



## Computing Bayes factors to measure evidence from experiments: An extension of the BIC approximation

Thomas J. Faulkenberry

Department of Psychological Sciences, Tarleton State University, USA,  
e-mail: faulkenberry@tarleton.edu

### SUMMARY

Bayesian inference affords scientists powerful tools for testing hypotheses. One of these tools is the Bayes factor, which indexes the extent to which support for one hypothesis over another is updated after seeing the data. Part of the hesitance to adopt this approach may stem from an unfamiliarity with the computational tools necessary for computing Bayes factors. Previous work has shown that closed-form approximations of Bayes factors are relatively easy to obtain for between-groups methods, such as an analysis of variance or  $t$ -test. In this paper, I extend this approximation to develop a formula for the Bayes factor that directly uses information that is typically reported for ANOVAs (e.g., the  $F$  ratio and degrees of freedom). After giving two examples of its use, I report the results of simulations which show that even with minimal input, this approximate Bayes factor produces similar results to existing software solutions.

**Key words:** Bayes factors, Bayesian inference, analysis of variance, hypothesis testing

### 1. Introduction

Hypothesis testing is the primary tool for statistical inference across much of the biological and behavioral sciences. As such, most scientists are trained in classical null hypothesis significance testing (NHST). The scenario for testing a hypothesis is likely familiar to most readers of this journal. Suppose one wants to test a specific research hypothesis (e.g., some treatment has an effect on some outcome measure). NHST works by first assuming a *null hypothesis* (e.g., the treatment has *no effect*) and then computing some test statistic for a sample of data. This sample test statistic is then compared

with a hypothetical distribution of test statistics that would arise if the null hypothesis were true. If the sample's test statistic is in the tail of the distribution (that is, it should occur with low probability), the scientist decides to *reject* the null hypothesis in favor of the alternative hypothesis. Further, the  $p$ -value, which indicates how surprising the sample would be if the null hypothesis were true, is often taken as a measure of evidence: the lower the  $p$ -value, the *stronger* the evidence.

While orthodox across many disciplines, NHST does have philosophical criticisms (Wagenmakers, 2007). Also, the  $p$ -value is prone to misinterpretation (Gigerenzer, 2004; Hoekstra et al., 2014). Finally, NHST is ideally suited to providing support for the alternative hypothesis, but the procedure does not work in the case where one wants to measure support for the null hypothesis. That is, we can reject the null, but we cannot *accept* the null. To overcome this limitation, we can use an alternative method for testing hypotheses that is based on Bayesian inference: the Bayes factor.

## 2. The Bayes factor

### 2.1. Background

Bayesian inference is a method of measurement that is based on the computation of  $P(H | D)$ , which is called the *posterior probability* of a hypothesis  $H$ , given data  $D$ . Bayes' theorem casts this probability as

$$P(H | D) = \frac{P(D | H) \cdot P(H)}{P(D)}. \quad (1)$$

One may think of Equation 1 in the following manner: before observing data  $D$ , one assigns a *prior probability*  $P(H)$  to hypothesis  $H$ . After observing data, one can then update this prior probability to a *posterior probability*  $P(H | D)$  by multiplying the prior  $P(H)$  by the likelihood  $P(D | H)$ . This product is then rescaled to a probability distribution (i.e., total probability = 1) by dividing by the *marginal probability*  $P(D)$ .

Bayes' theorem provides a natural way to test hypotheses. Suppose we have two competing hypotheses: an alternative hypothesis  $H_1$  and a null hypothesis  $H_0$ . We can directly compare the posterior probabilities of  $H_1$  and  $H_0$  by computing their ratio; that is, we can compute the *posterior odds* in favor of  $H_1$  over  $H_0$  as  $P(H_1 | D)/P(H_0 | D)$ . Using Bayes' theorem (Equation 1), it is trivial to see that

$$\underbrace{\frac{P(H_1 | D)}{P(H_0 | D)}}_{\text{posterior odds}} = \underbrace{\frac{P(D | H_1)}{P(D | H_0)}}_{\text{Bayes factor}} \cdot \underbrace{\frac{P(H_1)}{P(H_0)}}_{\text{prior odds}}. \quad (2)$$

