

Convergence of Posterior Odds¹

Richard A. Levine² George Casella³

Cornell University

BU-1260-MB

August 1994

Revised October 20, 1995

Appearing in the *Journal of Statistical Planning and Inference*

ABSTRACT

The problem of approximating an interval null or imprecise hypothesis test by a point null or precise hypothesis test under a Bayesian framework is considered. In the literature, some of the methods for solving this problem have used the Bayes factor for testing a point null and justified it as an approximation to the interval null. However, many authors recommend evaluating tests through the posterior odds, a Bayesian measure of evidence against the null hypothesis. It is of interest then to determine whether similar results hold when using the posterior odds as the primary measure of evidence. For the prior distributions under which the approximation holds with respect to the Bayes factor, it is shown that the posterior odds for testing the point null hypothesis does not approximate the posterior odds for testing the interval null hypothesis. In fact, in order to obtain convergence of the posterior odds, a number of restrictive conditions need to be placed on the prior structure. Furthermore, under a non-symmetrical prior setup, neither the Bayes factor nor the posterior odds for testing the imprecise hypothesis converges to the Bayes factor or posterior odds respectively for testing the precise hypothesis. To rectify this dilemma, it is shown that constraints need to be placed on the priors. In both situations, the class of priors constructed to ensure convergence of the posterior odds are not practically useful, thus questioning, from a Bayesian perspective, the appropriateness of point null testing in a problem better represented by an interval null. The theories developed are also applied to an epidemiological data set from White, LaFaunce, and Mohammed (1989) in order to illustrate and study priors for which the point null hypothesis test approximates the interval null hypothesis test.

¹Key Words: Hypothesis testing, Bayes, Bayes Factors

AMS 1990 Subject Classification: Primary 62F15; Secondary 62A15

²Research supported by the Office of Naval Research Graduate Fellowship and NIH Training Grant No. ES07261

³Research supported by National Science Foundation Grant No. DMS9305547

1 Introduction

Imagine we own a factory that produces nuts and bolts. Every month we wish to determine whether the equipment is manufacturing nuts of the incorrect size in terms of diameter. Consequently, we take a random sample of nuts forged at the plant and measure the diameter of each. Assume the optimal diameter for these nuts is θ_0 . Hence, to check whether the machinery is molding nuts of optimal diameter, we can evaluate the point null or precise hypothesis test

$$H_0 : \theta = \theta_0 \text{ vs. } H_A : \theta \neq \theta_0 \quad (1)$$

that, say, the unknown mean diameter length of nuts produced, θ , is equal to the optimal diameter length, θ_0 .

However, even if the nuts are constructed with slightly too small or too large diameters, they will still fit snugly around the companion bolts. Hence, we may not want to worry about fixing a machine that produces nuts with diameters very close to the optimal. A more appropriate analysis may then include a test of the interval null or imprecise (as called by Verdinelli and Wasserman (1993)) hypothesis

$$H_0^\epsilon : |\theta - \theta_0| \leq \epsilon \text{ vs. } H_A^\epsilon : |\theta - \theta_0| > \epsilon, \quad (2)$$

where $\epsilon > 0$ is “small.” Unfortunately, it may be difficult to specify ϵ . How far off from optimal can our nut diameters be while still fitting tightly around the bolts? Intuitively, a small ϵ may lead one to conclude that H_0^ϵ in (2) is essentially the same as the point null hypothesis H_0 in (1). Hence, the precise hypothesis test might be a good approximation of the imprecise hypothesis test. Through this potential approximation we can avoid the difficulty of determining ϵ by using a point null hypothesis as a surrogate to model a situation better fit by an interval null hypothesis.

Let us formulate this question in a general statistical setting. We wish to choose between two hypotheses, denoted H_0 and H_A , about a parameter, say θ , from the space $\Theta = \mathbf{R}$. To this end, assume a random variable X with density $f(x | \theta)$, dependent on the unknown θ , is observed. Furthermore, suppose we have some information about θ prior to running the experiment. For example, in the nuts and bolts situation, we possess data on a particular machine from previous months. By incorporating this information into our analysis, we may be able to get a better evaluation of the machinery. In a Bayesian framework, this prior knowledge is typically formulated in terms of a probability distribution over Θ since the information can seldom be precisely determined (Berger (1985)). Upon collecting data, the prior opinions about a situation are updated and presented as a posterior distribution, the probability of θ given the data x . The posterior distribution can be used to decide which of H_0 or H_A is “correct” based on prior beliefs and data. To accomplish this task the posterior probability under H_0 , $P(\Theta_0 | x)$, and under H_A , $P(\Theta_A | x)$, are calculated. The odds in favor (against) H_0 can be determined by calculating the ratio $P(\Theta_0 | x)/P(\Theta_A | x)$, termed the *posterior odds* and denoted PO from here onward. For example, if $PO = k$, then the null hypothesis, H_0 , is k times more likely than the alternative hypothesis, H_A . This Bayesian measure of evidence is conceptually appealing because it is the odds in favor of one hypothesis given both prior information and data (Berger (1985) section 4.3.3).

Similar to the posterior odds, the prior odds, $P(\Theta_0)/P(\Theta_A)$, denoted PII, determines the odds in favor (against) H_0 prior to data collection. Another measure used to evaluate evidence in a hypothesis testing problem is the odds ratio PO / PII . This odds ratio is termed the *Bayes factor*, denoted BF, and is interpreted as the increased (decreased) odds in favor of H_0 upon collecting data as compared to the odds in favor of H_0 prior to data collection.

