Statistics and Probability Letters 92 (2014) 121–124

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Simple relation between Bayesian order-restricted and point-null hypothesis tests

Richard D. Morey[a,*], Eric-Jan Wagenmakers[b]

[a] University of Groningen, Department of Psychometrics and Statistics, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
[b] University of Amsterdam, Department of Psychological Methods, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form 14 May 2014
Accepted 14 May 2014
Available online 24 May 2014

Keywords:
Bayes factor
Order-restriction
Sign-restriction
One-sided test
t test
Hypothesis testing

ABSTRACT

One of the main challenges facing potential users of Bayes factors as an inferential technique is the difficulty of computing them. We highlight a useful relationship that allows certain order-restricted and sign-restricted Bayes factors, such as one-sided Bayes factor tests, to be computed with ease.

1. Introduction

Consider an encompassing model $\mathcal{M}_e$ with nuisance parameters θ and parameter of interest $\boldsymbol{\delta}$ of length K with marginal prior distribution $p(\boldsymbol{\delta})$. Two restrictions of $\mathcal{M}_e$ can be considered: the null hypothesis $\mathcal{M}_0$ states that $\boldsymbol{\delta} = \mathbf{0}$, and $\mathcal{M}_r$ is an order-restricted hypothesis that the $\boldsymbol{\delta}$ parameters have a specific ordering. If R is the set of all vectors $\boldsymbol{\delta}$ that meet the specified restriction, then $\mathcal{M}_r$ states that $\boldsymbol{\delta} \in R$. If $K = 1$ and δ is a scalar parameter, then $\mathcal{M}_r$ is a sign hypothesis that δ is either positive or negative. We use the general term "order-restriction" to refer both the $K = 1$ case and the $K > 1$ case. Suppose that $p(\boldsymbol{\delta})$ is such that all orderings are equally-likely a priori, as will occur if the prior distributions on the $K\delta$ parameters are identical and mutually conditionally independent. The Bayes factor $B_{r0} = p(\mathbf{y} \mid \mathcal{M}_r)/p(\mathbf{y} \mid \mathcal{M}_0)$ quantifies the evidence that the data $\mathbf{y}$ provide for $\mathcal{M}_r$ versus $\mathcal{M}_0$ (Jeffreys, 1961; Kass and Raftery, 1995). This Bayes factor is of practical interest because researchers often have strong prior expectation about the direction of an effect or the ordering of means under the assumption that the null hypothesis is false. Unfortunately, B_{r0} is often not available in closed form because almost all tests have been developed for the two-sided scenario B_{e0}. In addition, the computation of B_{r0} is made difficult by the fact that the prior and posterior distributions under model $\mathcal{M}_r$ are bounded at 0 and therefore may not be members of familiar families of distributions. Hence, the calculation of $p(\mathbf{y} \mid \mathcal{M}_r)$ can be a non-trivial task that appears to require general procedures such as reversible jump Markov chain Monte Carlo (Green, 1995) that applied researchers may find challenging to implement.

However, Pericchi et al. (2008) proposed a general and simple solution to the computation of the one-sided Bayes factor B_{r0}, avoiding the need for integration over the parameter space when the two-sided Bayes factor B_{e0} is already in hand.

[*] Corresponding author.
E-mail addresses: r.d.morey@rug.nl (R.D. Morey), ej.wagenmakers@gmail.com (E.-J. Wagenmakers).

http://dx.doi.org/10.1016/j.spl.2014.05.010
0167-7152/© 2014 Elsevier B.V. All rights reserved.

Theorem 1. *Let L be*

$$L = \begin{cases} 2 & K = 1 \\ K! & K > 1. \end{cases}$$

Then

$$B_{r0} = Lp(\boldsymbol{\delta} \in R \mid \mathbf{y}, \mathcal{M}_e)B_{e0}.$$

Proof of Theorem 1. There are L specific order-restricted hypotheses on $\boldsymbol{\delta}$. Under a proper prior $p(\boldsymbol{\delta})$ in which all orderings on K means are equally likely, each ordering has an a priori probability of $1/L$. Because the total prior probability of all orderings is 1, the prior odds of $\mathcal{M}_r$ against the encompassing model $\mathcal{M}_e$ are thus $(1/L)/1 = 1/L$. The corresponding posterior odds are $p(\boldsymbol{\delta} \in R \mid \mathbf{y}, \mathcal{M}_e)$ (Klugkist et al., 2005). Because the Bayes factor is the ratio of the posterior odds to the prior odds,

$$\begin{aligned} B_{re} &= \frac{p(\boldsymbol{\delta} \in R \mid \mathbf{y}, \mathcal{M}_e)}{1/L} \\ &= Lp(\boldsymbol{\delta} \in R \mid \mathbf{y}, \mathcal{M}_e). \end{aligned}$$

