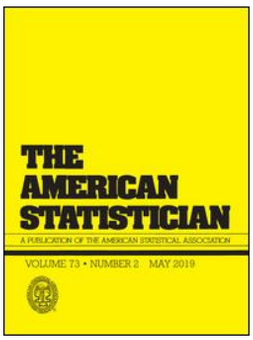

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Connecting and Contrasting the Bayes Factor and a Modified ROPE Procedure for Testing Interval Null Hypotheses

J. G. Liao, Vishal Midya & Arthur Berg

To cite this article: J. G. Liao, Vishal Midya & Arthur Berg (2020): Connecting and Contrasting the Bayes Factor and a Modified ROPE Procedure for Testing Interval Null Hypotheses, The American Statistician, DOI: 10.1080/00031305.2019.1701550

To link to this article: https://doi.org/10.1080/00031305.2019.1701550

THE AMERICAN STATISTICIAN
2020, VOL. 00, NO. 0, 1–9: General
https://doi.org/10.1080/00031305.2019.1701550

Taylor & Francis
Taylor & Francis Group

Connecting and Contrasting the Bayes Factor and a Modified ROPE Procedure for Testing Interval Null Hypotheses

J. G. Liao, Vishal Midya, and Arthur Berg

Division of Biostatistics and Bioinformatics, Penn State University College of Medicine, Hershey, PA

ABSTRACT

There has been strong recent interest in testing interval null hypotheses for improved scientific inference. For example, Lakens et al. and Lakens and Harms use this approach to study if there is a prespecified meaningful treatment effect in gerontology and clinical trials, instead of a point null hypothesis of any effect. Two popular Bayesian approaches are available for interval null hypothesis testing. One is the standard Bayes factor and the other is the region of practical equivalence (ROPE) procedure championed by Kruschke and others over many years. This article connects key quantities in the two approaches, which in turn allow us to contrast two major differences between the approaches with substantial practical implications. The first is that the Bayes factor depends heavily on the prior specification while a modified ROPE procedure is very robust. The second difference is concerned with the statistical property when data are generated under a neutral parameter value on the common boundary of competing hypotheses. In this case, the Bayes factors can be severely biased whereas the modified ROPE approach gives a reasonable result. Finally, the connection leads to a simple and effective algorithm for computing Bayes factors using draws from posterior distributions generated by standard Bayesian programs such as BUGS, JAGS, and Stan.

ARTICLE HISTORY
Received March 2019
Accepted November 2019

KEYWORDS
Bayes factor; Interval null hypothesis; Markov chain Monte Carlo; ROPE; Stan

1. Introduction

Hypothesis testing is widely used in scientific research. For a statistical model parameterized by θ, the general formulation is

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

where Θ_0 and Θ_1 are disjoint by assumption. Traditionally, a point null hypothesis takes Θ_0 to be a single point θ_0 leading to

$$\Theta_0 = \{\theta_0\} \quad \text{and} \quad \Theta_1 = \{\theta : \theta \neq \theta_0\}.$$

However, many researchers, such as Meehl (1978) and Cohen (1994), have argued that such formulation is not appropriate in most scientific research because the parameter θ cannot be exactly specified to a single point θ_0. For example, when comparing a new treatment with a standard one, the difference in the mean of an outcome variable is almost never exactly 0. Therefore, a more appropriate formulation is

$$\Theta_0 = \{\theta : \|\theta - \theta_0\| \leq \delta\} \quad \text{and} \quad \Theta_1 = \{\theta : \|\theta - \theta_0\| > \delta\},$$

where $\theta \in \Theta_0$ represents practically negligible deviation from θ_0 for some $\delta > 0$. Lakens et al. (2018) and Lakens (2017) used this approach to study if there is a prespecified meaningful treatment effect in gerontology and in clinical trials. Morey and Rouder (2011) reviewed and develop interval null hypothesis testing in the context of psychological research. See Blume et al. (2019) for a thorough discussion of the advantages of interval hypothesis testing.

A frequentist approach to interval null hypothesis testing is an equivalence test (e.g., Rogers, Howard, and Vessey 1993; Wellek 2010; Lakens 2017), in which $H_1 : \|\theta - \theta_0\| > \delta$ is in fact the default hypothesis to reject while $H_0 : \|\theta - \theta_0\| \leq \delta$ serves as the alternative. The limitation of the frequentist hypothesis testing is well documented (Wasserstein and Lazar 2016). In particular, it can only quantify evidence against the default hypothesis but not the evidence for it. Recently, there has been renewed interests (Morey and Rouder 2011; Kruschke 2013; Harms and Lakens 2018; Lakens et al. 2018) in using a Bayesian approach for tackling interval null hypothesis problem so that H_0 and H_1 can be treated on a more symmetric basis. The standard Bayes approach uses Bayes factors to quantify the relative support of the data for one hypothesis over the other. Standard references for the Bayes factor include Kass and Wasserman (1995), Berger (2013), and Berger and Pericchi (2015). A special issue of *Journal of Mathematical Psychology* (Mulder and Wagenmakers 2016) provides in-depth and updated discussion of the Bayes factor.

Alternatively, Krushke and others (Freedman, Lowe, and Macaskill 1984; Hobbs and Carlin 2007; Carlin and Louis 2008; Edwards and Berry 2010; Kruschke 2011, 2013, 2014, 2018) have championed a procedure called region of practical equivalence (ROPE), in which $\Theta_0 = \{\theta : \|\theta - \theta_0\| \leq \delta\}$ is treated as a region of practical equivalence. In this procedure, a prior for θ is first specified on the whole parameter space $\Theta = \Theta_0 \cup \Theta_1$. The $(1-\alpha)$ highest density interval of the posterior distribution of θ is then constructed. If this interval falls completely in Θ_j, either $j = 0$ or $j = 1$, then H_j is selected. Otherwise, the selection between H_0 and H_1 is declared uncertain.

CONTACT J. G. Liao jg_liao@hotmail.com Division of Biostatistics and Bioinformatics, Penn State University College of Medicine, 500 University Drive, Hershey, PA 17033.

The Bayes factor and ROPE procedures have been treated as two different and distinctive approaches. Because they can produce very different results (see examples below), their appropriate use and the choice between them have been the subject of debate both in published literature and in Internet blog space. In particular, Bayes factor is a controversial method with some prominent statisticians for it and some against it. In this article, a formal connection between key quantities of the two approaches is given in Lemma 1. This connection allows us to contrast two major differences between them with important practical implications. The first difference is that the Bayes factor depends heavily on the prior specification while the ROPE procedure is very robust. The second difference concerns the behavior of the two approaches when data are generated under a neutral point on the common boundary of Θ_0 and Θ_1. The Bayes factors can be severely biased whereas the modified ROPE approach gives a reasonable result. Finally, the connection between the two approaches leads to a simple and effective algorithm for computing Bayes factors in a wide range of problems by using draws from the posterior distribution of θ generated by standard Bayesian software programs such as WinBUGS (Lunn et al. 2000), JAGS (Plummer 2003), and Stan (Carpenter et al. 2017). This circumvents the need for custom-made software to calculate marginal distributions needed for the Bayes factors.