Getting back to the problem at hand, we are interested in determining whether the point null or precise hypothesis test (1) can approximate the interval null or imprecise hypothesis test (2). This problem can be reformulated as whether PO_ϵ , the posterior odds for testing the interval null hypothesis, converges to PO , the posterior odds for testing the precise hypothesis, as ϵ approaches zero. Analogously, we can ask under what framework will BF_ϵ converge to BF as ϵ approaches zero. Many papers including Berger and Delampady (1987), Berger and Sellke (1987), Delampady (1989), Dickey (1976), Gomez-Villegas and Sanchez-Manzano (1992), and Verdinelli and Wasserman (1993) consider this problem; but each discusses solutions in terms of only BF or bounds on BF and posterior probabilities over classes of prior distributions.

However, according to the hypothesis testing methodology established in texts by Berger (1985), Casella and Berger (1990), DeGroot (1970), Phillips (1974), Pollard (1986), and Robert (1994) to name a few, the primary Bayesian measures for evaluating evidence in a testing problem are the posterior probabilities or posterior odds. For applications of Bayesian hypothesis testing evaluated through posterior odds in practice see Bernstein et al. (1989), Brown and Klein (1986), Fienberg (1990), Press (1989, sections 7.6, 7.7, 8.4 and 8.6), Reckhow (1990), Reynolds (1982), and Zellner (1984, p. 296-297) as examples. As for the Bayes factor, it can be shown that BF is sometimes relatively insensitive to reasonable choices of prior and represents the impact of the data on posterior probabilities. Therefore, Berger (1985) suggests reporting BF and eliciting the reader to calculate personal posterior opinions by multiplying BF by his/her prior odds. We thus question whether the approximation of the imprecise test by the precise test is valid in terms of the posterior odds.

In this article we first show that PO_ϵ does not converge to PO as ϵ approaches zero under a standard prior framework. Consequently, for this commonly applied

prior structure, the point null test does not approximate the interval null test. We then develop classes of priors for a variety of situations under which the approximation does hold. However, a researcher may be reluctant to represent his/her prior opinions by an element from the appropriate class of priors because these functions have some undesirable properties. Hence, in terms of PO, the approximation of the interval null test by the point null test may not be valid from a practical viewpoint. Additionally, we study the correspondence between the convergences of PO and BF. We prove that convergence of BF is a necessary consequence of convergence of PO. In some circumstances with which an experimenter may be confronted we find the approximation is not even valid in terms of BF. Therefore, in general from the Bayesian perspective, a researcher may be unwilling or ill-advised to test a point null hypothesis in the place of an interval null hypothesis.

In Section 2, we will present the basic prior setup and relevant formulas necessary in the ensuing parts of the paper. Section 3 contains the main results concerning classes of priors under which the posterior odds and Bayes factors converge in various testing scenarios. In Section 4 we apply the results from Section 3 to a data set aimed at determining the optimal time postbreeding for a pregnancy examination in dairy cattle. Section 5 contains a summary of the paper and suggests alternatives for scientists when their prior beliefs situate them in a position where testing a point null is inappropriate as an approximation to the more realistic interval null hypothesis test.

2 Preliminaries

The general hypothesis testing setup in this paper is similar to that defined in Gomez-Villegas and Sanchez-Manzano (1992). Suppose a random variable X has

density $f(x | \theta)$ where θ is an unknown parameter from the space $\Theta = \mathbf{R}$. Let $\theta_0 \in \Theta$ and assume $f(x | \theta)$ is continuous at $\theta = \theta_0$. The two hypothesis testing situations are defined by (1) and (2). Without loss of generality assume $\theta_0 = 0$ thus simplifying subsequent calculations.

In a Bayesian framework, for the testing problem (1), it is typical to consider a mixed prior distribution P which places mass $\pi_0 \in (0, 1)$ on the point $\{\theta = 0\}$ and spreads the remaining $(1 - \pi_0)$ probability on $\Theta \setminus \{0\}$ according to the continuous density $\pi(\theta)$. Mathematically, the prior distribution, P , can be expressed by

$$P(B) = (1 - \pi_0) \int_B \pi(\theta) d\theta + \pi_0 I(0 \in B), \forall B \in \mathcal{B}, \quad (3)$$

where \mathcal{B} is the collection of Borel sets of \mathbf{R} and $I(A)$ is the indicator function of the set A . For the testing problem (2), consider the same prior density $\pi(\theta)$ as used for problem (1), but over the entire parameter space Θ . The same prior density is used for testing H_0 and H_0^ϵ because the main goal is to determine when a point null hypothesis test could be used to approximate an interval null hypothesis test. Therefore, the prior information should be incorporated into both testing situations in a similar manner. To remain consistent with Gomez-Villegas and Sanchez-Manzano (1992) and for ease of presentation, let $I_\epsilon = \{\theta : |\theta| \leq \epsilon\}$ and I_ϵ^c be the complement of the set I_ϵ . Additionally, for theoretical completeness in the subsequent proofs, assume $f(x | \theta) \pi(\theta)$ is integrable (i.e., assume the marginal density of X , $m(x)$, is finite for all x).

