Objective Bayesian Precise Hypothesis Testing

Jeffrey A. Mills
Department of Economics
Lindner College of Business
University of Cincinnati
Cincinnati, Ohio 45221
jeffrey.mills@uc.edu

September, 2007, latest revision, March, 2018

Abstract

This paper develops and implements an alternative precise hypothesis testing procedure. The procedure involves forming a posterior odds ratio by evaluating the posterior density function at the value in the null hypothesis and at its supremum. This leads to a Bayesian hypothesis testing procedure in which the Jeffreys-Lindley-Bartlett paradox does not occur, and that is scientifically objective in the sense that noninformative reference priors can be used. Further, under the proposed procedure, the prior is invariant to the hypotheses to be tested, there is no need to assign non-zero mass on a particular point in a continuum, and the same hypothesis testing procedure applies for all continuous and discrete distributions. The resulting test procedure is uniformly most powerful, robust to reasonable variations in the prior, and easy to interpret correctly in practice. Several examples are given to illustrate the use and performance of the test.

1 Introduction.

Frequentist and Bayesian inference are most clearly differentiated by their approaches to precise null hypothesis testing. Even with very large samples, the frequentist and Bayesian conclusions from a precise test can be contradictory. It is possible to obtain a small frequentist p-value, strongly rejecting H_0, but a large posterior odds or Bayes factor in favor of H_0. This paper provides alternative Bayesian procedure that does not lead to contradictory results, so that Bayesian inference, frequentist testing, and Bayesian testing all provide comparable results in standard problems, and typically match closely when uninformative priors are employed, thus providing an objective Bayesian solution to precise hypothesis testing.

First, consider the following illustrative example (given by Stone, 1997). Suppose θ is the proportion of a specific type of particle counted in an experiment. The theory under consideration predicts that $\theta = 0.2$ exactly, so the null hypothesis is well defined, $H_0 : \theta = 0.2$, and there is no specific alternative, so $H_1 : \theta \neq 0.2$. Experimental evidence yields $s = 106,298$ of the specific type out of $n = 527,135$ total particles. What is the evidence against H_0?

Frequentist methods give $\hat{\theta} = 0.201652$, and standard error $= 0.0005526$, resulting in a p-value $= 0.0028$, indicating strong evidence against the null. The Bayesian physicist instead adopts a uniform prior, $p(\Theta = \theta|\Theta = [0,1]) = 1$, $0 \leq \theta \leq 1$, and computes the Bayes factor to be $B = 8.27$, indicating evidence in favor of H_0. The Bayesian posterior distribution however, is not in conflict with the p-value, since the posterior probability given the data, D, $P(\Theta > 0.2|D) = \Phi(2.9895) = 1 - p\text{-value}/2$, where Φ is the standard Gaussian cumulative distribution function. Any Bayesian using a uniform prior then, must have a strong posterior belief that the true value of θ is larger than 0.2. A 0.99 equal-tailed Bayesian credible interval for $\theta = (0.20023, 0.20308)$ is identical to the frequentist 99% confidence interval and excludes 0.2.

Why are Bayesian posterior odds in conflict with both frequentist hypothesis testing and Bayesian posterior inference? Methods of obtaining scientifically objective Bayesian posterior distributions for inference are widely accepted, usually involving noninformative or reference priors (Berger, 2006). The real problem appears to be the hypothesis testing framework when used to test a precise null hypothesis: "Noninformative prior Bayesian analyses unfortunately do not work well for testing a point null hypothesis making impossible an objective Bayesian solution" (Berger, 1985, p.153). Further, "most of the difficulties in

interpreting hypothesis tests arise from the artificial dichotomy that is required between $\Theta = \theta_0$ and $\Theta \neq \theta_0$. Difficulties related to this dichotomy are widely acknowledged from all perspectives of statistical inference" (Gelman *et al.*, 2004, p.250).

The most famous of these difficulties is the paradox discovered by Jeffreys (1939), Lindley (1957) and Bartlett (1957) (hereafter JLB paradox). This paradox arises when parameters from the prior distribution appear in the posterior odds ratio or Bayes factor, so that reasonable variations in the prior distribution (especially increasing or decreasing the prior variance) lead to substantial changes in the test results. This leads to difficulties in specifying scientifically objective prior distributions that could be widely accepted as appropriate for precise hypothesis testing. A large literature exists attempting to develop informative priors for precise hypothesis testing that mitigate the practical adverse effects of the JLB paradox (see Berger and Perichi, 2001, 2004, Perichi, 2005). That these effects can be large in practice, and no satisfactory solution has been found, is a serious problem for Bayesian hypothesis testing (Cousins, 2014, Villa & Walker, 2017).

This paper takes a different approach which eliminates the need to develop priors specifically for hypothesis testing. The approach involves a general reformulation of the alternative hypothesis used when testing a precise null hypothesis. This alternative testing procedure, with a few notable exceptions, appears to have been overlooked in the literature. Basu (1996a,b) examined a similar approach and compares the results obtained when a prior with nonzero mass on the point null is assigned. Basu also points out that the resulting posterior density ratio was used by Good (1965, 1967, 1976) for significance testing in multinomial distributions and contingency tables. Goodman (1999) suggests a similar approach for medical research, but without justification. More importantly, its usefulness in resolving the serious problems inherent in the standard testing framework, especially the JLB paradox, has not been previously recognized.

The main contributions of this paper are (i) to present an alternative procedure for hypothesis testing based on a reformulation of the alternative hypothesis, (ii) to demonstrate that the most serious problems faced by standard Bayesian testing procedures, including the JLB paradox, are resolved when the proposed alternative hypothesis testing procedure is adopted, (iii) to demonstrate and draw attention to some important previously unrecognized advantages that the proposed alternative testing procedure possesses over standard

methods, and (iv) examine the practical significance of these advantages.

To facilitate a clear understanding of the essential issues, the paper focuses on a few of the simplest canonical hypothesis testing problems in the applied statistics literature. Further applications currently available, demonstrating improved performance along with ease of use and interpretation of results, include: comparison of means in randomized controlled experiments (Strawn et al., 2017, Mills et al., 2018), meta analysis (Strawn et al., 2018), ANOVA testing (Mills and Namavari, 2016), unit root testing (Mills, 2015), cointegration testing (Mills and Namavari, 2017), and model selection (Cornwall and Mills, 2017).

The testing procedure is presented in section 2, and the resulting resolution of the JLB paradox is examined in section 3. Section 4 examines the case of a normally distributed variable with unknown mean and known variance. Section 5 relaxes the known variance assumption and considers testing for regression coefficients. Section 6 presents a simple (contrived) clinical trial example to illustrate differences in the proposed approach and standard practice. A brief discussion is given in Section 7. Section 8 draws conclusions.

