Bayesian conclusions from classical p-values

Brendan Kline

Abstract. This paper asks what conclusions a Bayesian can draw from classical p-values. Results are asymptotic approximations corresponding to p-values for the null hypothesis that $\theta = c$. One result relates p-values to the Bayesian posterior probability that the parameter θ is greater than (or less than) d, for any specific d. Another result relates p-values to the Bayesian posterior probability that the null hypothesis is "approximately" true, for a specific meaning of "approximately."

Keywords: frequentist, hypothesis, posterior, testing

JEL classification: C11, C12, C18

1. Introduction

Standard practice in empirical research, both in economics and related fields, is to report p-values associated with null hypotheses of the form $\theta = c$, where θ is a finite-dimensional parameter and c is a known constant (e.g., when $c = 0$ the null hypothesis is $\theta = 0$). In this paper, a p-value is standard: the repeated sampling probability that the (Wald) test statistic would exceed the observed value of the test statistic, if the null hypothesis were true.

The main question is: what can a Bayesian conclude based on observing such p-values, and other pieces of information that are standard to report in published research, particularly the classical estimator $\hat{\theta}_N$ of θ? Formally, the paper contains asymptotic results, which may be viewed as approximations in large samples. The analysis is done presuming the Bernstein-von Mises phenomenon. Under the conditions of the Bernstein-von Mises theorem(s), the

University of Texas at Austin
E-mail address: brendan.kline@austin.utexas.edu.
Date: April 2019. First distributed: October 2018. Thanks to a seminar audience at Cornell University for helpful comments and discussion. Any errors are mine.

Bernstein-von Mises phenomenon is that the classical ("frequentist") sampling distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is approximately the same as the Bayesian posterior distribution of $\sqrt{N}(\theta - \hat{\theta}_N)$.

In setups that do not exhibit the Bernstein-von Mises phenomenon, an existing literature (see below in Section 1.1) looks at whether p-values for the null hypothesis $\theta = c$ are equal to the Bayesian posterior probability that $\theta = c$. This existing literature shows that the p-value can be substantially smaller than "any" Bayesian posterior probability that $\theta = c$. This existing literature appears to generally be understood to suggest that p-values have limited or perhaps no value for a Bayesian. However, it is important to note that this literature has explored the (lack of) connection between p-values and Bayesian posterior probabilities in setups that do not exhibit the Bernstein-von Mises phenomenon. Specifically, this literature has been based around a prior that places positive prior probability on the hypothesis $\theta = c$ being literally true. Such a prior is (generically) incompatible with the Bernstein-von Mises phenomenon. Moreover, it is rare in economics that there is such a strong prior belief in the null hypothesis. This setup was necessary for the purposes of that literature, but perhaps is not the appropriate setup for how p-values are used in economics.[1]

Consequently, there is an important question: what conclusions relating to hypothesis testing can a Bayesian draw from a classical p-value, *in settings exhibiting the Bernstein-von Mises phenomenon*? The Bernstein-von Mises phenomenon holds quite generally in parametric and semiparametric models, as discussed in Remark 1. Results showing that p-values have limited use for a Bayesian in settings that *do not* exhibit the Bernstein-von Mises phenomenon are therefore silent on what happens in such models that *do* exhibit the Bernstein-von Mises phenomenon. And as it turns out, p-values are very useful for a Bayesian in settings that *do* exhibit the Bernstein-von Mises theorem. Such a result is missed in the existing literature that does not work in the setting of the Bernstein-von Mises phenomenon. Given the ubiquity

[1]In setups that do exhibit the Bernstein-von Mises phenomenon, the Bayesian posterior distribution is continuous and therefore assigns 0% posterior probability to the hypothesis that $\theta = c$, whereas the p-value is strictly positive. Therefore, absence of the Bernstein-von Mises phenomenon is a necessary condition for the p-value for $\theta = c$ to *possibly* equal the Bayesian posterior probability that $\theta = c$.

of p-values in published research, it is important to know that a Bayesian actually can draw useful conclusions from p-values.

With the Bernstein-von Mises phenomenon, the posterior probability that $\theta = c$ is necessarily 0%. So this paper does not ask about how the p-value informs the Bayesian about the posterior probability that $\theta = c$. Rather, this paper establishes that p-values are informative about the Bayesian posterior probability of hypotheses *other than* the null hypothesis that is ostensibly tested by the p-value.

One result in this paper relates p-values to the Bayesian posterior probability that the parameter θ is greater than (or less than) any desired d. In the scalar case, the result is that the p-value (along with $\hat{\theta}_N$) is enough information to find an asymptotic approximation to such a Bayesian posterior probability. For example, if the parameter is the effect of a treatment on an outcome, like a regression coefficient, then this result shows that p-values are enough information for a Bayesian to approximate the Bayesian posterior probability that $\theta > d$, the hypothesis that the treatment effect exceeds some chosen d. In general, the results are derived in this paper for multivariate θ. In the case of multivariate θ, the results concern asymptotic bounds on the Bayesian posterior probability that vector θ is element-wise greater than (or less than) any desired vector d. Unavoidably, there is some flexibility in choosing d, and so the results accommodate different choices of d. On the other hand, deciding on a meaningful hypothesis is inherently part of doing empirical research. For example, this could mean deciding on what it means for a treatment effect to be of a meaningfully positive magnitude. (And, to be clear, the results accommodate working with multiple candidate definitions of "meaningful.")

A second result in this paper establishes that p-values are connected to the Bayesian posterior probabilities of *"approximate"* null hypotheses. This result is partially a response to the literature that focused on the question of whether p-values are equal to the Bayesian posterior probability of the null hypothesis, and found negative results. The "approximate" null hypotheses will be (data-dependent) "interval" hypotheses that are "centered" at the point hypothesis $\theta = c$. For example, when $c = 0$, if the parameter is the effect of a treatment on an outcome, like a regression coefficient, then these results can be used to evaluate a p-value in terms of whether it implies that the treatment effect is "approximately" 0 or not even "approximately" 0. Unavoidably, there is some flexibility in choosing an "approximate" null hypothesis, and so the results accommodate different choices of "approximate." On the other hand, in economics and other fields, it seems rare that empirical research cares about whether the null hypothesis is literally ("exactly") true or false. Therefore, deciding on a meaningful hypothesis is inherently part of doing empirical research. For example, this could mean deciding on what it means for a treatment effect to be meaningfully non-zero. (And, to be clear, the results accommodate working with multiple candidate definitions of "approximate.")

This result helps to clarify the understanding of what classical inference statements imply about whether the null hypothesis is (approximately) "true" or "false." As an example, suppose there are two different parameters, representing estimates of different treatment effects. For each, the classical estimate of the treatment effect is calculated and a p-value for the null hypothesis of no effect is calculated. The result shows that information can be used to compute the Bayesian posterior probability that each of the treatments have "approximately zero" treatment effect. Moreover, the result shows that the relationship between that information and the Bayesian posterior probability is complicated and violates certain common intuitions about p-values. For example, suppose the p-value for the first treatment effect is 0.75 and the p-value for the second treatment effect is 0.05. This might suggest that the first treatment effect is "more likely" to be "approximately zero;" however, it can happen that actually the first treatment effect is *"less likely"* to be "approximately zero." Therefore, the result shows both that it is possible to compute these Bayesian posterior probabilities of "approximate" null hypotheses, and that these behave in non-intuitive ways, reinforcing the need for such results.

One remarkably simple special case of this second result concerns the p-value for the null hypothesis that $\theta = c$, with scalar θ, and when the p-value is some number less than 0.25. In that case, there is essentially 50% Bayesian posterior probability of the "approximate" null hypothesis that $\theta \in (c - |\hat{\theta}_N - c|, c + |\hat{\theta}_N - c|)$ and there is essentially 50% Bayesian posterior probability of the logically alternative hypothesis that $\theta \notin (c - |\hat{\theta}_N - c|, c + |\hat{\theta}_N - c|)$.

Overall, if a Bayesian is presented with a p-value, for example in published research, or based on original calculations, then the Bayesian may draw the above useful conclusions from that information. Given the ubiquity of p-values in research, it is important to know that a Bayesian can draw such useful conclusions from p-values.

Although not the main point of this paper, these result can also be used to compute p-values based on computing the corresponding Bayesian posterior probability; this may be computationally attractive in complex models where Bayesian computations may be easier than classical computations (see Remark 3).

