The American Statistician

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

Benjamin, D. J., and Berger, J. O. (2019), "Three Recommendations for Improving the Use of *p*-Values", *The American Statistician*, 73, 186–191: Comment by Foulley

Jean-Louis Foulley

THE AMERICAN STATISTICIAN
2020, VOL. 74, NO. 1, 101–102: Letter to the Editor
https://doi.org/10.1080/00031305.2019.1668850

LETTER TO THE EDITOR

Benjamin, D. J., and Berger, J. O. (2019), "Three Recommendations for Improving the Use of *p*-Values", *The American Statistician*, 73, 186–191: Comment by Foulley

Benjamin and Berger (henceforth referred to as BB) pertinently advocated the users of NHST to convert p-values into Bayes factors according to a very simple, if not magic, formula: $\mathrm{BFB} = (-ep\log p)^{-1}$ (see recommendation 0.2). This formula is based on the well-known property of p-values uniformly distributed under the null hypothesis H_0. Under the alternative hypothesis H_1, the authors assume that p belongs to the flexible class of standard power-function distributions, that is, $\mathrm{Beta}(\xi, 1), 0 < \xi \leq 1$ (Sellke, Bayarri, and Berger 2001).

BB's formula reduces to the generalized likelihood ratio $G_{10} = f(p|\hat{\xi})/f(p|\xi = 1)$ in favor of H_1 versus H_0 where $f(p|\xi = 1) = 1$, and $\hat{\xi} = \arg_\xi \max f(p|\xi)$ is the maximum likelihood of ξ given the value of p observed. BFB can also be viewed as the value taken at the observed value p by the marginal distribution of p-values putting a point mass function at $\hat{\xi}$. Thus, as stated by BB, it provides an upper bound of the Bayes factor obtained under any prior distribution of the parameter ξ (see also Rougier 2019). Being an upper-bound of BF, it might be too high and still overestimate the strength of evidence in favor of H_1 versus H_0 (Berger and Sellke 1987; Sellke, Bayarri, and Berger 2001).

Our purpose is to investigate the magnitude of this "bias" toward H_1 using a generic prior distribution $\pi(\xi)$ for the class of $\mathrm{Beta}(\xi, 1)$ distributions of p-values. One way to relax shift toward H_1, is to consider the class of nonincreasing pdfs from $\xi = 1$ (H_0) to $\xi \rightarrow 0$ (H_1). A natural candidate lies in the PC priors introduced by Simpson et al. (2015). These priors penalize deviation from the base model and display, in addition, some interesting properties: they are proper and invariant to reparameterization.

Here they can be written under the general form:

$$\pi(\xi) = \lambda \exp[-\lambda D(\xi)] \left|D'(\xi)\right|, \tag{1}$$

where $D(\xi) = \sqrt{2\mathrm{KL}(\xi)}$ with $\mathrm{KL}(\xi) = \log\xi + \xi^{-1} - 1$ is the Kullback–Leibler distance between the pdfs of the flexible $f(p|\xi) = \xi p^{\xi-1}$ and of the base $f(p) = 1$ models, respectively (coefficient 2 for convenience); $\left|D'(\xi)\right| = \xi^{-2}(\xi - 1)(2\mathrm{KL})^{-1/2}$ is the Jacobian of the transformation of $D(\xi)$ to ξ, and λ a positive scalar.

It can be shown that $\pi(\xi)$ is a monotonically increasing function on $(0, 1)$ for $\lambda \geq 4/3$ thus meeting the requirement mentioned previously. Notice in passing the great flexibility of (1) regarding values of λ (Figure 1). We chose the value $\lambda = 4/3$ as the one providing the least favorable pdf for H_0 in the class of increasing prior functions on $(0, 1)$. For smaller values, priors put more weight on H_1 while it is the contrary for larger values (Figure 1). The computation of the value of the Bayes factor, $\mathrm{BF} = \int_\varepsilon^{1-\varepsilon} \xi p^{\xi-1}\pi(\xi)d\xi$, can be carried out easily by plugging the function under the integral sign with $\lambda = 4/3$ and the value of p observed into a web function calculator, for example, WIMS *https://wims.auto.u-psud.fr/wims/* with $\varepsilon = 10^{-4}$.

