Lower Bounds on Bayes Factors for Interval Null Hypotheses

by

Mohan Delampady
Purdue University

Technical Report #86-35

Department of Statistics
Purdue University

August 1986

Lower Bounds on Bayes Factors for Interval Null Hypotheses *

Mohan Delampady
Purdue University and
University of British Columbia

August 26, 1986

Abstract

Expressions are derived for lower bounds on Bayes factors for interval null hypotheses. These are then studied as the interval length varies and are related to two interesting results, one for point null hypotheses and another for one-sided hypotheses. Examples and numerical results are also presented.

1 Introduction

Lower bounds on Bayes factors and posterior probabilities of point null hypotheses have been studied extensively by Edwards, Lindman and Savage(1963), Dickey (1977), Good(1950, 1958, 1976), Berger and Sellke(1987), Casella and Berger (1987), Berger and Delampady (1986), Delampady(1986a), Delampady (1986c), and Delampady and Berger (1986) among others. One of the major justifications for testing a point null hypothesis is that it can be considered as an approximation to an appropriate interval hypothesis in a large number of situations. Berger and Sellke (1987) study the case of point null hypotheses in depth and show that the lower bounds on Bayes factors over many reasonable classes of densities are at least an order of magnitude larger than the corresponding P–values. On the other hand Casella and Berger (1987) study one-sided testing situations where

*Research supported by the National Science Foundation, Grant DMS-8401996.

they show that the lower bounds are usually exactly equal to the P–values. It is clear that as a solution to testing interval null hypotheses these are two extremes. It is possible to directly investigate the case of interval nulls and we feel that such lower bounds are often more reasonable than either extremes.

1.1 Set Up

A random quantity X has the density $f(x|\theta)$, $\theta \in \Theta \subseteq R^1$. For each $a \geq 0$ consider the testing problem

$$H_{0a} : |\theta - \theta_0| \leq a \text{ against } \qquad H_{1a} : |\theta - \theta_0| > a, \tag{1}$$

where θ_0 is a specified quantity. Consider the following examples.
Example 1. In a statistical quality control situation θ may be assumed to be the size of a unit and acceptable units are with $\theta \in (\theta_0 - a, \theta_0 + a)$. Then one would like to test

$$H_0 : |\theta - \theta_0| \leq a.$$

In this problem the length of the interval, $2a$, can be explicitly specified. On the other hand, this is not the case in the following.
Example 2. Suppose we want to test the hypothesis,

$$H_0 : \text{ Vitamin C has no effect on the common cold, .}$$

Clearly this is not meant to be thought of as an exact point null; surely Vitamin C has <u>some</u> effect, though perhaps a very miniscule effect. Thus, in reality, this is the case of an interval null hypothesis which would be better represented as a point null hypothesis. We shall discuss the question of approximating pont null hypotheses in the later sections.

Assume that a classical significance test is based on some test statistic $T(X)$, large values of which provide evidence against the null hypothesis. For the test in (1) the P–value, or the observed significance level, of data x is defined to be

$$\alpha = \sup_{|\theta-\theta_0|\leq a} P_\theta \left(T(X) \geq T(x)\right). \tag{2}$$

Approaching the above testing problem from the Bayesian and likelihood viewpoint assume that π is a prior distribution on Θ with density g and prior probabilities $\pi_0 = P^\pi(H_{0a})$ and $1 - \pi_0 = P^\pi(H_{1a})$. Then g may be written as

$$g(u) = g_0(u) I\left\{|\theta - \theta_0| \leq a\right\} + g_1(u) I\left\{|\theta - \theta_0| > a\right\},$$

where g_0 and g_1 are the densities of π conditional on the hypotheses H_{0a} and H_{1a} respectively. The posterior distribution of θ given x is given by

$$\pi(\theta|x) = \frac{f(x|\theta)g(\theta)}{m(x)},$$

assuming that the marginal density of X at x,

$$m(x) = \int_{\Theta} f(x|\theta)g(\theta)\, d\theta > 0.$$

The quantities of interest are
(i) the Bayes factor of H_{0a} to H_{1a}:

$$B^{\pi}(x) \quad = \quad \frac{\int_{|\theta-\theta_0|\leq a} f(x|\theta)g_0(\theta)\, d\theta}{\int_{|\theta-\theta_0|>a} f(x|\theta)g_1(\theta)\, d\theta}; \tag{3}$$