This equation can also be interpreted in terms of the “updating” metaphor that was explained above. Specifically, the posterior odds are equal to the prior odds multiplied by an updating factor. This updating factor is equal to the ratio of likelihoods  $P(D | H_1)$  and  $P(D | H_0)$ , and is called the *Bayes factor* (Jeffreys, 1961). Intuitively, the Bayes factor can be interpreted as the *weight of evidence* provided by a set of data  $D$ . For example, suppose that one assigned the prior odds of  $H_1$  and  $H_0$  equal to 1; that is,  $H_1$  and  $H_0$  are *a priori* assumed to be equally likely. Then, suppose that after observing data  $D$ , the Bayes factor was computed to be 10. Now, the posterior odds (the odds of  $H_1$  over  $H_0$  *after* observing data) are 10:1 in favor of  $H_1$  over  $H_0$ . As such, the Bayes factor provides an easily interpretable measure of the evidence in favor of  $H_1$ .

In order to help with interpreting Bayes factors, various classification schemes have been proposed. One simple scheme is a four-way classification proposed by Raftery (1995), where Bayes factors between 1 and 3 are considered *weak* evidence; those between 3 and 20 constitute *positive* evidence; those between 20 and 150 constitute *strong* evidence; and those beyond 150 are considered *very strong* evidence.

Note that in the discussion above, there was no specific assumption about the order in which we addressed  $H_1$  and  $H_0$ . If instead we wanted to assess the weight of evidence in favor of  $H_0$  over  $H_1$ , Equation 2 could simply be adjusted by taking reciprocals. As such, implied direction is important when computing Bayes factors, so one must be careful to define notation when representing Bayes factors. A common convention is to define  $BF_{10}$  as the Bayes factor for  $H_1$  over  $H_0$ ; similarly,  $BF_{01}$  would represent the Bayes factor for  $H_0$  over  $H_1$ . Note that  $BF_{10} = 1/BF_{01}$ .

In summary, the Bayes factor provides an index of preference for one hypothesis over another that has some advantages over NHST. First, the Bayes factor tells us by how much a data sample should update our belief in one hypothesis over a competing one. Second, though NHST does not allow one to accept a null hypothesis, doing so within a Bayesian framework makes perfect sense. Given these advantages, it may be surprising that Bayesian

inference has not been used more often in the empirical sciences. One reason for the lack of more widespread adoption may be that Bayes factors are quite difficult to compute. We tackle this issue in the next section.

## 2.2. Computing Bayes factors

As an example, suppose we are interested in computing the Bayes factor for a null hypothesis  $H_0$  over an alternative hypothesis  $H_1$ , given data  $D$ . Recall from Equation 2 that this Bayes factor (denoted  $BF_{01}$ ) is equal to

$$BF_{01} = \frac{P(D | H_0)}{P(D | H_1)}.$$

While this equation may seem conceptually quite simple, it is computationally much more difficult. This is because in order to compute the numerator and denominator, one must parameterize the hypotheses (or *models*, to be more clear), and then each likelihood is computed by conditioning over all possible parameter values and summing over this set. Since these potential parameter values are often over a continuous parameter space, this computation requires integration, and thus the formula for the Bayes factor amounts to

$$BF_{01} = \frac{\int_{\theta \in \Theta_0} P(D | H_0, \theta) \pi_0(\theta) d\theta}{\int_{\theta \in \Theta_1} P(D | H_1, \theta) \pi_1(\theta) d\theta} \quad (3)$$

where  $\Theta_0$  and  $\Theta_1$  are the parameter spaces for models  $H_0$  and  $H_1$ , respectively, and  $\pi_0$  and  $\pi_1$  are the prior probability density functions of the parameters of  $H_0$  and  $H_1$ , respectively.

