# Bayes factors based on $p$-values and sets of priors with restricted strength

Brendan Kline

Abstract. This paper focuses on the minimum Bayes factor compatible with a $p$-value, considering a set of priors with restricted strength. The resulting minimum Bayes factor depends on both the strength of the set of priors and the sample size. The results can be used to interpret the evidence for/against the hypothesis provided by a $p$-value in a way that accounts for the strength of the priors and the sample size. In particular, the results suggest further lowering the $p$-value cutoff for "statistical significance."

Keywords: $p$-value, Bayes factor, posterior probability

## 1. Introduction

It is important to understand what Bayesian conclusions can be drawn from classical $p$-values. Specifically, this paper concerns testing hypotheses of the form $\theta = c$, where $\theta$ is the parameter and $c$ is a candidate value for the parameter.[1] Often in practice, but not necessary for the theory, $c = 0$. The use of $p$-values is currently subject to important discussion, as in Wasserstein and Lazar (2016).

The following setup is standard relative to the related literature. The data comes from $N$ i.i.d. realizations of a random variable $X$, denoted as $X^{(N)} = \{X_i\}_{i=1}^{N}$. Let $f(X^{(N)}|\theta = \theta^*)$ be the likelihood of the data, given any particular value $\theta^*$ of the parameter $\theta$. As usual, the likelihood model will be taken to be a normal likelihood with an unknown mean $\theta$ and known variance $\sigma^2$. By standard results,

University of Texas at Austin
*E-mail address*: brendan.kline@austin.utexas.edu.
Date: January 2021. Thanks to the editor, associate editor, and referees for helpful comments and suggestions. Any errors are mine.

[1]This follows an important literature described below, but does distinguish it from certain papers that look at different questions entirely. Some papers have looked at the relationship between frequentist and Bayesian inference for one-sided hypotheses (e.g., Casella and Berger (1987) and Kline (2011)). Other papers have looked at the relationship between frequentist and Bayesian inference in partially identified models (e.g., Moon and Schorfheide (2012) and Kline and Tamer (2016)).

the frequentist ("classical") $p$-value is $p_N = 2(1-\Phi(|t_N|))$, based on $t_N = \frac{\hat{\theta}_N - c}{\sqrt{\frac{\sigma^2}{N}}}$, where $\hat{\theta}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ is the sample average. $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\phi(\cdot)$ is the standard normal probability density function. This is not based on an asymptotic approximation, given the setup. Anywhere in the statements of this paper $\Phi^{-1}(1-\frac{p_N}{2})$ can be replaced by $|t_N|$.

For the Bayesian analysis, the prior consists of two parts: a positive prior probability that $\theta = c$ and a prior on $\theta \neq c$. Let $\Pi_c = \Pi(\theta = c)$ be the prior probability that $\theta = c$ and let $g(\theta)$ be the density of the prior on $\Theta\backslash\{c\}$, where $\Theta$ is the parameter space for $\theta$. It is presumed throughout that $\Pi_c \in (0,1)$. The cases of $\Pi_c = 0$ and $\Pi_c = 1$ are not particularly interesting. Let $\mathcal{G}$ be a *set* of priors for $\Theta\backslash\{c\}$.

Let $\Pi(\theta = c|X^{(N)}; \Pi_c, g)$ be the posterior probability of the hypothesis $\theta = c$ based on data $X^{(N)}$, and the prior $\Pi_c$ and $g$. Let $B_c(g, X^{(N)})$ be the Bayes factor for the hypothesis $\theta = c$ based on the data $X^{(N)}$, and the prior $g$. $B_c(g, X^{(N)})$ is defined so that $\frac{\Pi(\theta=c|X^{(N)};\Pi_c,g)}{1-\Pi(\theta=c|X^{(N)};\Pi_c,g)} = B_c(g, X^{(N)})\frac{\Pi_c}{1-\Pi_c}$, and hence $B_c(g, X^{(N)}) = \frac{f(\hat{\theta}_N|\theta=c)}{\int_{\Theta\backslash\{c\}} f(\hat{\theta}_N|\theta)g(\theta)d\theta}$. The Bayes factor does not depend on $\Pi_c$.

As in Edwards, Lindman, and Savage (1963), Berger and Delampady (1987), and Berger and Sellke (1987) and a related literature, this paper focuses on the minimum Bayes factor over a set of priors $\mathcal{G}$, denoted as $\min_{g\in\mathcal{G}} B_c(g, X^{(N)})$, that is compatible with a given $p$-value. Because of the relationship $\Pi(\theta = c|X^{(N)}; \Pi_c, g) = \left(1 + \frac{1}{B_c(g,X^{(N)})}\frac{1-\Pi_c}{\Pi_c}\right)^{-1}$, there is a corresponding minimum posterior probability of the hypothesis $\theta = c$, $\left(1 + \frac{1}{\min_{g\in\mathcal{G}} B_c(g,X^{(N)})}\frac{1-\Pi_c}{\Pi_c}\right)^{-1}$ that is compatible with a given $p$-value. Hereafter, the minimum Bayes factor is known as minBF and the corresponding minimum Bayesian posterior probability of the hypothesis is known as minBPP. For any prior in $\mathcal{G}$, the Bayes factor (respectively, posterior probability of the hypothesis) is at least as great as the minBF (respectively, minBPP).

In this paper, the focus is on the consequence of restricting the strength of priors in $\mathcal{G}$ on the corresponding minBF. Important earlier related results can be viewed as having focused on polar special cases of this setup, in which the strength of the prior and/or sample size was implicitly taken to be at one extreme or the other. In particular, earlier results on minBFs had no restriction on the strength of priors in $\mathcal{G}$. Restricting the strength of priors in $\mathcal{G}$ has a substantial impact on the corresponding minBF.

The paper focuses on the case that $\mathcal{G} = \mathcal{C}_{\overline{\tau}} = \{g : g = \mathcal{N}(c, s^2) \text{ and } s^2 \geq \overline{\tau}^{-1}\}$ where $\overline{\tau} \in (0, \infty)$ is a statistician-specified upper bound on the precision of the priors. $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean $\mu$ and variance $\sigma^2$. Correspondingly, the precision is $\tau = \frac{1}{\sigma^2}$. $\mathcal{C}_{\overline{\tau}}$ is the set of normal priors centered at the null hypothesis $c$, with variance at least $\overline{\tau}^{-1}$. Equivalently, $\mathcal{C}_{\overline{\tau}}$ is the set of normal priors centered at the null hypothesis $c$, with precision no greater than $\overline{\tau}$. The $\mathcal{C}$ notation refers to normal priors that are *centered* at $c$. Priors with smaller variances (higher precision) are interpreted to be stronger priors. Thus, $\overline{\tau}$ can be interpreted as an upper bound on the strength of the priors. Relatively larger $\overline{\tau}$ correspond to a set of priors $\mathcal{C}_{\overline{\tau}}$ that includes relatively stronger priors. Conversely, relatively smaller $\overline{\tau}$ correspond to a set of priors $\mathcal{C}_{\overline{\tau}}$ that includes only relatively weaker priors. Moreover, by an appropriate scaling, $\overline{\tau}$ can be interpreted in terms of the "prior sample size" that measures the number of pseudo-observations associated with the prior, or equivalently, the "virtual sample size" associated with the prior. Specifically, if the precision of a normal prior is $\frac{M}{\sigma^2}$, then $M$ can be interpreted as the "prior sample size" or equivalently the size of the virtual sample associated with the prior.[2] Therefore, $\mathcal{C}_{\frac{M}{\sigma^2}}$ can be interpreted to be associated with virtual samples with virtual sample size no greater than $M$. The size of the virtual sample gives an interpretation of the strength of the prior. Any $\overline{\tau}$ can be "translated" to an equivalent, possibly non-integer, virtual sample size by multiplying by $\sigma^2$.

The minBF depends on $\frac{M}{N}$, which is the *relative* maximal strength of the priors compared to the data sample size $N$. This is the result of Section 2. It is also possible to compute related $p$-value cutoffs. These cutoffs can be used to define "statistical significance" based on a $p$-value, in terms of the corresponding minimum Bayes factor. For any given Bayes factor $b$, the corresponding "$p$-value cutoff" is the level of the $p$-value such that the minimum Bayes factor corresponding to that $p$-value is equal to $b$. This is the result of Section 3.

[2]Recall the following standard arguments. Consider a normal prior for $\theta$ with mean $m$ and variance $s^2$. Then, by standard Bayesian arguments, the posterior for $\theta$ based on a *virtual* sample of $M$ i.i.d. observations would have mean $\frac{\frac{M}{\sigma^2}\hat{\theta}_M + \frac{1}{s^2}m}{\frac{M}{\sigma^2}+\frac{1}{s^2}}$ and variance $\frac{1}{\frac{M}{\sigma^2}+\frac{1}{s^2}} = \frac{\sigma^2}{M+\frac{\sigma^2}{s^2}} = \frac{\sigma^2}{M(1+\frac{\sigma^2}{s^2}\frac{1}{M})}$, where $\hat{\theta}_M$ is the virtual sample average. If $\frac{\sigma^2}{s^2}\frac{1}{M}$ is small, which can arise when $s^2$ is large (a weak "initial" prior) and/or when $M$ is large (a large virtual sample), the posterior for $\theta$ is normal with variance approximately $\frac{\sigma^2}{M}$. Using this "posterior" from a *virtual* sample as the prior for the "actual" analysis would therefore result in a normal prior with variance $\frac{\sigma^2}{M}$ where $M$ is the sample size of the virtual sample.

