Journal of Modern Applied Statistical Methods

Volume 18 | Issue 1 Article 14

4-1-2020

Conflicts in Bayesian Statistics Between Inference Based on Credible Intervals and Bayes Factors

Miodrag M. Lovric
Radford University, mlovric@radford.edu

Conflicts in Bayesian Statistics Between Inference Based on Credible Intervals and Bayes Factors

Cover Page Footnote

This research was supported by the Artis College Faculty Research Grant, Radford University, USA.

Journal of Modern Applied Statistical Methods
May 2019, Vol. 18, No. 1, eP3320.
doi: 10.22237/jmasm/1556670540

בס"ד

ISSN 1538 − 9472

INVITED ARTICLE

Conflicts in Bayesian Statistics Between Inference Based on Credible Intervals and Bayes Factors

Miodrag M. Lovric
Radford University
Radford, Virginia

In frequentist statistics, point-null hypothesis testing based on significance tests and confidence intervals are harmonious procedures and lead to the same conclusion. This is not the case in the domain of the Bayesian framework. An inference made about the point-null hypothesis using Bayes factor may lead to an opposite conclusion if it is based on the Bayesian credible interval. Bayesian suggestions to test point-nulls using credible intervals are misleading and should be dismissed. A null hypothesized value may be outside a credible interval but supported by Bayes factor (a Type I conflict), or contrariwise, the null value may be inside a credible interval but not supported by the Bayes factor (Type II conflict). Two computer programs in R have been developed that confirm the existence of a countable infinite number of cases, for which Bayes credible intervals are not compatible with Bayesian hypothesis testing.

Keywords: Point-null hypothesis, Bayes Factor, credible intervals, HPD interval, Type I and Type II conflicts

Introduction

Bayesian arguments made substantial progress from the 1950s when they were labeled as dangerous (see Gelman & Robert, 2013) and when David (1949) claimed "the application of the [Bayes'] theorem in statistical method is wholly fallacious except under very restrictive conditions" (p. 71). As pointed out by Fienberg (2006a), today, Bayesian methods are "integrated into both the fabric of statistical thinking within the field of statistics and the methodology used in a broad array of applications" (p. 3). As emphasized by Bernardo (2010, p. 108), the Bayesian approach provides a complete coherent paradigm for both statistical inference and

doi: 10.22237/jmasm/1556670540 | Accepted: September 21, 2019; Published: April 01, 2020.
Correspondence: Miodrag M. Lovric, mlovric@radford.edu

decision making under uncertainty; it constitutes a scientific revolution in Kuhn's sense and is firmly based on axiomatic foundations. Similarly, Jaynes (2003) noted "orthodox statistics" (p. 510) is not a coherent body of theory but a loose collection of independent ad hoc devices, invented and advocated by many different people on many different intuitive grounds, often in sharp disagreement with each other. It can be argued, however, a similar image can be attributed to the Bayesian statistics. For example, Good (1971) opined there are "46,656 Varieties of Bayesians" (p. 65), depending on their positions in regard to eleven facets. That number was larger than the number of professional statisticians at the time.

In the similar vein Fienberg (2006b) stated "today there seem to be at least this many varieties of objective Bayesians, with each seeking out his or her own method for arriving at the perfect objective prior and then allowing for other idiosyncrasies" (p. 431). The pursuit for the objective prior distributions that reflect ignorance was compared to the search for the Holy Grail. Fienberg concluded this goal is intangible, fruitless, and ultimately diverting energies from computing quality statistics.

The objective of this study is to prove the Bayesian approach to inference is not coherent. Potentially, in a countably infinite number of cases, conclusions obtained by Bayesian credible intervals collide with conclusions from Bayesian hypothesis testing. Without loss of generality, the discussion is confined to the point null hypothesis testing under the normal probability model. In addition, normal conjugate priors are considered for estimation and Jeffreys mixed prior model for testing.

Testing point null hypotheses using credible or HPD intervals

In many problems involving the normal distribution, there is no collision between frequentist confidence intervals and Bayesian credible intervals established for a certain class of noninformative priors constructed by some formal rule expressing ignorance. Certainly, the philosophical bases and interpretations are quite different. Particularly, in making inferences about the single normal mean, $(1-\alpha)$ shortest confidence intervals and HPD (Highest Posterior Density) intervals are numerically identical (see, for example, DeGroot, 1989, p. 409; Berger, 1985, p. 141; Lindley, 1965, pp. 13-42). As Box and Tiao (1992, p. 85) pointed out, when σ is assumed known, this is because intervals are based on sufficient statistic $\bar{x}$, and in both approaches the pivotal quantity $z=\sqrt{n}\left(\mu-\bar{x}\right)/\sigma$ is distributed as $N(0,1)$.

When σ is unknown, Bernardo (2005, p. 57) showed in the simple normal problems for all sample sizes, posterior reference credible intervals for μ will numerically be identical to frequentist confidence intervals based on the sampling distribution of t.

Within frequentist framework significance, testing and confidence intervals are closely related. Any point-null hypothesis, $H_0 : \theta = \theta_0$, can be tested using corresponding confidence intervals. Specifically, we just invert the equation for the test statistic to give us the boundary values of the confidence intervals. Accordingly, a confidence interval is derived from an inverted hypothesis test. Therefore, a confidence interval is composed of all possible values of θ_0 for which the matching test would not reject the hypotheses. If the null value is within a $(1 - \alpha)$% confidence interval, frequentist tests will fail to reject it, and conversely, when the hypothesized value does not belong to that interval, we conclude that p-value $< \alpha$, and reject that value at α% significance level. This rationale is being taught even at the introductory statistics courses. The same principle must be applied in Bayesian statistics to provide a coherent paradigm. If not, it is an approach prone to the contradictions in terms.