1.1. **Results and relation to literature.** The literature has focused on the question of whether the classical p-value is also the "Bayesian posterior probability" of $\theta = c$, and has come to negative conclusions.[2] The literature has considered a few different ways of defining the "Bayesian posterior probability," essentially by considering different (classes of) priors. Lindley's paradox (e.g., Jeffreys (1939); Lindley (1957)) is a type of result in which the p-value for the null hypothesis is substantially less than the Bayesian posterior probability of the null hypothesis. The setup for the Lindley's paradox is that the Bayesian assigns a positive prior probability to the null hypothesis being exactly true (i.e., there is positive prior

[2]Whereas this paper and the other cited literature considers the p-value for the null hypothesis $\theta = c$, Casella and Berger (1987) consider a p-value for the null hypothesis that $\theta \leq 0$ versus the alternative $\theta > 0$, when θ is a scalar/location parameter, and find that for that null hypothesis, and for certain classes of priors (and certain realizations of the data that are "evidence against" the null hypothesis), the p-value can exactly equal the *infimum* Bayesian posterior probability across that class of priors, for that null hypothesis. This result does not extend (at least in certain ways) to multivariate parameters, see Kline (2011).

probability that $\theta = c$), and that the Bayesian considers a single fixed (uniform) prior over the rest of the parameter space (i.e., there is a prior on θ for $\theta \neq c$). Edwards, Lindman, and Savage (1963), Berger and Delampady (1987), and Berger and Sellke (1987) are important representative papers of a literature that analyzes a *class* of priors for θ on $\theta \neq c$ and shows that the p-value is smaller than even the *infimum* Bayesian posterior probability that the null hypothesis is true, where the infimum is over a class of priors.[3] Sellke, Bayarri, and Berger (2001) use a model of p-values directly, and suggest the corresponding smallest Bayesian posterior probability for the null hypothesis associated with a given p-value is a "calibration" for p-values. A complete review of this literature can be found in Held and Ott (2018). Hence, the American Statistical Association statement on p-values (Wasserstein and Lazar (2016)) includes the principle that "P-values do not measure the probability that the studied hypothesis is true [...]."

The results in the above cited literature (and the literature that follows in the same framework) are based on a framework that places positive prior probability on the null hypothesis.[4] Often, the prior is that there is 50% prior probability that $\theta = c$ exactly. In other cases, there is some other positive prior probability that $\theta = c$ exactly. However, it seems uncommon in economics and related fields that the empirical researcher actually has such a strong prior belief in the null hypothesis. Therefore, the results of this paper are based on a different framework: a "continuous prior" that is compatible with the Bernstein-von Mises theorem (see Assumption 1 and surrounding discussion).

[3]These papers often work with a Bayes factor, but of course there is a direct connection between a Bayes factor and a Bayesian posterior probability.

[4]The literature has used a prior that places positive prior probability on $\theta = c$ because in a Bayesian analysis with a continuous prior, there is zero prior probability and zero posterior probability assigned to the hypothesis that $\theta = c$. Therefore, a prior with a point mass at the null hypothesis evidently gives the maximal "opportunity" for the p-value to equal the Bayesian posterior probability that the null hypothesis is true. Hence, since the literature has focused on the question of whether the p-value is the Bayesian posterior probability of the hypothesis $\theta = c$, the literature almost necessarily was based on such a non-continuous prior on θ.

The related literature shows the very strong result that even the infimum Bayesian posterior probability over a class of priors is greater than the p-value. On the other hand, this paper aims to show that p-values can be used to draw Bayesian conclusions. Therefore, this paper does not consider the infimum over a class of priors. It might be hard to know what to do if it is only an *infimum* Bayesian posterior probability that has a known connection to p-values. Such results do not indicate what the posterior would be for any particular Bayesian (i.e., for any particular prior), beyond the fact that Bayesians with priors in a certain class of priors will have posterior probabilities of the null hypothesis greater than the corresponding infimum posterior probability for that class of priors. Those results do not say much about whether any two Bayesians will have approximately the same posterior probability of the null hypothesis, or whether the posterior probabilities are close or far from the corresponding infimum posterior. Along these lines, Sellke, Bayarri, and Berger (2001, p 71) note that the use of an infimum over a class of priors results in a mapping that "can be criticized for being biased against the null hypothesis." In the asymptotic framework in this paper, or in other words in the large sample approximation framework in this paper, the idea is to use p-values to approximate Bayesian posterior probabilities. These approximations are valid for any prior in the class of priors compatible with the assumptions, in large samples.

1.2. **Notation.** The $m \times m$ identity matrix is $I_{m\times m}$. The L^1 norm of $x \in \mathbb{R}^m$ is denoted by $||x||_1$. The Euclidean (L^2) norm of $x \in \mathbb{R}^m$ is denoted by $||x||_2$. The L^∞ norm of $x \in \mathbb{R}^m$ is denoted by $||x||_\infty$. Let $\delta_{\Sigma,2}(x, y) = ||\Sigma^{-\frac{1}{2}}(x - y)||_2$ be a metric for a specified positive definite Σ, and let $\delta_2(x, y) = \delta_{I_m,2}(x, y) = ||x - y||_2$ be a metric. The total variation distance (e.g., as defined in Van der Vaart (1998, Section 2.9)) between random variables (or random vectors) X and Y is $||X - Y||_{TV}$. The induced L^p matrix norm of $A \in \mathbb{R}^{m\times m}$ is denoted by $||A||_p$. The Frobenius matrix norm of $A \in \mathbb{R}^{m\times m}$ is denoted by $||A||_F$. The sign of a scalar A, sign(A), is 1, 0, or -1, depending respectively on whether A is positive, zero, or negative.

The indicator $1[E]$ of some logical statement E is 1 when E is true and 0 when E is false. As a notational convention, it is understood that $z1[\text{false}] = 0$ for all z; so, for example, $\frac{x}{y}1[y \neq 0] = 0$ when $y = 0$.

A multivariate normal random variable with mean μ and covariance Σ is denoted by $\mathcal{N}(\mu, \Sigma)$. A (central) chi-squared random variable with m degrees of freedom is denoted by χ^2_m. A noncentral chi-squared random variable with m degrees of freedom and non-centrality parameter λ is denoted by $\chi^2_{m,\lambda}$. Note that there are multiple parameterizations for the noncentrality parameter of a noncentral chi-squared random variable. This paper uses the parameterization from Johnson, Kotz, and Balakrishnan (1995), where the average of the non-central chi-squared distribution is the degrees of freedom plus the noncentrality parameter. For a random variable (or random vector) with distribution X, the cumulative distribution function is denoted by $F_X(\cdot)$ and the quantile function for a scalar random variable X is denoted by $Q_X(\cdot)$. The cumulative distribution function of a standard normal random variable is $\Phi(\cdot)$. For a random variable with distribution X, the complementary cumulative distribution function (e.g., the survival function) is denoted by $\overline{F}_X(\cdot) \equiv 1 - F_X(\cdot)$. For a random variable (or random vector) with distribution X, and for a Borel set $\mathcal{B}$, $X(\mathcal{B})$ is the probability of the set $\mathcal{B}$ under the distribution X.

A sequence of random variables (or random vectors) X_n indexed by $n = 1, 2, \ldots$ converges in probability to a random variable (or random vector) X if, for all $\epsilon > 0$, $P(||X_n - X||_2 \geq \epsilon) \to 0$, as $n \to \infty$. If the sequence X_n converges in probability to X, then that is denoted by $X_n \to^p X$. The notation $X_n = o_p(1)$ represents that the sequence of random variables X_n converges in probability to 0, and similarly, the notation $o_p(1)$ represents a sequence of random variables that converges in probability to 0. The notation $X_n \leq o_p(1)$ represents that the sequence of scalar random variables X_n satisfies the condition that there is a sequence of random variables X'_n with $X_n \leq X'_n$ a.s. and $X'_n = o_p(1)$. Note that $X_n \leq o_p(1)$ is equivalent to the condition that, for all $\epsilon > 0$, $P(X_n \geq \epsilon) \to 0$, as $n \to \infty$.[5] And of course $X_n \leq Y_n + o_p(1)$ is the same as $X_n - Y_n \leq o_p(1)$. The notation $X_n \geq o_p(1)$ is analogous, and is equivalent to the condition that, for all $\epsilon > 0$, $P(X_n \leq -\epsilon) \to 0$, as $n \to \infty$. A sequence of random variables (or random vectors) X_n indexed by $n = 1, 2, \ldots$, with cumulative distribution functions $F_n(\cdot)$, converges in distribution to a random variable (or random vector) X, with cumulative distribution function $F(\cdot)$, if $F_n(t) \to F(t)$ as $n \to \infty$ for all t such that $F(\cdot)$ is continuous at t. If the sequence X_n converges in distribution to X, then that is denoted by $X_n \to^d X$.

2. Setup and main assumption

The analysis concerns a finite-dimensional parameter θ. The parameter space is $\Theta \subseteq \mathbb{R}^m$, so that θ is m-dimensional. The true value of the parameter is θ_0. The true data generating process is P_0. The data is a sample of N i.i.d. observations from P_0, so that the data is $X^{(N)} = \{X_i\}_{i=1}^N$ where $X_i \sim^{iid} P_0$. All probability limits are understood to be with respect to P_0.

It is presumed that there is a classical ("frequentist") estimator $\hat{\theta}_N$ of θ, based on the data $X^{(N)}$, which is the basis for the construction of the p-value. The p-value will be the p-value from the null hypothesis significance test of the null hypothesis that $\theta = c$ for some specified c. It is also presumed that there is a Bayesian posterior for θ, denoted $\Pi_{\theta|X^{(N)}}(\cdot)$ so that $\Pi_{\theta|X^{(N)}}(A)$ is the posterior probability that θ is in the set A conditional on the observed data $X^{(N)}$. The analysis of this paper will concern what conclusions about the Bayesian posterior for θ may be concluded from the p-value. Note that the Bayesian posterior probabilities considered in this paper are the ones that condition on the entire dataset. So, the question is what a Bayesian can conclude about the "usual" Bayesian posterior probabilities based only on observing p-values, and other pieces of information that are standard to report in published research, specifically the classical estimator $\hat{\theta}_N$.