2. Connecting the Bayes Factor Approach and the ROPE Procedure

Section 2 and 3 will address statistical test of the form:

$$H_0 : \theta \in \Theta_0 = [\theta_0 - \delta, \theta_0 + \delta] \quad \text{versus} \qquad (1)$$
$$H_1 : \theta \in \Theta_1 = (-\infty, \theta_0 - \delta) \cup (\theta_0 + \delta, \infty),$$

where θ is a scalar parameter and $\delta > 0$ is prespecified. We start with the Bayes factor approach. Let π_0 be a density defined with support only on Θ_0 and π_1 be a density defined with support only on Θ_1. Suppose that θ can be generated from either π_1 (with prior probability η_1) or π_0 (with probability $\eta_0 = 1 - \eta_1$). Let $f(y \mid \theta)$ be the likelihood. The marginal distribution of y under π_j is then

$$f(y \mid \pi_j) \equiv \int f(y \mid \theta)\pi_j(\theta)d\theta, \quad \text{for } j = 0, 1.$$

The notation $f(y \mid \pi_j)$ emphasizes the dependence of marginal density on prior π_j. It follows directly from Bayes' theorem that

$$\frac{\Pr(\theta \sim \pi_1 \mid y)}{\Pr(\theta \sim \pi_0 \mid y)} = \frac{f(y \mid \pi_1)}{f(y \mid \pi_0)} \frac{\eta_1}{\eta_0}, \qquad (2)$$

where $\Pr(\theta \sim \pi_j \mid y)$ denotes the posterior probability that θ is generated from π_j. In this equation, $\frac{\eta_1}{\eta_0}$ is the prior odds of $\theta \sim \pi_1$ over $\theta \sim \pi_0$ and $\frac{\Pr(\theta \sim \pi_1 \mid y)}{\Pr(\theta \sim \pi_0 \mid y)}$ is the posterior odds. The Bayes factor $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ modifies the prior odds to posterior odds and is the relative support for π_1 over π_0 in data y. It is easy to see that $\frac{f(y|\pi_1)}{f(y|\pi_0)} = \frac{\Pr(\theta \sim \pi_1 \mid y)}{\Pr(\theta \sim \pi_0 \mid y)}$ when $\eta_1 = \frac{1}{2}$. Note that most authors use $\frac{f(y|\pi_0)}{f(y|\pi_1)}$ as the Bayes factor, which is the relative support of π_0 over π_1.

An alternative approach is to directly specify a single global prior distribution π for θ with support on Θ and base the inference on the posterior density $f(\theta \mid y) = c(y)f(y \mid \theta)\pi(\theta)$, where $c(y)$ is a normalizing constant that does not depend on θ. A prominent method using this approach is the ROPE procedure as summarized in Kruschke (2018):

> Consider a ROPE around a null value of a parameter. If the 95% HDI [Highest Density Interval] of the parameter distribution falls completely outside the ROPE, then one should reject the null value, because the 95% most credible values of the parameter are all not practically equivalent to the null value. If the 95% HDI of the parameter distribution falls completely inside the ROPE, then one should accept the null value for practical purposes, because the 95% most credible values of the parameter are all practically equivalent to the null value. If the 95% HDI is neither completely outside nor completely inside the ROPE, then one should remain undecided, because some of the most credible values are practically equivalent to the null but others are not.

In the ROPE procedure, the region of practical equivalence around θ_0 serves as Θ_0. The region outside Θ_0 is then Θ_1. The ROPE decision rule chooses Θ_0 over Θ_1, or chooses Θ_1 over Θ_0, or declares inconclusive if the 95% HDI of the posterior distribution of θ is contained inside Θ_0 or inside Θ_1 or is split between Θ_0 and Θ_1. A similar method of utilizing confidence or credible interval is the second generation of p-values discussed in Lakens and Delacre (2018). Procedures based on $f(\theta \mid y)$ are natural to vast number of statisticians familiar with posterior distribution and Bayesian parameter estimation. It also parallels the frequentist equivalence testing (e.g., Rogers, Howard, and Vessey 1993; Wellek 2010; Lakens 2017) in which a $(1 - 2\alpha)$ confidence interval is used in place of the $(1 - \alpha)$ HDI in ROPE.

In our development below, however, a slightly modified ROPE procedure is used in which inference is based on the posterior probability $\Pr(\theta \in \Theta_j \mid y) = \int_{\Theta_j} f(\theta \mid y)d\theta$, or equivalently the posterior odds $\frac{\Pr(\theta \in \Theta_1 \mid y)}{\Pr(\theta \in \Theta_0 \mid y)}$, in spite of the fact that the ROPE procedure championed by Kruschke relies on the $(1-\alpha)$ HDI interval. This modification is made because the optimal decision rule in choosing between H_0 and H_1 under the general weighted $0-1$ losses uses $\Pr(\theta \in \Theta_1 \mid y)$ (Robert 2007). In addition, the HDI depends on a particular parameterization of θ while $\Pr(\theta \in \Theta_j \mid y)$ generally does not. Conceptually, $\Pr(\theta \in \Theta_1 \mid y)$ is more natural in the Bayesian framework although HDI facilitates comparison with confidence interval in frequentist inference. Of course, the HDI and $\Pr(\theta \in \Theta_j \mid y)$ are closely related. In particular, the $(1-\alpha)$ HDI being contained in Θ_j implies $\Pr(\theta \in \Theta_j \mid y) \geq 1 - \alpha$ and therefore is a stronger requirement.

One can easily convert one formulation into the other. The Bayes factor approach can be formulated in the ROPE notation by combining π_0, π_1 and η_1 into a global prior π with support on $\Theta \equiv \Theta_0 \bigcup \Theta_1$:

$$\pi(\theta) = \eta_0\pi_0(\theta) + \eta_1\pi_1(\theta).$$

On the other hand, the ROPE approach can be formulated in the Bayes factor notation by truncating the global prior π separately

on Θ_0 and Θ_1. Following Robert (2007, chap. 5), we have

$$\begin{aligned}\pi_0(\theta) &= \frac{\pi(\theta)}{\int_{\Theta_0}\pi(\theta)d\theta} \quad \text{for} \quad \theta \in \Theta_0,\\ \pi_1(\theta) &= \frac{\pi(\theta)}{\int_{\Theta_1}\pi(\theta)d\theta} \quad \text{for} \quad \theta \in \Theta_1,\\ \eta_1 &= \int_{\Theta_1}\pi(\theta)d\theta.\end{aligned} \tag{3}$$

The Bayes factor $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ under any π_0 and π_1 and the posterior odds $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ in our modified ROPE approach under a global prior $\pi = \eta_0\pi_0 + \eta_1\pi_1$ have the following relationship:

Lemma 1. Assume that prior π_0 is defined with support on Θ_0 only and prior π_1 with support on Θ_1 only. Suppose that Θ_0 and Θ_1 are disjoint, that is, $\Theta_0 \bigcap \Theta_1 = \emptyset$. Then

$$\frac{\Pr(\theta \in \Theta_1 \mid y)}{\Pr(\theta \in \Theta_0 \mid y)} = \frac{f(y \mid \pi_1)}{f(y \mid \pi_0)}\frac{\eta_1}{\eta_0}. \tag{4}$$

Proof.

$$\begin{aligned}\Pr(\theta \in \Theta_j \mid y) &= c(y)\int_{\Theta_j} f(y \mid \theta)\left\{\eta_0\pi_0(\theta) + \eta_1\pi_1(\theta)\right\} d\theta\\ &= c(y)\eta_j\int_{\Theta_j} f(y \mid \theta)\pi_j(\theta)\, d\theta\\ &= c(y)\eta_j f(y \mid \pi_j), \quad j = 0, 1.\end{aligned}$$

Equation (4) then follows directly. □

The first equality of the proof comes from the definition of posterior density. The second follows from the assumption that $\pi_0(\theta) = 0$ for $\theta \in \Theta_1$ and $\pi_1(\theta) = 0$ for $\theta \in \Theta_0$. Note, however, Equation (4) may not be true if priors π_0 and π_1 have positive densities on a common parametric domain of Θ, which is quite common in the general formulation of Bayes factors. Equation (2), however, is generally true.