When the prior knowledge is vague and the prior distribution in the neighborhood is reasonably smooth, Lindley (1965, p. 61) answered positively. It was proposed credibility of the null hypothesized value can be tested by checking whether or not it belongs to a chosen Bayesian credible interval. A parameter value is declared not to be credible if it lies outside the 95% HDI of the posterior distribution of that parameter, and vice versa. Although this procedure certainly bypasses the well-known Jeffreys-Lindley paradox and is similar to the corresponding frequentist procedure, its major drawback is that it cannot attach a posterior probability to the null value. Arguably, this procedure might be suggested to lecturers in teaching their students the elements of the Bayesian analysis to counterbalance controversies in point-null hypothesis testing. As such, a decision rule similar to the following has been put forward by some Bayesians, including Kim (1991), Ghosh et al. (2006, p. 49), Drummond & Rambaut (2007), Thulin (2014), and Koch (2007, pp. 82-83):

> "A parameter value is declared to be not credible if it lies outside the 95% HDI of the posterior distribution of that parameter. If a parameter value lies within the 95% HDI, it is said to be among the credible values" (Kruschke, 2010, p. 240).

Similarly, Bolstad and Curran (2017) claimed when a Bayesian tests the two-sided hypothesis $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ using continuous prior, the posterior probability of the null hypothesis would equal zero because probability of any particular value of a continuous random variable always equals zero. Hence, in this case Bayesians should never perform any statistical test directly but calculate $(1 - \alpha)\times 100\%$ credible interval for the unknown population mean. In other words, the rationale should be the same as with the frequentist procedures. If the hypothesized value lies inside the interval we fail to reject the null hypothesis, and if it is outside the credible interval, it could be concluded it "does not have credibility as a possible value, and we will reject the null hypothesis" (Bolstad and Curran, 2017, p. 249).

Conversely, Berger and Delampady (1987, p. 319) argued testing point null hypotheses using credible intervals is wrong since it ignores the supposedly special prior believability in θ_0. They stressed that a hypothetical parameter value may be outside a credible interval, yet not strongly contraindicated by the data. Their view is that only Bayes factor or posterior probabilities $P(H_0 \mid x)$ can indicate the strength of the evidence against a particular hypothesized value. Nevertheless, whenever testing a special point value, they recommended reporting both the Bayes factor and a confidence or credible intervals. A similar view was given by Hoekstra et al. (2014). The aim of the present study is to refute the abovementioned decision rule.

Bayesian point null hypothesis testing in case of normal probability model

Consider testing a point null hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where θ is an unknown element with values in parameter space Θ, based on observing random variable, X, with density $f(x \mid \theta)$. Let's assign prior probabilities π_0 and $\pi_1 = (1 - \pi_0)$ to the null and alternative hypothesis, respectively. Suppose further that $g(\theta)$ is a continuous prior probability density, conditional on H_1 being true, that is on $\{\theta \neq \theta_0\}$. As proposed by Jeffreys (1961), a standard Bayesian solution is to allocate probability mass π_0 to the single point indicated by the null hypothesis $\theta = \theta_0$ and distributing the remainder, $(1 - \pi_0)$, according to the continuous density $g(\theta)$ over $\theta \neq \theta_0$. This results in a mixed prior distribution sometimes called spike-and-smear configuration. It has the form:

$$P(\theta) = \pi_0 \delta_{\{\theta=\theta_0\}} + (1-\pi_0) g(\theta)_{I\{\theta\neq\theta_0\}} \tag{1}$$

Jeffreys' mixed prior distribution is a fusion of two components, a discrete part and a continuous part. The main component in the discrete part is $\delta_{\{\theta=\theta_0\}}$. It represents Paul Dirac's mass, or delta function, at θ_0. Furthermore, as advocated by Berger and Delampady (1987, p. 318), $\pi_0 = ½$ is the objective choice for the prior probability of H_0.

When using normal conjugate priors, the procedure boils down to the following. Suppose that $(x_1, x_2, \ldots, x_n)$ is a random sample from a normal distribution of the mean θ and known variance, $\sigma^2 > 0$. Using (1) assign the mass π_0 to the null point $\theta = \theta_0$, and spread the remaining mass out on H_1 according to the conjugate prior density $g(\theta) = N(\mu_0, \sigma_0^2)$, where μ_0 is the prior mean, and σ_0^2 is the prior variance. As pointed out by Berger and Sellke (1987, p. 112), this prior closely follows Jeffreys recommendation for testing a point null.

In this study Bayesian hypothesis testing is based on the Bayes factor. Utilization of Bayes factor as a measure for quantifying evidence has become increasingly popular in recent years in many branches of science. Bayes factor is advocated by Bayesians as the "primary tool used in Bayesian inference for hypothesis testing and model selection" (Berger, 2006, p. 378). Bayes factor for comparing a null hypothesis to the alternative can be defined as the ratio of the posterior odds in favor of the null hypothesis to the prior odds in favor of the null:

$$BF_{H_0H_1} = BF_{01} = \frac{\text{Posterior odds in favor of the null hypothesis}}{\text{Prior odds in favor of the null hypothesis}} = \frac{P(H_0 \mid Data) \,/\, P(H_1 \mid Data)}{P(H_0) \,/\, P(H_1)} \tag{2}$$

According to Kass and Raftery (1995, p. 777), "Bayes factor is a summary of the evidence provided by the data in favor of one scientific theory… as opposed to another." A more accurate notion of the Bayes factor is given by Lavine and Schervish (1999 p. 120): "What the Bayes factor actually measures is the change in the odds in favor of the hypothesis when going from the prior to the posterior". Intuitively, as pointed by Bernardo and Smith (1994, p. 390), "the Bayes factor provides a measure of whether the data **x** have increased or decreased the odds on H_i relative to H_j."