Bayes factors are ratios of the corresponding marginal likelihoods, and thus

$$\begin{aligned} B_{r0} &= \frac{p(\mathbf{y} \mid \mathcal{M}_r)}{p(\mathbf{y} \mid \mathcal{M}_0)} \\ &= \frac{p(\mathbf{y} \mid \mathcal{M}_r)}{p(\mathbf{y} \mid \mathcal{M}_e)} \times \frac{p(\mathbf{y} \mid \mathcal{M}_e)}{p(\mathbf{y} \mid \mathcal{M}_0)} \\ &= B_{re}B_{e0} \end{aligned}$$

and the result follows. ■

The term B_{e0} is from the familiar two-sided test, and the term B_{re} equals the ratio between the marginal posterior and the marginal prior mass consistent with the restriction (Klugkist et al., 2005). If B_{re} is not available analytically, it can be easily obtained to any desired degree of approximation using numerical methods such as Markov chain Monte Carlo (e.g., Morey et al., 2011).

One application of Theorem 1 is the one-sided tests that arise when $K = 1$. In such one-sided tests, Theorem 1 implies that the one-sided test B_{r0} equals the two-sided test B_{e0} only when the posterior $p(\delta \mid \mathbf{y}, \mathcal{M}_e)$ is symmetric around 0. In addition, the use of a one-sided test can increase the evidence against $\mathcal{M}_0$ by a factor of 2 at most, which happens when almost the entire posterior distribution is consistent with the order-restriction. When the data are inconsistent with the sign-restriction $\delta > 0$ this means that $p(\delta > 0 \mid \mathbf{y}, \mathcal{M}_e)$ is lower than 0.5, and the use of a one-sided test increases the evidence for $\mathcal{M}_0$. In fact, when the data are wildly inconsistent with the order-restriction it may happen that B_{r0} is extremely low (indicating that $\mathcal{M}_0$ should be retained) and that, at the same time, B_{e0} is extremely high (indicating that $\mathcal{M}_0$ should be rejected). This underscores the relative nature of the Bayes factor as a measure of evidence.

The relevance of order-restricted tests is particularly acute for the replication research and for clinical trials, where compelling evidence for $\mathcal{M}_0$ may be obtained when the effect goes in the direction opposite to what was expected. The effects become more pronounced when more parameters are subject to test. Suppose $\mathcal{M}_e$ is a one-way model with $K = 4$ condition means for which the analyst has a strong a priori commitment to the orderings of the K means, if the null hypothesis were false. For instance, if the conditions arose from a manipulation of a single independent variable, such as dosage or difficulty, then the analyst may wish to test the specific ordering that implies a monotone relationship. If the posterior probability $p(\boldsymbol{\delta} \in R \mid \mathbf{y}, \mathcal{M}_e)$ in favor of the restriction is maximal, then increase in the evidence from B_{e0} to B_{r0} from properly restricting the test will be $4! = 24$, a substantial change in the evidence.

In special cases where the posterior probability can be easily approximated by the p value, such as in one- and two-sample tests, the correction factor can be easily computed using the output of a standard classical analysis. In the one-sample case, $L = 2$ and the correction needed to obtain the sign-restricted test equals $2 \times p(\delta > 0 \mid \mathbf{y}, \mathcal{M}_e)$ if the desired sign restriction is that $\delta > 0$. Exploiting the fact that for the test of location parameters the classical one-sided p value approximates $p(\delta < 0 \mid \mathbf{y}, \mathcal{M}_e) = 1 - p(\delta > 0 \mid \mathbf{y}, \mathcal{M}_e)$ (Casella and Berger, 1987; Lindley, 1965; Pratt, 1965), we obtain:

$$B_{r0} \approx \begin{cases} (2 - p) \times B_{e0} & \text{if } \hat{\delta} > 0, \\ p \times B_{e0} & \text{if } \hat{\delta} \le 0, \end{cases} \tag{1}$$

where p is two-sided, and $\hat{\delta} > 0$ indicates that the observed effect is consistent with the sign-restriction. When $\hat{\delta} > 0, B_{r0} > B_{e0}$, with a maximum of $B_{r0} = 2 \times B_{e0}$ when $p \to 0$. When $\hat{\delta} < 0$ (i.e., the observed effect goes in the opposite direction), $B_{r0} < B_{e0}$. In sum, (1) shows how the sign-restricted Bayes factor can be approximated by the product of two familiar terms, one involving the two-sided p value, and one involving the two-sided Bayes factor.