[5]If $X_n \leq o_p(1)$, then it is immediate that $P(X_n \geq \epsilon) \to 0$. If $P(X_n \geq \epsilon) \to 0$, take $X_n' = X_n 1[X_n \geq 0]$ to satisfy $X_n \leq o_p(1)$.

The analysis does not require a specific form of $\hat{\theta}_N$, nor a specific form of the Bayesian posterior for θ, nor does the analysis require much about the nature of θ. Rather, the analysis makes high-level assumptions on the classical estimator and Bayesian posterior of θ. Nevertheless, it may be worth keeping in mind a special case of the results, which is the case where there is an i.i.d. sample of N observations from a fully parametric model P_θ, so that the data is $X^{(N)} = \{X_i\}_{i=1}^N$ where $X_i \sim^{iid} P_{\theta_0}$. In that case, $\hat{\theta}_N$ can (generally) be taken to be the maximum likelihood estimator of θ based on $X^{(N)}$. Remark 1 shows the analysis applies much more broadly, including to semiparametric models. The results also apply to cases when the actual parameter of interest is a *function* of the parameter of a statistical model, based on Delta theorem arguments in Remark 2.

The main assumption concerns the relationship between the sampling distribution of $\hat{\theta}_N$ and the Bayesian posterior distribution for θ.

Assumption 1. *It holds that:*

(1) *The classical estimator $\hat{\theta}_N$ is asymptotically normal, in the sense that $\sqrt{N}(\hat{\theta}_N - \theta_0) \to^d \mathcal{N}(0, \Sigma_0)$ where Σ_0 is nonsingular. Moreover, there is a consistent estimator $\hat{\Sigma}_N$ of Σ_0.*

(2) *The Bayesian posterior for θ is asymptotically normal, in the sense that $||\Pi_{\sqrt{N}(\theta - \hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0, \Sigma_0)||_{TV} \to^p 0$.*

Part 1 requires that $\hat{\theta}_N$ has an asymptotically normal sampling distribution. Part 2 requires that the Bayesian posterior for θ is asymptotically normal. In particular, part 2 involves the condition that the Bayesian posterior does not depend on the prior asymptotically. Also, Assumption 1 restricts attention to a point identified parameter of interest θ. The relationship between classical and (quasi-)Bayesian inference for a partially identified parameter of interest is different (e.g., Moon and Schorfheide (2012), Kline and Tamer (2016), Chen, Christensen, and Tamer (2018), Giacomini and Kitagawa (2018), and Liao and Simoni (2019)).

Assumption 1 is a high-level assumption that has been proved to hold in many cases. In particular, it is the result of a class of theorems commonly known as the "Bernstein-von Mises theorem(s)," and therefore Assumption 1 is satisfied under the conditions of the "Bernstein-von Mises theorem(s)." The following remark collects some of the results from the literature that imply Assumption 1. The analysis of p-values in this paper therefore holds in any of the following settings.

Remark 1 (Sufficient conditions for Assumption 1)**.** Assumption 1 holds under any of the following settings:

(1) **Data from parametric model**: The classical setting for the Bernstein-von Mises theorem is the case where there is an i.i.d. sample of N observations from a fully parametric distribution P_θ, with true data generating process P_{θ_0}. The data is $\{X_i\}_{i=1}^N$ where $X_i \sim^{iid} P_{\theta_0}$, where P_θ is a parametric model with parameter θ. Often, $\hat{\theta}_N$ is taken to be the maximum likelihood estimate of θ, and Σ_0 the inverse Fisher information matrix. More generally, $\hat{\theta}_N$ can often be taken to be an asymptotically linear and efficient estimator of θ, and Σ_0 the inverse Fisher information matrix. The prior for θ must have a continuous and positive density on a neighborhood of θ_0. Under a few further regularity conditions, the parametric Bernstein-von Mises results from the literature show that Assumption 1 can be satisfied when the underlying statistical model is fully parametric (e.g., Le Cam (1986, Chapter 12), Van der Vaart (1998, Chapter 10), Le Cam and Yang (2000, Chapter 8)).

(2) **Data from semi-parametric model**: The Bernstein-von Mises theorem also holds in other cases. One such case is when the data is i.i.d. and the underlying statistical model is semi-parametric, in the sense that the model is parametrized by a finite-dimensional parameter of interest θ and an infinite-dimensional nuisance parameter η. The data is $\{X_i\}_{i=1}^N$ where $X_i \sim^{iid} P_{\theta_0,\eta_0}$, where $P_{\theta,\eta}$ is a semi-parametric model with finite-dimensional parameter θ and infinite-dimensional nuisance parameter η.

Often, $\hat{\theta}_N$ can be taken to be an asymptotically linear and efficient estimator of θ, and Σ_0 the inverse efficient Fisher information matrix. The marginal prior for θ must have a continuous and positive density on a neighborhood of θ_0. Under a few further regularity conditions, the semi-parametric Bernstein von-Mises theorem results from the literature show that Assumption 1 can be satisfied when the underlying statistical model is semi-parametric (e.g., Shen (2002), Bickel and Kleijn (2012), Castillo (2012), Castillo and Rousseau (2015)).

Remark 1 is not a comprehensive review of the literature on the Bernstein-von Mises theorem. For example, in addition to the above, there are also Bernstein-von Mises type results in the literature that cover certain cases of nonparametric models (e.g., Castillo and Nickl (2013)). In addition, Bernstein-von Mises type results have been specifically explored in the context of important models, including limited information and moment condition models (e.g., Kwan (1999), Kim (2002), and Chib, Shin, and Simoni (2018)), various kinds of linear and partially linear regressions (e.g., Bickel and Kleijn (2012) and Norets (2015)) proportional hazard models (e.g., Kim (2006)), and models based around quasi-posteriors (e.g., Chernozhukov and Hong (2003)). Research on the Bernstein-von Mises theorem is ongoing. The results of this paper simply require Assumption 1 to hold. The literature on the Bernstein-von Mises theorem shows Assumption 1 holds in many cases. Particularly the part of Assumption 1 concerning matching covariances of the Bayesian distribution and the classical distribution may not hold if the model is misspecified as in, for example, Kleijn and Van der Vaart (2012) and Müller (2013).

Under Assumption 1, the results of this paper concern what a Bayesian can conclude from p-values asymptotically. Equivalently, the results may be viewed as concerning approximations to Bayesian posterior probabilities in large samples, based on observing p-values (and classical estimates $\hat{\theta}_N$). The results hold for priors in the class of priors that are compatible with Assumption 1. As noted in Remark 1, a main condition is that the prior for θ has a continuous and positive density on a neighborhood of θ_0. The results will imply that, for any prior in that class of priors compatible with Assumption 1, in a large enough sample, there are connections between p-values and certain Bayesian posterior probabilities.

Assumption 1 is directly on the parameter of interest θ. Per the standard Delta theorem calculations in Remark 2, if Assumption 1 holds for θ but $\tilde{\theta} = f(\theta)$ is the parameter of interest, Assumption 1 also holds for $\tilde{\theta}$ under regularity conditions.

Remark 2 (Parameter of interest is a function of θ)**.** Suppose that Assumption 1 holds, but the parameter of interest is $\tilde{\theta} \in \mathbb{R}^r$ with $\tilde{\theta} = f(\theta)$ for some known function $f(\cdot) : \mathbb{R}^m \to \mathbb{R}^r$ that is continuously differentiable at θ_0, and $r \leq m$. Suppose that the matrix of first derivatives of $f(\cdot)$ is $F(\cdot)$, so that $F(\cdot)$ is $r \times m$. Then, the standard Delta theorem method (e.g., Van der Vaart (1998, Chapter 3), Wasserman (2004)) implies that Assumption 1 also holds for the parameter of interest, with $\hat{\theta}_N$ in Assumption 1 replaced by $f(\hat{\theta}_N)$, and θ_0 in Assumption 1 replaced by $f(\theta_0)$, and θ in Assumption 1 replaced by $f(\theta)$, and Σ_0 in Assumption 1 replaced by $F(\theta_0)\Sigma_0 F(\theta_0)'$, as long as $F(\theta_0)\Sigma_0 F(\theta_0)'$ is nonsingular, which holds if $F(\theta_0)$ has full (row) rank. In order to conserve on notation in the rest of the paper, it is therefore always understood that the results could apply to a parameter of interest that is such a continuously differentiable function of a parameter that satisfies Assumption 1.

3. Bayesian conclusions from p-values

Let $\theta = (\theta_1, \theta_2, \ldots, \theta_t, \ldots, \theta_m)$. Results are given in general for a multivariate parameter ($m > 1$). The important special case of a scalar parameter ($m = 1$) is given special attention in the results. Given the asymptotically normal sampling distribution of $\hat{\theta}_N$ characterized by Assumption 1, the Wald test is the classical ("frequentist") method for testing hypotheses of the form $\theta_t = c_t$, for $t \in \{1, 2, \ldots, m\}$. Such hypotheses concern the scalar components of θ. The Wald test statistic for the t-th component is

$$W_{N,t} = N(\hat{\theta}_{N,t} - c_t)'(\hat{\Sigma}_{N,tt})^{-1}(\hat{\theta}_{N,t} - c_t). \tag{1}$$

Note that $(\hat{\Sigma}_{N,tt})^{-1}$ is the inverse of the t-th diagonal element of $\hat{\Sigma}_N$, which in general is not the same as the t-th diagonal element of $\hat{\Sigma}_N^{-1}$. Because of Assumption 1, if the null hypothesis $\theta_t = c_t$ is true, then $W_{N,t}$ converges in distribution to χ_1^2, a chi-squared distribution with 1 degree of freedom. Therefore, the p-value for the null hypothesis that $\theta_t = c_t$, based on the Wald test statistic, is

$$p_{N,t} = P(\chi_1^2 > W_{N,t}) = \overline{F}_{\chi_1^2}(W_{N,t}). \tag{2}$$

In other words, the p-value is the probability that a random variable with χ_1^2 distribution exceeds $W_{N,t}$. The p-value is data-dependent, because it depends on $W_{N,t}$.