(ii) the posterior probability of H_{0a} given x:

$$\begin{aligned} P^{\pi}\left(|\theta-\theta_0| \leq a|\, x\right) &= \int_{|\theta-\theta_0|\leq a} \pi(\theta|x)\, d\theta \\ &= \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{B^{\pi}(x)}\right]^{-1}. \end{aligned}$$

Thus $B^{\pi}(x)$ is the factor by which the data changes prior odds of H_0 to H_1 to posterior odds. By considering $B^{\pi}(x)$, one thus considers the impact of the data; the prior probabilities of the hypotheses can be ignored. A likelihoodist is also interested in $B^{\pi}(x)$, since it is the ratio of the likelihood of H_0 to the average or weighted likelihood of H_1, the averaging being with respect to the "weight function" g.

Specification of g is natural and important to a Bayesian, but is resisted by others. Of interest is that lower bounds on $B^{\pi}(x)$ (and hence $P^{\pi}(H_0|x)$) can be found for important classes of densities g, and that these lower bounds tend to be surprisingly large. If G is a class of densities g under consideration, we will consider the lower bounds

$$\underline{B}_G(x) \quad = \quad \inf_{g\in G} B^{\pi}(x), \tag{4}$$

and

$$\underline{P}_G(H_0|x) = \inf_{g\in G} P^\pi(H_0|x)$$
$$= \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{\underline{B}_G}\right]^{-1}. \quad (5)$$

For the most part, we will only present results in terms of $\underline{B}_G$, since $\underline{P}_G(H_0|x)$ requires the additional specification of π_0.

1.2 Choice of G

A lower bound, such as $\underline{B}_G(x)$, is particularly useful when G is large enough to include all densities which are plausible, but is not so large as to include unreasonable densities. If reasonable densities are omitted from G, one could argue that $\underline{B}_G$ is not actually a valid lower bound. If G contains unreasonable distributions, on the other hand, then the lower bounds may be driven too low to be useful. Note, in particular, that minimizing $B^\pi(x)$ over g has the effect of finding that $g \in G$ which is most favorable to H_1.

All these lower bounds thus contain a potential substantial bias towards H_1, and it is obviously desirable to minimize this bias; this can best be done by restricting G in as many ways as are considered reasonable.

A Bayesian might restrict G to a single distribution, g_0. A robust Bayesian might restrict g to a small class of densities, say, those in a neighborhood of some g_0 (cf. Berger and Berliner (1986) and Sivaganesan and Berger (1986)). But any such restrictions require specific subjective input. Of interest to Bayesians and non-Bayesians alike are choices of G which require only general shape specifications concerning G. One such possibility is

$$G_{US} = \{\text{unimodal } g, \text{ symmetric about } \theta_0\}. \quad (6)$$

The appeal of this class of densities is that it seems to be an "objective" class. It acknowledges the central role of θ_0, and seeks to spread mass about θ_0 in a way that is not biased towards particular alternatives. Lower bounds derived from such G could be termed "objective lower bounds," and are thus of interest when subjective input is unavoidable or being avoided.

Berger and Sellke (1987) consider the class G_S of all symmetric densities about θ_0. We feel that this class has densities which give undesirable concentration of mass to particular parameter points. This can be felt by noting that the minimizing density is a discrete mass function, assigning mass to two points on either sides of θ_0. Further, as the length of the interval (specified by H_0) increases to ∞ the lower bounds tend to 0.

We shall show that the class G_{US} has no such undesirable feature and that, as the interval length increases, it shows a behaviour relating our results with those of Casella and Berger (1987).

Section 2 will discuss an example and related results of Casella and Berger (1985). In Section 3 the expression for the lower bound on Bayes factor will be derived citing the necessary conditions. Finally in Section 4 conclusions will be given.

