

Supplement to Bayesian Inference in Numerical Cognition: A Tutorial Using JASP

Thomas J. Faulkenberry¹, Alexander Ly^{2,3}, & Eric-Jan Wagenmakers²

¹ Tarleton State University

² University of Amsterdam

³ Centrum Wiskunde & Informatica

Theory of the Bayesian t -test

Recall that a Bayes factor compares two models, say, \mathcal{M}_1 and \mathcal{M}_0 , and suppose that the vector of free parameters of these models are denoted by θ_1 and θ_0 respectively. For instance, we write $\theta_1 = (\mu, \sigma)$ and $\theta_0 = \sigma$ for the one-sample t -test.

To construct a Bayes factor, the statistician needs to select two priors –one for each model, which are denoted by $\pi_1(\theta_1)$ and $\pi_0(\theta_0)$ respectively (e.g., Bayarri, Berger, Forte, & Garcia-Donato, 2012), and provide its users with a computational method. We only introduced the notation $\pi_1(\theta_1)$ and $\pi_0(\theta_0)$, chosen by the statistician, to highlight their difference from prior model probabilities, which are denoted with a capital P , chosen by the field experts.

The Bayesian t -test is based on the priors π_1 , which puts an “uninformative” prior on the nuisance parameter σ (e.g., Ly, Marsman, Verhagen, Grasman, & Wagenmakers, 2017) in conjunction with a Cauchy prior on the test-relevant parameter represented as a population effect size $\delta = \mu/\sigma$, and the prior π_0 with the same “uninformative” prior on the nuisance parameter σ . With these priors, the resulting Bayes factor has a number of desirable properties:

1. **Measurement invariant.** The Bayes factor as a measure of evidence does not change if the units of measurements are changed (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009). For instance, if the data were to consist of temperature measurements, the Bayes factor would be the same regardless of the data being entered as Celsius or Fahrenheit.
2. **Predictively matched.** The Bayes factor reflects perfectly indifference, i.e., $\text{BF}_{10} = 1$ when the data are completely uninformative. A Bayes factor of one implies that the

Correspondence concerning this article should be addressed to Thomas J. Faulkenberry, Department of Psychological Sciences, Box T-0820, Tarleton State University, Stephenville, TX 76401. E-mail: faulkenberry@tarleton.edu

posterior model probabilities equal the prior model probabilities, and nothing has been learned; see Jeffreys (1961), for further details.

3. **Information consistent.** For a sufficiently large enough fixed sample size n , the Bayes factor overwhelmingly provides support for the alternative, whenever the observed t -value goes to infinity; see Gronau, Ly, & Wagenmakers (2017), Ly, Verhagen, & Wagenmakers (2016) for further details.
4. **(Asymptotically) consistent.** As the Bayes factor takes in an increasing number of samples of data points generated under the null model, then as n grows, the Bayes factor BF_{10} tends to zero to convey support for the data generating model \mathcal{M}_0 . On the other hand, if n grows with data generated under the alternative, the Bayes factor BF_{10} grows unbounded; see for instance Liang, Paulo, Molina, Clyde, & Berger (2008).
5. **Robust to changes in the sample size.** With the uninformative prior on σ , this Bayes factor allows for optional stopping, that is, early stopping when there is enough evidence for the alternative (e.g., Hendriksen, Heide, & Grünwald, 2018) without inflating the type I error. On top of this, the Bayesian t -test also allows for optional continuation, which implies that one can build on top of previous results; see for instance Faulkenberry (2019), Grünwald, Heide, & Koolen (2019), Ly, Etz, Marsman, & Wagenmakers (2019). Note that properties (4), (5), and (6) imply that the Bayesian t -test, first derived by Jeffreys (1948), also has good frequentist properties.

With these priors at hand, the Bayes factor is defined as a ratio of marginal likelihoods, that is,

$$\text{BF}_{10} = \frac{\int_{\Theta_1} f(d | \theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_0} f(d | \theta_0) \pi_0(\theta_0) d\theta_0}, \quad (1)$$

which practitioners do not need to compute themselves, as JASP does the computations for them in the background.

Theory of Bayesian linear regression

Linear regression models. We first elaborate on linear regression models, before we discuss the Bayesian implementation.

We write Y_i for the continuous outcome variable of the i th participant, and $x_{i1}, x_{i2}, \dots, x_{ip}$ for the i th participant's covariates. For the example at hand, Y_i represents a participant's mathematical fluency score, and we have $p = 2$ covariates consisting of the i th participant's measurements $x_{i1} = \text{symbolicDE}_i$, and $x_{i2} = \text{nonsymbolicDE}_i$.