Under this setup, the desired posterior odds for testing the point null hypothesis,

$$PO = \frac{P(\Theta_0)}{P(\Theta_A)} = \frac{\pi_0 f(x | \theta_0)}{(1 - \pi_0) \int_{\{\theta \neq \theta_0\}} f(x | \theta) \pi(\theta) d\theta} \quad (4)$$

and the posterior odds for testing the interval null hypothesis (denoted by a subscript “ ϵ ”),

$$PO_\epsilon = \frac{P(\Theta_0^\epsilon)}{P(\Theta_A^\epsilon)} = \frac{\int_{- \epsilon}^{\epsilon} f(x | \theta) \pi(\theta) d\theta}{\int_{I_\epsilon^c} f(x | \theta) \pi(\theta) d\theta} \quad (5)$$

can be easily calculated. Here, Θ_0^ϵ and Θ_A^ϵ denote the subsets of the parameter space under the null and alternative hypotheses of the imprecise test (2). Additionally, by noting that the corresponding definitions of the prior odds are

$$P\Pi = \frac{P(\Theta_0)}{P(\Theta_A)} = \frac{\pi_0}{1 - \pi_0}$$

and

$$P\Pi_\epsilon = \frac{P(\Theta_0^\epsilon)}{P(\Theta_A^\epsilon)} = \frac{\int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta}{\int_{I_\epsilon^c} \pi(\theta) d\theta},$$

BF can be calculated in each instance (BF and BF_ϵ) using the definition

$$BF = PO / P\Pi.$$

3 Convergence of Posterior Odds

As mentioned in Section 1, the objective of this paper is to determine and study conditions on $\pi(\theta)$, in the general setup of Section 2, under which the posterior odds from testing an imprecise hypothesis will converge to the posterior odds from testing a point null hypothesis as the interval length, ϵ , approaches 0 in the interval null test. Underlying this question is the problem of whether convergence of the posterior odds corresponds to convergence of the Bayes factors. The following theorem explains part of the story.

Theorem 3.1 *Under the setup in section 2, $\lim_{\epsilon \rightarrow 0} PO_\epsilon = PO$ implies $\lim_{\epsilon \rightarrow 0} BF_\epsilon = BF$.*

Proof. This theorem clearly follows from the calculations of the posterior odds and Bayes factors in Section 2.

However, the converse does not necessarily hold.

Theorem 3.2 Assume $\pi(\theta)$, the prior density in the setup of Section 2, is continuous and the corresponding distribution is denoted by P . Then $\lim_{\epsilon \rightarrow 0} PO_\epsilon = 0$.

Proof. Notice from equation (5) that, as ϵ approaches zero, the denominator of PO_ϵ will approach $\int_{\Theta} f(x | \theta)\pi(\theta) d\theta$, the denominator of PO in equation (4). In addition, since $f(x | \theta)$ is continuous at $\theta = 0$ and $\pi(\theta)$ is a continuous probability density function (pdf),

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} f(x | \theta)\pi(\theta) d\theta &\leq \lim_{\epsilon \rightarrow 0} \sup_{|\theta| \leq \epsilon} f(x | \theta) \cdot \int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta \\ &= f(x | 0) \cdot 0 = 0, \end{aligned}$$

thus proving the theorem.

These two theorems indicate that the approximation of the point null hypothesis test by the interval null hypothesis test is not valid in terms of the posterior odds under the current setup. By Theorem 3.1, we can restrict attention to PO, but we need to change the definition of our prior function $\pi(\theta)$. From the proof of Theorem 3.2, the problem reduces to finding a prior function $\pi(\theta)$ such that

$$\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta = \frac{\pi_0}{1 - \pi_0}.$$

This integral is the area under the curve $\pi(\theta)$ from $\theta = -\epsilon$ to $\theta = \epsilon$. Consequently, by the crude approximation $\int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta \approx \pi(0) \cdot 2\epsilon$, it seems that the function $\pi(\theta)$ needs to approach infinity at the same rate ϵ approaches zero in order for $\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta$ to equal a nonzero constant (in particular $\frac{\pi_0}{1 - \pi_0}$). Additionally, if $\pi(\theta)$ is not restricted to be a function of ϵ , then the “base” of the rectangle, from $-\epsilon$ to ϵ , (Figure 1) may approach zero too quickly as ϵ approaches zero resulting in zero area. Alternatively, $\pi(\theta)$ may explode too fast as θ approaches zero (i.e., at a faster rate than ϵ approaching zero) resulting in infinite area. Hence, $\pi(\theta)$ as a function of

ϵ appears to be a requirement for $\pi(\theta)$ to approach infinity at the same rate that ϵ approaches zero, and thus force $\int_{-\epsilon}^{\epsilon} \pi(\theta) d\theta$ to converge to a nonzero constant. This dependence on ϵ is denoted by writing “ $\pi_{\epsilon}(\theta)$ ” instead of $\pi(\theta)$ for the prior function.