3. Contrasting the Bayes Factor and the Modified ROPE Procedure

As discussed in Section 2, the Bayes factor approach specifies π_0 on Θ_0, π_1 on Θ_1 and η_1 separately and explicitly and uses the Bayes factor, $\frac{f(y|\pi_1)}{f(y|\pi_0)}$, to quantify the relative support of $\theta \sim \pi_1$ over $\theta \sim \pi_0$ in data y. In the modified ROPE approach, a global prior π is first specified on Θ. Choosing between H_1 and H_0 then depends on the posterior probabilities $\Pr(\theta \in \Theta_1 \mid y)$ and $\Pr(\theta \in \Theta_0 \mid y) = 1 - \Pr(\theta \in \Theta_1 \mid y)$ under this π, or equivalently on posterior odds $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$.

The Bayes factor and ROPE approaches, however, differ in two significant ways. First, in the ROPE approach, the derived π_0, π_1, and η_1 in Equation (3) are linked together through π. In particular, a more diffuse π with support on Θ leads to a more diffuse π_1 with support on Θ_1 but also a larger η_1. Second, the specified π in the ROPE is usually continuous at all points in Θ. For the Bayes factor, however, the global prior $\pi = (1-\eta_1)\pi_0 + \eta_1\pi_1$ is generally discontinuous at $\theta = \theta_0 - \delta$ and $\theta = \theta_0 + \delta$ (the two common boundary points of Θ_0 and Θ_1), even though π_0 is continuous on Θ_0 and π_1 continuous on Θ_1. We now show that these two differences have a substantial implication in terms of their performance.

To facilitate the asymptotic argument below, we now make the model more explicit. Let $y = (y_1, \ldots, y_n) \sim f(y \mid \theta)$ be a collection of n observations, where each y_i is an independent and identically distributed univariate random variable. The asymptotic arguments will be made with respect to $n \to \infty$. We shall also assume the following two mild technical conditions:

1. The likelihood function $f(y \mid \theta)$ is differentiable in quadratic means for $\theta \in \Theta$ and allows for a nonsingular Fisher information matrix I_θ.
2. For $\theta^* = \theta_0 - \delta$ or $\theta^* = \theta_0 + \delta$, it is possible to separate θ^* from θ in $||\theta - \theta^*|| > \epsilon$ using a statistical test, for any $\epsilon > 0$.

Condition 1 is needed for standard asymptotic likelihood expansion such that when n is large, the likelihood function $f(y \mid \theta)$ can be approximated by the normal distribution $\mathcal{N}(\hat\theta, \hat\sigma^2)$, where $\hat\theta$ is the MLE and $\hat\sigma^2 = I_{\hat\theta}^{-1}$. Conditions 1 and 2 together make it possible to apply the Bernstein–von Mises theorem in the proof of Theorem 1. For detailed discussion of these technical conditions see Chapter 10 "Bayes Procedures" in Vaart (1998).

3.1. Contrast 1: Sensitivity to Prior Specification

To explore the first difference in the context of hypothesis testing in (1), specify a global π as uniform distribution with support on $[\theta_0 - T, \theta_0 + T]$ as in the ROPE approach, where T is much larger than δ. Let $\Theta_{1,T} = (\theta_0 - T, \theta_0 - \delta) \cup (\theta_0 + \delta, \theta_0 + T)$. It then follows that $\pi_0(\theta) = \frac{1}{2\delta}$ with support on $\Theta_0 = [\theta_0 - \delta, \theta_0 + \delta]$, $\pi_1(\theta) = \frac{1}{2(T-\delta)}$ with support on $\Theta_{1,T} \subset \Theta_1$ and $\eta_1 = \frac{T-\delta}{T}$.

Now consider some y whose corresponding MLE $\hat\theta$ is (say) larger than $\theta_0 + \delta + 10\hat\sigma$, where $\hat\sigma$ is the asymptotic standard deviation of $\hat\theta$. Intuitively this represents strong evidence for H_1 over H_0 because $\hat\theta$ is $10\hat\sigma$ away from $\theta_0 + \delta$, the upper boundary of Θ_0. Choose T so that $\theta_0 + T$ is much larger than $\hat\theta + 10\hat\sigma$. In this way, the likelihood function $f(y \mid \theta)$ peaks at $\theta = \hat\theta$ and rapidly diminishes when θ moves to the two boundaries of $(\theta_0 + \delta, \theta_0 + T)$. We can then obtain a Laplace approximation (Kass and Raftery 1995)

$$f(y \mid \pi_1) = \frac{\hat\sigma f(y|\hat\theta)\sqrt{2\zeta}}{2(T-\delta)}(1 + O(n^{-1})),$$

where $\zeta = 3.14159$ is the constant Pi. It then follows

$$\frac{f(y \mid \pi_1)}{f(y \mid \pi_0)} = \frac{\delta}{T-\delta}\frac{\hat\sigma f(y|\hat\theta)\sqrt{2\zeta}}{\int_{\Theta_0} f(y|\theta)d\theta}(1 + O(n^{-1})).$$

In the above expression, $f(y \mid \pi_0) = \delta^{-1}\int_{\Theta_0} f(y \mid \theta)d\theta$ does not depend on T. And $\hat\sigma$ and $f(y \mid \hat\theta)$ depend only on y and the likelihood function $f(y \mid \theta)$. Note also that the accuracy of this Laplace approximation is determined by how well the density of $\mathcal{N}(\hat\theta, \hat\sigma^2)$ approximates likelihood function $f(y \mid \theta)$ around $\hat\theta$ for the given y such that the Laplace approximation remains valid if T is further increased. We see that $\frac{f(y|\pi_1)}{f(y|\pi_0)} \to 0$ as $T \to \infty$. This is clearly an undesirable property of the Bayes

factor and is an illustration of Lindley's paradox (Robert 2014). Intuitively, when π_1 is very diffuse, marginal density $f(y \mid \pi_1)$ becomes very small for any y because π_1 puts the majority of its probability distribution in areas far away from $\hat{\theta}$, where $f(y \mid \theta)$ is almost 0. On the other hand,

$$\frac{\Pr(\theta \in \Theta_1|y)}{\Pr(\theta \in \Theta_0|y)} = \frac{\eta_1}{1-\eta_1}\frac{f(y \mid \pi_1)}{f(y \mid \pi_0)} = \frac{\hat{\sigma} f(y|\hat{\theta})\sqrt{2\zeta}}{\int_{\Theta_0} f(y|\theta)d\theta}(1+O(n^{-1})),$$

which is very robust to different values of T. This is because, as $T \to \infty$, the additional factor $\frac{\eta_1}{1-\eta_1}$, which approaches ∞, negates the effect of $\pi_1(\theta) \to 0$.