2 An alternative testing procedure.

For some unknown quantity Θ, suppose we wish to test the precise hypothesis $\Theta = \theta_0$. The standard procedure is to specify the competing hypotheses as $H_0 : \Theta = \theta_0$ and $H_1 : \Theta \neq \theta_0$, then compute the posterior odds or Bayes factor given a suitable choice of prior. However, this involves comparing two sets with fundamentally different properties, one being either finite (containing only one element) or a narrow interval (an $\varepsilon-$neighborhood around θ_0), and the other is typically infinite or at least contains a relatively large number of elements (since $H_1 = \Theta \setminus \theta_0$).

The alternative approach proposed herein is to replace the null and alternative hypotheses with a set of ε-neighborhood interval hypotheses, by partitioning the parameter space,

$$H_0 : |\theta - \theta_0| < \varepsilon,\ H_z : |\Theta - z| < \varepsilon| < \varepsilon,\ z \in \{\Theta : z \neq \theta\}. \tag{1}$$

where $z \in \Theta$ define a partition, $\mathbb{P}_z$ of Θ, and we consider what happens to the probabilities of each hypothesis as the partition is extended so that the number of elements of $\mathbb{P}_z$, $N_z \to \infty$.

Following Jaynes (2003), concerning the resolution of the Borel-Kolmogorov paradox, we form the posterior odds ratio before passing to the limit, which gives,

$$O_{z0} = \frac{P(|\theta - z| \leq \varepsilon | D)}{P(|\theta - \theta_0| \leq \varepsilon | D)}. \tag{2}$$

As $\varepsilon \to 0, \forall z \in \Theta, z \neq \theta_0, O_{z0} \to p(\Theta = z|D)/p(\Theta = \theta_0|D)$. The maximum O_{z0} is then,

$$O = \sup_z O_{z0} = \sup_z \left[\frac{p(\Theta = z|D)}{p(\Theta = \theta_0|D)} \right]. \tag{3}$$

Since $\sup_z p(\Theta = z|D)$ is the mode of the posterior density, we can determine the maximum odds against the precise null hypothesis, H_0, by calculating (3). This obviates the need to calculate O_{z0} for any other values of θ unless they are of particular interest, because the odds against the null are no greater for any other possible value of θ. It is worth noting that this method of deriving O_{z0} is important because it emphasizes the fact that the entire parameter space is considered, and so explicitly addresses the criticism raised by Edwards *et al.* (1963) that only two points on the parameter space are given consideration when using O_{z0}. Further, we could also follow the recommendation of Gelman and Stern (2017) of "saying No to binary conclusions" and evaluate the posterior odds over a set of alternative values, providing a function rather than a discrete 'true' or 'false' answer.

If we wish to be scientifically objective and select a prior that represents only the background information available, then according to posterior inference $\bar{\theta} = \arg\max_\theta p(\Theta = \theta|D)]$ is the most likely value against the set of all other possible values of θ. Further, this procedure does not involve specifying a prior with nonzero mass on the point value in the null hypothesis. Once we have specified a reasonable prior for inference, $\pi(\theta)$, over the parameter space, the prior weight given to each hypothesis is already implied by this prior, and it does not need to be modified depending on the hypotheses under consideration. If we use a diffuse or reference prior giving equal weight to every possible value of θ, the prior ratio, $\pi(\bar{\theta})/\pi(\theta_0)$ will cancel out of (3) so that O is the objective posterior odds or 'Bayes factor' and is equal to the likelihood ratio $p(D|\bar{\theta})/p(D|\theta_0)$, and hence has a frequentist interpretation. Note also that the procedure does not require the existence of a unique mode; if the posterior density is multimodal, or has a flat region around the mode, any value in the set of modes can be selected since any point in this set will give the same value for O.

According to the Neyman-Pearson lemma, when both hypotheses are simple the likelihood ratio test rejects the null hypotheses when $p(D|H_1)/p(D|H_0)$ exceeds a constant. This test is uniformly most powerful (UMP) because it maximizes the power (the probability of rejecting the null when it is false) among tests with its significance level (probability of rejecting the null when it is true). It is well known that if one or both hypotheses are not precise, then there is typically no UMP test (DeGroot and Schervish, 2002). If we use the posterior odds, O, given by equation (3), comparing only precise hypotheses, then under fairly general conditions we have a UMP test (see also Johnson, 2013). Further, if we choose to reject the null when O exceeds a constant, then rather than fixing the size of the test and minimizing the type II error, the test will minimize a linear combination of the type I and type II errors so that both errors approach zero as the sample size increases. In this way we are not bound to a fixed nominal significance level as with frequentist testing procedures (DeGroot and Schervish, 2002). The 'UMP Bayesian Test' Bayes factors developed in Johnson (2013) often have close correspondence to the objective posterior odds ratio in many settings.

Exploring the robustness of the resulting objective posterior odds to variations in the prior is straightforward. The prior shows up as the ratio $\pi(\theta = \theta_z)/\pi(\theta = \theta_0)$ in the posterior odds comparing $H_0 : \theta = \theta_0 vs. H_z : \theta = \theta_z$. Let $\bar{\theta}$ equal the maximum a posteriori (MAP) value of θ. Any reasonably uninformative prior that is dominated by the likelihood will result in $\pi(\theta = \bar{\theta})/\pi(\theta = \theta_0) \approx 1$. Further, the posterior odds, O, is the robust upper bound in the sense of Berger (1984, 1994), Berger and Delampady (1987), and Berger and Sellke (1987). That is, according to Theorem 1 of Berger and Sellke (1987), the upper bound for the most general set of priors $GA = \{\text{all distributions}\}$ is $O = \sup_\theta O_z$ as given by equation (3).

To summarize, instead of one precise alternative, we can consider a (possibly infinite) set of alternative hypotheses that partitions the entire parameter space. We evaluate $H_0 : |\theta - \theta_0| < \varepsilon$ against each and every other possible neighborhood of θ, $|\theta - z| < \varepsilon$ represented by the partition $\mathbb{P}_z$ of the parameter space. Over the entire set Θ, the value of θ that gives the strongest evidence against H_0 will be $\bar{\theta} = \arg\max_\theta p(\theta|D)$. In theory then, we consider every possible alternative value of θ and find that the posterior odds against H_0 are greatest for $\bar{\theta}$. In practice, we need only calculate $O = \sup_\theta p(\theta|D)/p(\theta_0|D)$ as given in equation (3). For a noninformative prior such that $\pi(\theta = \bar{\theta})/\pi(\theta = \theta_0) \approx 1$, equation (3) will be the objective posterior odds, providing the maximum evidence against

the null hypothesis according to the data and any background prior information incorporated in the likelihood, $p(D|\theta)$.