The difficulty here is that with p covariates there are 2^p models. For instance with

Table 1

List of all models with the two potential predictors and the prior model probability of each model. A "1" indicates that the predictor is active in model, whereas a "0" implies that the predictor is inactive.

Model	symbolicDE	nonsymbolicDE	$P(\mathcal{M})$
\mathcal{M}_0	0	0	0.25
\mathcal{M}_1	1	0	0.25
\mathcal{M}_2	0	1	0.25
\mathcal{M}_3	1	1	0.25

$p = 2$, we have $2^p = 4$ models, which are given by

$$\begin{aligned}\mathcal{M}_0 : Y_i &= \beta_0 + \varepsilon_i, \\ \mathcal{M}_1 : Y_i &= \beta_0 + \beta_1 x_{i1} + \varepsilon_i, \\ \mathcal{M}_2 : Y_i &= \beta_0 + \beta_2 x_{i2} + \varepsilon_i, \\ \mathcal{M}_3 : Y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,\end{aligned}$$

where β_0 represents the intercept, β_1 the effect of symbolicDE on Y_i , β_2 the effect of nonsymbolicDE on Y_i , and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed error. Each model can be thought of as an operationalisation of a hypothesis. Note that all models include the intercept and error variance σ^2 parameters, which are related to the outcome variable Y_i . The models $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ have 2, 3, 3 and 4 parameters, respectively. In general, the largest model will have $p + 2$ parameters.

All models decompose Y_i into two parts: a part that is structural, and a part that is simply noise. For the null model, all the structure is given by the intercept β_0 , whereas in $\mathcal{M}_1, \mathcal{M}_2$ and \mathcal{M}_3 the intercept β_0 competes with the covariates to predict Y_i . Hence, despite terms being denoted with the same symbol, such as the intercepts β_0 , they have a different impact on Y_i depending on the context given by the model. As such, the question whether symbolicDE has an effect on fluency depends on the particular model that is being compared.

Bayesian linear regression. To avoid the dependence on which model is being compared, we exploit the flexibility of Bayesian methods, and address questions such as whether symbolicDE has an effect on fluency by *averaging* across all models considered. To model average we require weights, and for simplicity we assume that all models are equally probable a priori. In other words, we put a uniform prior on the models, and because we have four models each model gets a prior model probability of $1/4 = 0.25$. Table ?? provides a summary of this information.

The a priori probability of the relevance of the covariate symbolicDE on Y_i is given by aggregating the probabilities of the models \mathcal{M}_1 and \mathcal{M}_3 , thus, 50%. Similarly, the a priori relevance of the covariate nonsymbolicDE is also 50%, whereas the a priori probability of both symbolicDE and nonsymbolicDE having an impact on Y_i is only given by $P(\mathcal{M}_3) = 0.25$.

To compute the relevance of each covariate **after** data observation, we simply replace the prior model probabilities in the table by the posterior model probabilities. These are

calculated using a generalisation of Equation 1 which specifies that the posterior model probability for model $k = 0, 1, \dots, p - 1 = 3$ is given by

$$P(\mathcal{M}_k \mid d) = \frac{\text{BF}_{k0}(d)P(\mathcal{M}_k)}{\sum_{k=0}^{p-1} \text{BF}_{k0}(d)P(\mathcal{M}_k)}, \quad (2)$$

where the comparison between the null model and itself is defined as $\text{BF}_{00}(d) = 1$ regardless of the data. Now, suppose that the model \mathcal{M}_k has $l \leq p$ active covariates. Then for the Bayes factor BF_{k0} , we need to choose a prior on $l + 2$ parameters for \mathcal{M}_k and a prior on 2 parameters for \mathcal{M}_0 . Zellner & Siow (1980) proposed to use essentially the same priors Jeffreys used for his Bayesian t -test. This connection is made explicit by exploiting the assumption of normally distributed errors to rewrite the linear regression models as the following normal means models:

$$\mathcal{M}_0 : Y_i \mid x_i \sim \mathcal{N}(\beta_0, \sigma^2), \quad (3)$$

$$\mathcal{M}_1 : Y_i \mid x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_{i1}, \sigma^2), \quad (4)$$

$$\mathcal{M}_2 : Y_i \mid x_i \sim \mathcal{N}(\beta_0 + \beta_2 x_{i2}, \sigma^2), \quad (5)$$

$$\mathcal{M}_3 : Y_i \mid x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \sigma^2), \quad (6)$$

where $Y_i \mid x_i$ refers to the outcome variable given the covariates x_i .