Note that the exhibited sensitivity characteristics remain true if the uniform distribution π on $[\theta_0 - T, \theta_0 + T]$ above is replaced by normal distribution $\mathcal{N}(\theta_0, T)$ on Θ, and the uniform distribution π_1 on $\Theta_{1,T}$ and π_0 on Θ_0 are replaced by $\mathcal{N}(\theta_0, T)$ truncated on Θ_1 and Θ_0, respectively.

3.2. Contrast 2: Behavior at the Common Boundary of Θ_0 and Θ_1

To explore the second difference, first note that, for fixed π_0 and π_1, Bayes factors are consistent under quite general conditions (Kass and Raftery 1995). In particular, $\frac{f(y|\pi_1)}{f(y|\pi_0)} \to \infty$, as $n \to \infty$, if y is generated under $f(y \mid \theta)$ with $\pi_0(\theta) = 0$ and $\pi_1(\theta) > 0$. Similarly, $\frac{f(y|\pi_1)}{f(y|\pi_0)} \to 0$ when $\pi_0(\theta) > 0$ and $\pi_1(\theta) = 0$. It then follows from Lemma 1 that $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ is also consistent.

We now discuss the property of $\Pr(\theta \in \Theta_1 \mid y)$ and $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ when $y \sim f(y \mid \theta^*)$, where $\theta^* = \theta_0 - \delta$ or $\theta^* = \theta_0 + \delta$, the two common boundary points between Θ_0 and Θ_1. We call these two points the "neutral points" between Θ_0 and Θ_1. We shall assume that π_0 and π_1 are both defined on $\theta_0 - \delta$ and $\theta_0 + \delta$ as the appropriate one-sided limits. For example, for $\theta^* = \theta_0 + \delta$, $\pi_0(\theta^*) = \lim_{\theta\uparrow\theta^*} \pi_0(\theta)$ and $\pi_1(\theta^*) = \lim_{\theta\downarrow\theta^*} \pi_1(\theta)$.

Theorem 1. Let y be a random sample form $f(y \mid \theta^*)$, where θ^* is either $\theta_0 - \delta$ or $\theta_0 + \delta$, the two neutral points. Assume $\pi(\theta)$ is positive and absolute continuous in the neighborhood of θ^*, then as $n \to \infty$,

$$\begin{aligned} \Pr(\theta \in \Theta_1 \mid y) &\xrightarrow{d} \mathcal{U}\text{niform}(0,1), \\ \log\frac{\Pr(\theta \in \Theta_1|y)}{\Pr(\theta \in \Theta_0|y)} &\xrightarrow{d} \mathcal{L}\text{ogistic}(0,1). \end{aligned} \tag{5}$$

For absolutely continuous prior densities π_0 and π_1 with $\pi_0(\theta^*) > 0$ and $\pi_1(\theta^*) > 0$,

$$\log\frac{f(y|\pi_1)}{f(y|\pi_0)} \xrightarrow{d} \log\frac{\pi_1(\theta^*)}{\pi_0(\theta^*)} + \mathcal{L}\text{ogistic}(0,1).$$

In the above formulation, $\mathcal{L}\text{ogistic}(0,1)$ has density $\frac{e^{-x}}{(1+e^{-x})^2}$, mean 0 and variance $\frac{\zeta^2}{3}$, where ζ is constant Pi. The proof of the Theorem 1 is in the Appendix.

To see the implication of Theorem 1, note that it is reasonable to expect that H_0 and H_1 be somewhat equally supported when data y is generated under one of these two neutral points. This is true asymptotically for $\Pr(\theta \in \Theta_1 \mid y)$ and $\log\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ in our modified ROPE approach because their distributions are symmetric between Θ_1 and Θ_0. In the Bayes factor specification of π_0 and π_1, however, $\pi_1(\theta^*)$ is often much smaller than $\pi_0(\theta^*)$. The asymptotic mean of $\log\frac{f(y|\pi_1)}{f(y|\pi_0)}$ is then much smaller than 0, favoring H_0 over H_1. A similar bias is exhibited in using Bayes factors to compare three competing genetic association models in Liao, Liao, and Berg (2016).

Table 1. Numerical comparison between the Bayes factor and the ROPE procedure.

T	$\Pr(\theta \in \Theta_1\|y)$	$\frac{\Pr(\theta\in\Theta_1\|y)}{\Pr(\theta\in\Theta_0\|y)}$	$\frac{f(y\|\pi_1)}{f(y\|\pi_0)}$
50	0.9998	4913.5778	49.63
200	0.9998	4913.5780	12.31
800	0.9998	4913.5788	3.07
3200	0.9998	4913.5820	0.768

3.3. Numerical Example

As an example to illustrate the two contrasts, let $y = (y_1, \ldots, y_n)$, where $y_i \sim \mathcal{N}(\theta, 1)$ independently. Let $\delta = \frac{1}{2}$ and let the global prior π be a uniform distribution with support on $[-T, T]$. Let π_0 and π_1 be implied local prior as given in Equation (3). For contrast 1 in Section 3.1, consider some y with $n = 50$ and $\bar{y} = \hat{\theta} = 1$. Table 1 gives the value of $\Pr(\theta \in \Theta_1 \mid y)$, $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ and $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ for several values of T. It is observed that $\Pr(\theta \in \Theta_1 \mid y)$, $\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ are very stable for different values of T while $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ is inversely proportional to T, leading to vastly different values for different specifications of T.

For contrast 2 in Section 3.2, we generated 1000 replications of y under $y \sim f(y \mid \theta_0 + \delta)$. For each replication, the three quantities in Table 1 are calculated. Their distributions from these 1000 replications are plotted in Figure 1 for $n = 50$ and $n = 100$ under $T = 200$. The histogram of $\Pr(\theta \in \Theta_1 \mid y)$ in the first subfigure, indeed approximates $\mathcal{U}\text{niform}(0,1)$ well. The distribution of $\log\frac{\Pr(\theta\in\Theta_1|y)}{\Pr(\theta\in\Theta_0|y)}$ in the second subfigure is mostly symmetric around 0, representing equal support for both H_1 and H_0. The $\log\frac{f(y|\pi_1)}{f(y|\pi_0)}$ in the third and fourth subfigures, however, is distributed predominantly below 0, which strongly favors H_0 over H_1. The results here are numerical computation using exact formula for reported quantities, not based on Laplace approximation.

3.4. Discussion and Recommendation

The above comparison between the Bayes factor and the modified ROPE procedure clearly favors the ROPE approach and raises serious concern about the suitableness of the Bayes factor in these examples. However, since the Bayes factor is derived from Bayes' theorem, it seems hard to dismiss it as an invalid statistical tool. To resolve this conflict, note that the Bayes factor is derived under the condition that θ is generated from either π_0 or π_1. The Bayes factor $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ then gives the right index of evidence from Bayes' theorem. In practical application, however, π_0 and especially π_1 are often unknown and their misspecification can lead to very misleading Bayes factors as shown above. Our recommendation is therefore:

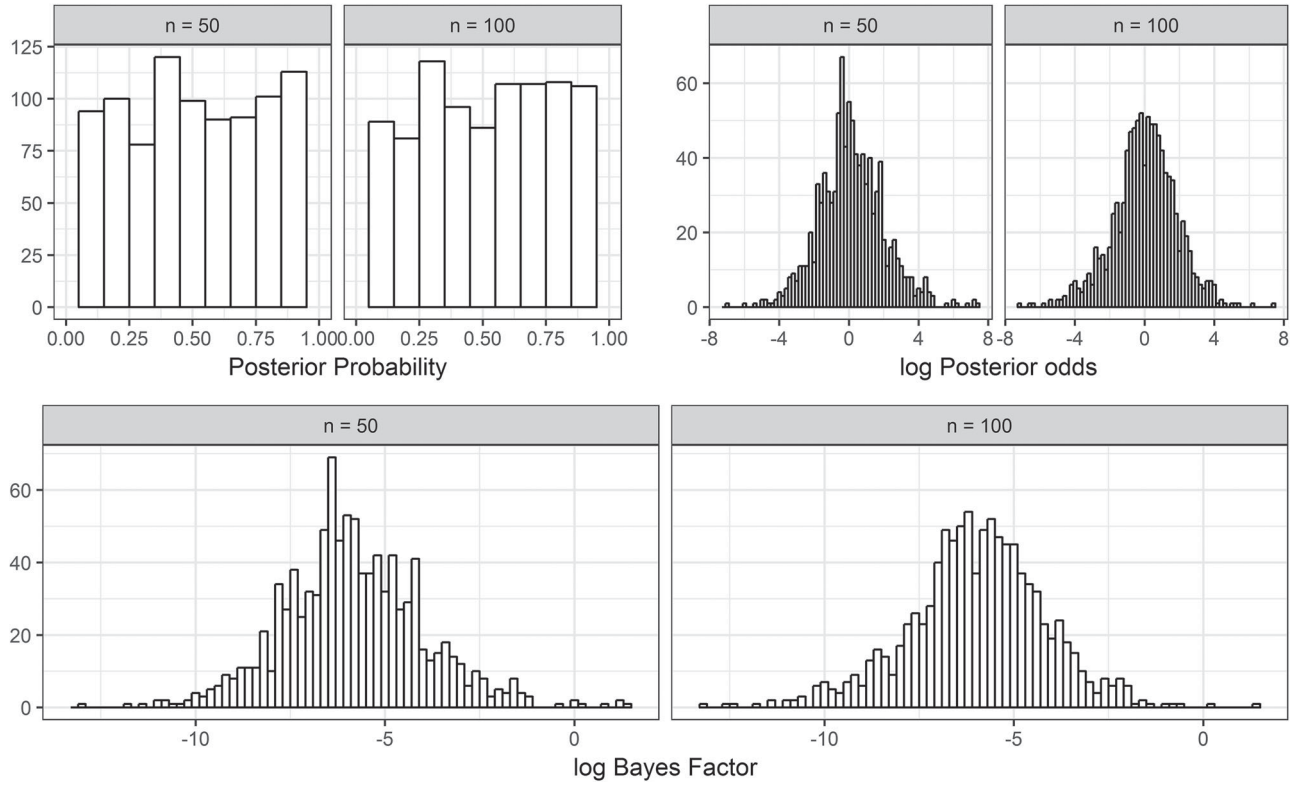

Figure 1. Distribution of posterior probability, log posterior odds, and log Bayes factors at the neutral point.

1. Use the Bayes factor $\frac{f(y|\pi_1)}{f(y|\pi_0)}$ if π_0 and π_1 can be specified with some confidence; either they are derived from a theoretical framework, or they reflect the effect size θ of particular practical interest, or they are formulated from past data. In this way, the two exposed deficiencies of the Bayes factor no longer apply.
2. One the other hand, the Bayes factor is not suitable if one cannot clearly quantify the size of θ under H_0 and H_1 with prior π_0 and π_1. In the case of $\Theta_0 = [\theta_0 - \delta, \theta_0 + \delta]$ and $\Theta_1 = (-\infty, \theta_0 - \delta) \cup (\theta_0 + \delta, \infty)$, one π_1 uniform on $\Theta_{1,T} = (\theta_0 - T, \theta_0 - \delta) \cup (\theta_0 + \delta, \theta_0 + T)$ for a smaller T and the other π_1 for a very large T lead to very different Bayes factors because they represent two different hypotheses even though both π_1 are defined with support on Θ_1 and both are compatible with H_1. In this case, the modified ROPE procedure can be a better choice because its robustness against different specifications of π is important practically. Its asymptotically equal support at the neutral points between Θ_0 and Θ_1 is one form of unbiasedness and nicely separates Θ_0 and Θ_1.

Finally, the asymptotic distributions of $\Pr(\theta \in \Theta_1 \mid y)$ and $\log \frac{\Pr(\theta \in \Theta_1 | y)}{\Pr(\theta \in \Theta_0 | y)}$ when y is generated under one of the neutral points can serve as a useful reference distribution in interpreting their observed values. They can play the same role that $\mathcal{U}$niform$(0, 1)$ plays for p-values. For example, even at a neutral point, there is 10% probability of $\Pr(\theta \in \Theta_1 \mid y) > 0.90$. So observing $\Pr(\theta \in \Theta_1 \mid y) = 0.90$ cannot be interpreted as too convincing an evidence for H_1. More research is needed, however, to fully understand the utility of results in Theorem 1.

4. Computing Bayes Factors Using MCMC Draws From the Posterior Distribution

We now extend Lemma 1 to models with nuisance parameter. Let $f(y|\theta, \lambda)$ be the likelihood, where θ is a scalar parameter of interest and λ is a nuisance parameter. As before, let π_0 be prior of θ with support on Θ_0, π_1 be prior with support on Θ_1, and $\pi = \eta_0\pi_0 + \eta_1\pi_1$ be the global prior with support on Θ. Additionally, let π_λ be the prior of λ on parameter space Λ. For simplicity, we shall assume that the prior of θ and λ are independent. It follows that, for $j = 0, 1$,

$$f(y \mid \pi_j, \pi_\lambda) \equiv \int_{\Theta_j} \left\{ \int_\Lambda f(y \mid \theta, \lambda)\pi_\lambda(\lambda)d\lambda \right\} \pi_j(\theta)d\theta.$$

The Bayes factor is then $\frac{f(y|\pi_1,\pi_\lambda)}{f(y|\pi_0,\pi_\lambda)}$. The posterior distribution of (θ, λ) under product prior $\pi(\theta)\pi_\lambda(\theta)$ is

$$f(\theta, \lambda \mid y) = c(y)f(y \mid \theta, \lambda)\pi(\theta)\pi_\lambda(\lambda),$$

where $c(y)$ is a normalizing constant depending only on y but not on (θ, λ). It follows that

$$\begin{aligned}
\Pr(\theta \in \Theta_j \mid y) &= c(y) \int_{\Theta_j} \left\{ \int_\Lambda f(y \mid \theta, \lambda)\pi_\lambda(\lambda)d\lambda \right\} \pi(\theta)d\theta \\
&= c(y)\eta_j \int_{\Theta_j} \left\{ \int_\Lambda f(y \mid \theta, \lambda)\pi_\lambda(\lambda)d\lambda \right\} \pi_j(\theta)d\theta \\
&= c(y)\eta_j f(y \mid \pi_j, \pi_\lambda).
\end{aligned}$$

We then have

$$\frac{f(y \mid \pi_1, \pi_\lambda)}{f(y \mid \pi_0, \pi_\lambda)} = \frac{\eta_0}{\eta_1} \frac{\Pr(\theta \in \Theta_1 \mid y)}{\Pr(\theta \in \Theta_0 \mid y)},$$

which is the same as Equation (4) in a slightly different format. Note that this identity is true for any $\eta_1 \in (0,1)$.