2 Methodology–Normal Example

Suppose $\overline{X} \sim N(\theta, \sigma^2/n)$, σ^2 known, and it is desired to test

$$H_0 : |\theta - \theta_0| \leq a \text{ against } H_1 : |\theta - \theta_0| > a.$$

Let $T = \sqrt{n}(\overline{X} - \theta_0)/\sigma$, and t_α be the critical value such that

$$\alpha = P_{\theta_0+a}(|T(X)| \geq t_\alpha).$$

Note that t_α depends on a. Were we to observe $T = t_\alpha$, we would report α as the P–value. To compare this with the posterior probability when t_α is observed, consider priors π for θ which (i) are symmetric about θ_0; (ii) are nonincreasing in $|\theta - \theta_0|$; and (iii) give prior probability .5 to H_0. Denote this class by G_{US}. Then

$$\underline{P}(H_0|x) \equiv \inf_{\pi \in G_{US}} P^\pi(H_0|x) \tag{7}$$

$$= \left(1 + \sup_r \frac{\Phi(r - t_\alpha) - \Phi(-r - t_\alpha) - \Phi(a^* - t_\alpha) + \Phi(-a^* - t_\alpha)}{2(r - 2a^*)\phi(t_\alpha) + \Phi(a^* - t_\alpha) - \Phi(-a^* - t_\alpha)}\right)^{-1} \tag{8}$$

where $b = \frac{\Phi(a^*-t_\alpha)-\Phi(-a^*-t_\alpha)}{2\phi(t_\alpha)}$, $a^* = \frac{a\sqrt{n}}{\sigma}$, and ϕ and Φ are the standard normal density and c.d.f., respectively. Figure 1 presents $\underline{P}(H_0|x)$ as a function of a^*, when the P–value is fixed at $\alpha = .01$, .05 and .10. Here "LENGTH" stands for the standardized length a^* of the half interval, and "BOUND" denotes $\underline{P}(H_0|x)$.

The expression in (7) is computationally quite attractive. In addition, the following iterative formula is available for the maximizing r of (7) (starting with $r_0 = t_\alpha$):

$$r_{i+1} = t_\alpha + \left(2\log\left(\frac{r_i - 2a^* + b}{\Phi(r_i - t_\alpha) - 2b\phi(t_\alpha)}\right) - 1.838\right)^{\frac{1}{2}}.$$

Convergence is quickly achieved for small values of a^* together with $t \geq 1.645$. For $a^* = 0$ this iterative formula reduces to that given in Berger and Sellke(1984). Figure 1 presents $\underline{P}(H_0|x)$ as a function of a^*, when the P–value is fixed at $\alpha = .01$, .05 and .10. Here "LENGTH" stands for the standardized length a^* of the half interval, and "BOUND" denotes $\underline{P}(H_0|x)$.

An interesting observation from Figure 1. is that, for a fixed value of the P–value, as a^* increases the lower bound on the posterior probability approaches the P–value. This can be rigorously established as follows: If T has a Normal distribution with unit variance, then, for each a, the P–value of the data t under H_{0a} is

$$\begin{aligned} \alpha &= \sup_{|\theta|\leq a} P_\theta(|T| > t) \\ &= P_a(|T| > t) \\ &= 1 - P_a(-t \leq T \leq t) \\ &= 1 - \Phi(t-a) + \Phi(-t-a). \end{aligned}$$

Let z_α be such that $\Phi(-z_\alpha) = \alpha$. Then data of the magnitude of z_α has the P–value of 2α under H_0. For each $a \geq 0$ let $t_a(\alpha)$ satisfy

$$P_a\left(|T| \geq t_a(\alpha)\right) = \alpha. \tag{9}$$

Then it is clear that $a + z_\alpha < t_a(\alpha) < a + z_{\alpha/2}$. Further it is seen from Berger and Sellke (1987) that the interesting cases have α that satisfy $z_{\alpha/2} > 1$, which means that $t = t_a(\alpha) > a + z_{\alpha/2} > a + 1$. For large a, $z_\alpha + a$ is a good approximation for $t_a(\alpha)$ (the value of t which obtains the P–value of α under H_{0a}).