Let

$$\pi_{\epsilon}(\theta) = \begin{cases} g_{\epsilon}(\theta) & \text{if } \theta \in (-\epsilon, \epsilon) \\ g_{\epsilon}^A(\theta) & \text{if } \theta \notin (-\epsilon, \epsilon) \end{cases} \quad (6)$$

be the prior on Θ for the interval hypothesis testing problem (2) where $g_{\epsilon}(\theta)$ and $g_{\epsilon}^A(\theta)$ are nonnegative continuous functions such that

$$\int_{\Theta} f(x | \theta) \pi_{\epsilon}(\theta) d\theta < \infty.$$

The superscript “A” on $g_{\epsilon}^A(\theta)$ signifies the use of $g_{\epsilon}^A(\theta)$ as the prior function over the Alternative space $\{\theta : |\theta| > \epsilon\}$. Additionally let

$$\pi(\theta) = \begin{cases} \pi_0 & \text{if } \theta = 0 \\ (1 - \pi_0)g(\theta) & \text{if } \theta \neq 0 \end{cases} \quad (7)$$

be the prior for the point null testing problem (1) where $\pi_0 \in (0, 1)$ and $g(\theta)$ is a continuous function such that

$$\lim_{\epsilon \rightarrow 0} g_{\epsilon}^A(\theta) = g(\theta).$$

In other words, the prior functions over the alternative spaces H_A and H_A^{ϵ} converge as the interval length, ϵ , approaches 0. Therefore the denominator of PO_{ϵ} still converges to the denominator of PO (recall equations (5) and (4)). The following theorem considers further constraints on $\pi_{\epsilon}(\theta)$ that allow the PO to converge.

Theorem 3.3 *Assume $g_{\epsilon}(\theta)$ is nonnegative, continuous, bounded in a deleted neighborhood of 0, and of the form $g_{\epsilon}(\theta) = h(\epsilon) \cdot p(\theta)$ for some functions p and h . Further, suppose one of*

$$a) \lim_{\epsilon \rightarrow 0} h(\epsilon) = \infty \text{ and } \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} p(\theta) d\theta = 0; \text{ or}$$

b) $\lim_{\epsilon \rightarrow 0} h(\epsilon) = 0$ and $\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} p(\theta) d\theta = \infty$

holds. If $\lim_{\epsilon \rightarrow 0} \left(\frac{1}{h(\epsilon)} \right)'$ exists and equals a real number $L \neq 0$, and $2p(0) = \frac{\pi_0}{1-\pi_0} \cdot L$, $\pi_0 \in (0, 1)$, then $\lim_{\epsilon \rightarrow 0} PO_{\epsilon} = PO$.

Proof. Let $k(\epsilon) = 1/h(\epsilon)$. Therefore

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} \pi_{\epsilon}(\theta) d\theta &= \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} h(\epsilon) p(\theta) d\theta \\ &= \lim_{\epsilon \rightarrow 0} \frac{\int_{-\epsilon}^{\epsilon} p(\theta) d\theta}{k(\epsilon)} \\ &\stackrel{L}{=} \lim_{\epsilon \rightarrow 0} \frac{p(\epsilon) + p(-\epsilon)}{k'(\epsilon)} \\ &= \frac{2p(0)}{L} = \frac{\pi_0}{1-\pi_0} \quad (L > 0), \end{aligned}$$

where $\stackrel{L}{=}$ shows an application of L'Hospital's rule. The second to last equality is due to the continuity of $g_{\epsilon}(\theta)$ for all θ (and hence continuity of $p(\theta)$). By the new definitions of the priors,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} PO_{\epsilon} &= \lim_{\epsilon \rightarrow 0} \frac{\int_{-\epsilon}^{\epsilon} f(x | \theta) \pi_{\epsilon}(\theta) d\theta}{\int_{I_{\epsilon}^c} f(x | \theta) \pi_{\epsilon}(\theta) d\theta} \\ &= \frac{f(x | 0) \cdot \frac{\pi_0}{1-\pi_0}}{\int_{\Theta} f(x | \theta) g(\theta) d\theta} \\ &= PO. \end{aligned}$$

A few remarks are necessary. Firstly, restricting $g_{\epsilon}(\theta)$ to be bounded inside some deleted neighborhood of zero (i.e., $\exists M > 0$ such that for some $\eta > 0$, $g_{\epsilon}(\theta) \leq M \forall \theta \in [-\eta, 0) \cup (0, \eta]$) prevents the continuous function from oscillating to infinity infinitely often inside an interval about zero. Although this assumption may seem constraining, recall that $\pi_{\epsilon}(\theta)$ represents the prior information about some parameter θ in an experiment. It seems unrealistic for an experimenter to describe prior

knowledge with a function that oscillates unboundedly and infinitely often in any interval of the parameter space. Secondly, assuming $g_\epsilon(\theta) = h(\epsilon) \cdot \pi(\theta)$ simplifies the prior form and clearly displays how prior information about θ is incorporated into the function. Finally, the remaining sufficient conditions restrict the limiting behavior of $g_\epsilon(\theta)$ as ϵ approaches zero and defines a value of the prior function at $\theta = 0$ to force convergence of the posterior odds.

Theorem 3.3 may not be useful for building a class of priors under which PO_ϵ converges to PO . The following corollary makes such a construction simpler.

Corollary 3.1 *If $g_\epsilon(\theta) = \frac{p(\theta)}{\epsilon}$ where $p(\theta)$ is a continuous probability density function (pdf) with $p(0) = \frac{\pi_0}{2(1-\pi_0)}$ and priors defined as in (6) and (7), then $\lim_{\epsilon \rightarrow 0} PO_\epsilon = PO$.*

Proof. Analogous to the proof of Theorem 3.3, since p is a pdf, the corollary follows after an application of L'Hospital's rule to $\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} (p(\theta)/\epsilon) d\theta$.