It can be proven (see, for example, Migon et al, 2014, p. 238) the Bayes factor in favor of the null over the alternative, in the aforementioned normal model, can be expressed as

$$BF_{01}(x)=\left[\left(\sigma^2+n\sigma_0^2\right)/\sigma^2\right]^{\frac{1}{2}}\exp\left\{\frac{n}{2}\left[\frac{\left(\bar{x}-\mu_0\right)^2}{\left(\sigma^2+n\sigma_0^2\right)}-\frac{\left(\bar{x}-\theta_0\right)^2}{\sigma^2}\right]\right\} \tag{3}$$

where $\bar{x}$ is the sufficient statistic for θ. To make a comparison between Bayesian credible intervals and Bayesian testing, using the exact same conditions as Berger and Sellke (1987) and Berger and Delampaday (1987), we will center prior density $g(\theta)$ over the hypothesized mean value, that is $\mu_0 = \theta_0$, and equate prior variance with the known variance, that is, $\sigma_0^2=\sigma^2$. Then (3) reduces to

$$BF_{01}(x)=\left(1+n\right)^{\frac{1}{2}}\exp\left\{-\frac{z^2}{2}\frac{n}{n+1}\right\} \tag{4}$$

where $z=\dfrac{\sqrt{n}\left(\bar{x}-\theta_0\right)}{\sigma}$ is a well-known Z test statistic. Hence, using (4), the posterior probability of H_0 is simply obtained as

$$\pi\left(H_0\,|\,x\right)=\left[1+\frac{\left(1-\pi_0\right)}{\pi_0}\frac{1}{BF_{01}}\right]^{-1}=\left[1+\frac{\left(1-\pi_0\right)}{\pi_0}\left(1+n\right)^{-\frac{1}{2}}\exp\left\{\frac{z^2}{2}\frac{n}{n+1}\right\}\right]^{-1} \tag{5}$$

Conflicts between Bayesian interval estimation and point-null hypothesis testing

Consider the Bayesian interval estimation of the mean of a normal distribution with known variance. Suppose $(x_1, x_2, \ldots, x_n)$ is a random sample from a normal distribution with the mean μ and known variance σ^2,$N(\mu,\sigma^2)$. As indicated by Murphy (2007, p. 2), because the likelihood is proportional to $N(\bar{x}\,|\,\mu,\sigma^2/n)$, the natural conjugate prior density for μ has the form $\mu\sim N\left(\mu_0,\sigma_0^2\right)$. Without loss of generality, assume σ^2 is known and that priors μ_0 and σ_0^2 are specified. It can be

easily seen (see, for example, Jackman, 2009, p. 516, or Murphy, 2007) that the posterior distribution of μ is normal with $\mu \mid \mathrm{x} \sim N\left(\mu_P, \sigma_P^2\right)$, with a posterior mean

$$\mu_p = \frac{n\sigma_0^2 \bar{x} + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2} = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0 \tag{6}$$

Hence, posterior mean μ_p is a weighted average of the sample mean and the posterior variance is given by

$$\sigma_P^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \tag{7}$$

Let $z_{\alpha/2}$ stand for the upper $\alpha/2$ quantile of the standard normal distribution. Then, the HPD (Highest Posterior Density) $100(1-\alpha)\%$ interval is given by

$$C_{1-\alpha}^{HPD} = \left[\mu_P - z_{\alpha/2}\sigma_p, \mu_P + z_{\alpha/2}\sigma_p \right] \tag{8}$$

where σ_p is a posterior standard deviation. HPD interval is the region of points containing $100(1-\alpha)\%$ of the posterior probability, with the following two main properties (as discussed by Box and Tiao, 1992, p. 123):

(a) Every point within the interval has posterior density at least as large as every point outside of it, and

(b) For a given credibility level $100(1-\alpha)\%$ it yields the posterior region of the shortest length among all credible intervals .

When a prior variance σ_0^2 becomes sufficiently large ($\sigma_0^2 \to \infty$) or when n is large, HPD intervals for the normal mean will yield essentially equal results as frequentist confidence intervals, even though they have different interpretations.

As pointed out by Ghosh et al. (2006, p. 49), and Carlin and Louis (2009, p. 50), for a unimodal symmetric posterior distribution (e.g., normal distribution), the equal tail credible intervals coincide with corresponding HPD intervals or, equivalently, Highest Density Intervals (HDI). In other words, (8) is simultaneously a HPD (HDI) interval and a credible interval. Therefore, the following discussion and conclusions are pertinent for credible intervals as well.

In Bayesian inference different types of priors are being used for constructing credible intervals and for testing via Bayes factors. As a consequence, the two following inconsistencies can be proven between these two Bayesian inferential methods:

I. A null hypothesized value may be outside the credible interval and therefore initially regarded as not credible but Bayes factor supports the null (Type I conflict)

II. The null hypothesized value belongs to the credible interval and is initially regarded as credible but Bayes factor reveals positive evidence against H_0 (Type II conflict).

Consider two theorems that set conditions for occurrences of Type I and II conflicts. The proofs are provided in the Appendix A.

Theorem 1 (Type I conflict theorem). Consider testing a point-null hypothesis of a normal mean $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. Let us further assume that the variance is known to be one, $\sigma^2 = 1$. Then, whenever Bayesian credible interval does not contain the point-null value (zero), and the following condition is satisfied

$$z_{\alpha/2}\frac{\sqrt{n+1}}{n} < \overline{x} < \frac{\sqrt{(n+1)\ln(n+1)}}{n} \tag{9}$$

Bayes factor and resultant posterior probability will support hypothesized null hypothesis.