The Appendix has results for three other classes of priors: the set of normal priors that are not restricted to be centered at the null hypothesis (Appendix A), the set of all priors with a density (Appendix B), and the set of normal priors restricted to be centered at the null hypothesis with both an upper bound and lower bound on the precision (Appendix C). Those results are summarized and compared to the main results at the end of the Conclusions. Appendix D contains the technical details of the proofs.

## 2. Minimum Bayes factors

Theorem 1 establishes the relationship between a $p$-value and the minBF for the set of priors $\mathcal{C}_{\frac{M}{\sigma^2}}$. As noted above, $M$ can be interpreted as the largest allowed size of the "virtual sample" associated with the prior.

**Theorem 1** (Results for $\mathcal{C}_{\frac{M}{\sigma^2}}$). *The minimum Bayes factor over the set of priors $\mathcal{C}_{\frac{M}{\sigma^2}}$ is*

$$\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \begin{cases} \frac{\sqrt{1+\frac{N}{M}}\phi\left(\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)} & \text{if } 1 \geq \frac{M}{N}([\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 - 1) \\ \frac{\phi(\Phi^{-1}\left(1-\frac{p_N}{2}\right))|\Phi^{-1}\left(1-\frac{p_N}{2}\right)|}{\phi(1)} & \text{if } 1 < \frac{M}{N}([\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 - 1) \end{cases}. \tag{1}$$

*The above representation implies also that:*

*(1) For any given $p_N$, $\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly decreasing function of $\frac{M}{N}$.*

*(2) For any given $\frac{M}{N}$, $\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly increasing function of $p_N$.*

Per Theorem 1, the minBF depends on the ratio of the largest allowed size of the "virtual sample" associated with the priors compared to the actual data sample size. As summarized in Corollary 1, existing related results can be recovered as special cases when $\frac{M}{N} \to \infty$ or $\frac{M}{N} \to 0$.

**Corollary 1** (Results for $\mathcal{C}_{\frac{M}{\sigma^2}}$). *When $\mathcal{C}_{\frac{M}{\sigma^2}}$ includes very strong priors relative to sample size,*

$$\lim_{\frac{M}{N} \to \infty} \min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \begin{cases} 1 & \text{if } 1 \geq [\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 \\ \frac{\phi(\Phi^{-1}\left(1-\frac{p_N}{2}\right))|\Phi^{-1}\left(1-\frac{p_N}{2}\right)|}{\phi(1)} & \text{if } 1 < [\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 \end{cases} \tag{2}$$

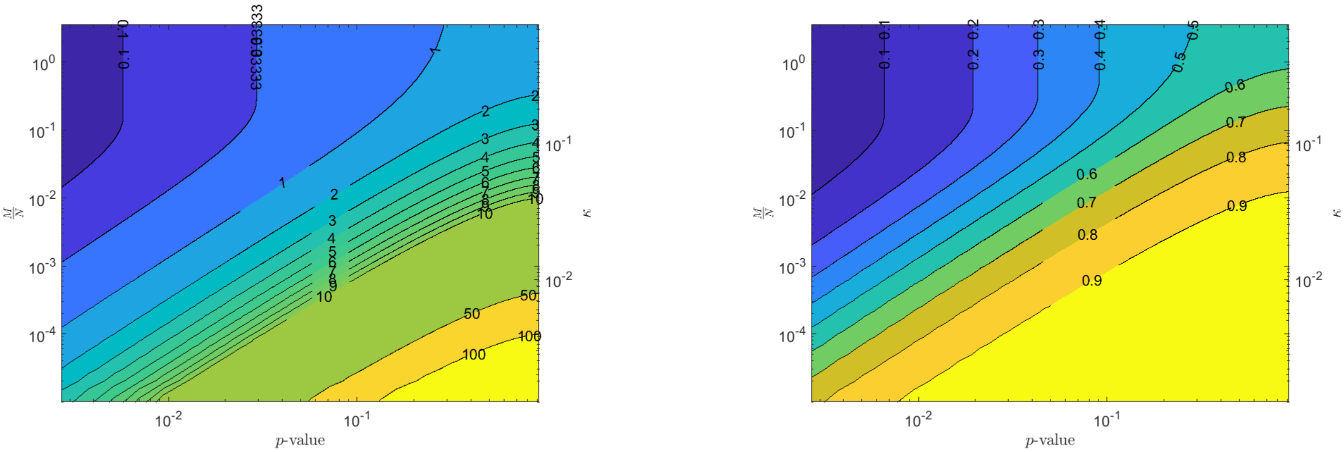
*When $\mathcal{C}_{\frac{M}{\sigma^2}}$ only includes very weak priors relative to sample size,*

$$\lim_{\frac{M}{N}\to 0} \min_{g\in\mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \infty \tag{3}$$

The results of Corollary 1 in Equation 2 concern the case of no restriction on the strength of the prior relative to sample size, and match the related results from Edwards, Lindman, and Savage (1963) and Berger and Sellke (1987) with normal priors centered at the null hypothesis. This means that those results can be understood to concern the polar special case of allowing for extremely strong priors relative to sample size. The result of Corollary 1 in Equation 3 match the results from the "Jeffreys (1939)-Lindley (1957) paradox." In other words, the Jeffreys (1939)-Lindley (1957) results are effectively the polar special case of extremely weak priors relative to sample size. The Sellke, Bayarri, and Berger (2001) approach is not a special case of the approach taken in this paper, but numerically those results are reasonably similar to results for $\mathcal{C}_{\frac{M}{\sigma^2}}$ with strong priors relative to sample size (i.e., large $\frac{M}{N}$). Along similar lines, Held and Ott (2016) propose a sample-size adjusted minBF for a linear model with $g$-priors. The adjustment for sample size results in minBFs somewhat smaller than those from Sellke, Bayarri, and Berger (2001), especially when sample size is less than 20 or so.[3] Indeed, the $g$-priors used by Held and Ott (2016) are reasonably similar to these normal priors centered at the null hypothesis. Similar to the derivation in this paper, the "$g$" could be restricted, very much like restricting the precision of the priors here, thereby resulting in a minBF based on $g$-priors with restricted strength.

Figure 1a plots the minBF as a function of $\frac{M}{N}$ and the $p$-value. Figure 1b plots the minBPP as a function of $\frac{M}{N}$ and the $p$-value, for $\Pi_c = 0.50$. In both figures, the axes are both on a logarithmic scale. On the left axis of these figures is $\frac{M}{N}$. For purposes of references from the Appendix, on the right axis of these figures is the equivalent "$\kappa$" which is relevant for other classes of priors considered in the Appendix. Specifically, $\kappa = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{M}{N}}$. Tables 1a and 1b tabulate the minBFs and minBPP, for

[3]The way the Held and Ott (2016) minBF depends on $N$ is qualitatively different from the current paper. As $N \to \infty$, the Held and Ott (2016) minBF approaches the finite Sellke, Bayarri, and Berger (2001) minBF. In the current paper, as $N \to \infty$ (for fixed strength of priors), the minBF approaches $\infty$.




(A) Minimum Bayes factor for $\mathcal{C}_{\frac{M}{\sigma^2}}$

(B) Minimum posterior probability for $\mathcal{C}_{\frac{M}{\sigma^2}}$, with $\Pi_c = 0.50$

FIGURE 1. Minimum Bayes factors and minimum posterior probabilities

selected values of the $p$-value and $\frac{M}{N}$. As with the figures, the tables display the equivalent $\kappa$ for each $\frac{M}{N}$. Readers not interested in those other classes of priors can ignore $\kappa$.

The minBF is a decreasing function of $\frac{M}{N}$, and consequently is a decreasing function of $M$ and an increasing function of $N$. In particular, the minBF from the literature that has no restriction on the strength of the priors is *smaller* than the minBF in this paper with a restriction on the strength of the priors. When minBFs are used as a "calibration" of $p$-values, this essentially means that the minBFs from the literature that have no restriction on the strength of priors can "overstate" the evidence against the null hypothesis. The results show that the minBF depends substantially on the relative strength of the set of priors compared to the data sample size.