It can be argued that the scientific method requires the identification of alternative hypotheses and comparison of the proposed hypothesis with these alternatives; it is a survival of the fittest hypothesis competition. To not define the alternative hypotheses is to be nonscientific.

> For example, if you ask a scientist, 'How well did the Zilch experiment support the Wilson theory?', you may get an answer like this: 'Well, if you had asked me last week I would have said that it supports the Wilson theory very handsomely; Zilch's experimental points lie much closer to Wilson's predictions than to Watson's. But just yesterday I learned that this fellow Woffson has a new theory based on more plausible assumptions, and his curve goes right through the experimental points. So now I'm afraid I have to say that the Zilch experiment pretty well demolishes the Wilson theory.' [Jaynes, 2003, p.135]

In practice, there is usually a set of well-defined alternative hypotheses in mind. Should we reject H_0, we will take one or more of the alternatives as our working hypothesis until something better comes along.

That this testing procedure resolves the JLB paradox is demonstrated in the next section, and some examples illustrating the use of this procedure in comparison with the standard approach are provided below. But first let's apply this procedure to the example given in the introduction. From equation (3), we obtain objective posterior odds $O = 87.26$, so over $87:1$ *against* the null hypothesis. Recall that the standard Bayes factor is 8.27 *in favor* of the null, whereas the two sided p-value $= 0.0028$, and a 99% frequentist confidence interval, which is equivalent to a 0.99 equal-tailed Bayesian probability interval for $\theta = (0.20023, 0.20308)$. The objective posterior odds therefore matches the conclusion indicated by p-value and probability interval calculations.

3 The Jeffreys-Lindley-Bartlett (JLB) paradox resolved.

Suppose $x \sim N(\theta, \sigma^2)$, σ^2 known, and we wish to test $H_0 : \Theta = \theta_0$ vs. $H_1 : \Theta \neq \theta_0$. Lindley (1957) assigns the 'slab and spike' prior $\pi(H_0) = q$, for some fixed

$q, 0 < q < 1$, and $g(\theta|H_1) = kI_N$, where I_N is an indicator function, $I_N = 1$ if $|\theta| < N, I_N = 0$ otherwise, with N large enough to contain θ_0 and $\bar{x}$, the observed sample mean of x, well within the interval, i.e. $g(\theta|H_1)$ is uniform over a wide interval of possible values for θ. The standard posterior odds, B_{01}, evaluated with the null hypothesis in the numerator, is then

$$B_{01} = (q/(1-q))(\sigma^{-n}\sqrt{N}/\sqrt{2\pi})\exp[-n(\bar{x}-\theta_0)^2/2\sigma^2] \tag{4}$$

Lindley notes that as $N \to \infty, B_{01} \to \infty$. Bartlett (1957) discovered a similar paradox: for the prior $g(\theta|H_1) = N(0, K)$, as $K \to \infty, B_{01} \to \infty$. Jeffreys (1939) had adopted a Cauchy prior distribution and obtained this same paradoxical result.

To see the generality of the JLB paradox, consider the case of testing $H_0 : \Theta = \theta_0$ vs. $H_1 : \Theta \neq \theta_0$, or more correctly $H_0 : |\theta - \theta_0| < \varepsilon$ vs. $H_0 : |\theta - \theta_0| \geq \varepsilon$, for an arbitrary likelihood function $p(x|\theta)$. This leads to the standard posterior odds in favor of H_0,

$$B_{01} = \pi_0 p(x|\Theta = \theta_0)/m_g(x), \tag{5}$$

where $\pi_0 = p(H_0) = p(\Theta = \theta_0)$ is the prior probability assigned to the null hypothesis (typically 0.5 in the literature), and $m_g(x) = (1-\pi_0)\int_\Theta g(\theta|H_1)d\theta$, and $g(\theta|H_1)$ is the prior density for θ conditional on H_1 being true. Any sensible specification of g, for example the usual improper uniform prior or a normal prior with large variance, will lead to paradoxical problems of the JLB type. Parameters in the numerator and denominator of (5) do not cancel because of the integral in the denominator, which is there because of the ill-defined alternative hypothesis. Since the JLB paradox generally arises when the prior variance parameter appears outside the kernel of the posterior density, we can state the problem as follows.

Many of the posterior density functions used in practice can, for expository purposes, be written solely as a function of the prior variance, leading to the following form,

$$p(\theta|V_\theta) = c(V_\theta)f(\theta|V_\theta), \tag{6}$$

where θ is an unknown quantity, V_0 is the prior variance, $f(\theta|V_0)$ is the kernel of the posterior density, $c(V_0)$ is the normalizing constant, and all conditioning factors other than the prior variance V_0, i.e. the data, other prior parameters, and parameters of the likelihood function, are suppressed to emphasize the role

of the prior variance. As can be seen from (4), the posterior leading to the Lindley paradox can be written in the form of (6). This is also true for the Normal and Cauchy examples examined by Bartlett and Jeffreys.

The following two conditions are sufficient for the JLB paradox to occur: (i) the posterior density function for θ can be written in the form of equation (6), and (ii) the hypothesis test is precise versus imprecise, such as $H_0 : \Theta = \theta_0$ vs. $H_1 : \Theta \neq \theta_0$. If condition (i) holds, since $\int_\Theta p(\theta|V_0)d\theta = 1$,

$$\int_\Theta f(\theta|V_0)d\theta = c^{-1}(V_0).$$

If condition (ii) holds, since for a continuous distribution $p(\Theta = \theta_0|V_0) = 0$, we assign a prior probability distribution, $p(\theta)$, typically giving equal weight to each hypothesis and hence nonzero mass, π_0 to the point $\Theta = \theta_0$,

$$p(\theta) = \begin{cases} \pi_0, & \Theta = \theta_0 \\ (1 - \pi_0)g(\theta), & \Theta \neq \theta_0. \end{cases} \tag{7}$$

with $g(\theta)$ a continuous prior density. Standard practice is to set $\pi_0 = 1/2$ (see Gomez-Villegas *et al.*, 2002, for examination of alternative values for π_0). Therefore, under H_0, $p(\Theta = \theta_0|V_0) = 1/2c(V_0)f(\Theta = \theta_0|V_0)$, and under H_1, $p(\Theta \neq \theta_0|V_0) = 1/2g()c(V_0)c^{-1}(V_0)$. The standard posterior odds in favor of H_0 is then,

$$B = \frac{c(V_0)}{g(\theta)} f(\Theta = \theta_0|V_0) \tag{8}$$

Since B contains $c(V_0)$, it exhibits the JLB paradox.