Theorem 2 (Type II conflict theorem). Assume the same setup as in the Theorem 1. Whenever a Bayesian credible interval includes the hypothesized value (zero), and the next condition is met

$$\frac{\sqrt{(n+1)\ln(n+1)}}{n} < \overline{x} < z_{\alpha/2}\frac{\sqrt{n+1}}{n} \tag{10}$$

corresponding Bayes factor and posterior probability will indicate evidence against the null hypothesis.

Undoubtedly, Bayesian procedures in both cases are in direct conflict. This confirms testing point-null hypotheses using credible or HPD intervals as advocated by some statisticians including Lindley, Bolstad & Curran, and others (as considered earlier) should be immediately abandoned. Conflicting Type I errors will lead those who put their faith in credible intervals to reject H_0, but if they rely on Bayes factor (or posterior probabilities) they will support H_0. Conflicting Type II errors, however, will give such credible intervals that support H_0, but Bayes factors give enough counterevidence to reject H_0. These types of anomalies do not occur in the frequentist approach to Statistics.

The algorithms provided in Appendix B & C can generate countably infinite number of contradictory statements (conflicts Type I and II) in Bayesian Statistics. Using the simulation results obtained by these programs two interesting remarks may be derived: (1) for Type I conflict, as the sample size increases, corresponding Bayes factors become increasingly larger, thus providing more decisive evidence to support null hypothesis, and (2) Type II conflicts occur for a relatively small sample. Simulation results are illustrated in Table 1 and Table 2.

Table 1. Null values are outside the credible intervals but Bayes factors support the nulls

	Case 1A	**Case 1B**
Interval	95%	95%
Sample size n	100	1,000
$\overline{x}$	0.196974	0.062011
$C_{95\%}^{HPD}$	(0.0000000568, 0.390047468)	(0.0000005142, 0.1238975877)
BF	1.4723060	4.6349070
Posterior Probability	0.5955193	0.8225348

	Case 1C	**Case 1D**
Interval	95%	95%
Sample size n	1,000,000	100,000,000
$\overline{x}$	0.00196	0.000196
$C_{95\%}^{HPD}$	(0.000000035, 0.003919961)	(0000000036, 0.0003919964)
BF	146.4901	1464.897
Posterior Probability	0.9932199	0.9993178

Table 1 exemplifies four simulation results for which Bayes factors suggest that point nulls are supportable, but HPD (and credible intervals) do not include these hypothesized values. As stated in the Type I conflict theorem, in all situations

we were considering null hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$, and the variance was assumed to be 1. Furthermore, out of all observed Type I conflicts, we have displayed only results for the largest Bayes factors and posterior probabilities.

Following Jeffreys' scale (1961, p. 432) of interpretation of obtained values of the Bayes factor, it can be concluded all four null hypotheses are supported by the evidence. Bayes factors should seemingly encapsulate all the data have to say about these testing problems. However, in all four cases, hypothesized values are outside the Bayesian HPD interval. Hence, all four point null hypotheses should be rejected. However, unexpectedly, Bayes factor supports the nulls. The same inharmonious conclusions are derived when considering posterior probabilities. As the sample size is getting larger, posterior probabilities are approaching 1, thus giving the extreme evidence that null value should be favored. We maintain that these conflicts are inconsistent with the accepted scientific practice.

Exemplified in Table 2 are several Type II conflicting cases. Simulation results show Bayes factors will provide stronger evidence against null values with the increasing of the credibility level. For example, in case 2A, when the credible level is 99.99%, there is extremely strong evidence against the null (BF = 0.0063475708, and posterior probability is 0.0063075333), although the null value belongs to the credible interval. Figure 1 illustrates the notion of Type I and Type II conflicts in Bayesian inference.

Table 2. Bayes factors do not support null values that belong to the credible intervals (one million iterations)

	Case 2A	**Case 2B**	**Case 2C**
Interval	99.99%	99%	95%
Observed number of Type II Conflicts	135,224	42,968	4,864
Sample size n	150	100	10
$\overline{x}$	0.318722	0.258867	0.650046
	$C^{HPD}_{99.99\%}$: (-0.0000006192, 0.6332231)	$C^{HPD}_{99\%}$: (-0.0000006349, 0.5126086)	$C^{HPD}_{95\%}$: (-0.0000004672, 1.181902)
BF	0.0063475708	0.3642657456	0.4858872214
Posterior Probability	0.0063075333	0.2670049782	0.3270014133

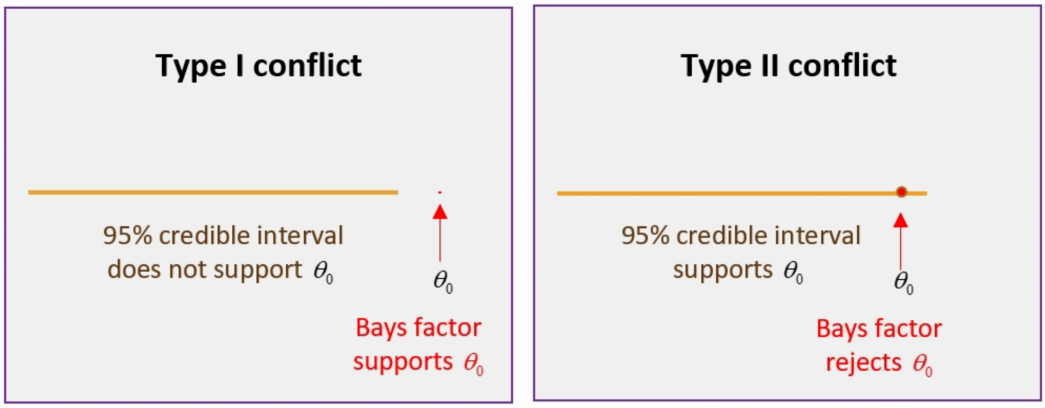

Figure 1. Type I and II conflicts in Bayesian inference

Conclusion

There are disturbing conflicts between Bayesian credible intervals and Bayes factor. They are caused by the unnatural configuration of the Jeffreys mixed prior distribution given in (1). Obviously, in some instances these priors are incompatible with priors that are used for establishing credible priors. These intrinsic disagreements within the foundation of Bayesian statistics raise elemental questions and deserve further intense study. At least in the Bayesian statistics, the current dominant way of testing null hypotheses using point null values is not suitable as an expression of uncertainty about the world in the 21st century.