Usually customized software is needed to calculate Bayes factors by computing the marginal distributions $f(y \mid \pi_1, \pi_\lambda)$ and $f(y \mid \pi_0, \pi_\lambda)$ through numerical integration. Coding such software can be a tedious and error-prone task. The above equation, however, provides a simple method to compute Bayes factors from draws of the posterior distribution $f(\theta \mid y) = \int f(\theta, \lambda \mid y) d\lambda$. Such draws can be generated using standard Bayesian programs such as Stan (Carpenter et al. 2017), JAGS (Plummer 2003), and WinBUGS (Lunn et al. 2000). The method is explained through the following steps.

1. Choose $\eta_1 \in (0,1)$ and $\eta_0 = 1 - \eta_1$ to reduce, as much as possible, the discontinuity of the global density $\pi = \eta_0\pi_0(\theta) + \eta_1\pi_1(\theta)$ at the two boundary points $\theta - \delta$ and $\theta + \delta$. For example, when π_0 and π_1 are both symmetric around θ_0, π becomes continuous at $\theta - \delta$ and $\theta + \delta$ with the choice of
$$\frac{\eta_1}{1-\eta_1} = \frac{\pi_0(\theta_0 + \delta)}{\pi_1(\theta_0 + \delta)}.$$
This choice of η_1 improves the computational efficiency in Step 2.
2. Generate Monte Carlo draws $\theta^{(1)}, \ldots, \theta^{(N)}$ from posterior distribution $p(\theta \mid y)$ under prior $\pi = \eta_0\pi_0 + \eta_1\pi_1$, where N is the Monte Carlo sample size. This step can be accomplished by one of the popular general software mentioned above.
3. It follows from MCMC theory (Gelman et al. 2013, chap. 11) that, under quite general conditions, the left side is a consistent estimator of the right side in the following two equations:
$$\frac{1}{N}\sum_{i=1}^{N} 1_{\theta^{(i)} \in \Theta_1} \xrightarrow{P} \Pr(\theta \in \Theta_1 \mid y)$$
and
$$\frac{\eta_0}{\eta_1} \frac{\frac{1}{N}\sum_{i=1}^{N} 1_{\theta^{(i)} \in \Theta_1}}{1 - \frac{1}{N}\sum_{i=1}^{N} 1_{\theta^{(i)} \in \Theta_1}} \xrightarrow{P} \frac{f(y \mid \pi_1, \pi_\lambda)}{f(y \mid \pi_0, \pi_\lambda)}. \tag{6}$$

There are several advantages of this approach of computing Bayes factors over custom software. First, it reduces programming cost considerably. Second, packages such as Stan (Carpenter et al. 2017), JAGS (Plummer 2003), and WinBUGS (Lunn et al. 2000) are well-tested, high quality software routines that are familiar to most statisticians performing Bayesian analyses. So it is more likely to get correct answers quickly. Third, people have gained considerable experience and understanding of the posterior distribution, and this approach facilitates the adoption of the Bayes factor as an integrated part of Bayesian inference. The disadvantage is that it can be computationally less efficient than custom software specifically designed for computing Bayes factors.

There has been a long line of research in using MCMC draws to compute Bayes factors. They include, for example, the harmonic mean approximation (Meng and Wong 1996; Weinberg 2012); reversible jump Markov chain in Green (1995), an improved Savage–Dickey method to compute Bayes factors for nested models in Gelfand and Smith (1990), Chib (1995), and Morey et al. (2011). Recently, Kypraios and O'Neill (2018) proposed a supermodel approach in which the competing models are components of a mixture distribution. These specialized methods can be numerically unstable or computationally very involved. Our procedure here, by using the special structure of interval hypotheses, is much simpler: Θ_0 and Θ_1 are naturally combined into a single parameter space Θ for easier MCMC draws of $\theta^{(i)}$ and a simple proportion of $\theta^{(i)} \in \Theta_j$ is needed in processing $\theta^{(1)}, \ldots, \theta^{(N)}$.

Table 2. Blood pressure change data from Lyle et al. (1987).

Calcium	7	−4	18	17	−3	−5	1	10	11	−2	
Placebo	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

4.1. Example 1: Two Sample t-Test

Here we consider a simple dataset from Lyle et al. (1987), which was also analyzed in Gönen et al. (2005) and Wang and Liu (2016). The data (reproduced in Table 2) consist of changes in systolic blood pressure over 12 weeks for 21 African-American men, 10 of whom took calcium supplements and the remaining 11 took placebo supplements. Testing for equality of means between these two groups with a two-sample t-test (assuming equality of variances) yields a t-statistic of 1.63 and a p-value of 0.12. We now proceed to compute the Bayes factor for interval null hypotheses. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_X, \sigma^2)$, and let $Y_1, \ldots, Y_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$. We are interested in evaluating the standardized effect size
$$\theta = \frac{\mu_Y - \mu_X}{\sigma}.$$
In particular, we consider the following hypothesis:
$$\Theta_0 = \{\theta : |\theta| \le \delta\} \quad \text{versus} \quad \Theta_1 = \{\theta : |\theta| > \delta\},$$
for some prespecified $\delta > 0$. To compute the Bayes factor, we assume the following priors
$$\begin{aligned}
&\mu_X \sim \mathcal{N}(0, 100^2),\\
&\sigma^2 \sim \mathcal{I}\text{nverse-}\mathcal{G}\text{amma}(10^{-2}, 10^{-2}),\\
&\theta \mid H_0 \sim \pi_0(\theta) \equiv \mathcal{U}\text{niform}(-\delta, \delta),\\
&\theta \mid H_1 \sim \pi_1(\theta) \equiv \mathcal{N}(0, \tau^2) I_{|\theta| > \delta},\\
&\mu_Y = \mu_X + \theta \times \sigma,
\end{aligned}$$
where τ in $\theta \in H_1$ is a parameter to be specified: a larger τ makes π_1 more deviated from π_0 and more diffuse on Θ_1. In this problem, nuisance parameter λ includes μ_X and σ. We implement the three-step algorithm above in Stan with its no-U-turn sampler (Hoffman and Gelman 2014).

We calculated the log Bayes factor, that is, $\log \frac{f(y|\pi_1,\pi_\lambda)}{f(y|\pi_0,\pi_\lambda)}$ for $\delta = 0.1$, which is generally considered as a small effect size, over a range of values of τ in π_1. The result is plotted in Figure 2. All these log Bayes factors are between $(0.2, -1.3)$, indicating mostly slight more support for H_0 over H_1. It is also seen that the log Bayes factor decreases as τ increases, similar to our example in Section 3.3. Computationally, we ran four Monte Carlo Markov chains in Stan with 5,00,000 iterations in each chain for every calculated Bayes factor. Standard MCMC diagnostics

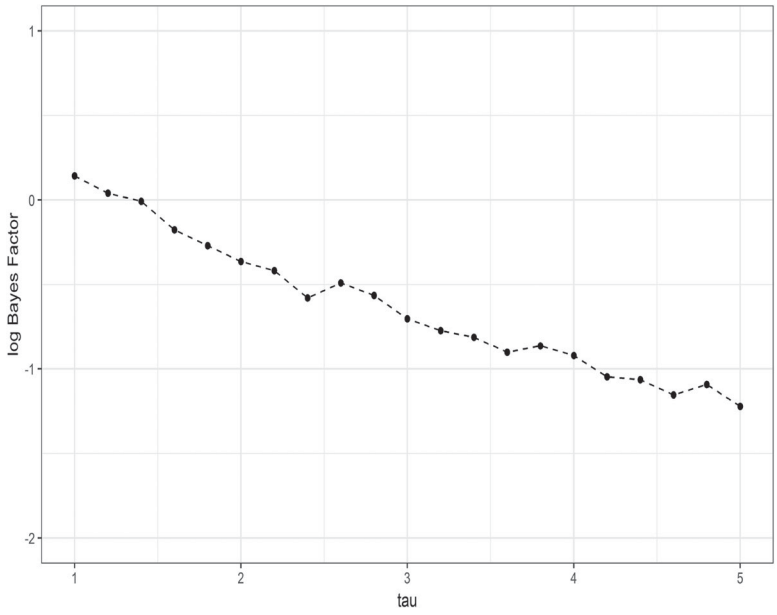

Figure 2. Log Bayes factors for the blood pressure data.