Thus, in order to compute  $BF_{01}$ , one must specify the priors  $\pi_0$  and  $\pi_1$  for  $H_0$  and  $H_1$ . Further, the integrals usually do not have closed-form solutions, so numerical integration techniques are necessary. These requirements cause a computation of the Bayes factor to be inaccessible to all but the most ardent researchers who have at least a more-than-modest amount of mathematical training.

Fortunately, there are an increasing number of solutions that avoid a direct encounter with computations of the above type. Recently, researchers have proposed *default* priors for standard experimental designs such as  $t$ -tests (Rouder et al., 2009; Morey and Rouder, 2011) and ANOVA (Rouder et al., 2012). These default priors are implemented in software packages such as the R package BayesFactor (Morey and Rouder, 2015), and as such, have provided a user-friendly method for researchers to compute Bayes factors

without the computational overhead needed in Equation 3. While these software solutions work quite well for computing Bayes factors from raw data, they are somewhat limited in the following context. Suppose that in the course of reading some published literature, a researcher comes across a result that is presented as “nonsignificant”, with associated test statistic  $F(1, 23) = 2.21$ ,  $p = 0.15$ . In an NHST context, this nonsignificant result does not provide evidence *for* the null hypothesis; rather, it just implies that we *cannot reject* the null. A natural question would be what, if any, support does this result provide for the null hypothesis? Of course, a Bayes factor would be useful here, but without the raw data, we cannot use the previously mentioned software solutions. To this end, it would be advantageous if there were some easy way to compute a Bayes factor directly from the reported test statistic.

It turns out that this computation is indeed possible, at least in certain cases. In the following, I will show how one particular method for computing Bayes factors (the BIC approximation; Raftery, 1995) can be adapted to solve this problem, thus allowing researchers to compute approximate Bayes factors from summary statistics alone (with no need for raw data). Further, I will show through simulations that this method compares well with the default Bayes factors for ANOVA developed by Rouder et al. (2012).

### 3. The BIC approximation of the Bayes factor

Wagenmakers (2007) demonstrated a method (based on earlier work by Raftery, 1995) for computing approximate Bayes factors using the BIC (Bayesian Information Criterion). For a given model  $H_i$ , the BIC is defined as

$$\text{BIC}(H_i) = -2 \log L_i + k_i \cdot \log n,$$

where  $n$  is the number of observations,  $k_i$  is the number of free parameters of model  $H_i$ , and  $L_i$  is the maximum likelihood for model  $H_i$ . He then showed that the Bayes factor for  $H_0$  over  $H_1$  can be approximated as

$$BF_{01} \approx \exp(\Delta\text{BIC}_{10}/2), \quad (4)$$

where  $\Delta\text{BIC}_{10} = \text{BIC}(H_1) - \text{BIC}(H_0)$ . Further, Wagenmakers (2007) showed that when comparing an alternative hypothesis  $H_1$  with a null hypothesis  $H_0$ ,

$$\Delta\text{BIC}_{10} = n \log \left( \frac{SSE_1}{SSE_0} \right) + (k_1 - k_0) \log n. \quad (5)$$

In this equation,  $SSE_0$  and  $SSE_1$  represent the sum of squares for the error terms in models  $H_0$  and  $H_1$ , respectively. Both Wagenmakers (2007) and Masson (2011) give excellent examples of how to use this approximation to compute Bayes factors, assuming one is given information about  $SSE_0$  and  $SSE_1$ , as is the case with most statistical software. However, we will now consider the situation where one is given the statistical summary (i.e.,  $F(1, 23) = 2.21$ ), but not the ANOVA output.

Suppose we wish to examine an effect of some independent variable with associated  $F$ -ratio  $F(df_1, df_2)$ , where  $df_1$  represents the degrees of freedom associated with the manipulation, and  $df_2$  represents the degrees of freedom associated with the error term. Then,  $F = \frac{SS_1/df_1}{SS_2/df_2} = \frac{SS_1}{SS_2} \cdot \frac{df_2}{df_1}$ , where  $SS_1$  and  $SS_2$  are the sum of squared errors associated with the manipulation and the error term, respectively.