If instead of condition (ii) the hypothesis testing procedure examines precise versus precise hypotheses, $H_0 : \Theta = \theta_0$ vs. $H_z : \Theta = z$, then the posterior odds in favor of H_0 is

$$O_{0z} = \frac{c(V_0)f(\Theta = \theta_0|V_0)}{c(V_0)f(\Theta = z|V_0)} = \frac{f(\Theta = \theta_0|V_0)}{f(\Theta = z|V_0)} \tag{9}$$

which does not involve $c(V_0)$ and so does not exhibit the JLB paradox.

Examining (8) indicates the usual approach to resolving the JLB paradox of attempting to avoid condition (i). This has led to a large literature seeking prior distributions for the alternative hypothesis that contain a component similar to $c^{-1}(V_0)$, so as to make the posterior odds less sensitive to the prior variance. This generally results in a prior that is either very informative or is dependent on the null hypothesis to be tested, as is the case with (7), or both.

The alternative proposed herein is to avoid condition (ii) by adopting a set of precise alternative hypotheses, so that (9) can be used and the paradox is resolved. In particular, for the Lindley (1957) paradox given by (4), we partition the parameter space, replace the ill-specified $H_1 : \Theta = \theta_0$ with the set of alternative hypothesis $H_z : \Theta = z, z \in \mathbb{P}_z$, and obtain prior density values $g(\Theta = \theta_0)$ and $g(\Theta = z)$ from the prior distribution on θ. Examining all possible alternative hypotheses, the maximum odds ratio against H_0 is, instead of equation (4),

$$O = \exp\left(\frac{n(\bar{x} - \theta_0)^2}{2\sigma^2}\right) \frac{g(\Theta = \bar{x})}{g(\Theta = \theta_0)} \tag{10}$$

where $\bar{x}$ is the sample mean. In this case, the closer $\bar{x}$ is to θ_0, the closer O is to unity, and the further $\bar{x}$ is from θ_0 the larger O will be. With a prior giving equal prior weight to each hypothesis, $g(\Theta = \bar{x}) = g(\Theta = \theta_0)$, the objective Bayes factor implies odds $O : 1$ against H_0 and the result is independent of the prior variance, resolving the JLB paradox.

4 Gaussian unknown mean, known variance

Berger and Delampady (1987) consider the same problem as Lindley (1957): suppose $\bar{x} \sim N(\theta, \sigma^2/n)$, σ^2 known. To test $H_0 : \Theta = \theta_0$ vs. $H_1 : \Theta \neq \theta_0$, they derive the compare a Bayes factor with the standard p-value $= 2[1\Phi(|z|)]$, where $z = \sqrt{n}(\bar{x} - \theta_0)/\sigma$, and Φ is the standard normal cumulative distribution function. Berger and Delampady set the prior probability for the null hypothesis, $p(H_0) = p(\Theta = \theta) = 1/2$, set the prior precision $\tau = \sigma$ in order to avoid some of the problems arising from the JLB paradox, and derive the Bayes factor,

$$B = \sqrt{(1 + \rho^{-2})} \exp\left(-\frac{z^2}{2(1 + \rho^2)}\right), \tag{11}$$

where $\rho = \sigma\sqrt{n\tau}$. Note that B depends on n and so still exhibits similar paradoxical behavior to that seen in the JLB paradox. Suppose H_0 is true, then as $n \to \infty, t \to 0$ since $\bar{x}$ is a consistent estimator of θ, so $B \to 1$. That is, as we obtain more data in support of the null hypothesis, B provides weaker evidence in support of the null. Further, since B also depends on the prior precision, the JLB paradox holds and as τ increases, $B \to 1$ regardless of the evidence provided by the data.

Now consider the same problem partitioning the parameter space into a set of alternative hypotheses and calculating the maximum evidence against the

null from this set, given by $H_m : |\theta - \bar{x}| < \varepsilon$. The posterior odds is then $P(|\theta - \bar{x}| < \varepsilon|x)/P(|\theta_0| < \varepsilon|x)$. As $\varepsilon \to 0$ this becomes $p(\bar{x}|x)/p(\theta_0)|x) = \exp(z^2/2)g(\theta_0)/g(\bar{x})$. For an uninformative prior $g(\theta_0) = g(\bar{x})$, so the objective posterior odds ratio is,

$$O = p(\bar{x}|x)/p(\theta_0|x) = \exp(z^2/2), \tag{12}$$

which is independent of the prior variance, and n only has influence through z.

Table 1: Objective posterior odds and Bayes factor

z	p-value	O	B $n = 10$	B $n = 20$	B $n = 50$	B $n = 100$
1.645	0.10	3.9	0.89	1.27	1.86	2.57
1.96	0.05	6.8	0.59	0.72	1.08	1.50
2.576	0.01	27.6	0.16	0.19	0.28	0.27
3.291	0.001	224.8	0.02	0.03	0.03	0.05

Table 1 provides values for z, the p-value, objective posterior odds against the H_0, O, given by (12), and Bayes factor, B, given by (11), which gives odds in favor of the null. When z is 1.96 and the p-value is 0.05, the objective odds are 6.8:1 against H_0, whereas B provides weaker evidence against the null as n increases. For $n \geq 50$, B indicates evidence in favor of H_0 when the p-value is 0.05, a 0.95 credible interval does not contain 1.96, and the odds are 6.8:1 against H_0.

5 The unknown variance case: Regression coefficients.

One of the most common uses of precise hypothesis testing in practice is to test the statistical significance of a regression parameter (i.e. difference from zero). Consider the simple regression model, $y = X\beta + u$, $u \sim N(0, \sigma^2 I_n)$, where y and u are $(n \times 1)$ vectors, X is an $(n \times k)$ matrix, and β is a $(k \times 1)$ vector of regression coefficients. If $X\beta$ is replaced with a constant, μ, this is a simple extension of the problem above to the case of unknown variance.

Adopting the standard Jeffreys diffuse prior over the parameter space $p(\beta, \sigma^2) \propto 1/\sigma$, the marginal posterior distribution for an individual regression parameter,

$\beta_j \sim t(\hat{\beta}, s^2(X'X)^{-1}, v)$ with $v = n-k$, $\hat{\beta} = (X'X)^{-1}X'y$, β_j is the jth element of $\hat{\beta}$ and $(X'X)^{-1}_j$ is the jth diagonal element of $(X'X)^{-1}$.