In particular, for some values of $\frac{M}{N}$, there are values of $p_N$ that might be understood to suggest evidence *against* the hypothesis that are actually circumstances in which there is "consensus" evidence *for* the hypothesis. This paper says that there is "consensus" evidence for the null hypothesis when the minBF exceeds 1, which implies that the minBPP *exceeds* the prior probability. This condition means that there is a "consensus" that there is evidence for the null hypothesis, in the sense that the posterior probability of the null hypothesis exceeds the prior probability of the null hypothesis, *for every prior in the set of priors.* For example, consider a $p$-value of 0.05. If the strength of the priors

| | $\frac{M}{N}$ | $\kappa$ | $p$-value | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.001 | 0.0025 | 0.005 | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.4457 | 1.036 | 1.946 | 3.626 | 14.65 | 25.86 | 51.61 | 79.66 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.1417 | 0.3291 | 0.6179 | 1.151 | 4.644 | 8.19 | 16.34 | 25.21 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.04723 | 0.1089 | 0.2033 | 0.3764 | 1.501 | 2.633 | 5.22 | 8.023 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.02642 | 0.05899 | 0.1075 | 0.1945 | 0.7356 | 1.264 | 2.44 | 3.69 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.02417 | 0.05203 | 0.09231 | 0.1625 | 0.5786 | 0.9696 | 1.817 | 2.697 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.02417 | 0.05162 | 0.09003 | 0.1539 | 0.4734 | 0.7011 | 1.016 | 1.262 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.02417 | 0.05162 | 0.09003 | 0.1539 | 0.4734 | 0.7011 | 0.9786 | 1.003 |

(A) Minimum Bayes factors for $\mathcal{C}_{\frac{M}{\sigma^2}}$

| | $\frac{M}{N}$ | $\kappa$ | $p$-value | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.001 | 0.0025 | 0.005 | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.308 | 0.509 | 0.661 | 0.784 | 0.936 | 0.963 | 0.981 | 0.988 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.124 | 0.248 | 0.382 | 0.535 | 0.823 | 0.891 | 0.942 | 0.962 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.0451 | 0.0982 | 0.169 | 0.273 | 0.6 | 0.725 | 0.839 | 0.889 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.0257 | 0.0557 | 0.0971 | 0.163 | 0.424 | 0.558 | 0.709 | 0.787 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.0236 | 0.0495 | 0.0845 | 0.14 | 0.367 | 0.492 | 0.645 | 0.73 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.0236 | 0.0491 | 0.0826 | 0.133 | 0.321 | 0.412 | 0.504 | 0.558 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.0236 | 0.0491 | 0.0826 | 0.133 | 0.321 | 0.412 | 0.495 | 0.501 |

(B) Minimum posterior probability for $\mathcal{C}_{\frac{M}{\sigma^2}}$, when $\Pi_c = 0.50$

TABLE 1. Results for $\mathcal{C}_{\frac{M}{\sigma^2}}$

relative to the data sample size is unrestricted, then the minBF is 0.4734. That is equivalent to a minBPP of 0.321, given a 0.50 prior probability. This suggests evidence against the null hypothesis. However, if the priors are restricted to be no more than 1% as strong as the data sample size (i.e., $\frac{M}{N} = 0.01$), then the minBF is 1.501. That is equivalent to a minBPP of 0.60, given a 0.50 prior probability. In the latter case, there actually is "consensus" evidence *for* the null hypothesis. This is a very different conclusion compared to the former case, when there was no restriction on the strength of the priors.

More generally, the relative strength of the set of priors compared to the data sample size has an important impact on the associated minBF. For example, consider a $p$-value of 0.005. This is the lowered $p$-value cutoff for "statistical significance" proposed by Benjamin, Berger, et al. (2018). Based on a range of approaches relating $p$-values to Bayes factors, including the minBF considerations from Sellke, Bayarri, and Berger (2001), and in the same basic setup of a normal likelihood with unknown mean and known variance and the same test statistic that generates the $p$-value, the conclusion of Benjamin, Berger, et al. (2018) is that a $p$-value of 0.005 is associated with Bayes factors between approximately 0.04 and 0.07 and correspondingly is associated with posterior probabilities of the hypothesis between approximately 0.04 and 0.07, given a 0.50 prior probability. However, consider the minBF from Theorem 1 with $\frac{M}{N} = 0.01$. In that case, a $p$-value of 0.005 is associated with a minBF of 0.2033. This is meaningfully higher than even the high end of the range of Bayes factors suggested by Benjamin, Berger, et al. (2018). That is equivalent to a minBPP of 0.169, given a 0.50 prior probability. Thus, when the strength of priors is restricted in this way, a $p$-value of 0.005 is associated with meaningfully less evidence against the null hypothesis. This suggests, extending the reasoning of Benjamin, Berger, et al. (2018) to a setting with priors with restricted strength compared to data sample size, that the $p$-value cutoff for "statistical significance" should be even lower still in applications where $\frac{M}{N}$ is not large. This point is elaborated in Section 3.

Overall, a main feature of the results is the dependence on $\frac{M}{N}$. This might seem to introduce extra subjectivity into the results, compared to existing results on minBFs. Those existing results may seem less subjective, in the sense that those results may be perceived to avoid an (explicit) specification of $\frac{M}{N}$ and/or are perceived to minimize dependence on the set of priors. This could be perceived to be an advantage of those existing results, since a reasonable goal of (objective) Bayesian analysis is to minimize dependence on the prior. However, per Corollary 1 and surrounding discussion, those existing results do depend on the set of priors, specifically relying on the condition that $\frac{M}{N}$ is large. Implicitly relying on the condition that $\frac{M}{N}$ is large does not minimize the dependence on the set of priors, and indeed the results show precisely how those minBFs depend on allowing for strong priors relative to the data sample size. Hence, those results do not minimize dependence on the set of priors. In particular, as discussed above by numerical example, relying on the condition that $\frac{M}{N}$ is large can result in "overstating" the evidence against the null hypothesis if in fact $\frac{M}{N}$ is not large, for example in larger datasets and/or with novel empirical questions such that there is relatively little prior information. Of course, existing results capture the situation that arises in small datasets, and/or with strong priors. The situation is different in large datasets, and/or with weak priors.

Because $\mathcal{C}_{\overline{\tau}} = \mathcal{C}_{\frac{\overline{\tau}\sigma^2}{\sigma^2}}$, Theorem 1 also gives the results for any choice of $\overline{\tau}$ in $\mathcal{C}_{\overline{\tau}}$ by choosing $M = \overline{\tau}\sigma^2$. The analysis does not presume $M$ to be an integer. In that case, the minBF depends on the ratio $\frac{\overline{\tau}\sigma^2}{N} = \frac{\overline{\tau}}{\frac{N}{\sigma^2}}$, which concerns the ratio of the maximum precision of the priors compared to the (frequentist sampling) precision of the estimator $\hat{\theta}_N$.

## 3. $p$-VALUE CUTOFFS FOR "STATISTICAL SIGNIFICANCE"

It is possible to derive $p$-value cutoffs for "statistical significance" that account for how $p$-values relate to Bayes factors. In general, let $\underline{p}_{\mathcal{G}}(b)$ be the $p$-value such that $\min_{g \in \mathcal{G}} B_c(g, X^{(N)}) = b$, which is the $p$-value that results in a minBF equal to any specified $b$, given a specified set of priors $\mathcal{G}$. To find $\underline{p}_{\mathcal{G}}(b)$, it is enough to solve for $p_N$ in the equation that sets $\min_{g \in \mathcal{G}} B_c(g, X^{(N)})$ equal to $b$. For $\mathcal{G} = \mathcal{C}_{\frac{M}{\sigma^2}}$, this is a trivial computational exercise, given the characterization of $\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)})$ in Theorem 1. Because of the monotonicity result in Theorem 1, any $p$-value above this "cutoff" results in a minBF greater than $b$, and any $p$-value below this "cutoff" results in a minBF less than $b$.

Table 2a reports the cutoff $p$-values: $\underline{p}_{\mathcal{C}_{\frac{M}{\sigma^2}}}(b)$ for different $b$ and different $\frac{M}{N}$. The $p$-value being below $\underline{p}_{\mathcal{G}}(1)$ is necessary but not sufficient for "rejecting the hypothesis." If the minBF exceeds 1, then for any prior in $\mathcal{G}$, the posterior probability of the hypothesis exceeds the prior probability of the hypothesis. That can be interpreted to mean that the data provides consensus evidence for the hypothesis. Alternatively, if the minBF falls below 1, then for at least some priors in $\mathcal{G}$, the posterior probability of the hypothesis falls below the prior probability of the hypothesis. Not finding consensus evidence for the hypothesis is presumably not enough to "reject the hypothesis." As such, it may be useful to consider $p$-value cutoffs based on different $b$. According to Jeffreys (1939) and also Kass and Raftery (1995), a Bayes factor of approximately $\frac{1}{3}$ is the cutoff for finding "substantial" or "positive" evidence against the hypothesis. According to Jeffreys (1939), a Bayes factor of approximately $\frac{1}{10}$ is the cutoff for finding "strong" evidence against the hypothesis. According to Kass and Raftery (1995), a Bayes factor of approximately $\frac{1}{20}$ is the cutoff for finding "strong" evidence against the hypothesis.

| | $\frac{M}{N}$ | $\kappa$ | $b$ | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 0.33333 | 0.1 | 0.05 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.002405 | 0.000731 | 0.0002015 | 9.653e-05 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.008544 | 0.002535 | 0.0006865 | 0.0003263 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.03085 | 0.008714 | 0.002276 | 0.001064 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.07378 | 0.01897 | 0.004595 | 0.002068 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.1044 | 0.02456 | 0.005512 | 0.002383 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.239 | 0.02918 | 0.005714 | 0.002404 |
| $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.3161 | 0.02918 | 0.005714 | 0.002404 |

(A) Cutoff $p$-values

| | | | $b$ | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 0.33333 | 0.1 | 0.05 |
| $p$-value | 0.001 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 1.985e-05 | 0.000179 | 0.002033 | 0.008801 |
| | 0.0025 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.0001073 | 0.0009746 | 0.01211 | n/p |
| | 0.005 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.0003797 | 0.003514 | 0.0654 | n/p |
| | 0.01 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.001327 | 0.01305 | n/p | n/p |
| | 0.05 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.02405 | n/p | n/p | n/p |
| | 0.1 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 0.09153 | n/p | n/p | n/p |
| | 0.25 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | 1.246 | n/p | n/p | n/p |
| | 0.5 | $\mathcal{C}_{\frac{M}{\sigma^2}}$ | n/p ($\infty$) | n/p | n/p | n/p |

(B) Strength of priors relative to sample size that results in a given minimum Bayes factor, for given $p$-value. n/p, not possible (there is no strength of prior relative to sample size that results in that minimum Bayes factor for that $p$-value). n/p ($\infty$), not possible for finite values of $\frac{M}{N}$ but achievable "in the limit" as $\frac{M}{N} \to \infty$.