The analysis in this paper concerns what a Bayesian can conclude from seeing p-values, and other information commonly reported in published research, specifically the classical estimate $\hat{\theta}_N$. Results are available for different cases of what p-values are available to the Bayesian.

The setup covers the important special case of a p-value for a scalar parameter θ. In that case, $m = 1$ and $p_{N,1}$ is available to the Bayesian. The setup also covers the important special case of a p-value for a scalar component of a multivariate parameter. In that case, attention is focused on a scalar component θ_{t^*} for some particular $t^* \in \{1, 2, \ldots, m\}$, and p_{N,t^*} is available to the Bayesian. Note that if Assumption 1 holds for multivariate θ, then Assumption 1 holds for each scalar component θ_t of θ. Therefore, the case of a scalar component of a multivariate parameter can be analyzed exactly the same way as a scalar parameter. The "scalar θ case" will therefore be understood to cover both of the above two cases: a scalar parameter and a scalar component of a multivariate parameter.

The setup also covers the case of multiple p-values for the components of a multivariate parameter. In that case, $p_{N,t}$ for all $t \in \{1, 2, \ldots, m\}$ is available to the Bayesian. For example, this case covers the common case of p-values for hypotheses $\theta_t = 0$ being reported for one or more parameters in a multi-parameter model (e.g., coefficient(s) in a linear regression of an outcome on multiple explanatory variables).

Finally, when θ is multivariate, for the null hypothesis $\theta = c$ concerning the entire multivariate parameter, the Wald test again is the classical ("frequentist") method. The Wald test statistic is

$$W_N = N(\hat{\theta}_N - c)'\hat{\Sigma}_N^{-1}(\hat{\theta}_N - c). \tag{3}$$

The p-value for the null hypothesis that $\theta = c$, based on the Wald test statistic, is

$$p_N = P(\chi^2_m > W_N) = \overline{F}_{\chi^2_m}(W_N). \tag{4}$$

3.1. **p-values and Bayesian posteriors for one-sided hypotheses.** The result of this section is that p-values can be used to find asymptotic approximations to Bayesian posterior probabilities concerning hypotheses like $\theta > d$ and/or $\theta < d$, for specified d. This is formally stated in Theorem 1 and discussed afterwards. The proofs of the main results depend on Lemma 1, which is stated below, and proved in Appendix A. The proof of Theorem 1 is in Appendix A.

Particularly in the multivariate case, let the hypothesis of interest (from the Bayesian perspective) have the form

$$\mathcal{H}_{a,d} = \{\theta : a_1\theta_1 > a_1 d_1, a_2\theta_2 > a_2 d_2, \ldots, a_m\theta_m > a_m d_m\} \tag{5}$$

for some choice of $a = (a_1, a_2, \ldots, a_m) \in \{-1, 1\}^m$ and $d = (d_1, d_2, \ldots, d_m)$. In other words, the hypothesis is that each element of θ is either greater than (if the corresponding a is positive), or less than (if the corresponding a is negative), the corresponding element of d.

Lemma 1. *Under Assumption 1,* $||\Pi_{\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0, \hat{\Sigma}_N)||_{TV} \to^p 0$ *and* $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0, I_{m\times m})||_{TV} \to^p 0$.

Theorem 1. *Under Assumption 1, in particular when* $m = 1$,

$$\sup_d \left| \left(\Pi\left(\theta > d \,|\, X^{(N)}\right) - \Phi\left(\sqrt{Q_{\chi^2_1}(1-p_N)}\frac{\hat{\theta}_N - d}{|\hat{\theta}_N - c|}\right)\right) 1[p_N < 1] \right| = o_p(1) \tag{6}$$

and

$$\sup_d \left| \left(\Pi\left(\theta < d \,|\, X^{(N)}\right) - \Phi\left(\sqrt{Q_{\chi_1^2}(1-p_N)} \frac{d - \hat{\theta}_N}{|\hat{\theta}_N - c|} \right) \right) 1[p_N < 1] \right| = o_p(1). \tag{7}$$

Under Assumption 1, including when $m \geq 1$*,*

$$\left(\Pi\left(\theta \notin \mathcal{H}_{a,d} \,|\, X^{(N)}\right) - \sum_t \Phi\left(-a_t \sqrt{Q_{\chi_1^2}(1-p_{N,t})} \frac{\hat{\theta}_{N,t} - d_t}{|\hat{\theta}_{N,t} - c_t|} \right) \right) 1[p_{N,t} < 1 \textit{ for all } t] \leq o_p(1) \tag{8}$$

and

$$\left(\Pi\left(\theta \notin \mathcal{H}_{a,d} \,|\, X^{(N)}\right) - \max_t \left\{ \Phi\left(-a_t \sqrt{Q_{\chi_1^2}(1-p_{N,t})} \frac{\hat{\theta}_{N,t} - d_t}{|\hat{\theta}_{N,t} - c_t|} \right) \right\} \right) 1[p_{N,t} < 1 \textit{ for all } t] \geq o_p(1). \tag{9}$$

Recall the notation was defined in Section 1.2.

Suppose a Bayesian sees the p-value p_N from Equation 4 when θ is a scalar. In that case, as long as the p-value is not 1,[6] Equations 6 and 7 of Theorem 1 establish that the p-value for the null hypothesis $\theta = c$ along with $\hat{\theta}_N$ can be used to find asymptotic approximations to Bayesian posterior probabilities concerning the hypothesis that $\theta > d$ and/or $\theta < d$, for any d. Fundamentally, this means that p-values are useful to a Bayesian in settings exhibiting the Bernstein-von Mises phenomenon. As discussed in the introduction, this contrasts with the literature of results that focus on the case of settings that do not exhibit the Bernstein-von Mises phenomenon.

[6]In Equations 6 and 7, there is a "discontinuity" in the interpretation of the p-value when $p_N = 1$ compared to when $p_N < 1$, since $p_N = 1$ corresponds to $\hat{\theta}_N = c$ and $p_N < 1$ corresponds to $\hat{\theta}_N \neq c$, so there would be a division by 0 when $p_N = 1$. Since $p_N = 1$ is an unlikely event, this has essentially no practical impact. Discussion will focus on the case that $p_N < 1$. See Remark 4.

There are two further contrasts with the literature. Note that, as with all of the results in this paper, these conclusions are large sample approximations to the posterior probabilities, for any prior consistent with Assumption 1. Also note that, as with all of the results in this paper, these conclusions concern the classical p-value for the null hypothesis that $\theta = c$ and corresponding Bayesian conclusions concerning *other* hypotheses about θ.

Consider the p-value for the null hypothesis that $\theta = 0$ when θ is a scalar (and when the p-value is not 1). Suppose, as is often the case, that θ is the effect of a treatment on an outcome, like a regression coefficient. Equations 6 and 7 of Theorem 1 establish that the p-value can be used to find asymptotic approximations to Bayesian posterior probabilities concerning the hypothesis that the treatment effect is greater than some d, or less than some d. Unavoidably, there is some flexibility in choosing d, and so the results accommodate different choices of d. On the other hand, deciding on a meaningful hypothesis is inherently part of doing empirical research. For example, this could mean deciding on what it means for a treatment effect to be of a meaningfully positive magnitude. (And, to be clear, the results accommodate working with multiple candidate definitions of "meaningful.")

This result also applies to multivariate cases. Suppose a Bayesian sees the p-values $p_{N,t}$ from Equation 2 when θ is multivariate. In that case, Equations 8 and 9 of Theorem 1 mean that multiple p-values for scalar components of θ can be combined to conclude asymptotic bounds on the Bayesian posterior probability that all elements of θ are either greater than (or less than) the corresponding element of d. Per Equation 5, note that by choosing the elements of a to be positive or negative, the hypothesis can be that some elements of θ are greater than the corresponding element of d and other elements of θ are less than the corresponding element of d.

Remark 3 (Computing p-values)**.** There is another very different use of these results. In complicated models, the point is often made that often Bayesian computations are simpler than classical ("frequentist") computations. In particular, computing Bayesian posterior probabilities can be easier than computing classical p-values. Although certainly not the main point of this paper, it is possible to use the results to compute p-values in complicated models, by computing the (potentially much simpler) Bayesian posterior probability of the hypothesis associated with the p-value. So for example, it is possible to compute a p-value by computing a corresponding Bayesian posterior probability, according to Theorem 1.