The standard Bayes factor for evaluating the hypotheses $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ is given as follows (Koop, *et al.*, 2007, Zhou & Guan, 2017). Rewriting the null and alternative hypotheses in the equivalent model selection form,

$$H_0 : y = Z\beta_0 + u_0,$$

$$H_1 : y = X\beta + u,$$

where Z is the subset of X omitting the covariate associated with β_j, and β_0 is the subset of β omitting β_j, so that the model in H_0 is equivalent to the model in H_1 with the additional constraint $\beta_j = 0$. The Bayes factor is then,

$$BF = \left(\frac{|Z'Z|}{|X'X|}\right)^{-1/2} \left(\frac{(y - Z\hat{\beta}_0)'(y - Z\hat{\beta}_0)}{(y - X\hat{\beta})'(y - X\hat{\beta})}\right)^{-n/2}. \tag{13}$$

As is well known in the literature, this exhibits the JLB paradox: as $n \to \infty$, $B \to 0$, so the procedure is not consistent: the more data available, the more likely we are to reject the null hypothesis regardless of the truth. The standard Bayesian approach to this problem is to adopt the Zellner-g prior, which somewhat mitigates the practical impact of the JLB paradox, but does not eliminate the problem (Rouder *et al.*, 2009, Zhou and Guan, 2017).

If instead we test $H_0 : \beta_j = 0$ vs. $H_i : \beta_j = \beta_{jz}$, $\forall \beta_{jz} \neq 0$, the objective posterior odds is,

$$O = \left(1 + \frac{t^2}{v}\right)^{(v+1)/2}, \tag{14}$$

where $t = \hat{\beta}_j / \sqrt{s^2(X'X)^{-1}_j}$ is the usual t-statistic, so there is a one-to-one correspondence between O and the standard t-test. Table 2 illustrates this relationship for two-tailed p-values of 0.1, 0.05 and 0.01. As the degrees of freedom, v increase, O converges to the values given in Table 1 for the Gaussian distribution, as does the t-statistic, so that in this simple case, with an uninformative prior the Bayesian and frequentist testing matches if we employ O.

Table 2: Fixed p-value objective odds and t-statistic

	p-value=0.1		p-value=0.05		p-value=0.01	
v	Odds	t-value	Odds	t-value	Odds	t-value
5	5.5	1.94	11.27	2.447	64.7	3.71
10	4.7	1.80	8.931	2.201	43.7	3.11
20	4.3	1.72	7.845	2.080	35.1	2.83
50	4.1	1.70	7.498	2.040	32.4	2.74
100	4.0	1.66	7.225	2.008	30.4	2.68
500	3.9	1.65	7.024	1.984	29.0	2.63
1000	3.9	1.65	6.865	1.965	27.9	2.59
∞	3.9	1.65	6.829	1.960	27.6	2.58

In Table 2, a p-value of 0.1 corresponds with a t-value ≈ 1.7 and odds ≈ 4, a p-value of 0.05 corresponds to a t-value ≈ 2 and odds ≈ 7, and a p-value of 0.01 corresponds with a t-value ≈ 2.7 and odds ≈ 30. Suggesting rule of thumb critical odds values of 4, 7 and 30 for weak, substantial and strong evidence against the null hypothesis, which are quite similar to the well known Jeffreys (1939) guidelines (see Berger, 2008, Greenberg, 2013).

For a particular value of t and v determined by observed data, we can also examine the odds as a function of the values in the alternative hypothesis. The posterior odds for this are given by,

$$O(\beta) = \left(1 + \frac{t(\beta = 0)^2}{v}\right)^{(v+1)/2} / \left(1 + \frac{t(\beta)^2}{v}\right)^{(v+1)/2}, \tag{15}$$

where $t(\beta) = (\beta - \hat{\beta}_j)/\sqrt{s^2(X'X)^{-1}_j}$. Figure 1 provides a plot of $O(\beta)$ for $\hat{\beta}_j = 1.0$, $\text{var}(\beta_j) = 0.16, \nu = 30$.

6 When p-values and odds do not match: a clinical trial example

Consider the following example from Johnson (2013), involving a clinical trial of a new drug. The current treatment (or placebo) is known (with very close to certainty) to have a probability of success of 0.25. A new additional treatment is given to $n = 30$ subjects and $s = 12$ of the 30 are cured. Does the new treatment improve upon the current treatment (or placebo)?

The sampling distribution (Fig. 2) is a binomial $= \binom{n}{s}\theta^s(1-\theta)^{n-s}$, $\theta = 0.25$. The p-value for observed successes of 12 is then prob$(s \geq 12|\theta = 0.25) = 0.0215$ and so is statistically significant at the 5% level, rejected the null hypothesis.

[Figures 2 and 3 here]

Employing a uniform prior, $\theta \sim U[0,1]$, the posterior in this case is a Beta$(s+1, n-s+1)$ distribution. The posterior density with $s = 12, n = 30$, along with the histogram for 10,000 pseudo-random Monte Carlo (MC) draws from the density, are shown in Fig. 3. A 0.95 highest posterior density (HPD) interval from this posterior is (0.244, 0.580), which contains 0.25 suggesting there is not (quite) enough evidence to reject the null. The objective posterior odds, given by

$$O = \frac{\bar{\theta}^s(1-\bar{\theta})^{n-s}}{\theta_0^s(1-\theta_0)^{n-s}}, \tag{16}$$

equals 5.07:1 (matching the computation in Johnson, 2013, equation (1)), providing only weak evidence against the null hypothesis and matching the conclusion from the posterior interval. The standard Bayes factor,

$$B = \frac{\binom{n}{s}\theta_0^s(1-\theta_0)^{n-s}}{\int_0^1 \binom{n}{s}\theta^s(1-\theta)^{n-s}d\theta} = \binom{n}{s}\theta_0^s(1-\theta_0)^{n-s}, \tag{17}$$

equals 34.4:1 odds against the null hypothesis, suggesting strong evidence against the null.

A common misconception in the literature is that a posterior probability for a precise null can be calculated by solving the standard Bayes factor for the value in the numerator. This is not a legitimate calculation as the value in the numerator must equal zero. The probability that $\theta = 0.25$ *exactly* is zero, since that is a point on a continuum. It is only by passing to the limit *after* forming the ratio that allows a valid odds ratio to be computed for a precise hypothesis. We can however, compute something that could be labeled a Bayesian p-value, i.e. the posterior probability $p(\theta \leq 0.25|n = 20, s = 12) = 0.029$, matching the posterior interval calculation, but somewhat larger than the frequentest p-value (because one calculation employs a continuous distribution, while the other uses a discrete distribution, restricting the precision of the calculation).

Table 3: Evidence against $H_0 : \theta = 0.25$ for s successes in 30 trials

s	SBF	OBF	p-value*
1	559.96	209.50	0.9981
2	115.86	32.45	0.9894
3	37.24	8.79	0.9624
4	16.55	3.47	0.9023
5	9.55	1.83	0.7966
6	6.87	1.23	0.6505
7	6.02	1.02	0.4838
8	6.28	1.02	0.3251
9	7.70	1.21	0.1957
10	11.01	1.68	0.1050
11	18.16	2.72	0.0503
12	34.41	5.07	0.0215
13	74.55	10.86	0.0083
14	184.17	26.66	0.0028
15	517.99	74.83	0.0009

* one-sided.