Table 3. Data format for the *i*th study.

	Death	Non-death	Total
Beta-blocker	$y_{i,1}$	$n_{i,1} - y_{i,1}$	$n_{i,1}$
Placebo	$y_{i,0}$	$n_{i,0} - y_{i,0}$	$n_{i,0}$

Table 4. Log Bayes factors for different values of δ.

δ	0.10	0.15	0.20	0.25	0.30	0.35	0.40
$\log \frac{f(y\|\pi_1,\pi_\lambda)}{f(y\|\pi_0,\pi_\lambda)}$	2.72	1.51	0.24	−0.95	−2.18	−3.60	−5.19

tools such as Rhat show that the MCMC chains converged to the target stationary distribution. The effective sample size of each MCMC chains is about 250,000.

It is noted that the Bayes factor for two-sample t-test setting with point null $H_0 : \mu_X - \mu_Y = 0$ has been extensively studied (Gönen et al. 2005; Rouder et al. 2009; Wang and Liu 2016) including the R package `BayesFactor` by Morey and Rouder (2018). The focus of this section, however, is on the computational algorithm for the Bayes factor.

4.2. Example 2: Meta-Analysis With a Hierarchical Bayes Model

Here we consider meta-analysis of a dataset originally published in Table 10 of Yusuf et al. (1985), which contains mortality data across 22 studies of patients who were treated with either a beta-blocker or placebo after experiencing a heart attack. The dataset is also reproduced and analyzed in Gelman et al. (2013, sec. 5.6). The data from each study forms a 2×2 table with the ith study as in Table 3.

For the ith study, let $p_{i,j}, j = 0, 1$, be the underlying probability of death for the placebo and beta-blocker group, respectively, and let θ_i be the corresponding log odds ratio, that is,

$$\theta_i = \log\left(\frac{p_{i,1}}{1-p_{i,1}} \Big/ \frac{p_{i,0}}{1-p_{i,0}}\right) = \text{logit}(p_{i,1}) - \text{logit}(p_{i,0}).$$

We assume $\theta_i \sim \mathcal{N}(\theta, \sigma^2)$, where θ, the mean of individual θ_i, is the focus of our inference. In particular, we compare $\Theta_0 = \{\theta : |\theta| \le \delta\}$ versus $\Theta_1 = \{\theta : |\theta| > \delta\}$. Our hierarchical Bayesian model is specified as follows:

$$\begin{aligned} y_{i,j} &\sim \mathcal{B}\text{inomial}(n_{i,j}, p_{i,j}), \\ \text{logit}(p_{i,0}) &\sim \mathcal{N}(0, 10^2), \\ \theta_i &\sim \mathcal{N}(\theta, \sigma^2), \\ \theta \mid H_0 &\sim \pi_0 \equiv \mathcal{U}\text{niform}[-\delta, \delta], \\ \theta \mid H_1 &\sim \pi_1 \equiv \mathcal{N}(0, \tau^2) I_{|\theta|>\delta}, \\ \sigma^2 &\sim \mathcal{I}\text{nverse-}\mathcal{G}\text{amma}(10^{-2}, 10^{-2}). \end{aligned}$$

For $\tau = 2.5\,\delta$, we use the three-step algorithm above to estimate the log Bayes factor. The results are given in Table 4 for five different values of effect size δ. Note that the log Bayes factor shows strong support for H_1 over H_0 for $\delta = 0.10$ whereas it presents weak support of H_1 for $\delta = 0.20$. When δ increases beyond 0.25, however, the log Bayes factor starts to show greater support of H_0 than H_1. Therefore, it seems that the most likely effect size δ is between 0.20 and 0.25. Computationally, we ran four Markov chains with 100,000 iterations in each chain with effective sample size about 38,000 for each chain. Based on Rhat, the MCMC chains converged to the target stationary distribution.

Appendix: Proof of Theorem 1

Proof. The following proof is for $\theta^* = \theta_0 - \delta$. The case of $\theta^* = \theta_0 + \delta$ can be shown similarly. We shall assume that mild conditions noted in Section 3 are satisfied. The Bernstein-Vonn Mises theorem then states that the posterior distribution $\sqrt{n}\theta \mid y$ can be approximated by normal distribution $\mathcal{N}\left(\sqrt{n}\hat{\theta}, \sigma^2\right)$, where $\sigma^2 = I_{\theta^*}^{-1}$ and I_{θ^*} is the individual

Fisher information matrix of $f(y \mid \theta)$ at $\theta = \theta^*$. It then follows that

$$\Pr(\theta < \theta^* \mid y) = \Pr(\sqrt{n}\theta < \sqrt{n}\theta^*) \mid y) = \Phi\left(\frac{\sqrt{n}(\theta^* - \hat{\theta})}{\sigma}\right) + o(1). \tag{A.1}$$

The asymptotic posterior probability on the right of Equation (A.1) depends on y through MLE $\hat{\theta}$. Further consider the variation of $y \sim f(y \mid \theta^*)$. The $\hat{\theta}$ itself has distribution $\sqrt{n}(\theta^* - \hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, where the convergence is in distribution. Applying this variation of $\hat{\theta}$ to the right side of Equation (A.1), we have that $\Pr(\theta < \theta^* \mid y)$, as a function of y, converges in distribution to $\mathcal{U}$niform$(0, 1)$. It can be further shown that, for $\theta^* = \theta_0 + \delta$, $\Pr(\theta \in \Theta_0 \mid y) = \Pr(\theta < \theta^* \mid y) + o(1)$ as $n \to \infty$. We arrive at the first conclusion of Theorem 1.

Now let π_0 and π_1 be the two priors specified in the Bayes factor approach. Let θ^* be a neutral point as before. Define

$$\eta_1^* = \frac{\pi_0(\theta^*)}{\pi_0(\theta^*) + \pi_1(\theta^*)}$$

and

$$\pi^*(\theta) = (1 - \eta_1^*)\pi_0(\theta) + \eta_1^*\pi_1(\theta).$$

Then $\pi^*(\theta)$ is then positive and absolute continuous in the neighborhood of θ^*. Let $y \sim f(y \mid \theta^*)$. It follows from the result above that

$$\log \frac{\Pr(\theta \in \Theta_1 \mid y)}{\Pr(\theta \in \Theta_0 \mid y)} \xrightarrow{d} \mathcal{L}\text{ogistic}(0, 1),$$

where $\Pr(\theta \in \Theta_j \mid y)$ is with respect to prior π^*. We then have, from Lemma 1, that

$$\log \frac{f(y \mid \pi_1)}{f(y \mid \pi_0)} \xrightarrow{d} \log \frac{1 - \eta_1^*}{\eta_1^*} + \mathcal{L}\text{ogistic}(0, 1),$$

where $\frac{1-\eta_1^*}{\eta_1^*} = \frac{\pi_1(\theta^*)}{\pi_0(\theta^*)}$.