Remark 4 (p-values that equal 1)**.** The discussion of the results focuses on the case of p-values that do not equal 1. This is not a practical limitation. Note that $1[p_N = 1] \to^p 0$ and $1[p_{N,t} = 1] \to^p 0$ for all t, under Assumption 1. Similarly, under a further assumption that $\hat{\theta}_N$ has a continuous sampling distribution in all sample sizes, it would follow that $P(1[p_N = 1] = 1) = 0$ and $P(1[p_{N,t} = 1] = 1) = 0$ for all t. Thus, in either of those senses, the case of a p-value that equals 1 has negligible practical relevance. Equivalently, $1[p_N < 1] \to^p 1$ and $1[p_{N,t} < 1] \to^p 1$ for all t, under Assumption 1. Under a further assumption that $\hat{\theta}_N$ has a continuous sampling distribution in all sample sizes, $P(1[p_N < 1] = 1) = 1$ and $P(1[p_{N,t} < 1] = 1) = 1$ for all t. Nevertheless, such terms are explicitly included in the formal statement of the results to be clear about the special case of a p-value of 1. For example, Equation 6 of Theorem 1 can be understood to imply that the p-value can be used to find an asymptotic approximation to the displayed Bayesian posterior probability in the event that $p_N < 1$, which asymptotically happens with probability approaching 1. Equation 6 is so written to be clear that it does not say anything about the displayed posterior probability in the exceptional event of $p_N = 1$. This is done because, although a p-value of 1 asymptotically happens with probability approaching 0, it might be tempting to "plug in" a p-value of 1 to check what the results say when a p-value is 1, if the results did not explicitly indicate that a p-value of 1 is an exceptional case. Similar remarks apply to other results in the paper.

3.2. **p-values and Bayesian posteriors for approximate null hypotheses.** This section establishes that p-values are connected to the Bayesian posterior probabilities of *"approximate"* null hypotheses. This result is partially a response to the literature that focused on the question of whether p-values are equal to the Bayesian posterior probability of the null hypothesis, and found negative results. The "approximate" null hypotheses will be "interval" hypotheses that are "centered" at the point hypothesis $\theta = c$.[7] Unavoidably, there is some flexibility in choosing an "approximate" null hypothesis, and so the results accommodate different choices of "approximate." Broadly, there are two closely related ways to connect p-values to an "approximate" null hypothesis, reflected in Theorem 2. The proof of Theorem 2 relies on Lemma 2, which is stated below, and proved and discussed in Appendix A. The proof of Theorem 2 also is in Appendix A. In particular, this result helps to clarify the understanding of what classical inference statements imply about whether the null hypothesis is (approximately) "true" or "false."

Lemma 2. *Under Assumption 1,*

$$p_N = \Pi(\delta_{\hat{\Sigma}_N,2}(\theta, c) \leq \sqrt{\frac{Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(p_N)}{N}} \mid X^{(N)})1[p_N < 1] + 1[p_N = 1] + o_p(1), \quad (10)$$

and

$$\sup_{0<q<1} \left| q - \Pi\left(\delta_{\hat{\Sigma}_N,2}(\theta, c) \leq \sqrt{\frac{Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)}{N}} \mid X^{(N)}\right)\right| = o_p(1). \quad (11)$$

Theorem 2. *Under Assumption 1, in particular when $m = 1$,*

$$p_N = \Pi\left(|\theta - c| \leq \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(p_N)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c| \mid X^{(N)}\right)1[p_N < 1] + 1[p_N = 1] + o_p(1), \quad (12)$$

[7]The existing literature (in the framework discussed in the introduction) has used the fact that "small" interval hypotheses can be suitably approximated by a point hypothesis (e.g., Berger and Sellke (1987, Section 4)), but here the approximate null hypotheses are not "small" enough to be approximately the same as the point null hypothesis. In other words, the point null hypothesis ($\theta = c$) has zero Bayesian posterior probability in this setup, whereas the "approximate" null hypotheses can be found to have Bayesian posterior probability asymptotically equal to the p-value, and the p-value need not be "close" to 0.

and

$$\sup_{0<q<1}\left|\left(q-\Pi\left(|\theta-c|\leq\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N-c|\mid X^{(N)}\right)\right)1[p_N<1]\right|=o_p(1). \quad (13)$$

Under Assumption 1, including when $m\geq 1$, for any $0<q<1$,

$$\left(\Pi\left(||\theta-c||_\infty\leq\min_t\left\{\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(q)}{Q_{\chi^2_1}(1-p_{N,t})}}|\hat{\theta}_{N,t}-c_t|\right\}\mid X^{(N)}\right)-q\right)1[p_{N,t}<1 \text{ for all } t]\leq o_p(1), \quad (14)$$

and

$$\left(\Pi\left(||\theta-c||_\infty\leq\max_t\left\{\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(\frac{m-1+q}{m})}{Q_{\chi^2_1}(1-p_{N,t})}}|\hat{\theta}_{N,t}-c_t|\right\}\mid X^{(N)}\right)-q\right)1[p_{N,t}<1 \text{ for all } t]\geq o_p(1). \quad (15)$$

The literature focused on the question of whether p-values are equal to the Bayesian posterior probability of the null hypothesis, and found negative results. With a related motivation, this result can be used to construct an "approximate" null hypothesis that has Bayesian posterior probability equal to the p-value. Subsequently, this result helps to clarify the understanding of what classical inference statements imply about whether the null hypothesis is (approximately) "true" or "false." These implications behave in non-intuitive ways, reinforcing the need for such results.

Suppose a Bayesian sees the p-value p_N from Equation 4 when θ is a scalar. In that case, as long as the p-value is not 1,[8] Equation 12 of Theorem 2 establishes that the

[8]In Equation 12, there is a "discontinuity" in the interpretation of the p-value when $p_N=1$ compared to when $p_N<1$, since Equation 12 concerns an "approximate" (and bounded) null hypothesis that has Bayesian posterior probability p_N, but no bounded set has Bayesian posterior probability 1 under the posterior for θ. Since $p_N=1$ is an unlikely event, this has essentially no practical impact. See Remark 4.

p-value is asymptotically equal to the Bayesian posterior probability that the distance between θ and c is less than $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(p_N)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|$. Thus, the p-value is asymptotically equal to the Bayesian posterior probability of this "approximate" null hypothesis. In other words, the p-value is asymptotically equal to the Bayesian posterior probability of the hypothesis that $\theta \in \left(c - \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(p_N)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|, c + \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(p_N)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|\right)$. This is an "approximate" null hypothesis because it is an interval centered at the null hypothesis value, c. This depends on the p-value (i.e., $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(p_N)}{Q_{\chi^2_1}(1-p_N)}}$) and the distance between $\hat{\theta}_N$ and c (i.e., $|\hat{\theta}_N - c|$). Therefore, this Bayesian use of classical p-values concerns a "data-dependent" hypothesis.

As a numerical example of Equation 12, if a Bayesian sees $p_N = 0.10$, then the hypothesis that $\theta \in (c - 0.2787|\hat{\theta}_N - c|, c + 0.2787|\hat{\theta}_N - c|)$ has Bayesian posterior probability asymptotically equal to 0.10. Hence, it is that "approximate" null hypothesis that has Bayesian posterior probability that is asymptotically equal to the p-value, when the p-value is 0.10.

This result can also be used to construct an "approximate" null hypothesis that has any specified Bayesian posterior probability. Equation 13 of Theorem 2 establishes that, given the p-value for the null hypothesis $\theta = c$, when θ is a scalar, it is possible to calculate an "approximate" null hypothesis that has $q\%$ Bayesian posterior probability, for any desired q. Specifically, there is asymptotically $q\%$ Bayesian posterior probability that the distance between θ and c is less than $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|$. In other words, there is asymptotically $q\%$ Bayesian posterior probability of the hypothesis that $\theta \in \left(c - \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|, c + \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|\right)$. This result can either be used to find an "approximate" null hypothesis that has a desired Bayesian posterior probability, or to compute the Bayesian posterior probability of a desired "approximate" null hypothesis. As a numerical example of Equation 13, suppose the Bayesian sees that $p_N = 0.10$, and that the Bayesian wants to know which "approximate" null hypothesis has 5% Bayesian posterior probability. Equation 13 establishes that the hypothesis that $\theta \in (c - 0.1450|\hat{\theta}_N - c|, c + 0.1450|\hat{\theta}_N - c|)$ has Bayesian posterior probability asymptotically equal to 0.05, when $p_N = 0.10$. In fact, for any given p-value a plot like those displayed in Figure 1 may be found. Such a plot can be used to find the Bayesian posterior probability that θ is within any given tolerance of c, being sure to remember the role of $|\hat{\theta}_N - c|$. Such a plot also can be used to find, for any desired Bayesian posterior probability, an "approximate" null hypothesis that has that desired Bayesian posterior probability.