Hence, any pdf can be easily transformed into a prior function under which the approximation of the interval null test by the point null test is valid. Note, though, that requiring a prior to be a function of ϵ (an unknown and potentially difficult to specify quantity) with limiting behavior as stated in Theorem 3.3 may not be realistic in practice. Additionally, the restriction placed on $p(0)$ constrains the scientist in his/her choice of prior. Thus, under these conditions, the usefulness of the approximation by a researcher is dubious. We will discuss this matter further in the next section.

In some situations an experimenter may want to describe prior beliefs about one part of the parameter space by a function g_1 and opinions about another part of the parameter space by a function g_2 . As before, we wish to find a class of priors under which the precise test approximates the imprecise test using BF or PO as the

measure of evidence. Mathematically, the prior density for testing the point null hypothesis in this “non-symmetric” prior framework can be written

$$\pi^0(\theta) = \begin{cases} \pi_0 & \text{if } \theta = \theta_0 \\ \pi_1 g_1(\theta) & \text{if } \theta > \theta_0 \\ \pi_2 g_2(\theta) & \text{if } \theta < \theta_0 \end{cases} \quad (8)$$

where $\pi_1 \neq \pi_2$, $\pi_1 + \pi_2 + \pi_0 = 1$, and $g_1(\theta)$ and $g_2(\theta)$ are nonnegative continuous functions of θ . Following a similar reasoning to that of the previous work, consider the prior

$$\pi(\theta) = \begin{cases} g_1(\theta) & \text{if } \theta > \theta_0 \\ g_2(\theta) & \text{if } \theta < \theta_0 \end{cases} \quad (9)$$

for testing the interval null where $g_1(\theta)$ and $g_2(\theta)$ are the same functions as those in (8). The next theorem follows directly from the definition of the Bayes factor and the assumption that $\pi(\theta)$ is a density.

Theorem 3.4 *Under the prior functions $\pi^0(\theta)$ and $\pi(\theta)$ described above for the point null (1) and interval null (2) tests respectively, if $\pi(\theta)$ is a continuous pdf, then $\lim_{\epsilon \rightarrow 0} BF_\epsilon \neq BF$.*

Hence, under a setup like that of Gomez-Villegas and Sanchez-Manzano (1992) in Section 2, BF_ϵ may not converge to BF (and thus by an argument similar to that in Theorem 3.1, PO_ϵ will not converge to PO). The failure of the point null approximation of the interval null in terms of both PO and BF is due to the assignment of specified prior mass on the positive and negative pieces of the parameter space Θ when testing H_0 . Following the intuition and development preceding Theorem 3.3, assume $\pi(\theta)$ is a function of ϵ . This supposition allows the priors in the interval null test (for which particular prior mass is not spread over subsets of Θ), as ϵ approaches zero, to account for the different prior masses in the point null test. In

other words, somehow, as ϵ approaches zero, BF_ϵ must “pick up” π_0 , π_1 , and π_2 . Let $\pi(\theta)$ depend on ϵ (denoted $\pi_\epsilon(\theta)$) such that

$$\pi_\epsilon(\theta) = \begin{cases} g_\epsilon(\theta) & \text{if } \theta \in (-\epsilon, \epsilon) \\ g_{1\epsilon}^A(\theta) & \text{if } \theta \geq \epsilon \\ g_{2\epsilon}^A(\theta) & \text{if } \theta \leq -\epsilon, \end{cases} \quad (10)$$

where $g_\epsilon(\theta)$, $g_{1\epsilon}^A(\theta)$, and $g_{2\epsilon}^A(\theta)$ are continuous nonnegative functions such that $\int_{\Theta} f(x | \theta) \pi_\epsilon(\theta) d\theta < \infty$.

Theorem 3.5 *Consider the priors $\pi^0(\theta)$ and $\pi_\epsilon(\theta)$ from equations (8) and (10) for the point null and interval null testing situations respectively. Assume there exists $h(\epsilon)$ and $p(\theta)$ such that $g_\epsilon(\theta) = h(\epsilon) \cdot p(\theta)$ and the continuous nonnegative functions $g_{1\epsilon}^A(\theta)$ and $g_{2\epsilon}^A(\theta)$ are such that $\lim_{\epsilon \rightarrow 0} g_{1\epsilon}^A(\theta) = \pi_1 g_1(\theta)$ and $\lim_{\epsilon \rightarrow 0} g_{2\epsilon}^A(\theta) = \pi_2 g_2(\theta)$. If $g_{1\epsilon}^A(\theta)$ is integrable over $\{\theta : \theta > 0\}$ and $g_{2\epsilon}^A(\theta)$ is integrable over $\{\theta : \theta < 0\}$, then $\lim_{\epsilon \rightarrow 0} BF_\epsilon = BF$. In addition, under the same assumptions as Theorem 3.3 with the exception that $2p(0) = \pi_0 \cdot L$, $\pi_0 \in (0, 1)$, $\lim_{\epsilon \rightarrow 0} PO_\epsilon = PO$.*

The proof of this theorem is similar to the proof of Theorem 3.3. The main difference is that we require more restrictions on the prior functions over the alternative space ($g_{1\epsilon}^A(\theta)$ and $g_{2\epsilon}^A(\theta)$ here) to force the denominators of PO_ϵ and PII_ϵ to converge to the denominator of PO and PII (recall Section 2 for the necessary definitions). Hence, the approximation of the interval null test by the point null test for a non-symmetrical prior set up holds under a class of priors analogous to the class described by Theorem 3.3. Again, though, this class may not be reasonable for a researcher thus challenging the practicality of the approximation.