From Equation 5, we see that

$$\begin{aligned} \Delta\text{BIC}_{10} &= n \log \left( \frac{SSE_1}{SSE_0} \right) + (k_1 - k_0) \log n \\ &= n \log \left( \frac{SS_2}{SS_1 + SS_2} \right) + df_1 \log n. \end{aligned}$$

This equality holds because  $SSE_1$  represents the sum of squares that is not explained by  $H_1$ , which is simply  $SS_2$  (the error term). Similarly,  $SSE_0$  is the sum of squares not explained by  $H_0$ , which is the sum of  $SS_1$  and  $SS_2$  (see Wagenmakers, 2007, p. 799). Finally, in the context of comparing  $H_1$  and  $H_0$  in an ANOVA design, we have  $k_1 - k_0 = df_1$ . Now, we can use algebra to re-express  $\Delta\text{BIC}_{10}$  in terms of  $F$ :

$$\begin{aligned} \Delta\text{BIC}_{10} &= n \log \left( \frac{SS_2}{SS_1 + SS_2} \right) + df_1 \log n = n \log \left( \frac{1}{\frac{SS_1}{SS_2} + 1} \right) + df_1 \log n \\ &= n \log \left( \frac{\frac{df_2}{SS_2} \cdot \frac{df_1}{df_1} + \frac{df_2}{df_1}}{\frac{df_2}{SS_2} \cdot \frac{df_1}{df_1} + \frac{df_2}{df_1}} \right) + df_1 \log n = n \log \left( \frac{\frac{df_2}{df_1}}{F + \frac{df_2}{df_1}} \right) + df_1 \log n \\ &= n \log \left( \frac{df_2}{F df_1 + df_2} \right) + df_1 \log n. \end{aligned}$$

Substituting this into Equation 4, we can compute:

$$\begin{aligned} BF_{01} &\approx \exp(\Delta\text{BIC}_{10}/2) = \exp \left[ \frac{1}{2} \left( n \log \left( \frac{df_2}{F df_1 + df_2} \right) + df_1 \log n \right) \right] \\ &= \exp \left[ \frac{n}{2} \log \left( \frac{df_2}{F df_1 + df_2} \right) + \frac{df_1}{2} \log n \right] = \left( \frac{df_2}{F df_1 + df_2} \right)^{n/2} \cdot n^{df_1/2} \\ &= \sqrt{\frac{df_2^n \cdot n^{df_1}}{(F df_1 + df_2)^n}} = \sqrt{\frac{n^{df_1}}{\left( \frac{F df_1}{df_2} + 1 \right)^n}}. \end{aligned}$$

Rearranging this last expression slightly yields the approximation:

$$BF_{01} \approx \sqrt{n^{df_1} \left(1 + \frac{F df_1}{df_2}\right)^{-n}} \quad (6)$$

Practically speaking, the approximation given in Equation 6 offers nothing new over the previous formulations of the BIC approximation given in Wagenmakers (2007) and Masson (2011). However, it does have two advantages over these previous formulations. First, one can directly take reported ANOVA statistics (e.g., sample size, degrees of freedom, and the  $F$ -ratio) and compute  $BF_{01}$  without having to compute  $SSE_0$  or  $SSE_1$ . We should note that Masson (2011) correctly mentions that  $SSE_1/SSE_0 = 1 - \eta_p^2$ , so if a paper reports  $\eta_p^2$ , the need for computing  $SSE_0$  and  $SSE_1$  is nullified. However, the method of Masson (2011) is still essentially a two-step process; one first computes  $\Delta BIC_{10}$ , which in turn is used to compute  $BF_{01}$ . In contrast, the expression derived in Equation 6 is a one-step process that can easily be implemented using a scientific calculator or a simple spreadsheet.

#### 4. Example computations

In this section, we will discuss two examples of using Equation 6 to compute Bayes factors. In the first example, I will show how to compute and interpret a Bayes factor for a reported null effect in the field of experimental psychology. In the second example, I will show how to modify Equation 6 to work with an independent samples  $t$ -test.