TABLE 2. Results for $\mathcal{C}_{\frac{M}{\sigma^2}}$

A main conclusion is that the $p$-value cutoff for "statistical significance" should be smaller when $\frac{M}{N}$ is smaller. Again, the reason is that $p$-values are associated with less evidence against the null hypothesis when $\frac{M}{N}$ is smaller. How much smaller the cutoff for "statistical significance" should be depends on the application-specific strength of the priors relative to sample size, and the Bayes factor that justifies "rejecting" the hypothesis. For example, in "big data" applications with large $N$, it might be reasonable to consider that $\frac{M}{N}$ is around 0.01 (or possibly smaller). And it might be reasonable to require a Bayes factor of around 0.1 to "reject" the hypothesis, reflecting "strong" evidence against the hypothesis, which roughly matches the suggestion of Benjamin, Berger, et al.

(2018). In that case, a $p$-value of less than around 0.0025 should be required to "reject" the hypothesis. Thus, this suggests considering further reducing the cutoff for "statistical significance" compared even to Benjamin, Berger, et al. (2018) for such applications. The analysis in Appendix A shows essentially the same conclusion would follow if the priors were normal priors that are not restricted to be centered at the null hypothesis, and the analysis in Appendix B does the same for general priors with a density of the same "strength."

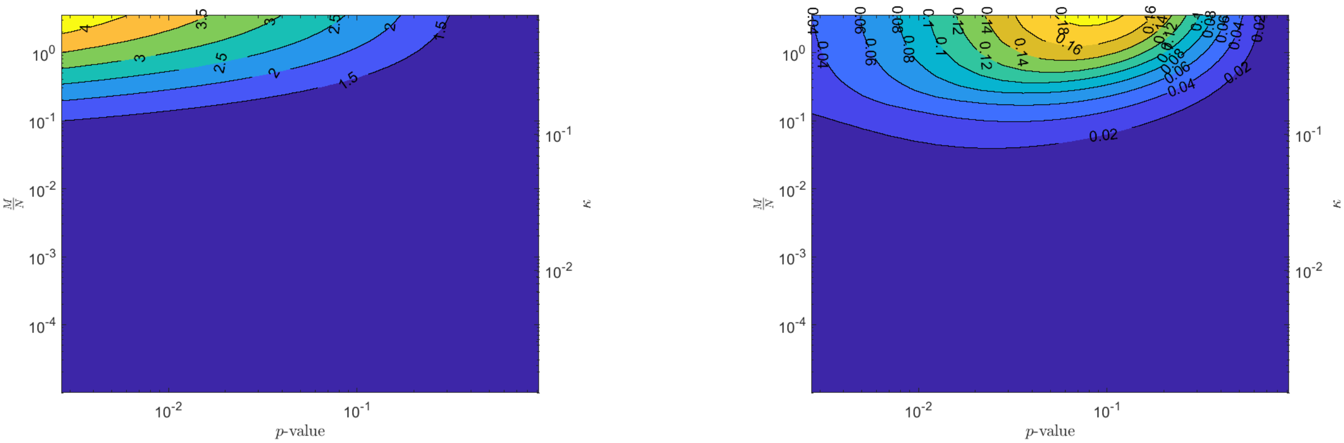
Although the difference between $p$-value cutoffs of 0.0025 and 0.005 might seem modest, Tables 1a and 1b show there is a meaningful difference in the associated minBF and minBPP. When $\frac{M}{N}$ is large, the results in this paper do not provide any reason to change the proposal made by Benjamin, Berger, et al. (2018). When $\frac{M}{N}$ is large, a $p$-value of 0.005 is associated with a minBF of around 0.1, thereby reflecting "strong" evidence against the hypothesis, as discussed in Benjamin, Berger, et al. (2018). But, when for example $\frac{M}{N} = 0.01$, a $p$-value of around 0.0025 is associated with a minBF of around 0.1.

Another way to use these results is to determine, given a particular $p$-value, how strong the set of priors would need to be relative to sample size in order to be able to draw a selected conclusion, in terms of the corresponding minBF. For example, to find the $\frac{M}{N}$ that results in a minBF equal to any specified $b$, given a specified $p$-value and based on set of priors $\mathcal{C}_{\frac{M}{\sigma^2}}$, it is enough to solve for $\frac{M}{N}$ in Equation 1, when $\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B(g, X^{(N)})$ is specified to be $b$. This is a trivial computational exercise, given the characterization in Theorem 1. The results for selected $p$-values are displayed in Table 2b. For example, for a $p$-value of 0.005 to result in a minBF of 0.1, it would be necessary that $\frac{M}{N} = 0.0654$, which requires the application be such that it is reasonable that there is a prior that is about 6.5% as strong as the data. This reinforces the argument that there is reason to lower the $p$-value cutoff for statistical significance beyond even the 0.005 proposed by Benjamin, Berger, et al. (2018), for applications such that $\frac{M}{N}$ is smaller.

## 4. Conclusions

This paper has derived the minimum Bayes factor compatible with a given $p$-value, based on priors with restricted strength. The resulting minimum Bayes factor depends on the relative strength of the set of priors compared to the data sample size (Section 2). These results can be used to derive cutoff $p$-values for "statistical significance" (Section 3). A main conclusion is that the $p$-value cutoff for "statistical significance" should in some applications be smaller than even the 0.005 proposed by Benjamin, Berger, et al. (2018), depending on the application-specific strength of priors relative to data sample size. Some notable existing results in the literature can be interpreted to be polar special cases of the results in this paper.

The Appendix has results for three other sets of priors: the set of priors with a density, the set of normal priors that are not restricted to be centered at the null hypothesis, and the set of normal priors restricted to be centered at the null hypothesis with both an upper bound and lower bound on the precision. The results for the set of normal priors that are not restricted to be centered at the null hypothesis (in Appendix A) and the results for the set of normal priors that are restricted to be centered at the null hypothesis (the main results of the paper) are basically the same, as long as $\frac{M}{N}$ is less than approximately 0.1. However, if $\frac{M}{N}$ is substantially larger than 0.1, then there can be a meaningful difference between the results for all normal priors and normal priors restricted to be centered at the null hypothesis. The results for the set of priors with a density (in Appendix B) and the results for the set of normal priors that are not restricted to be centered at the null hypothesis (in Appendix A) are basically the same, as long as the strength of the set of priors "matches" in a way formalized later on. Therefore, whether the set of priors includes only normal priors, or all priors with a density, does not make much of a difference. Finally, in Appendix C, for the set of normal priors centered at the null hypothesis with the addition of a lower bound on the precision, this paper considers the *maximum* Bayes factor in addition to the *minimum* Bayes factor. As an editorial decision, these results are relegated to the Appendix.




(A) Ratio of minimum Bayes factor for $\mathcal{C}_{\frac{M}{\sigma^2}}$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$

(B) Difference of minimum posterior probability for $\mathcal{C}_{\frac{M}{\sigma^2}}$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$, with $\Pi_c = 0.50$

FIGURE 2. $\mathcal{C}_{\frac{M}{\sigma^2}}$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$

## APPENDIX A. NORMAL PRIORS NOT RESTRICTED TO BE CENTERED AT THE NULL HYPOTHESIS

Let

$$\mathcal{N}_{\overline{\tau}} = \{g : g = \mathcal{N}(m, s^2) \text{ and } m \in \mathbb{R} \text{ and } s^2 \geq \overline{\tau}^{-1}\} \tag{4}$$

for $\overline{\tau} \in (0, \infty)$. $\mathcal{N}_{\overline{\tau}}$ is the set of normal priors with precision no greater than $\overline{\tau}$ (i.e., a variance no less than $\overline{\tau}^{-1}$). Unlike $\mathcal{C}_{\overline{\tau}}$, $\mathcal{N}_{\overline{\tau}}$ does not restrict the priors to be centered at the null hypothesis. Theorem 2 establishes the relationship between a $p$-value and the minBF for the set of priors $\mathcal{N}_{\frac{M}{\sigma^2}}$. As above, $M$ can be interpreted as the largest allowed size of the "virtual sample" associated with the prior.