If "a major goal of statistics (indeed science) is to find a completely coherent objective Bayesian methodology for learning from data" (Berger, 2006, p. 386), credible intervals and Bayes factors violate this goal; they are not complementary parts of the Bayesian vision when testing point null hypotheses. Furthermore, if frequentist statistics is not a coherent body of theory as stated by some Bayesians, Type I and II conflicts undoubtedly confirm that neither is Bayesian. Bayesians need to establish a new paradigm in statistical testing that will be consistent with interval estimation. This goal could be attained in different ways, including (a) finding the "statistical holy grail: prior distributions reflecting ignorance" (Fienberg, 2006a, p. 5), (b) using Jose Bernardo's integrated objective Bayesian estimation and hypothesis testing based on reference priors (2011), (c) developing a new paradigm of Bayesian testing such as in Kamary et. al. (2014), or (d) by abandoning

point null hypothesis and relying on tests based on practical significance as outlined in Rao and Lovric (2016).

Acknowledgements

This research was supported by the Artis College Faculty Research Grant, Radford University.

References

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd Ed. NY: Springer. doi: 10.1007/978-1-4757-4286-2

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3), 385–402. doi: 10.1214/06-ba115

Berger, J. O. & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association, 82*(397), 112-122. doi: 10.1080/01621459.1987.10478397

Berger, J. O. & Delampady, M. (1987). Testing Precise Hypotheses. *Statistical Science, 2*(3), 317-352. doi: 10.1214/ss/1177013238

Bernardo, J. & Smith, A. F. M. (1994). *Bayesian Theory*. New York: John Wiley. doi: 10.1002/9780470316870

Bernardo, J. M. (2005). Reference Analysis. In D. K. Dey and C. R. Rao (Eds.), *Handbook of Statistics, Volume 25: Bayesian Thinking, Modeling and Computation*. pp. 17-90. Amsterdam: North Holland. doi: 10.1016/s0169-7161(05)25002-2

Bernardo, J. M. (2010). Bayesian Statistics. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 107-133). Berlin: Springer.

Bernardo, J. M. (2011). Integrated objective Bayesian estimation and hypothesis testing. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.), Bayesian Statistics 9 (pp. 1-68). Oxford: Oxford: University Press. doi: 10.1093/acprof:oso/9780199694587.003.0001

Bolstad, W. M. & Curran J. M. (2017). *Introduction to Bayesian Statistics.* (3rd Ed.). NY: Wiley. doi: 10.1002/9781118593165

Box, G. E. P. & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis.* NY: Wiley. doi: 10.1002/9781118033197

Carlin, B. P. & Louis, T. A. (2009). *Bayesian Methods for Data Analysis.* (3rd Ed.). CRC Texts in Statistical Science. London: Chapman & Hall/CRC.

David, F. N. (1949). *Probability Theory for Statistical Methods*. Cambridge, UK: Cambridge University Press.

DeGroot, M. (1989). *Probability and Statistics*. (2nd Ed.). Boston, MA: Addison-Wesley.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology, 7*(1), 214. doi: 10.1186/1471-2148-7-214

Fienberg, S. E. (2006a). When Did Bayesian Inference Become "Bayesian"? *Bayesian Analysis, 1*(1), 1–40. doi: 10.1214/06-ba101

Fienberg, S. E. (2006b). Does it make sense to be an "objective Bayesian"? (Comment on articles by Berger and by Goldstein). *Bayesian Analysis, 1*(3), 429–432. doi: 10.1214/06-ba116c

Gelman, A. & Robert, C. P. (2013). Not Only Defended But Also Applied: The Perceived Absurdity of Bayesian Inference. *The American Statistician, 67*(1), 1-5. doi: 10.1080/00031305.2013.760987

Ghosh, J. K, Delampady, M. & Tapas, S. (2006). *An introduction to Bayesian analysis: theory and methods.* NY: Springer.

Good, I. (1971). 46,656 varieties of Bayesians. *American Statistician, 25*, 62–63.

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.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* NY: Wiley. doi: 10.1002/9780470686621

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511790423

Jeffreys, H. (1961). *Theory of Probability.* (3rd Ed.). Oxford, UK: Oxford University Press.

Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*(430), 773-795. doi: 10.1080/01621459.1995.10476572

Kamary, K., Mengersen, K., Robert, C. P. & Rousseau, J. (2014). Testing hypotheses via a mixture estimation model. (arXiv:1412.2044) Retrieved from https://arxiv.org/abs/1412.2044.

Kim, D. (1991). A Bayesian significance test of the stationarity of regression parameters. *Biometrika, 78*(3), 667–675. doi: 10.1093/biomet/78.3.667

Koch, K.-R. (2007). *Introduction to Bayesian Statistics.* Berlin, GER: Springer. doi: 10.1007/978-3-540-72726-2

Kruschke, J. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Cambridge, MA: Academic Press.