As expected BF and $\Pr(H_1|p)$ are lower than those shown in BB but the discrepancy is not so important practically in the range $0.1 \leq p \leq 0.005$ (Table 1). For example, a p-value of 0.05 corresponds to odds in favor of H_1 versus H_0 of 1.84 and 2.46 and corresponding $\Pr\left(H_1|p\right)$ of 0.65 and 0.71 for the PC(4/3) and BFB, respectively. At $p = 0.01$, there is still one chance among five of holding H_0. These figures again indicate the danger of misinterpreting the evidence provided by $p = 0.05$ against the null as stressed by BB and authors of the special issue of *The American Statistician* (2019).

An alternative would be to derive the BF using a conventional uniform prior on $\xi \in (0, 1)$ since Jeffrey's prior is here improper $\pi(\xi) \propto \xi^{-1}$. Integration leads to a simple algebraic expression for this BF:

$$\mathrm{BFU} = \left(1 + \frac{1-p}{p\log p}\right) / \log p. \tag{2}$$

Results are close to those obtained for $\lambda = 4/3$ (Table 1). Despite its astonishingly simple expression due to its derivation as a generalized likelihood ratio, the BB formula for BF turns out to be a good recommendation, at least for the short term, to prevent a misinterpretation of p-values toward disproving the null hypothesis. An alternative lies in using the PC or uniform priors as shown here for those more familiar to standard Bayesian procedures.

In the same vein, but using real data analyses, Bayarri et al. (2016) looked at the relationship between observed p-values from a range of scientific fields (epidemiology, genetic associations, ecology) and the corresponding BFB's (see Figure 3 of their paper). They concluded that "many of the estimated results lie fairly close" to the bound thus confirming empirically our own conclusions that BFB is theoretically well-founded.

As there is less evidence toward H_1 provided by fixed p-values when sample size (n) increases (Berger and Sellke 1987), it remains to see how to accommodate the formula for different n keeping in mind that one should preserve the consistency property of BF, that is, choosing the correct hypothesis as sample size increases. This problem has been addressed in great details by Held and Ott (2018) who proposed adjustments for n and also parameter dimension taking into account the kind of test-statistics used in calculating the p-value.