It can be reasonable to ask in particular about the "approximate" null hypothesis that has 50% Bayesian posterior probability. Beyond an intuitive appeal of a hypothesis that is "equally likely to be true or false," the form of this "approximate" null hypothesis is particularly simple. When $q = 0.5$, application of Equation 13 amounts to finding an "approximate" null hypothesis that asymptotically has 50% Bayesian posterior probability. Note that $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(0.5)}{Q_{\chi^2_1}(1-p_N)}} \approx 1$ for $p_N \leq 0.25$. This holds because the median of a non-central chi-square distribution with 1 degree of freedom is approximately equal to the non-centrality parameter, when the non-centrality parameter is not small. Therefore, when a Bayesian sees a p-value less than 0.25, the Bayesian can quickly conclude that there is asymptotically essentially a 50% Bayesian posterior probability that θ is in the interval $(c - |\hat{\theta}_N - c|, c + |\hat{\theta}_N - c|)$. In other words, there is essentially equal Bayesian odds of $\theta \in (c - |\hat{\theta}_N - c|, c + |\hat{\theta}_N - c|)$ and $\theta \notin (c - |\hat{\theta}_N - c|, c + |\hat{\theta}_N - c|)$, whenever $p_N \leq 0.25$.

These results give another use of a p-value for a Bayesian. For example, consider again the case that the parameter is the effect of a treatment on an outcome, like a regression coefficient, and consider the p-value associated with the null hypothesis that $\theta = 0$. These results can be used to evaluate the p-value in terms of whether it implies that the treatment effect is "approximately" 0 or not even "approximately" 0. Some numerical examples were derived previously. Obviously, there is some flexibility in choosing what "approximately" 0 means, and this seems unavoidable in cases where (as is usually the case in economics), the empirical analysis does not really care about whether the null hypothesis is literally true or false. In such cases, deciding on a meaningful hypothesis is inherently part of doing empirical research. For example, this could mean deciding on what it means for a treatment effect to be meaningful. Nevertheless, p-values are commonly reported and used in empirical research, and so it is important to know how to interpret them in useful ways. (And, to be clear, the results accommodate working with multiple candidate definitions of "approximate.")

This result helps to clarify the understanding of what classical inference statements imply about whether the null hypothesis is (approximately) "true" or "false." As an example, suppose there are many different parameters, representing different treatment effects. For each, the classical estimate of the treatment effect is calculated and a p-value for the null hypothesis of no effect is calculated. Respectively, the p-values and the classical estimates are: $(p = 0.05, \hat{\theta} = 0.15)$, $(p = 0.50, \hat{\theta} = 0.15)$, $(p = 0.75, \hat{\theta} = 0.15)$, and $(p = 0.05, \hat{\theta} = 0.12)$. Suppose that what "really" matters is whether the treatment effect is "approximately zero" or not. Suppose further that an "approximately zero" treatment effect is any treatment effect in $(-0.1, 0.1)$. As noted above, the results accommodate different choices of what "approximately zero" means. So, according to the classical estimate, none of the treatment effects are "approximately zero." But, considering Bayesian inference, which treatments are more likely to have an "approximately zero" treatment effect?

Compare the first and second treatments. The classical estimates of the treatment effects are the same. The first treatment effect has a much smaller p-value. This might suggest that the second treatment effect is "more likely" to have an "approximately zero" treatment effect. And indeed, the Bayesian posterior probability that the second treatment effect is "approximately zero" is greater than the probability that the first treatment effect is "approximately zero:" 28.1% compared to 25.6%.

But now, compare the first and third treatments. The classical estimates of the treatment effects are the same. The first treatment effect has a much smaller p-value. This might suggest that the third treatment effect is "more likely" to have an "approximately zero" treatment effect. Yet, the Bayesian posterior probability that the first treatment effect is "approximately zero" is actually greater than the probability that the third treatment effect is "approximately zero:" 25.6% compared to 16.0%.

Or compare the third and fourth treatments. The classical estimate of the fourth treatment effect is somewhat closer to zero. This might suggest that the fourth treatment effect is "more likely" to have an "approximately zero" treatment effect. But, the fourth treatment effect has a much smaller p-value. This might suggest that the third treatment effect is "more likely" to have an "approximately zero" treatment effect. Perhaps, on net, the probability of an "approximately zero" treatment effect is about the same for these two treatments. In fact, the Bayesian posterior probability that the fourth treatment effect is "approximately zero" is actually much greater than the probability that the third treatment effect is "approximately zero:" 37.2% compared to 16.0%.

Therefore, the result shows both that it is possible to compute these Bayesian posterior probabilities of "approximate" null hypotheses, and that these behave in non-intuitive ways, reinforcing the need for such results. For example, these examples show that a higher p-value does not necessarily mean that the null hypothesis is more likely to be (approximately) true. Rather, it can be that a higher p-value means that the null hypothesis is *less likely* to be approximately true, essentially because a higher p-value can suggest that values of the parameter *other than* the null hypothesis value are even more likely.

The results of this section also apply to multivariate cases.[9] In that case, the result provides asymptotic upper and lower bounds on the Bayesian posterior probabilities of "approximate" null hypotheses. Suppose a Bayesian sees the p-values $p_{N,t}$ from Equation

[9] In the case of scalar θ, the combination of Equations 14 and 15 show that there is asymptotically $q\%$ Bayesian posterior probability that $|\theta - c| \leq \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c|$, which is the result of Equation 13.

2 when θ is multivariate. In that case, as long as none of the p-values are 1,[10] Equation 14 of Theorem 2 establishes that an "approximate" null hypothesis can be computed that asymptotically has at most q% Bayesian posterior probability, for any desired q. For example, suppose that $m = 2$, $c = (0, 0)$, $p_{N,1} = 0.05$, $p_{N,2} = 0.05$, $\hat{\theta}_{N,1} = 0.1$, and $\hat{\theta}_{N,2} = 0.1$. Then $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-0.05)}}(0.5)}{Q_{\chi^2_1}(1-0.05)}} = 1.0001$ from Table 1. So, it may be approximated (asymptotically, i.e., when N is large) that there is no more than 50% Bayesian posterior probability that $||\theta||_\infty \leq 0.10001$.

Also in this case, Equation 15 of Theorem 2 establishes that an "approximate" null hypothesis can be computed that asymptotically has at least q% Bayesian posterior probability, for any desired q. For example, suppose again that $m = 2$, $c = (0, 0)$, $p_{N,1} = 0.05$, $p_{N,2} = 0.05$, $\hat{\theta}_{N,1} = 0.1$, and $\hat{\theta}_{N,2} = 0.1$. Then, for $q = 0.50$, $\frac{m-1+q}{m} = 0.75$, and therefore $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(\frac{m-1+q}{m})}{Q_{\chi^2_1}(1-p_{N,t})}} = 1.3441$ from Table 1. So, it may be approximated (asymptotically, i.e., when N is large) that there is at least 50% Bayesian posterior probability that $||\theta||_\infty \leq 0.13441$.

Remark 5 (Numerical values)**.** Table 1 calculates the numerical values of $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p)}}(q)}{Q_{\chi^2_1}(1-p)}}$ for different values of p and q. Figure 1 plots q (along the vertical) against $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p)}}(q)}{Q_{\chi^2_1}(1-p)}}$ (along the horizontal) for different values of p.

An important remark in interpreting these results, for example when reading Table 1 or Figure 1, is that $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}$ *multiplies* $|\hat{\theta}_N - c|$ in Equations 12 and 13. Therefore, the "magnitude" of $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}$ is not independently determinative of the "size" of the "approximate" null hypothesis that has q% Bayesian posterior probability.

[10]The indicator $1[p_{N,t} < 1 \text{ for all t}]$ means that Equations 14 and 15 are informative about the displayed posterior probability when $p_{N,t} < 1$ for all t, which asymptotically happens with probability approaching 1. See Remark 4.

		q					
		0.01	0.05	0.10	0.25	0.50	0.75
p	0.01	0.1227	0.3623	0.5026	0.7382	1.0000	1.2619
	0.05	0.0435	0.2032	0.3573	0.6568	1.0001	1.3441
	0.10	0.0295	0.1450	0.2787	0.5981	1.0008	1.4101
	0.25	0.0211	0.1055	0.2105	0.5195	1.0218	1.5903
	0.50	0.0233	0.1167	0.2338	0.5917	1.2432	2.0887
	0.75	0.0414	0.2070	0.4149	1.0520	2.2261	3.7938

TABLE 1. Numerical values of $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p)}}(q)}{Q_{\chi^2_1}(1-p)}}$ for different values of p and q, with $m = 1$

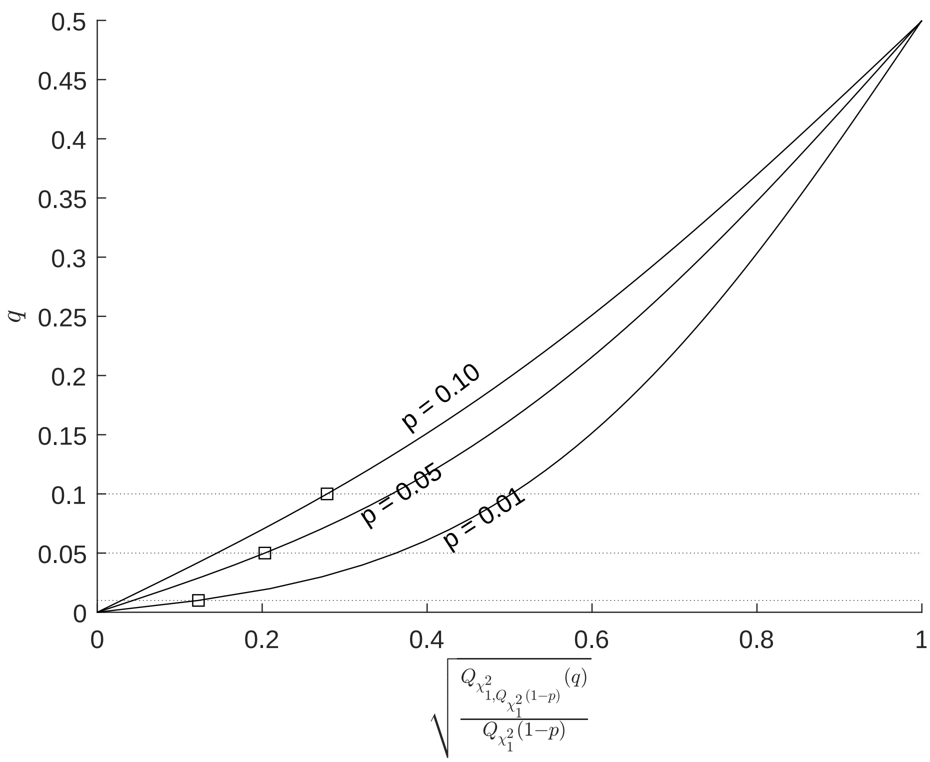

FIGURE 1. Plot of q against $\sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p)}}(q)}{Q_{\chi^2_1}(1-p)}}$, with $m = 1$