Lavine, M. & Schervish, M. J. (1999). Bayes Factors: What They Are and What They Are Not. *The American Statistician, 53*(2), 119-122. doi: 10.1080/00031305.1999.10474443

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Vol 2: Inference.* Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511662973

Migon, H. S., Gamerman, D. & Louzada, F. (2014). *Statistical Inference: An Integrated Approach.* (2nd Ed.). London: Chapman and Hall/CRC. doi: 10.1201/b17229

Murphy, K. P. (2007). Conjugate Bayesian analysis of the Gaussian distribution. (Technical report). Retrieved from: https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf

Rao, C. R. & Lovric, M. M. (2016). Testing Point Null Hypothesis of a Normal Mean and the Truth: 21st Century Perspective. *Journal of Modern Applied Statistical Methods, 15*(2) doi: 10.22237/jmasm/1478001660

Thulin, M. (2014). Decision-theoretic justifications for Bayesian hypothesis testing using credible sets. *Journal of Statistical Planning and Inference, 146*, 133-138. doi: 10.1016/j.jspi.2013.09.014

Appendix A: Proofs Of The Theorems

Proof of Theorem 1

Proof. First we will consider the upper bound in (9). Bayes factor as given in (4) supports the null hypothesis when it is larger than 1:

$$BF_{01}(x) = (1+n)^{\frac{1}{2}} \exp\left\{-\frac{z^2}{2}\frac{n}{n+1}\right\} > 1 \tag{A1}$$

When we take the logarithms of both sides of (A1) it becomes $\frac{1}{2}\ln(1+n) - \frac{z^2}{2}\frac{n}{n+1} > 0$, or $z^2 < (n+1)\ln(1+n)$. When $\theta_0 = 0$ and $\sigma^2 = 1$, this inequality reduces to

$$\bar{x}^2 < (n+1)\ln(1+n)/n^2 \tag{A2}$$

Since both sides of this inequality are positive, by taking the square root of both sides we obtain

$$-\frac{\sqrt{(n+1)\ln(1+n)}}{n} < \bar{x} < \frac{\sqrt{(n+1)\ln(1+n)}}{n} \tag{A3}$$

Now we turn our attention to the lower bound in (9). Since hypothesized value $\theta_0 = 0$ does not belong to the HPD (or credible) interval, consider the case when it is smaller than the lower bound of the interval

$$\mu_P - z_{\alpha/2}\sigma_p > 0 \tag{A4}$$

Using (6) and (7) we can easily expand (A4) to

$$\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 - z_{\alpha/2}\sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}} > 0 \tag{A5}$$

When $\theta_0 = 0$ and $\sigma^2 = 1$, by equating prior variance with the known variance ($\sigma_0^2 = \sigma^2 = 1$), and by setting the prior mean to zero ($\mu_0 = \theta_0 = 0$), (A5) reduces to

$$\frac{n}{n+1}\bar{x} - z_{\alpha/2}\sqrt{\frac{1}{n+1}} > 0$$

and finally to

$$\bar{x} > z_{\alpha/2}\frac{\sqrt{n+1}}{n} \tag{A6}$$

From (A2) and (A6) we obtain the bounds for the sample mean

$$z_{\alpha/2}\frac{\sqrt{n+1}}{n} < \bar{x} < \sqrt{\frac{(n+1)\ln(n+1)}{n}}$$

This completes the proof.

Proof of Theorem 2

Proof. First we will consider the upper bound in (10). Bayes factor as given in (4) favors alternative hypothesis when it is less than 1:

$$BF_{01}(x) = (1+n)^{\frac{1}{2}}\exp\left\{-\frac{z^2}{2}\frac{n}{n+1}\right\} < 1 \tag{A7}$$

By taking logarithms of both sides of (A7) it becomes $\frac{1}{2}\ln(1+n) - \frac{z^2}{2}\frac{n}{n+1} < 0$, or $z^2 > \frac{(n+1)}{n}\ln(n+1)$. When $\theta_0 = 0$ and $\sigma^2 = 1$, this inequality reduces to

$$\bar{x}^2 > \frac{(n+1)}{n^2}\ln(n+1) \tag{A8}$$

Since both sides of this equality are positive, by taking the square root of both sides we obtain

$$\bar{x} > \frac{\sqrt{(n+1)\ln(n+1)}}{n} \text{ and } \bar{x} < -\frac{\sqrt{(n+1)\ln(n+1)}}{n} \tag{A9}$$

Now we focus on the lower bound in (9). Since hypothesized value $\theta_0 = 0$ belongs to the HPD (or credible) interval, it follows that

$$\mu_P - z_{\alpha/2}\sigma_p < 0 < \mu_P + z_{\alpha/2}\sigma_p \tag{A10}$$

Using (6) and (7) we can easily transform (A10) to

$$\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 - z_{\alpha/2}\sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}} < 0,$$
$$\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + z_{\alpha/2}\sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}} > 0$$

When $\theta_0 = 0$ and $\sigma^2 = 1$, by equating prior variance with the known variance ($\sigma_0^2 = \sigma^2 = 1$), and by setting the prior mean to zero ($\mu_0 = \theta_0 = 0$), the bounds of this interval reduce to

$$\frac{n}{n+1}\bar{x} - z_{\alpha/2}\sqrt{\frac{1}{n+1}} < 0 < \frac{n}{n+1}\bar{x} + z_{\alpha/2}\sqrt{\frac{1}{n+1}} \tag{A11}$$

Now solving for $\bar{x}$ in the upper bound of (A11) we obtain

$$\bar{x} > -z_{\alpha/2}\frac{\sqrt{n+1}}{n} \tag{A12}$$

Similarly, solving for $\bar{x}$, the lower bound of (A11) gives

$$\bar{x} < z_{\alpha/2}\frac{\sqrt{n+1}}{n} \tag{A13}$$

Summarizing (A12) and (A13) we obtain

$$-z_{\alpha/2}\frac{\sqrt{n+1}}{n} < \bar{x} < z_{\alpha/2}\frac{\sqrt{n+1}}{n} \tag{A14}$$

From (A9) and (A14) we finally obtain the bounds for the sample mean

$$\frac{\sqrt{(n+1)\ln(n+1)}}{n} < \bar{x} < z_{\alpha/2}\frac{\sqrt{n+1}}{n}$$

This completes the proof.