The p value approximation is particularly useful when the posterior probability $p(\delta < 0 \mid \mathbf{y}, \mathcal{M}_e)$ is not immediately available. For instance, not all methods of estimating Bayes factors involve MCMC chains that can be used to estimate the required posterior probability, and even when they do the software may not report the chains. The widely-used JZS Bayes factor web calculator (Rouder et al., 2009; http://pcl.missouri.edu/bayesfactor), for instance, does not return posterior probabilities.

Table 1
Inference for three ESP replication attempts by Ritchie et al. (2012). DR% stands for differential recall percentage (for details see Bem, 2011); a positive DR indicates evidence for ESP. $p_>$ stands for the one-sided p value. $B^p_{r0} = 2 \times (1 - p_>) \times B^{\text{JZS}}_{e0}$.

	Mean DR% (SD)	t	$p_>$	B^{JZS}_{e0}	B^{JZS}_{r0}	B^p_{r0}
Repl. 1 ($n = 50$)	0.19% (12.63)	0.1100	0.4564	0.1547	0.1679	0.1682
Repl. 2 ($n = 50$)	−2.72% (12.23)	−1.5700	0.9386	0.4835	0.0640	0.0594
Repl. 3 ($n = 50$)	−0.58% (14.27)	−0.2900	0.6135	0.1601	0.1246	0.1238

With improper uniform prior distributions on the location parameters from the exponential family, the p value estimate of the posterior probability $p(\delta < 0 \mid \mathbf{y}, \mathcal{M}_e)$ is exact, see Lindley (1965), pp. 9–10, 13–15, 31–33 and Jaynes (1976), pp. 193, 199–200, 206. With non-uniform prior distributions the p value only approximates the posterior probability. It is important to emphasize, however, that we are not suggesting the use of an improper prior on δ; the p value is merely a useful approximation to the true posterior probability, which assumes a proper prior on δ.

The quality of the p value approximation depends on the data, the extent to which the prior distribution is non-uniform, and the number of observations. The default priors often used for Bayes factor hypothesis testing (e.g., unit-information priors, Cauchy priors) are relatively wide and symmetric around zero; in such cases, the data quickly overwhelm the prior. Hence, the posterior distribution is relatively robust to the prior specification (in this case, uniform priors versus default priors for hypothesis testing) and consequently the p value approximation to the posterior probability $p(\delta < 0 \mid \mathbf{y}, \mathcal{M}_e)$ will be very good. The example below illustrates this point.

2. Example: the *t* test

As an example, consider three experiments on extrasensory perception (ESP) conducted by Ritchie et al. (2012). These experiments were direct replications of an experiment by Bem (2011), in which participants were shown lists of words for later recall. Critically, some of the words from the study list were also presented after the test phase. According to Bem (2011), people can look into the future and take advantage of these additional post-test presentations to boost their recall performance. Hence, the crucial statistical analysis involves a t test between the control words, presented only during the study phase, and the post-test words, presented both during the study phase and also later, following the test phase.

Table 1 shows the results. Recall performance was quantified by differential recall percentage (DR; for details see Bem, 2011); a positive DR indicates evidence for ESP. Rouder et al. (2009) suggest a Bayes factor for one-sample designs, based on the g prior setting of Liang et al. (2008). Under $\mathcal{M}_0$, the t statistic has a Student t distribution with $N - 1$ degrees of freedom:

$$\mathcal{M}_0 : t \sim \text{Student } t_{N-1}.$$

Under the alternative $\mathcal{M}_e$, t has a noncentral t distribution:

$$\mathcal{M}_e : t \sim \text{Noncentral } t_{N-1}(\delta\sqrt{N})$$

where δ is the standardized effect size $\delta = \mu/\sigma$ and thus $\delta\sqrt{N}$ is the noncentrality parameter. A scaled Cauchy prior distribution is placed on the effect size δ:

$$\delta \sim \text{Cauchy}(r)$$

where r is the scale parameter. Model $\mathcal{M}_0$ is thus a restriction of $\mathcal{M}_e$ in which $\delta = 0$. Rouder et al. (2009) dub the resulting Bayes factor, the JZS Bayes factor, after Jeffreys (1961) and Zellner and Siow (1980). The two-sided Bayes factor B^{JZS}_{e0}, calculated using the R package `BayesFactor` (using Gaussian quadrature; Morey and Rouder, 2014) and shown in the fourth column, indicates that the evidence in each replication attempt favors the null hypothesis that $\delta = 0$ over an unrestricted alternative. Note that we used the `BayesFactor` R package's default scale $r = \sqrt{2}/2$ for the Cauchy prior on δ. The evidence ranges from a factor of 2 (Replication 2) to a factor of about 6.5 (Replication 1). The one-sided Bayes factor B^{JZS}_{r0}, shown in the fifth column, is arguably more appropriate in this situation, as it more closely reflects the directional hypothesis of retroactive facilitation of recall. The data from Replications 2 and 3, however, have the effect going slightly in the opposite of the predicted direction; consequently, the one-sided B^{JZS}_{r0} provides more evidence for the null hypothesis than did the two-sided B^{JZS}_{e0}. The evidence for the null hypothesis $1/B_{r0}$ now ranges from a factor of 6 (Replication 1) to a factor of about 16.6 (Replication 2). Replication 2, which provides the least evidence for the null when the alternative is unrestricted (B_{e0}), provides the most evidence for the null when the alternative is properly restricted (B_{r0}).

The sixth column in the table, labeled B^p_{r0}, contains the one-sided JZS Bayes factors computed using the p value approximation to the posterior probability $p(\delta < 0 \mid \mathbf{y}, \mathcal{M}_e)$. The approximation is quite good, as expected.

3. Concluding comments

We have outlined a straightforward and general method to derive one-sided Bayes factors from their two-sided counterparts. We integrate three earlier contributions; first, the Bayes factor product factorization by Pericchi et al. (2008); second, the encompassing prior technique by Klugkist et al. (2005); and third, the work by Lindley (1965) and others, showing that the one-sided p value closely approximates the mass of the posterior distribution on one side of zero. In the present paper, we combined these disparate ideas to form a simple expression for a one-sided test.

The expression for the one-sided test using the two-sided p value is simple and straightforward, yet has not been proposed previously. This is possibly due to the fact that the p value and Bayes factor are often seen as competitors: the p value as a classical measure of the evidence against $\mathcal{M}_0$, and the Bayes factor as the Bayesian measure of the relative evidence for $\mathcal{M}_0$ compared to an alternative $\mathcal{M}_1$. Those who find the p value useful generally do not report Bayes factors, and those who report Bayes factors have no use for p values (but see Berger, 2003). Thanks to the interpretation of the p value as a posterior probability, however, the two can be combined into a single, easy-to-compute measure of evidence for order-restricted hypotheses.

The present approach to calculate one-sided Bayes factors is very general. For instance, it also applies to a comparison of nonnested models, such as when we already have a Bayes factor between models $\mathcal{M}_\alpha$ and $\mathcal{M}_\beta$, but we seek a Bayes factor between models $\mathcal{M}_\alpha$ and a version of $\mathcal{M}_\beta$ in which one or more parameters are subject to order-restrictions. More importantly, the approach allows immediate one-sided extensions for model comparison methods such as BIC (Raftery, 1995; Schwarz, 1978), fractional Bayes factors (O'Hagan, 1995), intrinsic Bayes factors (Berger and Pericchi, 1996; Berger and Mortera, 1999), and Bayes factors calculated from test statistics (Johnson, 2005). The correction for order-restrictions through $2 \times p(\delta > 0 \mid \mathbf{y}, \mathcal{M}_e)$ may even prove useful for model comparison methods such as AIC and DIC that are not related to the Bayes factor. Note that the correction factor for the addition of order-restrictions is based on the posterior distribution; changes to the prior distributions that do not substantially affect the posterior distribution will also not substantially affect the correction factor. The ease with which the correction factor suggested here can be used to obtain one-sided model comparison statistics will hopefully encourage practical researchers to test statistical models that more accurately reflect substantive theories about the processes under investigation.