##### 4.1. Example 1

Sevos et al. (2016) performed an experiment to assess whether schizophrenics could internally simulate motor actions when perceiving graspable objects. The evidence for such internal simulation comes from a statistical interaction between response-orientation compatibility and the presence of an individual name prime. Sevovs et al. reported that in a sample of  $n = 18$  schizophrenics, there was no interaction between this compatibility and name prime,  $F(1, 17) = 2.584$ ,  $p = 0.126$ . Critically, Sevovs et al. claimed that this null effect was evidence for the absence of sensorimotor simulation, which, as they indicate, would imply that schizophrenics would have to rely on higher cognitive processes for even the most simple daily tasks. This claim is based on a null effect, which as pointed out earlier, is problematic

for a null hypothesis testing framework. I will now show how to compute a Bayes factor  $BF_{01}$  to assess the evidence for this null effect.

To this end, we use Equation 6 to compute

$$BF_{01} \approx \sqrt{n^{df_1} \left(1 + \frac{F df_1}{df_2}\right)^{-n}} = \sqrt{18^1 \left(1 + \frac{2.584 \cdot 1}{17}\right)^{-18}} = 1.187.$$

This Bayes factor can be interpreted as follows: after seeing the data, our belief in the null hypothesis should increase only by a factor of 1.19. In other words, this data is not very informative toward our belief in the null, which implies that the claim of a null effect in Sevos et al. (2016) may be somewhat optimistic. According the classification scheme of Raftery (1995), this result provides *weak* evidence for the null.

#### 4.2. Example 2

Borota et al. (2014) observed that with a sample of  $n = 73$  participants, those who received 200 mg of caffeine performed significantly better on a test of object memory compared with a control group of participants who received a placebo,  $t(71) = 2.0$ ,  $p = 0.049$ . Borota et al. (2014) claimed this result as evidence that caffeine enhances memory consolidation.

As before, we can measure the evidence provided from this data sample by computing a Bayes factor. However, we note that because Equation 6 casts the Bayes factor in terms of an  $F$ -ratio, it may not be immediately obvious whether we can use Equation 6 in this context. It turns out to be straightforward to modify Equation 6 to work for an independent samples  $t$ -test. All we need are two simple transformations: (1)  $F = t^2$ , and (2)  $df_1 = 1$ . Applying these to Equation 6, we get

$$BF_{01} \approx \sqrt{n^{df_1} \left(1 + \frac{F df_1}{df_2}\right)^{-n}} = \sqrt{n^1 \left(1 + \frac{t^2(1)}{df_2}\right)^{-n}} = \sqrt{n \left(1 + \frac{t^2}{df_2}\right)^{-n}}$$

We can now apply this equation to the reported results of Borota et al. (2014). We see that

$$BF_{01} \approx \sqrt{n \left(1 + \frac{t^2}{df_2}\right)^{-n}} = \sqrt{73 \left(1 + \frac{(2.0)^2}{71}\right)^{-73}} = 1.16$$

So, perhaps counterintuitively, the significant result reported in Borota et al. (2014) turns out to be weak evidence in support of the *null*! Such



results are an example of Lindley’s paradox (Lindley, 1957), where “significant”  $p$  values between 0.04 and 0.05 can actually imply evidence in favor of the null when analyzed in a Bayesian framework.

### 5. Simulations: BIC approximation versus default Bayesian ANOVA

At this stage, it is clear that Equation 6 provides a straightforward method for computing an approximate Bayes factor, especially in cases when one is given only minimal output from a reported ANOVA or  $t$  test. However, it is not yet clear to what extent this BIC approximation would result in the same decision if a Bayesian analysis of variance (e.g., Rouder et al., 2012) were performed on the raw data. To answer this question, I performed a series of simulations.