If instead we observe $s = 7$ successes in 30 trials, the frequentist p-value is 0.484, the posterior mean is 0.233, with the true value well within the 0.1 posterior probability interval, and the objective odds are only 1.02:1 against the null, so all are in agreement that the null hypothesis should not be rejected. The standard Bayes factor however, still suggests the evidence is 6.01:1 against the null hypothesis and, in fact, always gives odds against the null hypothesis for any observed value of s, as illustrated in Table 3. The standard fix for this is to adopt an informative prior that puts considerable probability on the null hypothesis being true, regardless of prior knowledge, and contradicting the prior used for inference. So SBF always gives odds against H_0 unless an informative prior is adopted that strongly favors the null hypothesis, shifting the odds in favor of H_0.

An alternative version of the standard Bayes factor is thus to employ a Beta(a, b) prior distribution (Niemi, 2018), giving,

$$BF(H_0 : H_1) = \frac{\theta_0 beta(a, b)}{beta(a + s, b + n - s)}, \tag{18}$$

where $beta(a, b)$ is the beta function, which is the normalizing constant of the

Beta density. Table 4 provides results for this version of SBF with a few different prior parameter values, all of which are relatively uninformative. While the posterior density remains very similar for all of these prior choices (Fig. 4), the Bayes factor is very sensitive to the choice of prior (Fig. 5). The priors are illustrated in Fig. 6. For $s = 12, n = 30$ the SBF in Table 4 is at most 1.6:1 against the null hypothesis, failing to reject the null at any sensible critical value. For large and small values of s, the SBF is very sensitive to the choice of prior, e.g. for $s = 15$ it ranges between 11:1 and 24:1 against the null hypothesis.

Table 4: SBF for various Beta(a, b) priors

	a = b = 0.5		a = b = 1		a = b = 2	
s	SBF	1/SBF	SBF	1/SBF	SBF	1/SBF
1	0.0342	29.207	0.0554	18.0634	0.1624	6.158
2	0.2168	4.6116	0.2676	3.7373	0.5413	1.8474
3	0.7951	1.2577	0.8325	1.2013	1.3081	0.7644
4	2.0067	0.4983	1.873	0.5339	2.4419	0.4095
5	3.7904	0.2638	3.2466	0.308	3.6628	0.273
6	5.6281	0.1777	4.5091	0.2218	4.5349	0.2205
7	6.7826	0.1474	5.1533	0.1941	4.7239	0.2117
8	6.7826	0.1474	4.9386	0.2025	4.199	0.2382
9	5.7187	0.1749	4.024	0.2485	3.2192	0.3106
10	4.1134	0.2431	2.8168	0.355	2.1462	0.466
11	2.5464	0.3927	1.7072	0.5858	1.2519	0.7988
12	1.3655	0.7324	0.901	1.1099	0.642	1.5576
13	0.6372	1.5693	0.4158	2.4047	0.2904	3.4431
14	0.2596	3.852	0.1683	5.9411	0.1162	8.6078
15	0.0925	10.81	0.0598	16.7093	0.0411	24.3

[Figures 4-6 here]

As Johnson (2013) points out, "in this case, the null hypothesis is rejected at the 5% level of significance even though the data support it." While it is not clear that the data actually support the null hypothesis in this case, since it is close the lower bound of a 0.95 probability interval and the objective odds are 5:1 against, the evidence against the null is much weaker than suggested by the frequentist significance test. The standard Bayes factor does not perform

well, exhibiting similar problems as with the JLB paradox, stemming from the poorly defined alternative hypothesis. Again, the correction for these problems is not to be found in developing unjustifiable informative priors that are not based on prior background information, but in reconsidering the definition of the alternative hypothesis.

7 HPD intervals and p-values as viable alternatives?

If p-values and odds are the same, why not just continue to use p-values, or use probability intervals instead? For testing only one parameter, intervals will work and lead to the same conclusions as the objective odds. Further, with a parameter that is bounded, both the odds and probability intervals can be employed because you can reject the null if the interval is strictly to one side of the value in the null hypothesis. However, p-values cannot be computed with a bounded parameter if the null hypothesis is at the boundary. Interval estimates work in this situation and are complementary to the odds ratio.

Testing a precise hypothesis that is on the boundary of the parameter space is a common situation. For example, waiting times, prices and interest rates may all have a minimum possible value of zero, and it is often of interest to test whether the minimum value in a particular case is greater than zero or not. In general, given data, y, that is strictly nonnegative, it is not possible to test $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$ with $\theta \geq 0$ using p-values, confidence or probability intervals, since we can never observe values below the lower bound of zero. If y is continuous, since $y \geq 0$, the p-value for a critical value of 0 is $\int_0^0 p(y|\theta)d\theta = 0$.

HPD intervals will also work with multimodal distributions, and when the distribution is skewed. They become problematic for joint hypothesis testing. Using confidence or probability intervals for testing suffers from the curse of dimensionality. For two parameters, probability ellipses must be evaluated. Every additional parameter in the null hypothesis adds a dimension to the probability region. The objective posterior odds for a joint hypothesis testing, involving more than one parameter, is a straightforward computation from an MC (or MCMC) sample from the posterior distribution. Using Rao-Blackwellization, an accurate evaluation of the posterior odds can be computed with one loop through the MC sample, regardless of the number of parameters involved (either in the null hypothesis, or nuisance parameters).

For example, the posterior odds ratio for testing $H_0 : R\theta = r$, where θ is a vector of parameters, and R and r define the linear restrictions on these parameters under the null hypothesis, is,

$$O = \frac{\int p(\bar{\theta}_R, \phi | R\theta - r = 0, D) d\phi}{\int p(\bar{\theta}, \phi | D) d\phi}, \tag{19}$$

where ϕ is a vector of nuisance parameters, $\bar{\theta}_R$ is a vector of modes of the joint posterior with the restrictions imposed, and $\bar{\theta}_R$ is a vector of modes for the unrestricted joint posterior density. This is computed as,

$$O = \frac{\sum_{i=1}^{M} p(\bar{\theta}_R, \phi^{(i)} | R\theta - r = 0, D)}{\sum_{i=1}^{M} p(\bar{\theta}, \phi^{(i)} | D)}, \tag{20}$$

where the sum is over a MC sample of size M, which can be made arbitrarily large to improve numerical accuracy (see Mills and Namavari, 2017, for a detailed examination of joint hypothesis testing).