4. Conclusions

The results of this paper show that p-values have Bayesian uses concerning hypotheses *other than* the null hypothesis of $\theta = c$ that is ostensibly tested by the p-values. The related literature has essentially concluded strongly against p-values. These results show that the p-value actually does have value for a Bayesian. In particular, these results show that a Bayesian can draw useful conclusions from p-values that are reported in the research of other researchers. The relationship between Bayesian posterior probabilities of "approximate" null hypotheses and p-values behaves in non-intuitive ways, reinforcing the need for such results to understand what classical inference statements imply about whether the null hypothesis is (approximately) "true" or "false."

Appendix A. Technical appendix: proofs and further discussions

Proof of Lemma 1. Because $||\mathcal{N}(0,\Sigma_1) - \mathcal{N}(0,\Sigma_2)||_{TV} \to 0$ if $\Sigma_1 \to \Sigma_2$, since the total variation distance is bounded above by the square root of the Kullback-Leibler divergence (e.g., DasGupta (2008, Chapter 2)), it follows that $||\mathcal{N}(0,\hat{\Sigma}_N) - \mathcal{N}(0,\Sigma_0)||_{TV} \to^p 0$ since $\hat{\Sigma}_N \to^p \Sigma_0$ by Assumption 1. Therefore, by Assumption 1, $||\Pi_{\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0,\hat{\Sigma}_N)||_{TV} \to^p 0$.

Let $h_N(\cdot) : \mathbb{R}^m \to \mathbb{R}^m$ be given by $h_N(w) = \hat{\Sigma}_N^{-\frac{1}{2}} w$. Since $h_N(\cdot)$ is continuous, it follows that $h_N^{-1}(\mathcal{B}) \subseteq \mathbb{R}^m$ is a Borel set for any Borel set $\mathcal{B} \subseteq \mathbb{R}^m$. Therefore, $|\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}(\mathcal{B}) - \mathcal{N}(0, I_{m\times m})(\mathcal{B})| = |\Pi_{\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}(h_N^{-1}(\mathcal{B})) - \mathcal{N}(0,\hat{\Sigma}_N)(h_N^{-1}(\mathcal{B}))|$. Therefore, because $||\Pi_{\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0,\hat{\Sigma}_N)||_{TV} \to^p 0$, it follows that $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}} - \mathcal{N}(0, I_{m\times m})||_{TV} \to^p 0$. □

Proof of Theorem 1. **Equations 6 and 7:** Note that $1 - p_N = F_{\chi^2_1}(W_N)$ so $Q_{\chi^2_1}(1 - p_N) = W_N = N(\hat{\Sigma}_N)^{-1}(\hat{\theta}_N - c)^2$, so $\hat{\Sigma}_N = \frac{N(\hat{\theta}_N - c)^2}{Q_{\chi^2_1}(1-p_N)}$ as long as $0 < p_N < 1$. Let $Z \sim \mathcal{N}(0, I_{1\times 1})$. Note that $\Pi_{\theta|X^{(N)}}((d,\infty)) = \Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}((\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N), \infty))$ and similarly $\Pi_{\theta|X^{(N)}}((-\infty, d)) = \Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}((-\infty, \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N)))$. By Assumption 1 and Lemma 1, using the notation that Borel set $\mathcal{B} \subseteq \mathbb{R}$, $\sup_{\mathcal{B}} |\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}(\mathcal{B}) - Z(\mathcal{B})| \to^p 0$. Therefore, in particular, $\sup_d |\Pi_{\theta|X^{(N)}}((d,\infty)) - \Phi(-\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N))| = \sup_d |\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}((\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N), \infty)) - Z((\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N), \infty))| \to^p 0$ and $\sup_d |\Pi_{\theta|X^{(N)}}((-\infty, d)) - \Phi(\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N))| = \sup_d |\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-\hat{\theta}_N)|X^{(N)}}((-\infty, \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N))) - Z((-\infty, \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N)))| \to^p 0$. After substitution, when $0 < p_N < 1$, $\Phi(-\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d -$

$\hat{\theta}_N)) = \Phi(-\frac{\sqrt{Q_{\chi_1^2}(1-p_N)}}{|\hat{\theta}_N - c|}(d - \hat{\theta}_N))$ and $\Phi(\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(d - \hat{\theta}_N)) = \Phi(\frac{\sqrt{Q_{\chi_1^2}(1-p_N)}}{|\hat{\theta}_N - c|}(d - \hat{\theta}_N))$, establishing Equations 6 and 7.

Equation 8: Note that the event that $\theta \notin \mathcal{H}_{a,d}$ is the same as the event that $\theta \in \bigcup_t \{\theta : a_t\theta_t \le a_t d_t\}$. By Equations 6 and 7, and subadditivity of $\Pi(\cdot|X^{(N)})$, $\Pi(\bigcup_t\{\theta : a_t\theta_t \le a_t d_t\} \mid X^{(N)})1[p_{N,t} < 1 \text{ for all } t] \le \sum_t \Pi(a_t\theta_t \le a_t d_t \mid X^{(N)})1[p_{N,t} < 1 \text{ for all } t] = \sum_t \left(\Phi\left(a_t \frac{\sqrt{Q_{\chi_1^2}(1-p_{N,t})}}{|\hat{\theta}_{N,t} - c_t|}(d_t - \hat{\theta}_{N,t})\right)\right) 1[p_{N,t} < 1 \text{ for all } t] + o_p(1)$, establishing Equation 8.

Equation 9: Note that $\Pi(\bigcup_t\{\theta : a_t\theta_t \le a_t d_t\} \mid X^{(N)}) \ge \max_t \Pi(a_t\theta_t \le a_t d_t \mid X^{(N)})$. By Equations 6 and 7, Equation 9 follows. □

Proof of Lemma 2. **Equations 10 and 11:** Note that $||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)||_2^2 = \sqrt{N}(\theta - c)'\hat{\Sigma}_N^{-1}\sqrt{N}(\theta - c) = \sqrt{N}(\theta - \hat{\theta}_N + \hat{\theta}_N - c)'\hat{\Sigma}_N^{-1}\sqrt{N}(\theta - \hat{\theta}_N + \hat{\theta}_N - c)$. Let $Z \sim \mathcal{N}(0, I_{m\times m})$. For any Borel set $\mathcal{B} \subseteq \mathbb{R}^m$, $\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)|X^{(N)}}(\mathcal{B}) = \Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - \hat{\theta}_N + \hat{\theta}_N - c)|X^{(N)}}(\mathcal{B}) = \Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - \hat{\theta}_N)|X^{(N)}}(\mathcal{B} - \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))$. By Assumption 1 and Lemma 1, $|\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - \hat{\theta}_N)|X^{(N)}}(\mathcal{B} - \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c)) - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))(\mathcal{B})| = |\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - \hat{\theta}_N)|X^{(N)}}(\mathcal{B} - \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c)) - Z(\mathcal{B} - \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))| \to^p 0$. Therefore, $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)|X^{(N)}} - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))||_{TV} \to^p 0$.

Consider $||\Pi_{||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)||_2^2|X^{(N)}} - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))'(Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))||_{TV}$, and let $h(\cdot) : \mathbb{R}^m \to \mathbb{R}$ be given by $h(z) = z'z$. Since $h(\cdot)$ is continuous, it follows that $h^{-1}(\mathcal{B}) \subseteq \mathbb{R}^m$ is a Borel set for any Borel set $\mathcal{B} \subseteq \mathbb{R}$. Therefore, $|\Pi_{||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)||_2^2|X^{(N)}}(\mathcal{B}) - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))'(Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))(\mathcal{B})| = |\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)|X^{(N)}}(h^{-1}(\mathcal{B})) - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))(h^{-1}(\mathcal{B}))|$. And therefore, because $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)|X^{(N)}} - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))||_{TV} \to^p 0$ by the above, it follows that $||\Pi_{||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)||_2^2|X^{(N)}} - (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))'(Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))||_{TV} \to^p 0$.