Each simulation consisted of 1000 randomly generated data sets under a  $2 \times 3$  factorial design. The choice of this design is to replicate experimental conditions that are common across many applications in the biological and behavioral sciences. Further, I simulated varying common levels of statistical power in these experiments by testing three different cell-size conditions:  $n = 20, 50$  or  $80$ . Specifically, each data set consisted of a vector  $\mathbf{y}$  generated as

$$y_{ijk} = \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$$

where  $i = 1, 2$ ,  $j = 1, 2, 3$ , and  $k = 1, \dots, n$ . The “effects”  $\alpha$ ,  $\tau$ , and  $\gamma$  were generated from multivariate normal distributions with mean 0 and variance  $g$ , yielding three different effect sizes obtained by setting  $g = 0, 0.05$ , and  $0.2$  (as in Wang, 2017). In all, there were nine different simulations, generated by crossing the three cell sizes ( $n = 20, 50, 80$ ) with the three effect sizes ( $g = 0, 0.05, 0.2$ ).

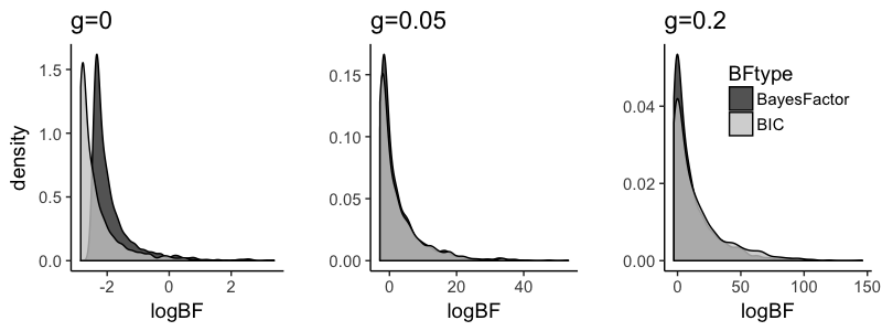
For each data set, I computed (1) a Bayesian ANOVA using the BayesFactor package in R (Morey and Rouder, 2015) and (2) the BIC approximation using Equation 6 from the traditional ANOVA. Bayes factors were computed as  $BF_{10}$  to assess evidence in favor of the alternative hypothesis over the null hypothesis. Similar to Wang (2017), I set the decision criterion to select the alternative hypothesis if  $\log(BF) > 0$ , and the null hypothesis otherwise. Because the different cell sizes resulted in similar outcomes, for brevity I only report the  $n = 50$  cell size condition in the summaries below. Also note that all BayesFactor models were fitted with a “wide” prior, which is roughly equivalent to the unit-information prior used by Raftery (1995) for the BIC approximation.

First, I will report the results of computing Bayes factors for the main effect  $\alpha$  in each of the effect size conditions  $g = 0$ ,  $g = 0.05$ , and  $g = 0.2$ . Five-number summaries for  $\log(BF)$  are reported for the  $n = 50$  simulation in Table 1, as well as the proportion of simulated data sets for which the Bayesian ANOVA and the BIC approximation from Equation 6 selected the same model.

As shown in Table 1, the BIC approximation from Equation 6 provides a similar distribution of Bayes factors compared to those computed from the BayesFactor package in R. Figure 1 shows this pattern of results quite clearly, as the kernel density plots for the two different types of Bayes factors exhibit a large amount of overlap for the  $g = 0.05$  and  $g = 0.2$  conditions. It is notable that the BIC approximation tended to underestimate the BayesFactor output in the  $g = 0$  case. However, as can be seen in the ‘‘Consistency’’ column of Table 1, regardless of effect size conditions, the two different types of Bayes factors resulted in the same decision in a large proportion of simulations (at least 98.4% of simulation trials).