8 Discussion

There are well known issues with frequentist p-values (see for example, Gelman and Carlin, 2017, Berry, 2017, and references therein). In the frequentist approach, the null hypothesis is assumed true to derive a test procedure, then the null can be rejected using this procedure, but logically that means the test procedure is no longer valid because it is conditional on a hypothesis that has been rejected (Jaynes, 2003, p.524). Further, as Jeffreys (1939) pointed out, in this case the null hypothesis is rejected based on values of the test statistic that were not observed (the tail area of the test statistic distribution). Even accepting all of this, two-tailed p-values are typically used, which includes the area in the tail of the distribution opposite to the observed test statistic value (e.g. we observe a positive test statistic value, but include the area in the negative tail of the distribution in our p-value calculation, which is the opposite side of the distribution relative to the value in the null hypothesis. So we can reject a null hypothesis of zero based on observing a positive test statistic by using the area in the negative tail of the distribution.

In the standard Bayesian Bayes factor approach, similar logical contradictions occur. The resulting Bayes factor is generally sensitive to the prior such that the less informative the prior, the stronger the evidence in favor of the null hypothesis regardless of the evidence. Probabilities are assigned to a point on a

continuum, which by the axioms of probability must have probability zero, then these assignments are used to compute a posterior probability for the point on a continuum and used as evidence in favor of a hypothesis.

However, in many standard hypothesis testing problems, despite the difficulty of interpretation, p-values work reasonably well. Any proposed testing procedure should perform as well in these situations. The standard Bayesian approach does not, whereas the objective testing procedure, based on a multiplicity of alternative hypotheses, does, and has been validated in a variety of simulation studies and practical applications.

When frequentist and Bayesian inference matches (i.e. the posterior density provides similar estimates and intervals to the frequentist values), then we should expect Bayesian and frequentist testing to match. If probability and confidence intervals match, we would expect hypothesis testing to match p-values in terms of indicated strength of evidence against the null hypothesis. When the intervals do not match, then we would expect Bayesian and frequentist testing to lead to different conclusions; one's that match the respective inferences. This is the case with the objective posterior odds, whereas the current Bayes factor approach leads to inconsistencies between intervals and testing results, even in standard problems.

While the objective posterior odds lead to the same conclusions as the frequentist p-value approach in situations where that approach obviously provides a sensible answer, the objective odds also provide an intuitively understandable value: "odds against the null hypothesis being true". The interpretation of the p-value is generally misleading: a p-value of 0.05 does not mean the probability of the null being true is ≤ 0.05 (see Gelman and Stern, 2017, and Berry, 2017).

It should also be noted that the objective odds are identical to the standard Bayes factor for composite vs. composite hypotheses, both of which closely match frequentist p-values in standard problems.

Lastly, a credible interval says there is probability $(1-\alpha)$ that the true value is in the interval, or probability α that the true value lies outside the range. This is important supplemental information when evaluating a hypothesis, as is visual inspection of the posterior density. Good practice when evaluating a hypothesis should not consist of considering only one perspective, such as the odds ratio or an interval estimate. All the evidence must be considered, and a formal test based on odds or an interval is only part of the case.

9 Conclusion

The objective Bayesian testing procedure proposed herein addresses several shortcomings with current Bayesian and frequentist testing procedures. The JLB paradox is resolved. The same priors that are valid for inference are resurrected as usable and valid for precise hypothesis testing, and the prior is specified independently of any tests to be performed. There is no need to violate the axioms of probability by assigning nonzero mass to a particular point on a continuous distribution. The posterior odds obtained can be interpreted as a likelihood ratio, and so has a frequentist interpretation as the uniformly most powerful test. The proposed testing procedure has a clear interpretation as the odds against the null hypothesis, so test results are easy to interpret in practice. The test procedure is fully Bayesian and so satisfies the likelihood principle. Bayesian robustness methods readily apply to the procedure. The advantages enumerated above, along with validation experiments using simulated data, suggest that this testing procedure provides a viable approach to objective Bayesian precise hypothesis testing.

The proposed testing procedure has been applied to comparison of means (Strawn et al., 2018*a*), ANOVA testing (Mills and Namavari, 2016), meta-analysis (Strawn et al. 2018*b*) unit root and cointegration testing (Mills, 2013, Mills and Namavari, 2016), Granger causality testing (Mills et al., 2017) and predictive model comparison (Cornwall and Mills, 2017). In all of these applications the procedure performs as well or better than frequentist methods, and does not exhibit any of the shortcomings of standard Bayes factors.

10 References

Berger, J. "A comparison of testing methodologies." PHYSTAT LHC Workshop on Statistical Issues for LHC Physics, PHYSTAT 2007 - Proceedings (December 1, 2008): 8-19.

Bartlett, M. S. (1957) A Comment on D. V. Lindleys Statistical Paradox. Biometrika, 44, 533-534.

Basu, S. (1996a) Bayesian hypothesis testing using posterior density ratios, Statistics and Probability Letters, 30, 79-86.

Basu, S. (1996b) A new look at Bayesian point null hypothesis testing, Sankhya, Series A, 58, 292-310.

Berger, J. O. (1984) The robust Bayesian viewpoint, Robustness of Bayesian

Analysis (J. B. Kadane, ed.). Amsterdam: North-Holland, 63124 (with discussion).

Berger, J. O. (1985) Statistical Decision Theory and Bayesian Analysis. Second Edition, Springer-Verlag, New York.

Berger, J. O. (1994) An Overview of Robust Bayesian Analysis, Test, 3, 5-124 (with discussion).

Berger, J.O. (2006) The Case for Objective Bayesian Analysis. Bayesian Analysis, 1:3, 385-402.

Berger, J. O. and M. Delampady (1987) Testing Precise Hypotheses. Statistical Science, 2, 317- 352.

Berger, J. O. and Perichi, L.R. (2001) Objective Bayesian methods for model selection: Introduction and comparison. In: Lahiri, P. (ed.), IMS Lecture Notes - Monograph Series, 38, 135-207.

Berger, J. O. and Perichi, L.R. (2004) Training Samples in objective Bayesian model selection. Annals of Statistics, 32:3, 841-869.

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, 112-122.

Berry, D. (2017) A p-Value to Die for. Journal of the American Statistical Association, 112:519, 895-897.

Cornwall, G. and Mills, J. (2017) Prediction Based Model Selection Criteria. University of Cincinnati.

Chib, S. (2001) Markov Chain Monte Carlo Methods: Computation and Inference, Handbook of Econometrics, 5, 3569-3649.

DeGroot, M. H. and M. J. Schervish (2002) Probability and Statistics, Third Edition. Addison-Wesley.

Edwards, W., Lindman, H. and Savage, L. (1963). Bayesian statistical inference for psychological research. Psychological Review, 70, 193242

Gelman, A. and Carlin, J. B. (2017) Some Natural Solutions to the p-Value Communication Problemand Why They Won't Work. Journal of the American Statistical Association, 112:519, 899-901.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (2004) Bayesian Data Analysis, Second Edition. Chapman and Hall.

George, E. I. and R. E. McCulloch (1993) Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88, 881-889.