By standard results, $C_N = (Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c))'(Z + \hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\theta}_N - c)) \sim \chi^2_{m,W_N}$. Therefore, $\sup_{0<q<1} |\Pi_{||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta - c)||_2^2|X^{(N)}}((-\infty, Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)]) - C_N((-\infty, Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)])| \to^p 0$. By construction, $W_N = Q_{\chi^2_m}(1-p_N)$, so $C_N((-\infty, Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)]) = C_N((-\infty, Q_{\chi^2_{m,W_N}}(q)]) = q$. By definition of $\delta_{\hat{\Sigma}_N,2}$, $\Pi(\delta^2_{\hat{\Sigma}_N,2}(\theta, c) \le \frac{Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)}{N} \mid X^{(N)}) =$

$\Pi_{||\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\theta-c)||_2^2|X^{(N)}}((-\infty, Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)])$, so the claim in Equation 11 follows, and the claim in Equation 10 follows for $q = p_N$ for $0 < p_N < 1$. □

Remark 6 (Discussion of Lemma 2)**.** Suppose that a Bayesian sees p_N from Equation 4. In that case, as long as the p-value is not 1,[11] Equation 10 of Lemma 2 establishes that the p-value is asymptotically equal to the Bayesian posterior probability that the $\delta_{\hat{\Sigma}_N,2}$ distance between θ and c is less than $\sqrt{\frac{Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(p_N)}{N}}$. Or in other words, the p-value is asymptotically equal to the Bayesian posterior probability that the null hypothesis $\theta = c$ is "approximately" true, where "approximately" true means that θ is within this specific certain tolerance of c. Similarly, Equation 11 of Lemma 2 establishes that, given the p-value, it is possible to calculate an "approximate" null hypothesis that asymptotically has q% Bayesian posterior probability, for any desired q. Specifically, there is asymptotically q% Bayesian posterior probability that the $\delta_{\hat{\Sigma}_N,2}$ distance between θ and c is less than $\sqrt{\frac{Q_{\chi^2_{m,Q_{\chi^2_m}(1-p_N)}}(q)}{N}}$. Such a result can either be used to find an "approximate" null hypothesis that has a desired Bayesian posterior probability, or to compute the Bayesian posterior probability of a desired "approximate" null hypothesis. Note that Equations 10 and 11 depend on $\hat{\Sigma}_N$. Hence, for these equations to be used to interpret published p-values, it is necessary that $\hat{\Sigma}_N$ be reported. Some publications may report $\hat{\Sigma}_N$, whereas others do not. Critically, the results in Theorem 2 do not have this requirement.

Proof of Theorem 2. **Equations 12 and 13:** When $m = 1$, note that $1 - p_N = F_{\chi^2_1}(W_N)$ so $Q_{\chi^2_1}(1 - p_N) = W_N = N\hat{\Sigma}_N^{-1}(\hat{\theta}_N - c)^2$, so $\hat{\Sigma}_N^{-\frac{1}{2}} = \frac{\sqrt{Q_{\chi^2_1}(1-p_N)}}{\sqrt{N}|\hat{\theta}_N-c|}$ as long as $|\hat{\theta}_N - c| \neq 0$ which is equivalent to $p_N < 1$. Further, as long as $|\hat{\theta}_N - c| \neq 0$, note that $\delta_{\hat{\Sigma}_N,2}(\theta, c) = \hat{\Sigma}_N^{-\frac{1}{2}}|\theta - c| = \frac{\sqrt{Q_{\chi^2_1}(1-p_N)}}{\sqrt{N}|\hat{\theta}_N-c|}|\theta - c|$. Equation 12 then follows from Equation 10 of Lemma 2 by substitution. And Equation 13 then follows from Equation 11 of Lemma 2 by substitution.

[11]The indicator $1[p_N < 1]$ means that Equation 10 is informative about the displayed posterior probability when $p_N < 1$, which asymptotically happens with probability approaching 1. See Remark 4.

Equation 14: Because $\{\theta : ||\theta - c||_\infty \leq K\} = \cap_t \{\theta : |\theta_t - c_t| \leq K\}$, for any K, and by the Fréchet inequalities, as long as $0 < p_{N,t} < 1$ for all t, it holds that

$$\Pi\left(||\theta - c||_\infty \leq \min_t \left\{ \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(q)}{Q_{\chi^2_1}(1-p_{N,t})}} |\hat{\theta}_{N,t} - c_t| \right\} \mid X^{(N)} \right) 1[p_{N,t} < 1 \text{ for all t}] \leq$$

$$\min_{t^*} \Pi\left(|\theta_{t^*} - c_{t^*}| \leq \min_t \left\{ \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(q)}{Q_{\chi^2_1}(1-p_{N,t})}} |\hat{\theta}_{N,t} - c_t| \right\} \mid X^{(N)} \right) 1[p_{N,t} < 1 \text{ for all t}] \leq$$

$$\min_{t^*} \Pi\left(|\theta_{t^*} - c_{t^*}| \leq \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t^*})}}(q)}{Q_{\chi^2_1}(1-p_{N,t^*})}} |\hat{\theta}_{N,t^*} - c_{t^*}| \mid X^{(N)} \right) 1[p_{N,t} < 1 \text{ for all t}] =$$

$$q1[p_{N,t} < 1 \text{ for all t}] + o_p(1),$$

by Equation 13.

Equation 15: By the Fréchet inequalities and Equation 13, as long as $0 < p_{N,t} < 1$ for all t,

$$\Pi\left(||\theta - c||_\infty \leq \max_t \left\{ \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(\frac{m-1+q}{m})}{Q_{\chi^2_1}(1-p_{N,t})}} |\hat{\theta}_{N,t} - c_t| \right\} \mid X^{(N)} \right) 1[p_{N,t} < 1 \text{ for all t}] \geq$$

$$\left(1 - m + \sum_{t^*} \Pi\left(|\theta_{t^*} - c_{t^*}| \leq \max_t \left\{ \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t})}}(\frac{m-1+q}{m})}{Q_{\chi^2_1}(1-p_{N,t})}} |\hat{\theta}_{N,t} - c_t| \right\} \mid X^{(N)} \right) \right) 1[p_{N,t} < 1 \text{ for all t}] \geq$$

$$\left(1 - m + \sum_{t^*} \Pi\left(|\theta_{t^*} - c_{t^*}| \leq \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_{N,t^*})}}(\frac{m-1+q}{m})}{Q_{\chi^2_1}(1-p_{N,t^*})}} |\hat{\theta}_{N,t^*} - c_{t^*}| \mid X^{(N)} \right) \right) 1[p_{N,t} < 1 \text{ for all t}] =$$

$$\left(1 - m + m\left(\frac{m-1+q}{m} \right) \right) 1[p_{N,t} < 1 \text{ for all t}] + o_p(1) = q1[p_{N,t} < 1 \text{ for all t}] + o_p(1). \qquad \square$$

References

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.

Bickel, P., and B. Kleijn (2012): "The semiparametric Bernstein–von Mises theorem," *The Annals of Statistics*, 40(1), 206–237.

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.

Castillo, I. (2012): "A semiparametric Bernstein–von Mises theorem for Gaussian process priors," *Probability Theory and Related Fields*, 152(1-2), 53–99.

Castillo, I., and R. Nickl (2013): "Nonparametric Bernstein–von Mises theorems in Gaussian white noise," *The Annals of Statistics*, 41(4), 1999–2028.

Castillo, I., and J. Rousseau (2015): "A Bernstein–von Mises theorem for smooth functionals in semiparametric models," *The Annals of Statistics*, 43(6), 2353–2383.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo confidence sets for identified sets," *Econometrica*, 86(6), 1965–2018.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

Chib, S., M. Shin, and A. Simoni (2018): "Bayesian estimation and comparison of moment condition models," *Journal of the American Statistical Association*, 113(524), 1–13.

DasGupta, A. (2008): *Asymptotic Theory of Statistics and Probability*. Springer, New York.

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

Giacomini, R., and T. Kitagawa (2018): "Robust Bayesian Inference for Set-identified Models."

Held, L., and M. Ott (2018): "On p-values and Bayes factors," *Annual Review of Statistics and Its Application*, 5, 393–419.

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

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995): *Continuous Univariate Distributions*, vol. 2. Wiley, New York, 2 edn.

Kim, J.-Y. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107(1-2), 175–193.

Kim, Y. (2006): "The Bernstein–von Mises theorem for the proportional hazard model," *The Annals of Statistics*, 34(4), 1678–1700.

Kleijn, B., and A. Van der Vaart (2012): "The Bernstein-von-Mises theorem under misspecification," *Electronic Journal of Statistics*, 6, 354–381.

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.

Kwan, Y. K. (1999): "Asymptotic Bayesian analysis based on a limited information estimator," *Journal of Econometrics*, 88(1), 99–121.

Le Cam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

Le Cam, L., and G. L. Yang (2000): *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York, 2 edn.

Liao, Y., and A. Simoni (2019): "Bayesian inference for partially identified smooth convex models."

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.

Müller, U. K. (2013): "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix," *Econometrica*, 81(5), 1805–1849.

Norets, A. (2015): "Bayesian regression with nonparametric heteroskedasticity," *Journal of Econometrics*, 185(2), 409–419.

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.

Shen, X. (2002): "Asymptotic normality of semiparametric and nonparametric posterior distributions," *Journal of the American Statistical Association*, 97(457), 222–235.

Van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.

Wasserman, L. (2004): *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York.

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.