**Table 1.** Summary of values for  $\log BF$  for main effect  $\alpha$  with cell size  $n = 50$

$g$	$BF$ type	Min	$Q_1$	Median	$Q_3$	Max	Consistency
0	BayesFactor	-2.40	-2.35	-2.14	-1.74	3.38	
	BIC	-2.85	-2.80	-2.59	-2.17	3.17	0.985
0.05	BayesFactor	-2.40	-1.83	-0.20	4.24	47.49	
	BIC	-2.85	-2.23	-0.41	4.44	53.29	0.984
0.2	BayesFactor	-2.40	-0.61	4.88	17.26	119.69	
	BIC	-2.85	-0.63	6.76	21.71	146.02	0.987

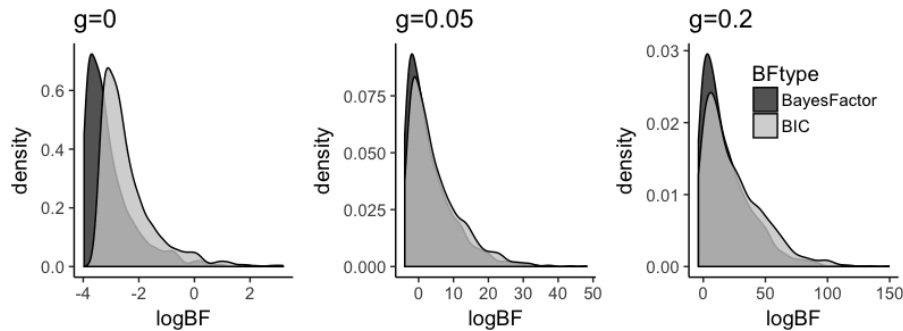


**Figure 1.** Kernel density plots of distributions of  $\log BF$  for main effect  $\alpha$ , presented as a function of Bayes factor type (BIC versus BayesFactor) and effect size ( $g = 0, 0.05, 0.2$ )

A similar picture emerges for the main effect  $\tau$ . As can be seen in Table 2 and Figure 2, the BIC approximation and the BayesFactor outputs are largely consistent and result in mostly the same model choice decisions. As with the results for main effect  $\alpha$ , there is some slight difference in the kernel density plots when simulating null effects (i.e., the condition  $g = 0$ ). However, both methods chose the same model on at least 92.7% of simulated trials, showing a good amount of consistency.

**Table 2.** Summary of values for  $\log BF$  for main effect  $\tau$  with cell size  $n = 50$

$g$	$BF$ type	Min	$Q_1$	Median	$Q_3$	Max	Consistency
0	BayesFactor	-3.97	-3.68	-3.31	-2.70	2.25	0.980
	BIC	-3.40	-3.10	-2.70	-2.05	3.19	
0.05	BayesFactor	-3.97	-1.89	1.15	6.28	43.55	0.927
	BIC	-3.40	-1.05	2.28	8.07	48.17	
0.2	BayesFactor	-3.92	3.19	12.00	28.51	134.04	0.952
	BIC	-3.34	6.03	17.15	36.58	149.39	

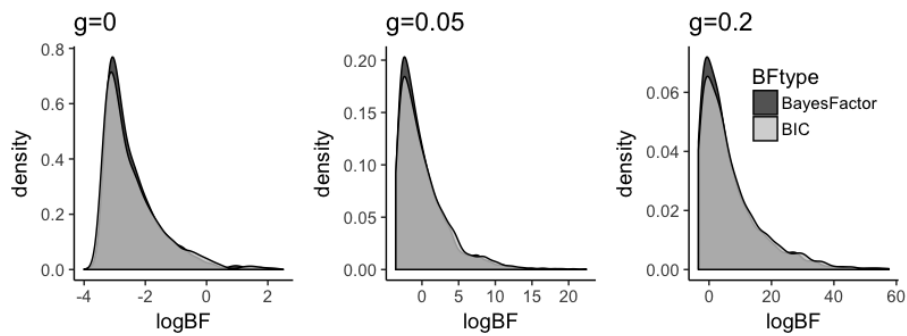


**Figure 2.** Kernel density plots of distributions of  $\log BF$  for main effect  $\tau$ , presented as a function of Bayes factor type (BIC versus BayesFactor) and effect size ( $g = 0, 0.05, 0.2$ )

Finally, we can see in Table 3 and Figure 3 that the BIC approximation closely mirrors the output of the BayesFactor package for the interaction effect  $\gamma$ . Indeed, the kernel density plots in Figure 3 show considerable overlap between the distributions of BIC values and the distributions of BayesFactor outputs, and this picture is consistent across all three effect sizes ( $g = 0, 0.05, 0.2$ ). As expected from this picture, both methods arrive at very similar model choices, picking the same model on at least 97% of trials.