A special case of the non-symmetrical prior occurs when an experimenter wishes to test against a one-sided alternative, $\{\theta : \theta > 0\}$, say, as opposed to the two-sided situation we have been considering. If we think about the one-sided test as

a two-sided test with prior mass zero on the portion of the alternative space not covered by the one-sided scenario (e.g., $\{\theta : \theta < 0\}$), then all the previous work can be applied. In other words, a class of priors can be constructed using Theorem 3.3 such that PO_ϵ converges to PO and the approximation of the interval null test by the point null test is valid. The comments following Corollary 3.1 concerning the applicability of these priors still holds however.

4 Application

We will consider the practical implications of the theory developed in Section 3. To accomplish this task, the previous theorems and corollaries are applied to a data set collected by White, LaFaunce, and Mohammed (1989). The data consists of the calving interval (number of days from artificial insemination to birth) and pregnancy examination interval (number of days from artificial insemination to pregnancy examination) for 1318 cows on five California dairies. Cows examined for pregnancy 36 to 68 days postbreeding are categorized into groups based on three day intervals: 36 - 39 days postbreeding, 49 - 51, ..., 66 - 68, and ≥ 69 days. Furthermore, cows bred 30 - 35 days prior to the pregnancy examination are grouped into one category. Thus, the data is divided into thirteen groups. The authors were trying to determine the effect of pregnancy examination interval on calving interval in dairy cattle in order to calculate the optimal time postbreeding for a pregnancy examination. The statistical analysis involved a one-way ANOVA with a comparison of group means to determine if any of the differences between category means were significant.

For the purposes of this paper, consider the hypothesis test

$$H_0 : \theta_1 = \dots = \theta_{13} \text{ vs. } H_A : \text{not all } \theta_i\text{s are equal,} \quad (11)$$

where θ_i is the true mean calving interval for each pregnancy examination group

i. The first step in evaluating this hypothesis test is to setup a system of priors (since the Bayesian methodology is being studied here). Consider the hierarchical modeling approach suggested by Berger and Deely (1988) where

$$\mathbf{X} \mid \boldsymbol{\theta}, \sigma^2 \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{I}) \quad (12)$$

$$\boldsymbol{\theta} \mid \beta, \sigma_\pi^2 \sim N(\mathbf{y}\beta, \sigma_\pi^2 \mathbf{I}), \quad (13)$$

$$\beta \sim U(-\infty, \infty) \quad (14)$$

$$\sigma_\pi^2 \sim P \quad (15)$$

and P is a mixed prior probability distribution on \mathbf{R}^+ (positive real numbers), as in (3). The observed mean calving interval, X_i , for each pregnancy examination interval $i = 1, \dots, 13$ (in the 13×1 data vector \mathbf{X}) are assumed independent and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{13})^T$ contains the actual group means. Furthermore, σ^2 is an unknown constant, β is the constant mean calving interval under H_0 ($\theta_1 = \dots = \theta_{13} = \beta$), \mathbf{I} is a 13×13 identity matrix, and $\mathbf{y} = (1, \dots, 1)^T$. This model fits nicely into the framework of the paper because under the prior P , the hypothesis test (11) is equivalent to

$$H_0 : \sigma_\pi^2 = 0 \text{ vs. } H_A : \sigma_\pi^2 > 0 \quad (16)$$

a one-sided point null test. However, an experimenter is typically interested in the interval null test

$$H_0^\epsilon : 0 \leq \sigma_\pi^2 < \epsilon \text{ vs. } H_A^\epsilon : \sigma_\pi^2 \geq \epsilon \quad (17)$$

where $\epsilon > 0$ is “small.” For example, one may consider a very small σ_π^2 to be negligible and hence still accept the null hypothesis that the mass is concentrated about the equal cell means.

As in the previous section, it is of interest to study two measures of evidence for Bayesian hypothesis testing, posterior odds and Bayes factor. In particular, can the

Table 1: Posterior odds and Bayes factors for the interval null and point null tests with the hierarchical prior structure defined by (12) - (15). $\pi_0 = 1/2$.

ϵ	PO_ϵ	BF_ϵ
0.1	0.00773	7.312
0.001	7.77×10^{-5}	7.330
1×10^{-6}	7.77×10^{-8}	7.331
point null	7.331	7.331

point null test approximate the interval null test with respect to either PO or BF.