Finally, a note of caution. Bayes factors quantify only relative model adequacy; when the Bayes factor strongly supports the inclusion of a predictor in a regression model, for instance, this does not mean that this model can be relied on to provide a satisfactory fit to the data. In order to draw valid conclusions from a model it is important to assess both relative and absolute model adequacy (Morey et al., 2013).

Acknowledgments

We wish to thank the action editor and an anonymous referee for their helpful comments during the review process. The R code for the examples can be found at http://drsmorey.org/oneSidedBF/. This research was supported by a consolidator grant from the European Research Council (ERC) Grant number: 283876.

References

Bem, D.J., 2011. Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. J. Personality Soc. Psychol. 100, 407–425.

Berger, J.O., 2003. Could Fisher, Jeffreys and Neyman have agreed on testing? Statist. Sci. 18, 1–32.

Berger, J.O., Mortera, J., 1999. Default Bayes factors for nonnested hypothesis testing. J. Amer. Statist. Assoc. 94, 542–554.

Berger, J.O., Pericchi, L.R., 1996. The intrinsic Bayes factor for model selection and prediction. J. Amer. Statist. Assoc. 91, 109–122.

Casella, G., Berger, R.L., 1987. Reconciling Bayesian and frequentist evidence in the one–sided testing problem. J. Amer. Statist. Assoc. 82, 106–111.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Jaynes, E.T., 1976. Confidence intervals vs Bayesian intervals. In: Harper, W.L., Hooker, C.A. (Eds.), Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, vol. II. D. Reidel Publishing Company, Dordrecht, Holland, pp. 175–257.

Jeffreys, H., 1961. Theory of Probability, third ed.. Oxford University Press, New York.

Johnson, V.E., 2005. Bayes factors based on test statistics. J. R. Statist. Soc., Ser. B 67, 689–701.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.

Klugkist, I., Laudy, O., Hoijtink, H., 2005. Inequality constrained analysis of variance: a Bayesian approach. Psychol. Methods 10, 477–493.

Liang, F., Paulo, R., Molina, G., Clyde, M.A., Berger, J.O., 2008. Mixtures of g-priors for Bayesian variable selection. J. Amer. Statist. Assoc. 103, 410–423. http://pubs.amstat.org/doi/pdf/10.1198/016214507000001337.

Lindley, D.V., 1965. Introduction to Probability & Statistics from a Bayesian Viewpoint. Part 2. Inference. Cambridge University Press, Cambridge.

Morey, R.D., Romeijn, J.-W., Rouder, J.N., 2013. The humble Bayesian: model checking from a fully Bayesian perspective. British J. Math. Statist. Psych. 66, 68–75. http://dx.doi.org/10.1111/j.2044-8317.2012.02067.x.

Morey, R.D., Rouder, J.N., 2014. BayesFactor 0.9.6. Comprehensive R Archive Network. http://cran.r-project.org/web/packages/BayesFactor/index.html.

Morey, R.D., Rouder, J.N., Pratte, M.S., Speckman, P.L., 2011. Using MCMC chain outputs to efficiently estimate Bayes factors. J. Math. Psychol. 55, 368–378.

O'Hagan, A., 1995. Fractional Bayes factors for model comparison. J. R. Stat. Soc. B 57, 99–138.

Pericchi, L.R., Liu, G., Torres, D., 2008. Objective Bayes factors for informative hypotheses: "completing" the informative hypothesis and "splitting" the Bayes factor. In: Hoijtink, H., Klugkist, I., Boelen, P.A. (Eds.), Bayesian Evaluation of Informative Hypotheses. Springer Verlag, New York, pp. 131–154.

Pratt, J.W., 1965. Bayesian interpretation of standard inference statements. J. R. Statist. Soc. Ser. B 27, 169–203.

Raftery, A.E., 1995. Bayesian model selection in social research. In: Marsden, P.V. (Ed.), Sociological Methodology. Blackwells, Cambridge, pp. 111–196.

Ritchie, S.J., Wiseman, R., French, C.C., 2012. Failing the future: three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. PLoS ONE 7, e33423.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t-tests for accepting and rejecting the null hypothesis. Psychon. Bull. & Rev. 16, 225–237. http://dx.doi.org/10.3758/PBR.16.2.225.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Zellner, A., Siow, A., 1980. Posterior odds ratios for selected regression hypotheses. In: Bernardo, J.M., DeGroot, M.H., Lindley D.V., Smith A.F.M. (Eds.), Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain), University of Valencia, pp. 585–603.