The proof of Theorem 1 is complete. □

Acknowledgments

The authors are deeply grateful to the three anonymous reviewers, an associate editor, and editor Dr. Jeske for their expert guidance in the revision of the article.

References

Berger, J. O. (2013), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer. [1]

Berger, J., and Pericchi, L. (2015), "Bayes Factors," *Wiley StatsRef: Statistics Reference Online*, DOI: 10.1002/9781118445112.stat00224.pub2. [1]

Blume, J. D., Greevy, R. A., Welty, V. F., Smith, J. R., and Dupont, W. D. (2019), "An Introduction to Second-Generation p-Values," *The American Statistician*, 73, 157–167. [1]

Carlin, B. P., and Louis, T. A. (2008), *Bayesian Methods for Data Analysis* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC. [1]

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76, 1–32. [2,6]

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321. [6]

Cohen, J. (1994), "The Earth Is Round ($p < .05$)," *American Psychologist*, 49, 997–1003. [1]

Edwards, J. R., and Berry, J. W. (2010), "The Presence of Something or the Absence of Nothing: Increasing Theoretical Precision in Management Research," *Organizational Research Methods*, 13, 668–689. [1]

Freedman, L., Lowe, D., and Macaskill, P. (1984), "Stopping Rules for Clinical Trials Incorporating Clinical Opinion," *Biometrics*, 40, 575–586. [1]

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [6]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2013), *Bayesian Data Analysis* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC. [6,7]

Gönen, M., Johnson, W. O., Lu, Y., and Westfall, P. H. (2005), "The Bayesian Two-Sample t Test," *The American Statistician*, 59, 252–257. [6,7]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [6]

Harms, C., and Lakens, D. (2018), "Making 'Null Effects' Informative: Statistical Techniques and Inferential Frameworks," *Journal of Clinical and Translational Research*, 3, 382–393. [1]

Hobbs, B. P., and Carlin, B. P. (2007), "Practical Bayesian Design and Analysis for Drug and Device Clinical Trials," *Journal of Biopharmaceutical Statistics*, 18, 54–80. [1]

Hoffman, M. D., and Gelman, A. (2014), "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, 15, 1593–1623. [6]

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [3,4]

Kass, R. E., and Wasserman, L. (1995), "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion," *Journal of the American Statistical Association*, 90, 928–934. [1]

Kruschke, J. K. (2011), "Bayesian Assessment of Null Values via Parameter Estimation and Model Comparison," *Perspectives on Psychological Science*, 6, 299–312. [1]

——— (2013), "Bayesian Estimation Supersedes the t Test," *Journal of Experimental Psychology: General*, 142, 573. [1]

——— (2014), *Doing Bayesian Data Analysis: A Tutorial With R, JAGS, and Stan* (2nd ed.), London: Academic Press. [1]

——— (2018), "Rejecting or Accepting Parameter Values in Bayesian Estimation," *Advances in Methods and Practices in Psychological Science*, 1, 270–280. [1,2]

Kypraios, T., and O'Neill, P. D. (2018), "Bayesian Nonparametrics for Stochastic Epidemic Models," *Statistical Science*, 33, 44–56, DOI: 10.1214/17-STS617. [6]

Lakens, D. (2017), "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses," *Social Psychological and Personality Science*, 8, 355–362. [1,2]

Lakens, D., and Delacre, M. (2018), "Equivalence Testing and the Second Generation P-Value," PsyArXiv, DOI: 10.31234/osf.io/7k6ay. [2]

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., and Dienes, Z. (2018), "Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests," *The Journals of Gerontology*, Series B, DOI: 10.1093/geronb/gby065 (in press). [1]

Liao, J. G., Liao, D., and Berg, A. (2016), "Calibrated Bayes Factors in Assessing Genetic Association Models," *The American Statistician*, 70, 250–256. [4]

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility," *Statistics and Computing*, 10, 325–337. [2,6]

Lyle, R. M., Melby, C. L., Hyner, G. C., Edmondson, J. W., Miller, J. Z., and Weinberger, M. H. (1987), "Blood Pressure and Metabolic Effects of Calcium Supplementation in Normotensive White and Black Men," *JAMA*, 257, 1772–1776. [6]

Meehl, P. E. (1978), "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology," *Journal of Consulting and Clinical Psychology*, 46, 806–834. [1]

Meng, X. L., and Wong, W. H. (1996), "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 6, 831–860. [6]

Morey, R. D., and Rouder, J. N. (2011), "Bayes Factor Approaches for Testing Interval Null Hypotheses," *Psychological Methods*, 16, 406. [1]

——— (2018), "BayesFactor: Computation of Bayes Factors for Common Designs," R Package Version 0.9.12-4.2 ed., available at *https://CRAN.R-project.org/package=BayesFactor*. [7]

Morey, R. D., Rouder, J. N., Pratte, M. S., and Speckman, P. L. (2011), "Using MCMC Chain Outputs to Efficiently Estimate Bayes Factors," *Journal of Mathematical Psychology*, 55, 368–378. [6]

Mulder, J., and Wagenmakers, E.-J. (2016), "Editors' Introduction to the Special Issue 'Bayes Factors for Testing Hypotheses in Psychological Research: Practical Relevance and New Developments'," *Journal of Mathematical Psychology*, 72, 1–5. [1]

Plummer, M. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, eds. K. Hornik, F. Leisch, and A. Zeileis (eds.), available at *https://www.r-project.org/conferences/DSC-2003/*. [2,6]

Robert, C. P. (2007), *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (2nd ed.), New York: Springer. [2,3]

——— (2014), "On the Jeffreys-Lindley Paradox," *Philosophy of Science*, 81, 216–232. [4]

Rogers, J. L., Howard, K. I., and Vessey, J. T. (1993), "Using Significance Tests to Evaluate Equivalence Between Two Experimental Groups," *Psychological Bulletin*, 113, 553–565. [1,2]

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009), "Bayesian *t* Tests for Accepting and Rejecting the Null Hypothesis," *Psychonomic Bulletin & Review*, 16, 225–237. [7]

Vaart, A. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press, pp. 138–152. [3]

Wang, M., and Liu, G. (2016), "A Simple Two-Sample Bayesian *t*-Test for Hypothesis Testing," *The American Statistician*, 70, 195–201. [6,7]

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on *p*-Values: Context, Process, and Purpose," *The American Statistician*, 70, 129–133. [1]

Weinberg, M. D. (2012), "Computing the Bayes Factor From a Markov Chain Monte Carlo Simulation of the Posterior Distribution," *Bayesian Analysis*, 7, 737–770. [6]

Wellek, S. (2010), *Testing Statistical Hypotheses of Equivalence and Noninferiority* (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC. [1,2]

Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985), "Beta Blockade During and After Myocardial Infarction: An Overview of the Randomized Trials," *Progress in Cardiovascular Diseases*, 27, 335–371. [7]