Appendix B

The following R program was used in this paper to empirically prove Theorem 1 and produce values in Table 1.

```
#########################################################################
# This program generates Bayes factors that support
# null hypothesized values that are outside Bayesian credible interval,
#
# More specifically, null hypothesized value is set to 0
# and credible intervals created for which 0 is less
# than its lower bound.
#
# Note: the larger sample size, the larger BF and posterior probability.
#
# This analysis is based on the normal conjugate model with known
# variance
#
# IMPORTANT: Remove the leading # in line 124 to display all results
# Developed by M. Lovric August 1-10, 2019
#########################################################################
# Initial data
# Conflict Type I

sigma_2 <- 1              # Known variance

mu_prior <- 0             # prior mean
sigma_2_prior <- 1        # prior variance

n <- 10000                # choose your own sample size
conf.level = 0.95         # confidence level
#########################################################################
# BF_interpret function interprets values of Bayes factor according to
# the paper "Bayes factors", by Robert Kass & Adrian Raftery (1995),
# JASA, Vol. 90, No. 430. pp. 773-795.
#
# BF in this program is BF_0_1, not BF_1_0
# hence the inverse values are taken
```

```
#
BF_interpret <- function(BF){
  if (1/BF > 150){
    res <- "Very strong evidence against Ho."
  }else if(1/BF > 20){
    res <-  "We have strong evidence against Ho."
  }else if(1/BF > 3){
    res <-  "We have positive evidence against Ho."
  }else if(1/BF > 1){
    res <-  "Not worth more than a bare mention evidence against Ho."
  }else{
    res <- "supports Ho."
  }
}
alpha <- 1 - conf.level
## Bayesian credible interval

# First, find the critical value from the standard normal distribution
z_alpha_2 <- qnorm(1- alpha/2, lower.tail= TRUE)  # conf level = 95%
Theta_0 = 0
Z <- 0
Theta_0_Critical <- c()
BF_Critical <- c()
Z_Critical <- c()
p_value_critical <- c()
Post_H0_two_sided_critical <- c()
x_bar <-0
# prior prob under H_0
pi_H_0 <- pnorm(mu_prior, mean=mu_prior, sd = sqrt(sigma_2_prior))
# prior prob under H_1
pi_H_1 <- 1 - pnorm(mu_prior, mean=mu_prior, sd = sqrt(sigma_2_prior))

N = 100000 # Number of iterations

# This loop generates as many Bayes factors as you wish that support
# null hypothesized values (BF > 1), that are outside Bayesian
# credible interval, Point null hypothesis is :
# H_0: Theta = 0, vs H_1: Theta <> 0.
```

```
#
Output_me = TRUE

for (i in 1:N) {
  # Calculate posterior mean
  mu_post <- (sigma_2/(n*sigma_2_prior+sigma_2))*mu_prior +
    (n*sigma_2_prior/(n*sigma_2_prior+sigma_2))*x_bar

  # Calculate posterior variance
  variance_post <- (sigma_2 * sigma_2_prior)/(n * sigma_2_prior +
sigma_2)

  mu_post  # Posterior mean

  # Posterior standard deviation
  stdev_post <- sqrt(variance_post)

    # Now calculate credible interval
  LL <- mu_post -  z_alpha_2 * stdev_post
  UL <- mu_post +  z_alpha_2 * stdev_post
  Z <- (sqrt(n)*(x_bar - Theta_0))/sqrt(sigma_2) ;
  p_value <- 2*(1-pnorm(Z)) ;
  BF <- (1 + n)^(1/2) *exp((-0.5*Z^2)*n/(n+1))   ;
  Post_H0_two_sided <- (1 + ((1 - pi_H_0)/pi_H_0)*(1 + n)^(-1/2)*
                          exp((0.5*Z^2)*n/(n+1)))^(-1);
  if (0 < LL){
    if (BF > 1 ) {
      if (x_bar > z_alpha_2* sqrt(n+1)/n) {
        if (x_bar < (sqrt((n+1)*log(n+1)))/n) {
          BF_Critical <- c(BF_Critical, BF)
            if (Output_me) {
               cat("\014");
               AnalysisTime <- Sys.time();
               myInfo <- Sys.info()[["user"]];
               output1 = cat("Analysis done by ", myInfo, "\n");
               print(Sys.time());
               cat("\n");
               cat("LL = ", sprintf("%.10f", LL), " UL = ",
```

```
                       sprintf("%.10f", UL), " BF = ", BF, " p-value =",
                p_value, "\n");
                cat("Initial data: ", "\n", "n = ", sprintf("%.0f", n),
                       "\n", "x bar: ", x_bar, "\n",
               "Confidence level: ", conf.level*100, "%", "\n",
               "Known variance: ", sigma_2, "\n",
               "Prior mean: ", mu_prior, "\n",
               "Prior variance: ", sigma_2_prior, "\n", "\n");

                cat("Theta_0 = 0", "does not belong to the Bayesian",
               "HPD interval: ",  "\n", sprintf("%.10f", LL), " ",
                       sprintf("%.10f", UL),  "\n");

                cat("However in testing", " H_0: Population mean =  ", 0,
               "versus", "H_1: Population mean != 0","\n") ;

                cat("   H_0 is supported by the Bayes factor since
                       BF = ", BF, "\n" );
                cat("   H_0 is also supported by the posterior
                       probability = ",
              Post_H0_two_sided, "\n") ;
              Output_me = FALSE;
}
#  cat("LL = ", LL, "   BF = ", BF, " p-value =", p_value, "\n")
}
}
}
}
  x_bar <- x_bar + 0.000001
}

cat("Number of observed type I conflicts:", length(BF_Critical))
```