**Theorem 2** (Results for $\mathcal{N}_{\frac{M}{\sigma^2}}$)**.** *The minimum Bayes factor over the set of priors $\mathcal{N}_{\frac{M}{\sigma^2}}$ is*

$$\min_{g \in \mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \sqrt{1 + \frac{N}{M}} \sqrt{2\pi}\phi\left(\Phi^{-1}\left(1 - \frac{p_N}{2}\right)\right). \tag{5}$$

*The above representation implies also that:*

*(1) For any given $p_N$, $\min_{g \in \mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)})$ is a decreasing function of $\frac{M}{N}$.*
*(2) For any given $\frac{M}{N}$, $\min_{g \in \mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)})$ is an increasing functions of $p_N$.*

| | $\frac{M}{N}$ | $\kappa$ | *p*-value | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.001 | 0.0025 | 0.005 | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.4455 | 1.036 | 1.945 | 3.625 | 14.65 | 25.85 | 51.6 | 79.66 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.1409 | 0.3276 | 0.6155 | 1.147 | 4.635 | 8.179 | 16.33 | 25.2 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.04477 | 0.1041 | 0.1955 | 0.3643 | 1.472 | 2.598 | 5.186 | 8.005 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.02041 | 0.04745 | 0.08915 | 0.1661 | 0.6713 | 1.185 | 2.365 | 3.65 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.01477 | 0.03434 | 0.06452 | 0.1202 | 0.4859 | 0.8574 | 1.711 | 2.642 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.0063 | 0.01464 | 0.02751 | 0.05126 | 0.2072 | 0.3656 | 0.7297 | 1.126 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.004477 | 0.01041 | 0.01955 | 0.03643 | 0.1472 | 0.2598 | 0.5186 | 0.8005 |

(A) Minimum Bayes factors for $\mathcal{N}_{\frac{M}{\sigma^2}}$

| | $\frac{M}{N}$ | $\kappa$ | *p*-value | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.001 | 0.0025 | 0.005 | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.308 | 0.509 | 0.66 | 0.784 | 0.936 | 0.963 | 0.981 | 0.988 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.124 | 0.247 | 0.381 | 0.534 | 0.823 | 0.891 | 0.942 | 0.962 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.0429 | 0.0943 | 0.164 | 0.267 | 0.596 | 0.722 | 0.838 | 0.889 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.02 | 0.0453 | 0.0819 | 0.142 | 0.402 | 0.542 | 0.703 | 0.785 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.0146 | 0.0332 | 0.0606 | 0.107 | 0.327 | 0.462 | 0.631 | 0.725 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.00626 | 0.0144 | 0.0268 | 0.0488 | 0.172 | 0.268 | 0.422 | 0.53 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.00446 | 0.0103 | 0.0192 | 0.0351 | 0.128 | 0.206 | 0.341 | 0.445 |

(B) Minimum posterior probability for $\mathcal{N}_{\frac{M}{\sigma^2}}$, when $\Pi_c = 0.50$

TABLE 3. Results for $\mathcal{N}_{\frac{M}{\sigma^2}}$

Figure 2a shows the *ratio* of the minBFs for $\mathcal{C}_{\frac{M}{\sigma^2}}$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$. Figure 2b shows the *difference* of the minBPP for $\mathcal{C}_{\frac{M}{\sigma^2}}$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$, for $\Pi_c = 0.50$. These sets of priors are based on the same virtual sample size, but differ in being restricted to being centered at $c$ or not. As long as $\frac{M}{N}$ is less than about 0.1, the minBF for $\mathcal{N}_{\frac{M}{\sigma^2}}$ is very close to the minBF for $\mathcal{C}_{\frac{M}{\sigma^2}}$. The same is true for the minBPP. However, when $\frac{M}{N}$ exceeds about 0.1, the additional restriction to normal priors centered at $c$ has an important impact. Tables 3a and 3b tabulate the minBFs and minBPPs. Table 4a reports the cutoff $p$-values for statistical significance, and Table 4b reports the strength of prior relative to sample size that results in a given minimum Bayes factor.

| | $\frac{M}{N}$ | $\kappa$ | $b$ | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 0.33333 | 0.1 | 0.05 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.0001 | 0.0039894 | 0.002406 | 0.0007314 | 0.0002017 | 9.661e-05 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.001 | 0.012616 | 0.008577 | 0.002548 | 0.0006908 | 0.0003285 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.01 | 0.039894 | 0.03169 | 0.009053 | 0.002393 | 0.001127 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.05 | 0.089206 | 0.08101 | 0.02205 | 0.005678 | 0.002647 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.1 | 0.12616 | 0.1215 | 0.03206 | 0.008137 | 0.003774 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 1 | 0.39894 | 0.4051 | 0.08911 | 0.02135 | 0.009725 |
| $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 100 | 3.9894 | 0.9205 | 0.1374 | 0.03169 | 0.01429 |

(A) Cutoff $p$-values

| | | | $b$ | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 0.33333 | 0.1 | 0.05 |
| $p$-value | 0.001 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 1.985e-05 | 0.0001786 | 0.001988 | 0.008001 |
| | 0.0025 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.0001072 | 0.0009659 | 0.01084 | 0.04481 |
| | 0.005 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.0003786 | 0.003418 | 0.03933 | 0.1784 |
| | 0.01 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.001315 | 0.01196 | 0.1512 | 1.107 |
| | 0.05 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.02193 | 0.2394 | n/p | n/p |
| | 0.1 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.07162 | 1.509 | n/p | n/p |
| | 0.25 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 0.3629 | n/p | n/p | n/p |
| | 0.5 | $\mathcal{N}_{\frac{M}{\sigma^2}}$ | 1.736 | n/p | n/p | n/p |

(B) Strength of priors relative to sample size that results in a given minimum Bayes factor, for given $p$-value. n/p, not possible (there is no strength of prior relative to sample size that results in that minimum Bayes factor for that $p$-value).

Table 4. Results for $\mathcal{N}_{\frac{M}{\sigma^2}}$

## Appendix B. Priors with bounded density

Let

$$\mathcal{D}_K = \{g : ||g||_\infty \leq K\}, \tag{6}$$

for $K \in (0, \infty)$. $\mathcal{D}_K$ is the set of priors that admit a density that is bounded above by the fixed constant $K$. Relatively larger $K$ correspond to a set of priors $\mathcal{D}_K$ that includes relatively stronger priors and relatively smaller $K$ correspond to a set of priors $\mathcal{D}_K$ that includes only relatively weaker priors. Specifically, for many priors that are themselves parameterized, $K$ can be interpreted as a bound on the parameters of the prior. Then these parameters can be interpreted in terms of the strength of the prior.

Thus, $K$ can be interpreted as a bound on the strength of the priors in $\mathcal{D}_K$. For example, generalized normal priors are contained in $\mathcal{D}_K$, with density $g(\theta) = \frac{s}{2c\Gamma(\frac{1}{s})} \exp(-(\frac{|\theta-m|}{c})^s) = \frac{1}{2c\Gamma(\frac{1}{s}+1)} \exp(-(\frac{|\theta-m|}{c})^s)$, with location $m$ and shape $s$ and scale $c$. As special cases, for different values of the parameters (including limiting cases), the class of generalized normal priors include Laplace priors ($s = 1$), normal priors ($s = 2$), and uniform priors ($s \to \infty$). The class of generalized normal priors also includes "intermediate" cases in between those named distributions. As such, the class of generalized normal priors can be motivated as a class of priors (e.g., Diananda (1949, Section 4), Box and Tiao (1973/1992, Section 3.2.1), and Goodman and Kotz (1973), among others). Laplace priors have a connection to lasso estimation of linear models (e.g, Tibshirani (1996) and Park and Casella (2008)). The maximal value of the density of a generalized normal distribution is $\frac{s}{2c\Gamma(\frac{1}{s})}$, regardless of the location of the prior. So, $\mathcal{D}_K$ contains all of the generalized normal priors with parameters satisfying $\frac{1}{2c\Gamma(\frac{1}{s}+1)} \leq K$. In particular, setting $K = \frac{\sqrt{M}}{\sqrt{2\pi\sigma^2}}$ results in the set of priors with maximal density that is no greater than that of a normal prior based on $M$ observations in the associated virtual sample.