For subsequent calculations we consider the prior density over σ_π^2

$$\pi_{2,2}(\sigma_\pi^2) = \pi_0 I(\sigma_\pi^2 = 0) + (1 - \pi_0) \pi_{2,2}^*(\sigma_\pi^2) \quad (18)$$

where

$$\pi_{2,2}^*(\sigma_\pi^2) = \frac{\sigma}{2(\sigma^2 + \sigma_\pi^2)^{3/2}} d\sigma_\pi^2, \quad (19)$$

the “conventional” proper prior suggested by Berger and Deely (1988). As in Section 2, the interval null test can be evaluated with prior density $\pi_{2,2}^*(\sigma_\pi^2)$ over the entire parameter space. In addition, σ is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{13} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{1318 - 12 + 2}$$

the minimum risk equivariant estimator (Lehmann (1983)) as recommended by Berger (1980). For this data $\hat{\sigma}^2 = 47.20$. Also, $\pi_0 = 0.5$ is chosen because it is considered “noninformative” (Berger and Deely (1988)). As expected from the theory developed thus far, Table 1 indicates that under this setup, using relevant formulas provided in the paper by Berger and Deely (1988), $\lim_{\epsilon \rightarrow 0} BF_\epsilon = BF$ but $\lim_{\epsilon \rightarrow 0} PO_\epsilon = 0$. In fact, $BF = 7.3$ ($\pi_0 = 0.5$ implies $BF = PO$ and $P(H_0 | x) = 0.88$) is interpreted as evidence in favor of H_0 whereas $PO_\epsilon \rightarrow 0$ indicates strong evidence against H_0^ϵ as ϵ approaches zero. Therefore the point null test is not a good approximation of the interval null test under this setup. As a side point, $BF = 7.3$ is

Table 2: Posterior odds and Bayes factors for the interval null and point null tests with the hierarchical prior structure defined by (12) - (15) and π_0 is defined by (21).

ϵ	PO_ϵ	BF_ϵ
0.1	0.00773	7.312
0.001	0.0776	7.330
1×10^{-6}	0.0777	7.331
point null	0.0777	7.331

typically interpreted as evidence provided by the data in favor of equal mean calving intervals for each pregnancy examination group since BF does not depend on the prior probability π_0 . This interpretation is different than that inferred from PO, a measure sensitive to the prior.

By Corollary 3.1, if we redefine the prior on σ_π^2 for testing the imprecise null as

$$\pi_{2,2}^{*\epsilon}(\sigma_\pi^2) = \begin{cases} \frac{\pi_{2,2}^*(\sigma_\pi^2)}{\epsilon} & \text{if } \sigma_\pi^2 \in [0, \epsilon) \\ \pi_{2,2}^*(\sigma_\pi^2) & \text{if } \sigma_\pi^2 \geq \epsilon, \end{cases} \quad (20)$$

then if $\pi_{2,2}^*(0) = \pi_0/(1 - \pi_0)$, the approximation of the PO for the interval null test by point null test PO will be valid. To obtain this last condition, we take

$$\pi_0 = \frac{1}{2\hat{\sigma}^2} \cdot \left(1 + \frac{1}{2\hat{\sigma}^2}\right)^{-1}. \quad (21)$$

Table 2 indicates that under the hierarchical model with second stage prior (20) and $\pi_0 = 0.01048$ from (21), $\lim_{\epsilon \rightarrow 0} \text{PO}_\epsilon = \text{PO}$. Furthermore, $P(H_A | x) = 0.93$ implying the hypothesis of equal mean calving intervals is rejected, unlike the interpretation obtained from BF under the prior (19) with $\pi_0 = 0.5$ (note that BF here is the same as BF under the previous framework since the same prior is used for testing the precise hypothesis in both situations). Of course, the discrepancy between the two inferences is partially due to the small prior mass placed on H_0 ($\pi_0 \approx 0.01$), a restriction forced by Corollary 3.1.

The authors may not be comfortable representing their prior knowledge by the function (20) though. Two drawbacks are immediately apparent. Firstly, though the authors probably believe a priori that the point null hypothesis is false, placing such small prior mass on H_0 introduces an undesirably large bias towards the null. Secondly, the authors might have originally chosen to model the situation by a point null hypothesis because it was difficult to set an appropriate interval length in (17). Therefore, requiring the prior function to depend on the unknown ϵ is not as intuitively appealing as the noninformative prior considered in (19). Taking these issues into account, the authors may not choose to compromise their prior opinions, through utilization of $\pi_{2,2}^{*\epsilon}(\sigma_\pi^2)$, for the sake of testing a point null hypothesis. We will mention alternative strategies in the next section.

5 Discussion

The point null or precise hypothesis test is one of the most common tests done by scientists and experimenters. Unfortunately, it is rare to model a situation with an exact hypothesis such as $\theta = \theta_0$ that is of scientific interest. Any biases due to the experimental equipment or the experimenter taking measurements to test this hypothesis will prevent the representation of a hypothesis as a point (Berger and Delampady (1987), Casella and Berger (1987)). A more realistic description of the problem is a test of the interval null hypothesis $\theta \in (\theta_0 \pm \epsilon)$. However, if the point null test can be approximated by the interval null test, then using the precise hypothesis to model the situation may not be so unreasonable.

Previously this problem was attacked utilizing BF as the primary measure of evidence. Gomez-Villegas and Sanchez-Manzano (1992) show that under a reasonable framework of symmetric priors as defined in Section 2, BF_ϵ , the Bayes factor

for testing the interval null hypothesis, converges to BF, the Bayes factor for testing the point null, as ϵ approaches zero. Thus for this prior structure, BF indicates that the point null test does approximate the interval null test. However, if we change the prior structure to impose some asymmetry in the distribution of prior mass over the parameter space as in Section 3, the Bayes factors do *not* converge. The validity of approximating the imprecise test by the precise test in terms of BF seems to be a consequence of the nice properties (symmetry for example) inherent in the class of priors we choose.

