             |             |           | $\sigma_2 = 2\sigma_1$ |        |          |        |            | $\sigma_2 = 3\sigma_1$ |        |          |        |            |
|                       | Very Strong | Strong      | Positive    | Weak        | Negative  | Very Strong            | Strong | Positive | Weak   | Negative   | Very Strong            | Strong | Positive | Weak   | Negative   |
| IP                    | 449 (89.8%) | 49 (9.8%)   | 2 (0.4%)    | 0 (0%)      | 0 (0%)    | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Robust                | 417 (83.4%) | 78 (15.6%)  | 5 (1%)      | 0 (0%)      | 0 (0%)    | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| TESS                  | 348 (69.6%) | 139 (27.8%) | 13 (2.6%)   | 0 (0%)      | 0 (0%)    | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Conjugate             | 0 (0%)      | 0 (0%)      | 183 (36.6%) | 268 (53.6%) | 49 (9.8%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 412 (82.4%) | 62 (12.4%)  | 26 (5.2%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Schwarz               | 0 (0%)      | 0 (0%)      | 452 (90.4%) | 31 (6.2%)   | 17 (3.4%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| ZS                    | 0 (0%)      | 0 (0%)      | 465 (93%)   | 25 (5%)     | 10 (2%)   | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| EIA                   | 0 (0%)      | 0 (0%)      | 451 (90.2%) | 32 (6.4%)   | 17 (3.4%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |

| $\mu_2 = 1.5\mu_1$    |             |             |             |            |             |                        |          |           |           |             |                        |        |          |        |            |
|-----------------------|-------------|-------------|-------------|------------|-------------|------------------------|----------|-----------|-----------|-------------|------------------------|--------|----------|--------|------------|
| $\sigma_1 = \sigma_2$ |             |             |             |            |             | $\sigma_2 = 2\sigma_1$ |          |           |           |             | $\sigma_2 = 3\sigma_1$ |        |          |        |            |
|                       | Very Strong | Strong      | Positive    | Weak       | Negative    | Very Strong            | Strong   | Positive  | Weak      | Negative    | Very Strong            | Strong | Positive | Weak   | Negative   |
| IP                    | 30 (6%)     | 141 (28.2%) | 186 (37.2%) | 63 (12.6%) | 80 (16%)    | 0 (0%)                 | 3 (0.6%) | 33 (6.6%) | 35 (7%)   | 429 (85.8%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Robust                | 21 (4.2%)   | 119 (23.8%) | 191 (38.2%) | 68 (13.6%) | 101 (20.2%) | 0 (0%)                 | 0 (0%)   | 22 (4.4%) | 39 (7.8%) | 439 (87.8%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| TESS                  | 8 (1.6%)    | 109 (21.8%) | 186 (37.2%) | 72 (14.4%) | 125 (25%)   | 0 (0%)                 | 0 (0%)   | 16 (3.2%) | 33 (6.6%) | 451 (90.2%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Conjugate             | 0 (0%)      | 0 (0%)      | 2 (0.4%)    | 32 (6.4%)  | 466 (93.2%) | 0 (0%)                 | 0 (0%)   | 0 (0%)    | 0 (0%)    | 500 (100%)  | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Jeffrey               | 0 (0%)      | 0 (0%)      | 19 (3.8%)   | 49 (9.8%)  | 432 (86.4%) | 0 (0%)                 | 0 (0%)   | 0 (0%)    | 0 (0%)    | 500 (100%)  | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Schwarz               | 0 (0%)      | 0 (0%)      | 36 (7.2%)   | 63 (12.6%) | 401 (80.2%) | 0 (0%)                 | 0 (0%)   | 0 (0%)    | 0 (0%)    | 500 (100%)  | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| ZS                    | 0 (0%)      | 0 (0%)      | 50 (10%)    | 73 (14.6%) | 377 (75.4%) | 0 (0%)                 | 0 (0%)   | 0 (0%)    | 0 (0%)    | 500 (100%)  | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| EIA                   | 0 (0%)      | 0 (0%)      | 35 (7%)     | 64 (12.8%) | 401 (80.2%) | 0 (0%)                 | 0 (0%)   | 0 (0%)    | 0 (0%)    | 500 (100%)  | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |

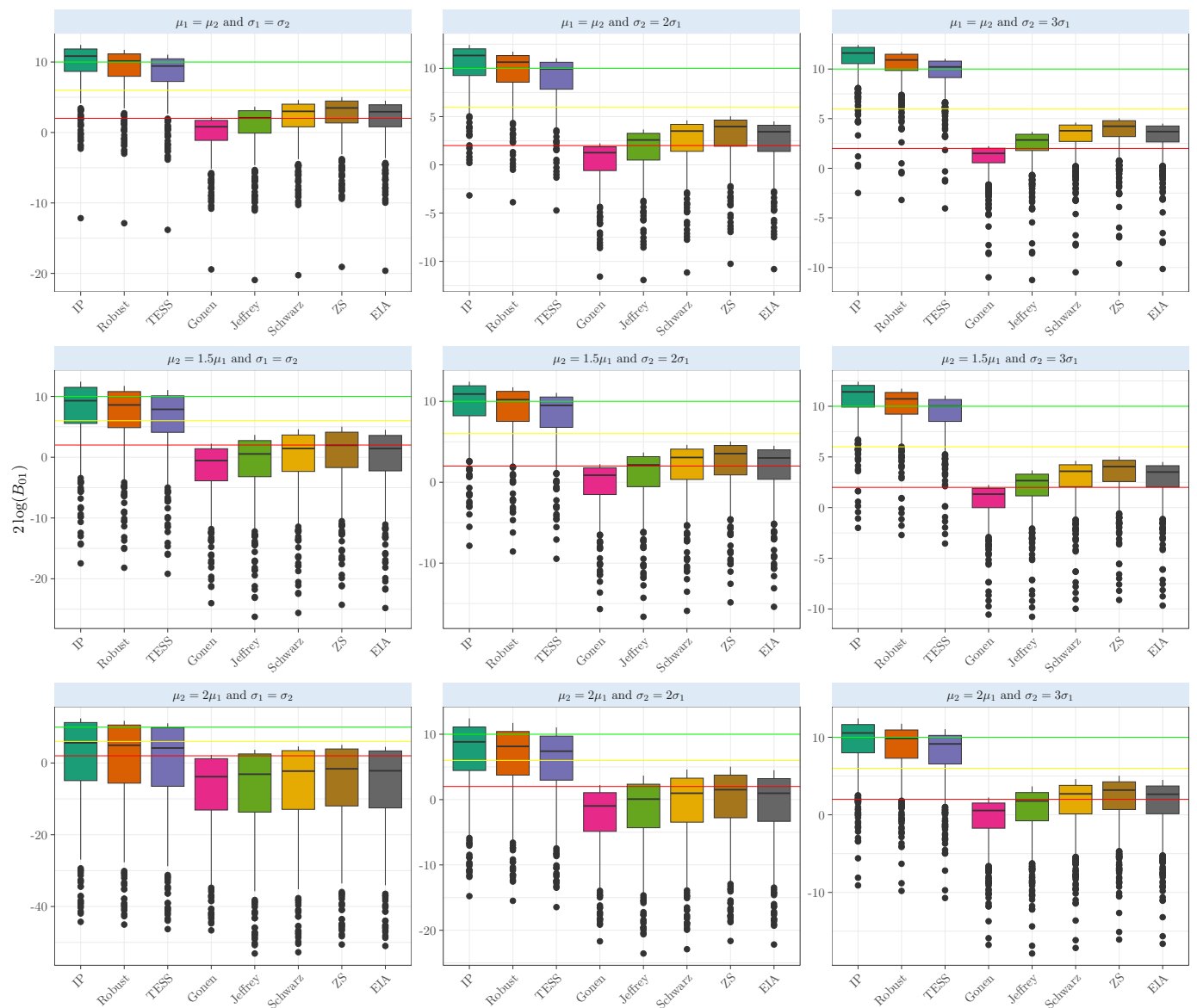
  

| $\mu_2 = 2\mu_1$      |             |           |             |            |             |                        |             |             |             |             |                        |        |          |        |            |
|-----------------------|-------------|-----------|-------------|------------|-------------|------------------------|-------------|-------------|-------------|-------------|------------------------|--------|----------|--------|------------|
| $\sigma_1 = \sigma_2$ |             |           |             |            |             | $\sigma_2 = 2\sigma_1$ |             |             |             |             | $\sigma_2 = 3\sigma_1$ |        |          |        |            |
|                       | Very Strong | Strong    | Positive    | Weak       | Negative    | Very Strong            | Strong      | Positive    | Weak        | Negative    | Very Strong            | Strong | Positive | Weak   | Negative   |
| IP                    | 0 (0%)      | 19 (3.8%) | 103 (20.6%) | 86 (17.2%) | 292 (58.4%) | 148 (29.6%)            | 218 (43.6%) | 103 (20.6%) | 17 (3.4%)   | 14 (2.8%)   | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Robust                | 0 (0%)      | 13 (2.6%) | 87 (17.4%)  | 73 (14.6%) | 327 (65.4%) | 111 (22.2%)            | 229 (45.8%) | 122 (24.4%) | 16 (3.2%)   | 22 (4.4%)   | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| TESS                  | 0 (0%)      | 6 (1.2%)  | 71 (14.2%)  | 66 (13.2%) | 357 (71.4%) | 73 (14.6%)             | 239 (47.8%) | 140 (28%)   | 21 (4.2%)   | 27 (5.4%)   | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Conjugate             | 0 (0%)      | 0 (0%)    | 0 (0%)      | 0 (0%)     | 500 (100%)  | 0 (0%)                 | 0 (0%)      | 28 (5.6%)   | 124 (24.8%) | 348 (69.6%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Jeffrey               | 0 (0%)      | 0 (0%)    | 0 (0%)      | 2 (0.4%)   | 498 (99.6%) | 0 (0%)                 | 0 (0%)      | 107 (21.4%) | 130 (26%)   | 263 (52.6%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| Schwarz               | 0 (0%)      | 0 (0%)    | 0 (0%)      | 5 (1%)     | 495 (99%)   | 0 (0%)                 | 0 (0%)      | 157 (31.4%) | 139 (27.8%) | 204 (40.8%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| ZS                    | 0 (0%)      | 0 (0%)    | 0 (0%)      | 7 (1.4%)   | 493 (98.6%) | 0 (0%)                 | 0 (0%)      | 200 (40%)   | 115 (23%)   | 185 (37%)   | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |
| EIA                   | 0 (0%)      | 0 (0%)    | 0 (0%)      | 5 (1%)     | 495 (99%)   | 0 (0%)                 | 0 (0%)      | 154 (30.8%) | 143 (28.6%) | 203 (40.6%) | 0 (0%)                 | 0 (0%) | 0 (0%)   | 0 (0%) | 500 (100%) |



**Figure A1.** Evidence in the  $2 \log(B_{01})$  scale when comparing the population means of two samples that arise from normal distributions with several means and variances with equal sizes  $n_1 = 50$  and  $n_2 = 500$ . Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].





**Figure A2.** Evidence in the  $2 \log(B_{01})$  scale when comparing the population means of two samples that arise from a Student's  $t$  distributions with several means and variances with equal sizes  $n_1 = 50$  and  $n_2 = 500$ . Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].



**Figure A3.** Evidence in the  $2 \log(B_{01})$  scale when comparing the population means of two samples that arise from gamma random samples with several means and variances with equal sizes  $n_1 = 50$  and  $n_2 = 500$ . Intrinsic priors (IP), Berger robust prior (Robust), BIC based on the effective sample size (TESS), conjugate, Jeffrey's, Schwarz, Zellner and Siow (ZS), and the expected arithmetic intrinsic prior of [24].