Theorem 2.

$$\lim_{a\longrightarrow\infty} \inf_{\pi\in G_{US}} P^\pi(H_{0a}|x) = \alpha.$$

Proof: Let

$$L(r,t) = \frac{\Phi(r-t) - \Phi(-r-t) - \Phi(a-t) + \Phi(-a-t)}{2(r-2a)\phi(t) + \Phi(a-t) - \Phi(-a-t)},$$

for $r > 2a$.

It can be seen that, for large a, $\frac{dL(r,t_a(\alpha))}{dr}$ is positive at $r = 2a$ but negative for $r \geq 3a$. Further, for $2a \leq r \leq 3a$, $L(r,t_a(\alpha))$ is close to $\frac{(1-\alpha)}{\alpha}$. Since,

$$\inf_{\pi\in G_{US}} P^\pi(H_0|x) = \left[1 + \sup_r L(r,t)\right]^{-1},$$

Figure 1: Lower Bounds for Interval Nulls

from (7), this explains the phenomenon. □

We have, thus, shown that

$$\lim_{a\to\infty} \underline{P}_{G_{US}}\left(H_{0a}|t_a(\alpha)\right) = \alpha,$$

which is very related to the result obtained by Casella and Berger(1987). There it was shown for one–sided testing that the lower bound on the posterior probability of H_0 is exactly equal to the P–value. Indeed, if a is large and $t_a(\alpha) = z_\alpha + a$ then the test of H_{0a} versus H_{1a} is similar to the one–sided test, $H_0 : \theta \le a$ against $H_1 : \theta > a$.

3 Lower Bounds on Bayes Factors

3.1 General Theory

We shall assume the following conditions. These are fairly general and are satisfied in many interesting situations. A specific illustration will be given in the next subsection. Define $b = \frac{\int_{-a}^{a} f(x|\theta)\, d\theta}{2f(x|0)}$.

1.

$$H(r) = \frac{1}{2r}\int_{-r}^{r} f(x|\theta)\, d\theta$$

has a minimum at $r = 0$ and a unique maximum at r_0, the solution of $\frac{dH(r)}{dr} = 0$, $r > 0$.

2. For each $a > 0$,

$$H_a(r) = \frac{1}{2(r-2a+b)f(x|0)}\left(\int_{-r}^{r} f(x|u)\, du - 2bf(x|0)\right)$$

has its unique maximum at the solution of $\frac{dH_a(r)}{dr} = 0$, $r > a$.

3. For each fixed a, the observed value of x is such that $r_0 > a$.

In the following discussion let G_{US} denote all unimodal symmetric densities that have mass π_0 in the interval $[-a, a]$. Also for the sake of convenience let $\theta_0 = 0$ in (1).

Theorem 1. *Under the conditions 1–3,*

$$\inf_{\pi\in G_{US}} B^\pi(x) = \left[\sup_r \frac{\int_{-r}^{r} f(x|\theta)\, d\theta - 2bf(x|0)}{2(r-2a+b)f(x|0)}\right]^{-1}.$$

Proof: The result will be proved in the following three steps. To start with, for all $g \in G_{US}$, we have

$$g(u) = \int_0^\infty \frac{1}{2r} I_{[-r,r]}(u)\, d\mu(r),$$

for some measure μ, since g is symmetric and unimodal. Since $\int_{-a}^{a} g(u)\, du = \pi_0$, we, further, have the following.