Gomez-Villegas, M. A., P. Main and L. Sanz (2002) A Suitable Bayesian Approach in Testing Point Null Hypothesis: Some Examples Revisited, Com-

munications in Statistics, 31, 201-217.

Good, I. J. (1965) The Estimation of Probabilities: An Essay on Modern Bayesian Methods, MIT Press, Cambridge, MA.

Good, I. J. (1967) The Bayesian significance test for multinomial distributions, Journal of the Royal Statistical Society, Series B, 29, 399-431.

Good, I. J. (1976) On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. Annals of Statistics, 4, 1159-1189.

Greenberg, E. (2008) Introduction to Bayesian Econometrics, Cambridge University Press.

Ishwaran, H. and J. S. Rao (2005) Spike and slab variable selection: frequentist and Bayesian strategies, Annals of Statistics, 33, 730-773.

Jaynes, E. T. (2003) Probability: The Logic of Science. Cambridge University Press, Cambridge.

Jeffreys, H. (1939, 1961) Theory of Probability. Oxford University Press, Oxford. (1st ed. 1939, 2nd ed. 1961).

Johnson, V.E. (2013) Uniformly Most Powerful Bayesian Tests. Annals of Statistics, 41:1, 1716-1741.

Koop, G., Poirier, D.J., and Tobias, J. L. (2007) Bayesian Econometric Methods. Cambridge University Press.

Lindley, D. V. (1957) A Statistical Paradox. Biometrika, 44, 187-192.

Mills, J.A. (2015) Objective Bayesian Unit Root Testing. University of Cincinnati.

Mills, J.A. Cornwall, G, Sauley, B., Weng, H. (2018) Bayesian Predictive Granger Causality Testing. University of Cincinnati.

Mills, J.A. and Namavari, H. (2016) Objective Bayesian ANOVA Testing. University of Cincinnati

Mills, J.A. and Namavari, H. (2017) Residual Based Objective Bayesian Cointegration Testing. University of Cincinnati.

Perichi, L.R. (2005) Model selection and hypothesis testing based on objective probabilities and Bayes factors. Handbook of Statistics, 25, 115-149.

Rouder *et al.* (2009) Bayesian t tests for accepting and rejecting the null hypothesis, Psychonomic Bulletin and Review, 16, p.231-

Stone, M. (1997) Discussion of Aitkin (1997). Statistics and Computing, 7, 263.264.

Strawn J.R., Mills J.A., Cornwall G.J., Mossman, S.A., Varney, S.T., Keeshin B.R., Croarkin, P.E. (2018*a*) Buspirone in Children and Adolescents with

Anxiety: A Review and Bayesian Analysis of Abandoned Randomized Controlled Trials. Journal of Child and Adolescent Psychopharmacology. 28:1, 2-9.

Strawn J.R., Mills, J.A., Sauley, B.A., Welge, J.A. (2018*b*) The Impact of Antidepressant Dose and Class on Treatment Response in Pedriatric Anxiety Disorders: A Meta-Analysis. Journal of the American Academy of Child and Adolescent Psychiatry, 57:4, 235-244.

Villa, C. and Walker, S. (2017) On the mathematics of the Jeffreys-Lindley paradox. Communications in Statistics: Theory and Methods, 46:24, 12290-12298.

Zellner, A. (1971) An Introduction to Bayesian Inference in Econometrics. Wiley & Sons, New York.

Zellner, A. (1984) Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results. In A. Zellner (1984) Basic Issue in Econometrics. University of Chicago Press, 275-305.

Zhou, Q. and Guan, Y. (2017) On the Null Distribution of Bayes Factors in Linear Regression. Journal of the American Statistical Association, DOI: 10.1080/01621459.2017.1328361

Figure 1: Posterior odds as a function of β

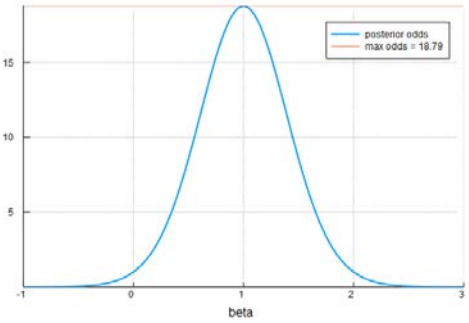

Figure 2: Binomial sampling density

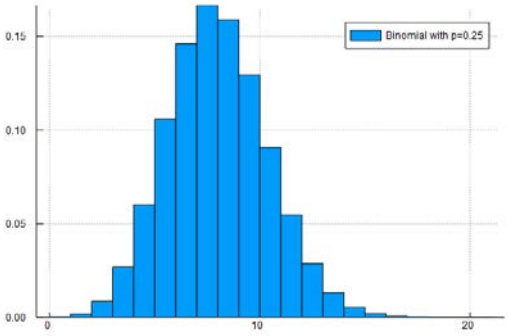

Figure 3: Beta posterior density

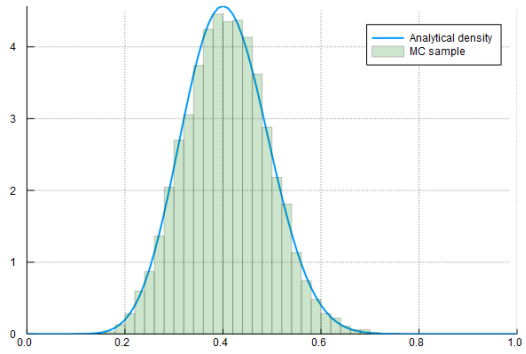

Figure 4: SBF

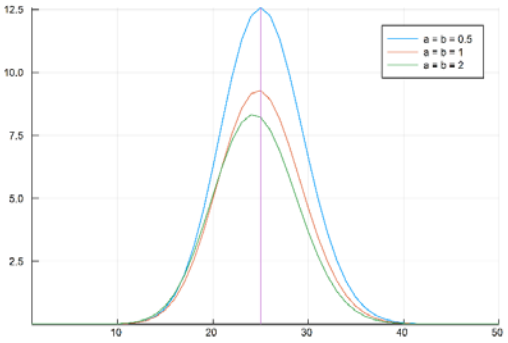

Figure 5: Posterior

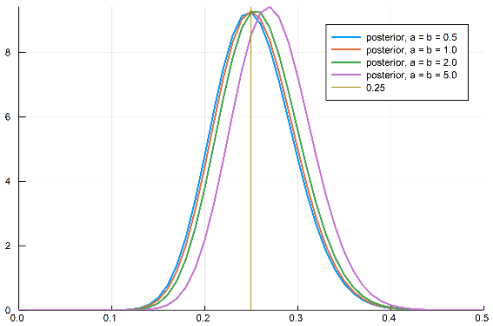

Figure 6: Priors