**Lemma 1** (Characterization of $\mathcal{D}_K$)**.** *$\mathcal{D}_K$ contains all normal priors with variance at least $\frac{1}{2\pi K^2}$ and no other normal priors; contains all uniform priors with support with Lebesgue measure at least $K^{-1}$ and no other uniform priors; contains all Laplace priors with variance at least $\frac{1}{2K^2}$ and no other Laplace priors. All priors in $\mathcal{D}_K$ have support with Lebesgue measure at least $K^{-1}$. It holds that $\mathcal{N}_{\overline{\tau}} \subsetneq \mathcal{D}_{\frac{\sqrt{\overline{\tau}}}{\sqrt{2\pi}}}$ for any $\overline{\tau} \in (0, \infty)$, and $\mathcal{N}_{\overline{\tau}} \not\subseteq \mathcal{D}_K$ if $K < \frac{\sqrt{\overline{\tau}}}{\sqrt{2\pi}}$.*

Theorem 3 establishes the relationship between a $p$-value and the minBF for $\mathcal{D}_K$. Let

$$\kappa := \kappa(K, N, \sigma^2) \equiv \frac{K}{\sqrt{N}}\sigma. \tag{7}$$

**Theorem 3** (Results for $\mathcal{D}_K$)**.** *The minimum Bayes factor over the set of priors $\mathcal{D}_K$ is*

$$\min_{g \in \mathcal{D}_K} B_c(g, X^{(N)}) = \frac{\phi\left(\Phi^{-1}\left(1 - \frac{p_N}{2}\right)\right)}{\kappa\left(2\Phi\left(\frac{1}{2\kappa}\right) - 1\right)}. \tag{8}$$

*The above representation implies also that:*

*(1) For any given $p_N$, $\min_{g \in \mathcal{D}_K} B_c(g, X^{(N)})$ is a decreasing function of $\kappa(K, N, \sigma^2)$.*

*(2) For any given $\kappa$, $\min_{g\in\mathcal{D}_K} B_c(g, X^{(N)})$ is an increasing functions of $p_N$.*

By Lemma 1, for any given $\kappa$ associated with $\mathcal{D}_K$, the largest $M$ such that $\mathcal{N}_{\frac{M}{\sigma^2}} \subseteq \mathcal{D}_K$ has $\frac{M}{N} = 2\pi\kappa^2$, so this formula is a mapping from $\kappa$ to the "analogous" $\frac{M}{N}$.[4] This mapping is how the figures have both $\kappa$ and $\frac{M}{N}$ on the axes. Based on numerical evaluation of Corollary 2, $\frac{\min_{g\in\mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g,X^{(N)})}{\min_{g\in\mathcal{D}_{\frac{\sqrt{M}}{\sqrt{2\pi\sigma^2}}}} B_c(g,X^{(N)})}$ is often very close to 1, and never exceeds around 1.15 over the relevant range of $\frac{M}{N}$. Thus, the minBF for the set of priors $\mathcal{D}_K$ is very close to the minBF for the set of priors $\mathcal{N}_{\frac{M}{\sigma^2}}$, when $\frac{M}{N} = 2\pi\kappa^2$ holds, or equivalently, when $M = 2\pi K^2\sigma^2$. Consequently, normality of the priors does not substantially influence the resulting minBF.

**Corollary 2** (Results for $\mathcal{D}_K$ compared to $\mathcal{N}_{\frac{M}{\sigma^2}}$)**.** *It holds that*

$$\frac{\min_{g\in\mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g,X^{(N)})}{\min_{g\in\mathcal{D}_{\frac{\sqrt{M}}{\sqrt{2\pi\sigma^2}}}} B_c(g,X^{(N)})} = \sqrt{1+\frac{M}{N}}\left(2\Phi\left(\sqrt{\frac{\pi}{2}}\sqrt{\frac{N}{M}}\right)-1\right). \tag{9}$$

The result of Corollary 3 in Equation 10 concerns the case of no restriction on the strength of the prior relative to data sample size, and matches the related results from Edwards, Lindman, and Savage (1963) and Berger and Sellke (1987). Therefore, those results can be understood to concern the polar special case of allowing for extremely strong priors relative to sample size. Similar to above, the result of Corollary 3 in Equation 11 matches the results from the "Jeffreys (1939)-Lindley (1957) paradox."

**Corollary 3** (Results for $\mathcal{D}_K$)**.** *When $\mathcal{D}_K$ includes very strong priors relative to sample size,*

$$\lim_{\frac{K}{\sqrt{N}}\to\infty} \min_{g\in\mathcal{D}_K} B_c(g,X^{(N)}) = \exp\left(-\frac{1}{2}[\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2\right) \tag{10}$$

*When $\mathcal{D}_K$ only includes very weak priors relative to sample size,*

$$\lim_{\frac{K}{\sqrt{N}}\to 0} \min_{g\in\mathcal{D}_K} B_c(g,X^{(N)}) = \infty \tag{11}$$

[4] $\mathcal{D}_K$ contains exactly those normal priors with precision no greater than $2\pi K^2 = 2\pi\frac{N}{\sigma^2}\kappa^2$. Therefore, the largest $M$ such that $\mathcal{N}_{\frac{M}{\sigma^2}} \subseteq \mathcal{D}_K$ is $(2\pi\frac{N}{\sigma^2}\kappa^2)\sigma^2 = 2\pi N\kappa^2$.

## Appendix C. Normal priors with an upper bound and lower bound on the precision

Let

$$\mathcal{C}_{\overline{\tau},\underline{\tau}} = \{g : g = \mathcal{N}(c, s^2) \text{ and } \underline{\tau}^{-1} \geq s^2 \geq \overline{\tau}^{-1}\} \tag{12}$$

for $0 < \underline{\tau} < \overline{\tau}$. Relative to $\mathcal{C}_{\overline{\tau}}$, $\mathcal{C}_{\overline{\tau},\underline{\tau}}$ adds a *lower bound* on the precision of the prior (i.e., an *upper bound* on the variance of the prior). Equivalently, this can be viewed as corresponding to a *lower bound* on the size of the virtual sample associated with the prior. Based on $\mathcal{C}_{\overline{\tau},\underline{\tau}}$, it is possible to derive a *maximum* Bayes factor and associated *maximum* posterior probability of the hypothesis. Hereafter, these quantities are known as maxBF and maxBPP, respectively. Having non-trivial bounds on the strength of the prior is key for the maxBF to be non-trivial.

**Theorem 4** (Results for $\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}$)**.** *The minimum Bayes factor over the set of priors* $\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}$ *is*

$$\min_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)}) = \begin{cases} \min_{g\in\mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) & \text{if } 1 \geq \frac{m}{N}([\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 - 1) \\ \frac{\sqrt{1+\frac{N}{m}}\phi\left(\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{m}}}\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)} & \text{if } 1 < \frac{m}{N}([\Phi^{-1}\left(1-\frac{p_N}{2}\right)]^2 - 1) \end{cases}. \tag{13}$$

*The maximum Bayes factor over the set of priors* $\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}$ *is*

$$\max_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)}) = \max\left\{\frac{\sqrt{1+\frac{N}{M}}\phi\left(\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}, \frac{\sqrt{1+\frac{N}{m}}\phi\left(\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{m}}}\Phi^{-1}\left(1-\frac{p_N}{2}\right)\right)}\right\}. \tag{14}$$

*The above representations imply also that:*

*(1) For any given $p_N$ and given $\frac{m}{N}$, $\min_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly decreasing function of $\frac{M}{N}$.*
*(2) For any given $p_N$ and given $\frac{M}{N}$, $\min_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly increasing function of $\frac{m}{N}$.*
*(3) For any given $p_N$ and given $\frac{m}{N}$, $\max_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly increasing function of $\frac{M}{N}$.*
*(4) For any given $p_N$ and given $\frac{M}{N}$, $\max_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly decreasing function of $\frac{m}{N}$.*
*(5) For any given $\frac{M}{N}$ and $\frac{m}{N}$, $\min_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ and $\max_{g\in\mathcal{C}_{\frac{M}{\sigma^2},\frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ are weakly increasing functions of $p_N$.*

## Appendix D. Proofs

The likelihood is $f(X^{(N)}|\theta=\theta^*) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^N(X_i-\hat{\theta}_N)^2)\exp(-\frac{1}{2\frac{\sigma^2}{N}}(\hat{\theta}_N-\theta^*)^2) \propto \exp(-\frac{1}{2\frac{\sigma^2}{N}}(\hat{\theta}_N-\theta^*)^2)$, where $\propto$ drops multiplicative terms that do not depend on $\theta^*$. Let $f(\hat{\theta}_N|\theta=\theta^*) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp(-\frac{(\hat{\theta}_N-\theta^*)^2}{2\frac{\sigma^2}{N}})$. Because $\Pi(\theta=c|X^{(N)};\Pi_c,g) = \frac{f(X^{(N)}|\theta=c)\Pi_c}{f(X^{(N)}|\theta=c)\Pi_c+\left(\int_{\Theta\setminus\{c\}}f(X^{(N)}|\theta)g(\theta)d\theta\right)(1-\Pi_c)}$, for any $\mathcal{G}$ the minBPP is $\frac{f(X^{(N)}|\theta=c)\Pi_c}{f(X^{(N)}|\theta=c)\Pi_c+\left(\sup_{g\in\mathcal{G}}\int_{\Theta\setminus\{c\}}f(X^{(N)}|\theta)g(\theta)d\theta\right)(1-\Pi_c)} = \left(1+\frac{\left(\sup_{g\in\mathcal{G}}\int_{\Theta\setminus\{c\}}f(\hat{\theta}_N|\theta)g(\theta)d\theta\right)}{f(\hat{\theta}_N|\theta=c)}\frac{1-\Pi_c}{\Pi_c}\right)^{-1}$ and the minBF is $\frac{f(\hat{\theta}_N|\theta=c)}{\left(\sup_{g\in\mathcal{G}}\int_{\Theta\setminus\{c\}}f(\hat{\theta}_N|\theta)g(\theta)d\theta\right)}$.

The proofs of Lemma 1 and Corollary 2 are trivial and therefore omitted.