Step 1. *μ is a mixture of 2–point measures.*

Proof:

$$\begin{aligned}
\int_{-a}^{a} g(u)\, du &= \int_{-a}^{a} \int_0^\infty \frac{1}{2r} I_{[-r,r]}(u)\, d\mu(r)\, du \\
&= \int_0^\infty \left(\int_{-a}^{a} \frac{1}{2r} I_{[-r,r]}(u)\, du \right) d\mu(r) \\
&= \int_0^a \left(\int_{-r}^{r} \frac{1}{2r}\, du \right) d\mu(r) + \int_a^\infty \left(\int_{-a}^{a} \frac{1}{2r}\, du \right) d\mu(r) \\
&= \mu[0,a] + \int_a^\infty \frac{a}{r}\, d\mu(r).
\end{aligned}$$

Therefore, we have,

$$\mu[0,a] + \int_a^\infty \frac{a}{r}\, d\mu(r) = \pi_0. \tag{10}$$

Now if we put $\mu[0,a] = c$, then the conditions on μ are,

$$\begin{aligned}
\int_a^\infty d\mu(r) &= 1 - c, \\
\int_a^\infty \frac{a}{r}\, d\mu(r) &= \pi_0 - c,
\end{aligned} \tag{11}$$

subject to $0 \le c \le \pi_0$. Substituting, $d\mu(r) = (1-c)\, d\nu(r)$, we get

$$\begin{aligned}
\int_a^\infty d\nu(r) &= 1, \\
\int_a^\infty \frac{1}{r}\, d\nu(r) &= \frac{(\pi_0 - c)}{a(1-c)}.
\end{aligned} \tag{12}$$

Putting, $dm(s) = d\nu(\frac{1}{s})$, these conditions become

$$\begin{aligned}
\int_0^{1/a} dm(s) &= 1, \\
\int_0^{1/a} s\, dm(s) &= \frac{(\pi_0 - c)}{a(1-c)} = \frac{1}{u}.
\end{aligned} \tag{13}$$

Since any mean 0 distribution is a mixture of 2–point mean 0 distributions, (Freedman(1971)), the measure m must be a mixture of 2–point mean $\frac{1}{u}$ measures. □

This step reduces the lower bound on the Bayes factors as follows:

Step 2.

$$\left[\inf_{\pi \in G_{US}} B^\pi(x)\right]^{-1} = \sup_{a \le u_1 \le u \le u_2 < \infty} \frac{1}{2(u-2a+b)f(x|0)}$$
$$\times \left[\frac{u_2-u}{u_2-u_1}\int_{-u_1}^{u_1} f(x|u)\,du + \frac{u-u_1}{u_2-u_1}\int_{-u_2}^{u_2} f(x|u)\,du - 2bf(x|0)\right].$$

Proof: Following (3),

$$B^\pi(x) = \frac{\int_{-a}^{a} f(x|\theta)g(\theta)\,d\theta}{\int_{-\infty}^{\infty} f(x|\theta)g(\theta)\,d\theta - \int_{-a}^{a} f(x|\theta)g(\theta)\,d\theta}. \tag{14}$$

Let A be the set of all measures μ satisfying (9) and G be the class of all unimodal symmetric densities which put π_0 mass between $-a$ and a. Then

$$\inf_{\pi \in G_{US}} B^\pi(x) = \inf_{g \in G} \frac{\int_{-a}^{a} f(x|u)g(u)\,du}{\int_{-\infty}^{\infty} f(x|u)g(u)\,du - \int_{-a}^{a} f(x|u)g(u)\,du}.$$

Using the representation

$$g(u) = \int_0^\infty \frac{1}{2r} I_{[-r,r]}(u)\,d\mu(r),$$

we get

$$\begin{aligned}
&\left[\inf_{\pi \in G_{US}} B^\pi(x)\right]^{-1} \\
&= \sup_{\mu \in A} \frac{1}{\int_0^a \left(\frac{1}{2r}\int_{-r}^{r} f(x|u)\,du\right)\,d\mu(r) + \int_a^\infty \left(\frac{1}{2r}\int_{-a}^{a} f(x|u)\,du\right)\,d\mu(r)} \\
&\quad \times \left[\int_a^\infty \left(\frac{1}{2r}\int_{-r}^{r} f(x|u)\,du\right)\,d\mu(r) - \int_a^\infty \left(\frac{1}{2r}\int_{-a}^{a} f(x|u)\,du\right)\,d\mu(r)\right] \\
&= \sup_{c, \mu \in A} \frac{\int_a^\infty \left(\frac{1}{2r}\int_{-r}^{r} f(x|u)\,du - 2bf(x|0)\right)\,d\mu(r)}{[c + (\pi_0 - c)(b/a)]\,f(x|0)},
\end{aligned} \tag{15}$$

where the supremum in the last expression is over all μ satisfying (10) and $0 \le c \le \pi_0$. We get the last equality by recalling (10) and observing that, for $0 \le r \le a$,