In any case, the literature on Bayesian inference suggests that a Bayesian uses the posterior odds rather than the Bayes factor for evaluating a hypothesis test (e.g., Berger (1985), Casella and Berger (1990), and Degroot (1970)). The Bayes factor is useful for reporting since it allows the reader to use his/her own prior odds $P\Pi$ to calculate the posterior odds ($PO = BF \times P\Pi$). Hence, we studied the approximation of the interval null test by the point null test in terms of PO. Under the symmetric prior structure discussed before, the posterior odds do not converge. We did develop classes of priors under which the approximation is valid for a variety of testing situations when considering PO as the primary measure of evidence (and by the work of Section 3, it follows that the approximation holds in terms of BF). However, this class of priors has some undesirable properties. In particular, a researcher most likely will not desire to express his/her prior beliefs through a function that depends on the unknown interval length ϵ and is either unbounded or has infinite area over the interval null as ϵ approaches zero. Therefore, the testing of a point null hypothesis from a practical Bayesian perspective is questionable.

Of course, these properties are only sufficient conditions for the convergence of posterior odds. There may exist other, more realistic, priors for which precise tests approximate imprecise tests. Or, perhaps, the conditions are not only sufficient but

also necessary for convergence of the posterior odds. These issues are good problems for future study. Be that as it may, under some realistic and commonly assumed prior settings, the precise test is not valid as an approximation to the imprecise test.

Other solutions are available for a statistician confronted with this hypothesis testing dilemma. We may think of ways to avoid the approximation altogether. For instance, if a researcher insists on doing a hypothesis test, he/she can directly test the desired imprecise null hypothesis and evaluate evidence provided by the data and by prior knowledge through the posterior odds. In a Bayesian framework, once the scientist's prior distributions are determined, such a test is not too difficult. Alternatively, a scientist can set up confidence intervals or credible sets instead of testing hypotheses, or simply calculate the posterior distribution of θ given the data x and make inferences about the unknown parameter through this distribution. These latter approaches will avert some of the difficulties encountered under the Bayesian hypothesis testing paradigm.

References

- [1] Berger, J. O. (1980). A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean. *Annals of Statistics* 8, 716-761.
- [2] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer-Verlag, New York.
- [3] Berger, J. O. and Deely, J. (1988). A Bayesian Approach to Ranking and Selection of Related Means with Alternatives to Analysis of Variance Methodology. *Journal of the American Statistical Association* 83, 364-373.

- [4] Berger, J. O. and Delampady, M. (1987). Testing Precise Hypotheses (with discussion). *Statistical Science* 2, 317-335.
- [5] Berger, J. O. and Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence (with discussion). *Journal of the American Statistical Association* 82, 112-122.
- [6] Bernstein, L. H., Good, I. J., Holtzman, I. G., Deaton, M. L., and Barb, J. (1989). Diagnosis of Acute Myocardial Infarction from Two Measurements of Creative Kinase Isoenzyme MB with Use of Nonparametric Probability Estimation. *Clinical Chemistry* 35 (3), 444-447.
- [7] Brown, S. J. and Klein, R. W. (1986). Model Selection in the Federal Courts: An Application of the Posterior Odds Ratio Criterion. In *Bayesian Inference and Decision Techniques*, P. K. Goel and A. Zellner (Eds.). Elsevier Science Publishers, New York.
- [8] Casella, G. and Berger, R. L. (1987). Comment on "Testing Precise Hypotheses," by J. O. Berger and M. Delampady. *Statistical Science* 2, 344-347.
- [9] Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Wadsworth, California.
- [10] Delampady, M. (1989). Lower Bounds on Bayes Factors for Interval Null Hypotheses. *Journal of the American Statistical Association* 84 120-124.
- [11] DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

- [12] Dickey, J. (1976). Approximate Posterior Distributions. *Journal of the American Statistical Association* 71, 680-689.
- [13] Fienberg, S. E. (1990). Legal Likelihoods and a Priori Assessments: What Goes Where? In *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner (Eds.). Elsevier Science Publishers, North Holland.
- [14] Gomez-Villegas, M. A. and Sanchez-Manzano, E. G. (1992). Bayes Factors in Testing Precise Hypotheses. *Communications in Statistics- Theory and Methods* 21, 1707-1715.
- [15] Lehmann, E. L. (1983). *Theory of Point Estimation*. Wadsworth, California.
- [16] Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. John Wiley and Sons, New York.
- [17] Reckhow, K. H. (1990). Bayesian Inference in Non-replicated Ecological Studies. *Ecology* 71 (6), 2053-2059.
- [18] Reynolds, A. (1982). Posterior Odds for the Hypothesis of Independence between Stochastic Regressors and Disturbances. *International Economics Review* 23, 479-490.
- [19] Robert, C. P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer-Verlag, New York.
- [20] Verdinelli, I. and Wasserman, L. (1993). Bayes Factors, Nuisance Parameters, and Imprecise Tests. Technical Report no. 570, Department of Statistics, Carnegie Mellon.

- [21] White, M. E., LaFaunce, N., and Mohammed, H. O. (1989). Optimal Time Postbreeding for Pregnancy Examination in Dairy Cattle. *Canadian Veterinary Journal* 30, 147-149.
- [22] Zellner, A. (1984). *Basic Issues in Econometrics*. University of Chicago Press, Chicago.

Figure 1: Crude approximation of area under the curve of $\pi(\theta)$ on $(-\epsilon, \epsilon)$