Appendix C

The following R program was used in this paper to empirically prove Theorem 2 and produce values in Table 2.

```
#########################################################################
# This program generates Bayes factors that do not support those
# null hypothesized values that are within Bayesian credible interval.
#
# More specifically, null hypothesised value is set to 0
# and credible intervals created for which 0 is within the interval.
#
# This analysis is based on the normal conjugate model with known
# variance
#
# Developed by M. Lovric August 1-10, 2019
#########################################################################
# Initial data
# Conflict Type II

sigma_2 <- 1              # Known variance

mu_prior <- 0             # prior mean
sigma_2_prior <- 1    # prior variance (not prior precision!)

n <- 10                       # choose your own sample size
conf.level = 0.95       # credible level
#########################################################################
# BF_interpret function interprets values of Bayes factor according to
# the paper "Bayes factors", by Robert Kass & Adrian Raftery (1995),
# JASA, Vol. 90, No. 430. pp. 773-795.
#
# BF in this program is BF_0_1, not BF_1_0
# hence the inverse values are taken
#
BF_interpret <- function(BF){
  if (1/BF > 150){
    res <- "Very strong evidence against Ho."
```

```
  }else if(1/BF > 20){
    res <-  "We have strong evidence against Ho."
  }else if(1/BF > 3){
    res <-  "We have positive evidence against Ho."
  }else if(1/BF > 1){
    res <-  "Not worth more than a bare mention evidence against Ho."
  }else{
    res <- "supports Ho."
  }
}

alpha <- 1 - conf.level
## Bayesian credible interval

# First, find the critical value from the standard normal distribution
z_alpha_2 <- qnorm(1- alpha/2, lower.tail= TRUE)  # conf level = 95%
Theta_0 = 0
Z <- 0
Theta_0_Critical <- c()
BF_Critical <- c()
Z_Critical <- c()
p_value_critical <- c()
Post_H0_two_sided_critical <- c()
x_bar <- -0.1
# prior prob under H_0
pi_H_0 <- pnorm(mu_prior, mean=mu_prior, sd = sqrt(sigma_2_prior))
# prior prob under H_1
pi_H_1 <- 1 - pnorm(mu_prior, mean=mu_prior, sd = sqrt(sigma_2_prior))

N = 1000000 # Number of iterations

# This loop generates as many Bayes factors as you wish that do not
# favor null hypothesized values (BF < 1), that are within a Bayesian
# credible interval, Point null hypothesis is :
# H_0: Theta = 0, vs H_1: Theta <> 0.
#
telliter <- 1000
```

```
for (i in 1:N) {
if( i %% telliter == 0 ) cat(paste("Iteration", i), "out of ",
sprintf("%.0f", N), "\n")
  # Calculate posterior mean
  mu_post <- (sigma_2/(n*sigma_2_prior+sigma_2))*mu_prior +
    (n*sigma_2_prior/(n*sigma_2_prior+sigma_2))*x_bar

  # Calculate posterior variance
  variance_post <- (sigma_2 * sigma_2_prior)/(n * sigma_2_prior +
                          sigma_2)

  # Posterior standard deviation
  stdev_post <- sqrt(variance_post)

  # Now calculate credible interval
  LL <- mu_post -  z_alpha_2 * stdev_post
  UL <- mu_post +  z_alpha_2 * stdev_post
  Z <- (sqrt(n)*(x_bar - Theta_0))/sqrt(sigma_2) ;
  p_value <- 2*(1-pnorm(Z)) ;
  BF <- (1 + n)^(1/2) *exp((-0.5*Z^2)*n/(n+1))   ;
  Post_H0_two_sided <- (1 + ((1 - pi_H_0)/pi_H_0)*(1 + n)^(-1/2)*
                          exp((0.5*Z^2)*n/(n+1)))^(-1);
  if ((LL < 0) & (UL > 0)){
    if (BF < 0.5 ) {
      if (x_bar < z_alpha_2* sqrt(n+1)/n) {
        if (x_bar > (sqrt((n+1)*log(n+1)))/n) {
              BF_Critical <- c(BF_Critical, BF, Post_H0_two_sided, LL,
                          UL, x_bar);
}
}
}
}
  x_bar <- x_bar + 0.000001
}

cat("\014");
AnalysisTime <- Sys.time();
myInfo <- Sys.info()[["user"]];
```

```
myInfo;
output1 = cat("Analysis done by ", myInfo, "\n");
print(Sys.time());
cat("\n");
cat("Sample size n = ", n, " x_bar: ", BF_Critical[length(BF_Critical)],
       "LL = ", sprintf("%.10f", BF_Critical[length(BF_Critical)-2]),
                          "UL = ",
       BF_Critical[length(BF_Critical)-1],  "\n", "BF = ",
                          sprintf("%.10f",
       BF_Critical[length(BF_Critical)-4]),
       " Posterior =", sprintf("%.10f", BF_Critical[length(BF_Critical)-
                          3]), "\n")

cat("Number of observed type II conflicts:", length(BF_Critical)/5)
```