**Table 3.** Summary of values for  $\log BF$  for interaction effect  $\gamma$  with cell size  $n = 50$ 

$g$	$BF$ type	Min	$Q_1$	Median	$Q_3$	Max	Consistency
0	BayesFactor	-3.97	-3.08	-2.72	-2.05	3.47	0.990
	BIC	-3.40	-3.12	-2.74	-1.98	4.03	
0.05	BayesFactor	-3.57	-2.30	-0.96	1.24	20.64	0.970
	BIC	-3.40	-2.27	-0.78	1.64	22.40	
0.2	BayesFactor	-3.37	-0.40	3.26	9.57	54.17	0.975
	BIC	-3.39	-0.20	3.84	10.72	57.61	

**Figure 3.** Kernel density plots of distributions of  $\log BF$  for the interaction effect  $\gamma$ , presented as a function of Bayes factor type (BIC versus BayesFactor) and effect size ( $g = 0, 0.05, 0.2$ )

## 6. Conclusion

The BIC approximation given in Equation 6 provides an easy-to-use estimate of Bayes factors for simple between-subject ANOVA and  $t$ -test designs. It requires only minimal information, which makes it well-suited for use in a meta-analytic context. In simulations, the estimates derived from Equation 6 compare favorably with Bayes factors computed using existing software solutions with raw data. Thus, the researcher can confidently add this BIC approximation to the ever-growing collection of Bayesian tools for scientific measurement.

## REFERENCES

- Borota D., Murray E., Keceli G., Chang A., Watabe J.M., Ly M., Yassa M.A. (2014): Post-study caffeine administration enhances memory consolidation in humans. *Nature Neuroscience* 17(2): 201–203. doi: 10.1038/nn.3623

- Gigerenzer G. (2004): Mindless statistics. *The Journal of Socio-Economics* 33(5): 587–606. doi:10.1016/j.socec.2004.09.033
- Hoekstra R., Morey R.D., Rouder J.N., Wagenmakers E.J. (2014): Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review* 21(5): 1157–1164. doi: 10.3758/s13423-013-0572-3
- Jeffreys H. (1961): *The Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Lindley D.V. (1957): A statistical paradox. *Biometrika* 44(1-2): 187–192. doi:10.1093/biomet/44.1-2.187
- Masson M.E.J. (2011): A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods* 43(3): 679–690. doi: 10.3758/s13428-010-0049-5
- Morey R.D., Rouder J.N. (2011): Bayes factor approaches for testing interval null hypotheses. *Psychological Methods* 16(4): 404–419. doi: 10.1037/a0024377
- Morey R.D., Rouder J.N. (2015): *BayesFactor: Computation of Bayes factors for common designs* [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=BayesFactor> (R package version 0.9.12-2)
- Raftery A.E. (1995): Bayesian model selection in social research. *Sociological Methodology* 25: 111–163. doi: 10.2307/271063
- Rouder J.N., Morey R.D., Speckman P.L., Province J.M. (2012): Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology* 56(5): 356–374. doi:10.1016/j.jmp.2012.08.001
- Rouder J.N., Speckman P.L., Sun D., Morey R.D., Iverson G. (2009): Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 16(2): 225–237. doi: 10.3758/pbr.16.2.225
- Sevos J., Grosselin A., Brouillet D., Pellet J., Massoubre C. (2016): Is there any influence of variations in context on object-affordance effects in schizophrenia? Perception of property and goals of action. *Frontiers in Psychology* 7: 1551. doi: 10.3389/fpsyg.2016.01551
- Wagenmakers E.-J. (2007): A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* 14(5): 779–804. doi: 10.3758/bf03194105
- Wang M. (2017): Mixtures of g-priors for analysis of variance models with a diverging number of parameters. *Bayesian Analysis* 12(2): 511–532. doi: 10.1214/16-ba1011