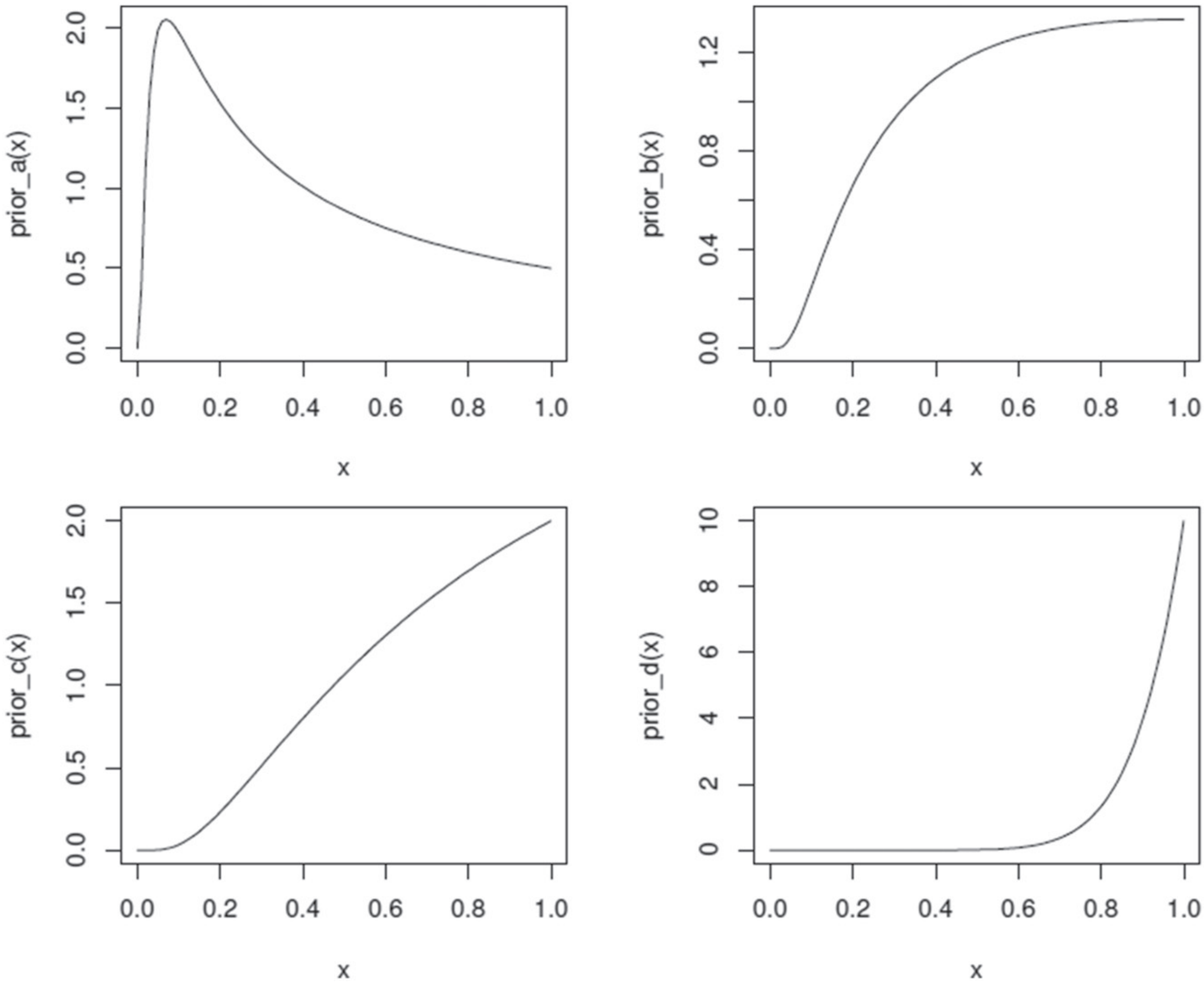

Figure 1. Examples of PC prior pdf's for the following λ values: $a = 0.5$, $b = 4/3$; $c = 2$, and $d = 10$.

Table 1. Bayes factors and corresponding probabilities of the alternative hypothesis $\Pr(H1|p)$ under different prior distributions.

p	0.10	0.05	0.01	0.005	0.001	0.0001	0.00001
BFB	1.60	2.46	7.99	13.9	53.3	399	3195
$\Pr(H1\|p)$	0.62	0.71	0.89	0.93	0.981	0.998	0.9997
BFP	1.36	1.84	4.04	5.89	15.3	71	370
$\Pr(H1\|p)$	0.57	0.65	0.80	0.85	0.939	0.986	0.9973
BFU	1.26	1.78	4.45	6.90	20.8	118	754
$\Pr(H1\|p)$	0.56	0.64	0.82	0.87	0.954	0.992	0.9987

NOTE: BFB: (upper) bound based on generalized likelihood ratio by Benjamin and Berger (2019). BFP: penalizing complexity prior with $\lambda = 4/3$. BFU: uniform prior.

Acknowledgments

The author is grateful to Prof. Christoph Leuenberger (Universität Freiburg, CH) for his hint on the proof of monotonicity of the prior function for $\lambda \geq 4/3$.

References

Bayarri, M. J., Benjamin, D. J., Berger, J. O., and Sellke, T. (2016), "Rejection Odds and Rejection Ratios: A Proposal for Statistical Practice in Testing Hypotheses," *Journal of Mathematical Psychology*, 72, 90–103. [101]

Benjamin, D. J., and Berger, J. O. (2019), "Three Recommendations for Improving the Use of *p*-Values", *The American Statistician*, 73, 186–191. [102]

Berger, J. O., and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of *P* Values and Evidence," *Journal of the American Statistical Association*, 82, 112–122. [101]

Held, L., and Ott, M. (2018), "On *p*-Values and Bayes Factors," *Annual Review of Statistics and Its Application*, 5, 393–519. [101]

Rougier, J. (2019), "*p*-Values, Bayes Factors, and Sufficiency," *The American Statistician*, 73, 148–151. [101]

Sellke, T., Bayarri, M. J., and Berger, J. O. (2001), "Calibration of *p* Values for Testing Precise Null Hypotheses," *The American Statistician*, 55, 62–71. [101]

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., Martins, T. G., and Sørbye, S. H. (2015), "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors," arXiv no. 1403.4630v4. [101]

Jean-Louis Foulley
IMAG, Université de Montpellier
Montpellier, France
foulleyjl@gmail.com