## References

1. Student. The probable error of a mean. *Biometrika* **1908**, *6*, 1–25. [\[CrossRef\]](#)
2. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical tests,  $p$  values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Wasserstein, R.L.; Lazar, N.A. The ASA statement on  $p$ -values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129–133. [\[CrossRef\]](#)
4. Vidgen, B.; Yasseri, T.  $p$ -values: Misunderstood and misused. *Front. Physics* **2016**, *4*, 6. [\[CrossRef\]](#)
5. Held, L.; Ott, M. On  $p$ -values and Bayes factors. *Annu. Rev. Stat. Its Appl.* **2018**, *5*, 393–419. [\[CrossRef\]](#)
6. Dienes, Z. How Bayes factors change scientific practice. *J. Math. Psychol.* **2016**, *72*, 78–89. [\[CrossRef\]](#)
7. Marden, J.I. Hypothesis testing: From  $p$  values to Bayes factors. *J. Am. Stat. Assoc.* **2000**, *95*, 1316–1320. [\[CrossRef\]](#)
8. Page, R.; Satake, E. Beyond  $p$  Values and Hypothesis Testing: Using the Minimum Bayes Factor to Teach Statistical Inference in Undergraduate Introductory Statistics Courses. *J. Educ. Learn.* **2017**, *6*, 254–266. [\[CrossRef\]](#)
9. Lavine, M.; Schervish, M.J. Bayes factors: What they are and what they are not. *Am. Stat.* **1999**, *53*, 119–122.

10. Berger, J.O.; Mortera, J. Default Bayes factors for nonnested hypothesis testing. *J. Am. Stat. Assoc.* **1999**, *94*, 542–554. [[CrossRef](#)]
11. Berger, J.; Pericchi, L. Bayes factors. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons: Hoboken, NJ, USA, 2014; pp. 1–14.
12. Jeffreys, H. *The Theory of Probability*; OUP Oxford: Oxford, UK, 1998.
13. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [[CrossRef](#)]
14. Gönen, M.; Johnson, W.O.; Lu, Y.; Westfall, P.H. The Bayesian two-sample t test. *Am. Stat.* **2005**, *59*, 252–257. [[CrossRef](#)]
15. Berger, J.O.; Pericchi, L.R.; Ghosh, J.; Samanta, T.; De Santis, F.; Berger, J.; Pericchi, L. *Objective Bayesian Methods for Model Selection: Introduction and Comparison*; Lecture Notes-Monograph Series; Institute of Mathematical Statistics: Beachwood, OH, USA, 2001; pp. 135–207.
16. Berger, J.O.; Pericchi, L.R. The intrinsic Bayes factor for linear models. *Bayesian Stat.* **1996**, *5*, 25–44.
17. Berger, J.O.; Pericchi, L.R. The intrinsic Bayes factor for model selection and prediction. *J. Am. Stat. Assoc.* **1996**, *91*, 109–122. [[CrossRef](#)]
18. Berger, J.O. Robust Bayesian analysis: Sensitivity to the prior. *J. Stat. Plan. Inference* **1990**, *25*, 303–328. [[CrossRef](#)]
19. Moreno, E. *Bayes Factors for Intrinsic and Fractional Priors in Nested Models. Bayesian Robustness*; Lecture Notes-Monograph Series; Institute of Mathematical Statistics: Beachwood, OH, USA, 1997; pp. 257–270.
20. Berger, J.O.; Berger, J. *Bayesian Analysis*; Springer: Berlin/Heidelberg, Germany, 1985.
21. Berger, J.; Bayarri, M.; Pericchi, L. The effective sample size. *Econom. Rev.* **2014**, *33*, 197–217. [[CrossRef](#)]
22. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
23. Zellner, A.; Siow, A. Posterior odds ratios for selected regression hypotheses. *Trab. Estad. Investig. Oper.* **1980**, *31*, 585–603. [[CrossRef](#)]
24. Kim, D.H.; Kang, S.G.; Lee, W.D. Intrinsic priors for testing two normal means with intrinsic bayes factors. *Commun. Stat. Methods* **2006**, *35*, 63–81. [[CrossRef](#)]
25. Cushny, A.R.; Peebles, A.R. The action of optical isomers: II. Hyoscines. *J. Physiol.* **1905**, *32*, 501. [[CrossRef](#)]
26. Senn, S.; Richardson, W. The first t-test. *Stat. Med.* **1994**, *13*, 785–803. [[CrossRef](#)]
27. Senn, S. A century of t-tests. *Significance* **2008**, *5*, 37–39. [[CrossRef](#)]
28. Falk, J.L.; Tang, M.; Forman, S. Schedule-induced chronic hypertension. *Psychosom. Med.* **1977**, *39*, 252–263. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.