More explicitly, the Bayes factor BF_{k0} , where \mathcal{M}_k has l active covariates, uses an “uninformative” prior on the nuisance parameter β_0 and σ in both models, and for the \mathcal{M}_k an additional l -dimensional multivariate Cauchy prior on the vector of population effect sizes δ . The vector of effect sizes is defined by rescaling the vector of active β coefficients with respect to σ , but also with respect to the covariates. More explicitly, when $l = 1$, say, only β_2 is active then the effect size parameter is given by $\delta = \beta_2 s_{x_2} / \sigma$, where s_{x_2} is observed standard error of the second covariate. If the second covariate is standardised, then $s_{x_2} = 1$, and we get the same definition of effect size as in the t -test, but with μ replaced by β_2 . Furthermore, with $l = 1$, the prior on δ is than a 1-dimensional Cauchy distribution, indeed, the same one we would use in the t -test.

With both the prior model probabilities and the Bayes factor at hand (2) now provides the posterior model probabilities. Once the latter are computed, the posterior relevance of each covariate can be assessed as we have done in Table ???. We elaborate on this procedure with the example.

References

- Bayarri, M. J., Berger, J. O., Forte, A., & Garcia-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40(3), 1550–1577.
- Faulkenberry, T. J. (2019). Estimating evidential value from analysis of variance summaries: A comment on ly et al.(2018). *Advances in Methods and Practices in Psychological Science*, 2515245919872960.
- Gronau, Q. F., Ly, A., & Wagenmakers, E.-J. (2017). Informed Bayesian *t*-tests. *arXiv Preprint arXiv:1704.02479*.
- Grünwald, P., Heide, R. de, & Koolen, W. (2019). Safe testing. *arXiv Preprint arXiv:1906.07801*.
- Hendriksen, A., Heide, R. de, & Grünwald, P. (2018). Optional stopping with Bayes factors: A categorization and extension of folklore results, with an application to invariant situations. *arXiv Preprint arXiv:1807.09077*.
- Jeffreys, H. (1948). *Theory of probability* (2nd ed.). Oxford, UK: Oxford University Press.
- Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481).
- Ly, A., Etz, A., Marsman, M., & Wagenmakers, E. (2019). Replication Bayes factors from evidence updating. *Behavior Research Methods*, 51(6), 2498–2508. <https://doi.org/https://doi.org/10.3758/s13428-018-1092-x>
- Ly, A., Marsman, M., Verhagen, A. J., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80, 40–55. <https://doi.org/https://doi.org/10.1016/j.jmp.2017.05.006>
- Ly, A., Verhagen, A. J., & Wagenmakers, E. (2016). 201601Harold Jeffreys’s default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72, 19–32. <https://doi.org/http://dx.doi.org/10.1016/j.jmp.2015.06.004>
- 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.
- Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. Smith (Eds.), *Bayesian statistics: Proceedings of the first international meeting held in Valencia* (Vol. 1, pp. 585–603). Springer.

- Bayarri, M. J., Berger, J. O., Forte, A., & Garcia-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40(3), 1550–1577.
- Faulkenberry, T. J. (2019). Estimating evidential value from analysis of variance summaries: A comment on ly et al.(2018). *Advances in Methods and Practices in Psychological Science*, 2515245919872960.
- Gronau, Q. F., Ly, A., & Wagenmakers, E.-J. (2017). Informed Bayesian *t*-tests. *arXiv Preprint arXiv:1704.02479*.
- Grünwald, P., Heide, R. de, & Koolen, W. (2019). Safe testing. *arXiv Preprint arXiv:1906.07801*.
- Hendriksen, A., Heide, R. de, & Grünwald, P. (2018). Optional stopping with Bayes factors: A categorization and extension of folklore results, with an application to invariant situations. *arXiv Preprint arXiv:1807.09077*.
- Jeffreys, H. (1948). *Theory of probability* (2nd ed.). Oxford, UK: Oxford University Press.
- Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481).
- Ly, A., Etz, A., Marsman, M., & Wagenmakers, E. (2019). Replication Bayes factors from evidence updating. *Behavior Research Methods*, 51(6), 2498–2508. <https://doi.org/https://doi.org/10.3758/s13428-018-1092-x>
- Ly, A., Marsman, M., Verhagen, A. J., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80, 40–55. <https://doi.org/https://doi.org/10.1016/j.jmp.2017.05.006>
- Ly, A., Verhagen, A. J., & Wagenmakers, E. (2016). 201601Harold Jeffreys’s default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72, 19–32. <https://doi.org/http://dx.doi.org/10.1016/j.jmp.2015.06.004>
- 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.
- Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. Smith (Eds.), *Bayesian statistics: Proceedings of the first international meeting held in Valencia* (Vol. 1, pp. 585–603). Springer.