*Proof of Theorem 1 and Theorem 2 and Theorem 4 and Corollary 1.* The expression for minBF with normal priors, when $g$ is a $\mathcal{N}(m,s^2)$ prior, depends on $\int_{\Theta\setminus\{c\}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{(\hat{\theta}_N-\theta)^2}{2\frac{\sigma^2}{N}}\right)\frac{1}{s}\phi\left(\frac{\theta-m}{s}\right)d\theta = \int_{\Theta\setminus\{c\}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{(\hat{\theta}_N-\theta)^2}{2\frac{\sigma^2}{N}}\right)\frac{1}{\sqrt{2\pi s^2}}\exp\left(-\frac{(\theta-m)^2}{2s^2}\right)d\theta = \frac{1}{\sqrt{\frac{\sigma^2}{N}+s^2}}\phi\left(\frac{\hat{\theta}_N-m}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right)$. This recognizes the integral as the density of a sum of independent $\mathcal{N}(0,\frac{\sigma^2}{N})$ and $\mathcal{N}(m,s^2)$ random variables, evaluated at $\hat{\theta}_N$, which is the density of a $\mathcal{N}(m,\frac{\sigma^2}{N}+s^2)$ random variable evaluated at $\hat{\theta}_N$.

The proof considers different sets of normal priors, with different restrictions on $m$ and $s^2$.

*Unrestricted $m$ and lower bound on $s^2$:* With lower bound on the variance of the prior, $s^2 \geq \underline{s}^2 > 0$, the maximum is achieved at $m=\hat{\theta}_N$ and $s^2=\underline{s}^2$. In that case, the minBF is

$$\frac{f(\hat{\theta}_N|\theta=c)}{\sup_{g\in\mathcal{G}}\int_{\Theta\setminus\{c\}}f(\hat{\theta}_N|\theta)g(\theta)d\theta} = \frac{\frac{1}{\sqrt{\frac{\sigma^2}{N}}}\phi\left(\frac{\hat{\theta}_N-c}{\sqrt{\frac{\sigma^2}{N}}}\right)}{\frac{1}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}\phi(0)} = \frac{\phi(t_N)}{\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}\frac{1}{\sqrt{2\pi}}} \tag{15}$$

Equation 5 of Theorem 2 follows because Equation 15 can be written, when $\underline{s}^2=\frac{\sigma^2}{M}$,

$$\min_{g\in\mathcal{N}_{\frac{M}{\sigma^2}}} B_c(g,X^{(N)}) = \sqrt{1+\frac{N}{M}}\sqrt{2\pi}\phi(t_N) \tag{16}$$

The rest of Theorem 2 follows immediately.

*Restricted $m$ and unrestricted $s^2$:* With the prior restricted to have $m = c$, the optimization problem becomes

$$\sup_{g\in\mathcal{G}} \frac{1}{\sqrt{\frac{\sigma^2}{N}+s^2}}\phi\left(\frac{\hat{\theta}_N - c}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right) = \sup_{g\in\mathcal{G}} \frac{1}{\sqrt{\frac{\sigma^2}{N}+s^2}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+s^2}}t_N\right) = \begin{cases} \frac{1}{\sqrt{\frac{\sigma^2}{N}}|t_N|}\phi(1) & \text{if } |t_N| \geq 1 \\ \frac{1}{\sqrt{\frac{\sigma^2}{N}}}\phi(t_N) & \text{if } |t_N| < 1 \end{cases} \tag{17}$$

The optimization result holds as follows. Using the fact that $\phi'(x) = -x\phi(x)$, the derivative of the objective with respect to $s^2$ is $-\frac{1}{2}\left(\frac{\sigma^2}{N}+s^2\right)^{-\frac{3}{2}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}t_N}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right)\left(1-\left(\frac{\sqrt{\frac{\sigma^2}{N}}t_N}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right)^2\right)$.

If $|t_N| \geq 1$, by solving the first order condition, this is maximized at $s^2 = \frac{\sigma^2}{N}t_N^2 - \frac{\sigma^2}{N} = \frac{\sigma^2}{N}(t_N^2-1)$. This is indeed a maximum because the derivative is of the form of a negative multiple of $\left(1-\left(\frac{\sqrt{\frac{\sigma^2}{N}}t_N}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right)^2\right)$. Note that $\left(\frac{\sqrt{\frac{\sigma^2}{N}}t_N}{\sqrt{\frac{\sigma^2}{N}+s^2}}\right)^2$ is a decreasing function of $s^2$. It equals 1 exactly at the claimed maximizer value for $s^2$, is less than 1 for greater values of $s^2$, and is greater than 1 for lesser values of $s^2$. Therefore, the derivative is negative for values of $s^2$ greater than the claimed maximizer, and is positive for values of $s^2$ less than the claimed maximizer.

If $|t_N| < 1$, the derivative is negative for all values of $s^2$ and the maximizing value is $s^2 = 0$.

*Restricted $m$ and lower bound on $s^2$:* Suppose that $\underline{s^2} > 0$ is a lower bound on the variance of the prior. From above, this lower bound is binding when $\underline{s^2} \geq \frac{\sigma^2}{N}(t_N^2 - 1)$ and $|t_N| \geq 1$ and is always binding when $|t_N| < 1$. When the lower bound is binding, because the minBF is decreasing for $s^2 > \frac{\sigma^2}{N}(t_N^2 - 1)$, the minBF is

$$\frac{f(\hat{\theta}_N|\theta = c)}{\sup_{g\in\mathcal{G}} \int_{\Theta\setminus\{c\}} f(\hat{\theta}_N|\theta)g(\theta)d\theta} = \frac{\frac{1}{\sqrt{\frac{\sigma^2}{N}}}\phi(t_N)}{\frac{1}{\sqrt{\frac{\sigma^2}{N}+\underline{s^2}}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s^2}}}t_N\right)} \tag{18}$$

Therefore, overall, the minBF is

$$\begin{cases} \dfrac{\phi(t_N)}{\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}t_N\right)} & \text{if } \underline{s}^2 \geq \frac{\sigma^2}{N}(t_N^2-1) \\ \frac{\phi(t_N)|t_N|}{\phi(1)} & \text{if } \underline{s}^2 < \frac{\sigma^2}{N}(t_N^2-1) \end{cases} \tag{19}$$

Equation 1 of Theorem 1 follows because Equation 19 can be written, when $\underline{s}^2 = \frac{\sigma^2}{M}$

$$\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \begin{cases} \dfrac{\sqrt{1+\frac{N}{M}}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}t_N\right)} & \text{if } 1 \geq \frac{M}{N}(t_N^2-1) \\ \frac{\phi(t_N)|t_N|}{\phi(1)} & \text{if } 1 < \frac{M}{N}(t_N^2-1) \end{cases} \tag{20}$$

As $\frac{M}{N} \to \infty$, $1 \geq \frac{M}{N}(t_N^2 - 1)$ exactly when $|t_N| \leq 1$. And $\frac{\sqrt{1+\frac{N}{M}}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}t_N\right)} \to 1$. Equation 2 of Corollary 1 follows. Equation 3 of Corollary 1 immediately follows.

The derivative of the minBF in Equation 20 with respect to $\frac{N}{M}$ is $\frac{\frac{1}{2}\phi(t_N)\left(1-\left(\frac{1}{\sqrt{1+\frac{N}{M}}}t_N\right)^2\right)}{(1+\frac{N}{M})^{\frac{1}{2}}\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}t_N\right)} \geq 0$ for $1 \geq \frac{M}{N}(t_N^2-1)$, precisely because $1 \geq \frac{t_N^2}{1+\frac{N}{M}}$. The derivative is 0 for $1 < \frac{M}{N}(t_N^2 - 1)$. Therefore, the minBF is a decreasing function of $\frac{M}{N}$.

Also correspondingly, Equation 20 can be written as

$$\min_{g \in \mathcal{C}_{\frac{M}{\sigma^2}}} B_c(g, X^{(N)}) = \begin{cases} \dfrac{\sqrt{1+\frac{N}{M}}}{\exp\left(-\frac{1}{2}t_N^2\left(\frac{1}{1+\frac{N}{M}}-1\right)\right)} & \text{if } 1 \geq \frac{M}{N}(t_N^2-1) \\ \frac{\exp(-\frac{1}{2}(t_N^2-\log(t_N^2)))}{\sqrt{2\pi}\phi(1)} & \text{if } 1 < \frac{M}{N}(t_N^2-1) \end{cases} \tag{21}$$

Therefore, since $\left(\frac{1}{1+\frac{N}{M}} - 1\right) < 0$, the minBF is an increasing function of $p_N$ when $1 \geq \frac{M}{N}(t_N^2 - 1)$. Further, the derivative of $t_N^2 - \log(t_N^2)$ with respect to $t_N^2$ is $1 - t_N^{-2}$, which is positive because $1 < \frac{M}{N}(t_N^2 - 1)$ implies that $t_N^2 > 1 + \frac{N}{M} > 1$. Therefore, the minBF is an increasing function of $p_N$ also when $1 < \frac{M}{N}(t_N^2 - 1)$.

*Restricted $m$ and upper and lower bound on $s^2$:* First consider minBF. If there is an upper bound on the variance $\overline{s}^2 \geq \frac{\sigma^2}{N}(t_N^2 - 1)$, then the above arguments deriving minBF are unaffected. If $\overline{s}^2 < \frac{\sigma^2}{N}(t_N^2 - 1)$, then it is binding in the optimization problem to find minBF, in which case the minBF is $\frac{\sqrt{1+\frac{N}{m}}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{m}}}t_N\right)}$ when $\overline{s}^2 = \frac{\sigma^2}{m}$. Equation 13 of Theorem 4 follows, and the rest of the properties of the minBF follow by arguments similar to above.

Now consider maxBF. By the same arguments as above, for any restricted range of variances of the prior $[\underline{s^2}, \overline{s^2}]$, the Bayes factor is maximized at either $\underline{s^2}$ or $\overline{s^2}$. Therefore, the optimization problem to achieve the maxBF is

$$\inf_{g \in \mathcal{G}} \frac{1}{\sqrt{\frac{\sigma^2}{N} + s^2}} \phi\left(\frac{\hat{\theta}_N - c}{\sqrt{\frac{\sigma^2}{N} + s^2}}\right) = \min\left\{\frac{1}{\sqrt{\frac{\sigma^2}{N} + \underline{s}^2}} \phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N} + \underline{s}^2}} t_N\right), \frac{1}{\sqrt{\frac{\sigma^2}{N} + \overline{s}^2}} \phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N} + \overline{s}^2}} t_N\right)\right\} \quad (22)$$

Therefore, overall, the maxBF is

$$\frac{\phi(t_N)}{\min\left\{\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\underline{s}^2}}t_N\right), \frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\overline{s}^2}}\phi\left(\frac{\sqrt{\frac{\sigma^2}{N}}}{\sqrt{\frac{\sigma^2}{N}+\overline{s}^2}}t_N\right)\right\}} \quad (23)$$

Equation 14 of Theorem 4 follows because Equation 23 can be written, when $\underline{s^2} = \frac{\sigma^2}{M}$ and $\overline{s^2} = \frac{\sigma^2}{m}$,

$$\max_{g \in \mathcal{C}_{\frac{M}{\sigma^2}, \frac{m}{\sigma^2}}} B_c(g, X^{(N)}) = \max\left\{\frac{\sqrt{1+\frac{N}{M}}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{M}}}t_N\right)}, \frac{\sqrt{1+\frac{N}{m}}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\frac{N}{m}}}t_N\right)}\right\} \quad (24)$$

Maximizing $\frac{\sqrt{1+\rho}\phi(t_N)}{\phi\left(\frac{1}{\sqrt{1+\rho}}t_N\right)}$ over an interval $[\underline{\rho}, \overline{\rho}]$ is such that increasing $\underline{\rho}$ results in a weakly smaller maximized value, and increasing $\overline{\rho}$ results in a weakly greater maximized value. Therefore, $\max_{g \in \mathcal{C}_{\frac{M}{\sigma^2}, \frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is a weakly increasing function of $\frac{M}{N}$ and weakly decreasing function of $\frac{m}{N}$. Because the maximum of increasing functions is increasing, and based on similar arguments to before from Equation 21, $\max_{g \in \mathcal{C}_{\frac{M}{\sigma^2}, \frac{m}{\sigma^2}}} B_c(g, X^{(N)})$ is an increasing function of $p_N$. □

*Proof of Theorem 3 and Corollary 3.* It holds that $\sup_{g\in\mathcal{D}_K}\int_{\Theta\setminus\{c\}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{(\hat\theta_N-\theta)^2}{2\frac{\sigma^2}{N}}\right)g(\theta)d\theta$ is achieved at $g^*(\theta)=\begin{cases}K & \theta\in[\hat\theta_N-\frac{1}{2K},\hat\theta_N+\frac{1}{2K}]\\ 0 & \text{otherwise}\end{cases}$, resulting in $\sup_{g\in\mathcal{D}_K}\int_{\Theta\setminus\{c\}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{(\hat\theta_N-\theta)^2}{2\frac{\sigma^2}{N}}\right)g(\theta)d\theta=$ $\int_{\hat\theta_N-\frac{1}{2K}}^{\hat\theta_N+\frac{1}{2K}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{(\hat\theta_N-\theta)^2}{2\frac{\sigma^2}{N}}\right)Kd\theta=\int_{-\frac{1}{2K}}^{\frac{1}{2K}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left(-\frac{\theta^2}{2\frac{\sigma^2}{N}}\right)Kd\theta=K\left(2\Phi\left(\frac{1}{2K\sqrt{\frac{\sigma^2}{N}}}\right)-1\right)$.

So, the minBF is $\frac{f(\hat\theta_N|\theta=c)}{\sup_{g\in\mathcal{D}_K}\int_{\Theta\setminus\{c\}}f(\hat\theta_N|\theta)g(\theta)d\theta}=\frac{\frac{1}{\sqrt{\frac{\sigma^2}{N}}}\phi(t_N)}{K\left(2\Phi\left(\frac{1}{2K\sqrt{\frac{\sigma^2}{N}}}\right)-1\right)}=\frac{\phi(t_N)}{\frac{K}{\sqrt{N}}\sigma\left(2\Phi\left(\frac{1}{2\frac{K}{\sqrt{N}}\sigma}\right)-1\right)}$. This establishes Equation 8 of Theorem 3.

The derivative of the denominator of the minBF with respect to $\kappa$ is $(2\Phi(\frac{1}{2\kappa})-1)+\kappa 2\phi(\frac{1}{2\kappa})\frac{1}{2}(-\kappa^{-2})=(2\Phi(\frac{1}{2\kappa})-1)-\kappa^{-1}\phi(\frac{1}{2\kappa})=\mathrm{erf}(\frac{1}{2\sqrt{2}\kappa})-\kappa^{-1}\phi(\frac{1}{2\kappa})\geq 0$. The inequality follows: by the representation that $\mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z\exp(-t^2)dt\geq\frac{2}{\sqrt{\pi}}z\exp(-z^2)$, $\mathrm{erf}(\frac{1}{2\sqrt{2}\kappa})\geq\frac{1}{\sqrt{2\pi}}\kappa^{-1}\exp(-\frac{(\frac{1}{2\kappa})^2}{2})=\kappa^{-1}\phi(\frac{1}{2\kappa})$. Therefore, the minBF is a decreasing function of $\kappa$.

As $\kappa=\frac{K}{\sqrt{N}}\sigma\to\infty$, by L'Hopital's rule, $\kappa\left(2\Phi\left(\frac{1}{2\kappa}\right)-1\right)=\frac{\left(2\Phi(\frac{1}{2\kappa})-1\right)}{\frac{1}{\kappa}}\to\left.\frac{2\phi(\frac{1}{2\kappa})\frac{-1}{\kappa^2}\frac{1}{2}}{\frac{-1}{\kappa^2}}\right|_{\kappa\to\infty}=\phi(0)$. So, as $\kappa\to\infty$, the minBF approaches $\frac{\phi(t_N)}{\phi(0)}=\frac{\phi(t_N)}{\frac{1}{\sqrt{2\pi}}}$. Similarly, as $\kappa\to 0$, $\kappa\left(2\Phi\left(\frac{1}{2\kappa}\right)-1\right)\to 0$. So, as $\kappa\to 0$, the minBF approaches $\infty$. This establishes Equations 10 and 11 of Corollary 3. $\square$

## References


Benjamin, D. J., J. O. Berger, et al. (2018): "Redefine statistical significance," *Nature Human Behaviour*, 2(1), 6–10.

Berger, J. O., and M. Delampady (1987): "Testing precise hypotheses," *Statistical Science*, 2(3), 317–335.

Berger, J. O., and T. Sellke (1987): "Testing a point null hypothesis: The irreconcilability of p values and evidence," *Journal of the American Statistical Association*, 82(397), 112–122.

Box, G. E., and G. C. Tiao (1973/1992): *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York.

Casella, G., and R. L. Berger (1987): "Reconciling Bayesian and frequentist evidence in the one-sided testing problem," *Journal of the American Statistical Association*, 82(397), 106–111.

Diananda, P. (1949): "Note on some properties of maximum likelihood estimates," *Mathematical Proceedings of the Cambridge Philosophical Society*, 45(4), 536–544.

Edwards, W., H. Lindman, and L. J. Savage (1963): "Bayesian statistical inference for psychological research," *Psychological Review*, 70(3), 193–242.

Goodman, I. R., and S. Kotz (1973): "Multivariate $\theta$-generalized normal distributions," *Journal of Multivariate Analysis*, 3(2), 204–219.

Held, L., and M. Ott (2016): "How the maximal evidence of p-values against point null hypotheses depends on sample size," *The American Statistician*, 70(4), 335–341.

Jeffreys, H. (1939): *Theory of Probability*. The Clarendon Press, Oxford.

Kass, R. E., and A. E. Raftery (1995): "Bayes factors," *Journal of the American Statistical Association*, 90(430), 773–795.

Kline, B. (2011): "The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean," *Journal of Statistical Planning and Inference*, 141(9), 3131–3141.

Kline, B., and E. Tamer (2016): "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 7(2), 329–366.

Lindley, D. V. (1957): "A statistical paradox," *Biometrika*, 44(1/2), 187–192.

Moon, H. R., and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80(2), 755–782.

Park, T., and G. Casella (2008): "The Bayesian lasso," *Journal of the American Statistical Association*, 103(482), 681–686.

Sellke, T., M. Bayarri, and J. O. Berger (2001): "Calibration of p values for testing precise null hypotheses," *The American Statistician*, 55(1), 62–71.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Wasserstein, R. L., and N. A. Lazar (2016): "The ASA's statement on p-values: context, process, and purpose," *The American Statistician*, 70(2), 129–133.