$\frac{1}{2r}\int_{-r}^{r} f(x|u)\,du$ has a minimum at $r=0$. Recall that μ is a mixture of 2–point measures satisfying

$$\frac{d(c)}{u_1} + \frac{(1-d(c))}{u_2} \;=\; \frac{1}{u}, \tag{16}$$

for some $0 \le d(c) \le 1$, if (u_1, u_2) is any point where it puts a weight. Here we can solve for $d(c)$ to obtain

$$d(c) \;=\; \frac{(u_2-u)}{(u_2-u_1)}\frac{u_1}{u}. \tag{17}$$

Now let $F(r) = \frac{1}{2r}\left(\int_{-r}^{r} f(x|u)\,du - 2bf(x|0)\right)$. Then from (10), (14), (15) and (16),

$$\begin{aligned}\left[\inf_{\pi\in G_{US}} B^{\pi}(x)\right]^{-1} &= \sup_{c,u_1,u_2} \frac{(1-c)\left[\frac{u_1(u_2-u(c))}{u(c)(u_2-u_1)}F(u_1) + \frac{u_2(u(c)-u_1)}{u(c)(u_2-u_1)}F(u_2)\right]}{[c+(\pi_0-c)(b/a)]\,f(x|0)}\\ &= \sup_{u,u_1,u_2} \frac{\left[\frac{u_1(u_2-u)}{(u_2-u_1)}F(u_1) + \frac{u_2(u-u_1)}{(u_2-u_1)}F(u_2)\right]}{2(u-2a+b)f(x|0)},\end{aligned}$$

recalling, $u = u(c) = \frac{(1-c)a}{(\pi_0)-c)}$, and thus,

$$\frac{c+(\pi_0-c)(b/a)}{(1-c)} = \frac{u}{(u-2a+b)}.$$

This proves step 2. □

Now define,

$$\begin{aligned} &g(u,u_1,u_2)\\ &= \frac{1}{2(u-2a+b)}\left[\frac{(u_2-u)}{(u_2-u_1)}u_1F(u_1) + \frac{(u-u_1)}{(u_2-u_1)}u_2F(u_2)\right]\\ &= \frac{1}{2(u-2a+b)}\Big\{\frac{(u_2-u)}{(u_2-u_1)}\int_{-u_1}^{u_1} f(x|v)\,dv\\ &\quad + \frac{(u-u_1)}{(u_2-u_1)}\int_{-u_1}^{u_1} f(x|v)\,dv - 2bf(x|0)\Big\}\end{aligned}$$

for $a \le u_1 \le u \le u_2 < \infty$, $u > 2a$. Finally we have the following Step which will prove Theorem 1.

Step 3.

$$\sup_{a\le u_1\le u\le u_2<\infty} g(u,u_1,u_2)$$
$$= \sup_r \frac{1}{2(r-2a+b)}\left[\int_{-r}^{r} f(x|v)\,dv - 2bf(x|0)\right].$$

Proof: If the maximum of g is at a point (u^*, u_1^*, u_2^*) which satisfies $u_1^* = u_2^*$, the above claim is proved because, then, we have $u_1^* = u^* = u_2^*$. Therefore assume $u_1 < u_2$. Then extreme points of g satisfy $\frac{\partial g}{\partial u} = 0$. However,

$$\frac{\partial g}{\partial u} = \frac{1}{2(u_2-u_1)(u-2a+b)^2}\left[(u_1-2a+b)u_2F(u_2) - (u_2-2a+b)u_1F(u_1)\right].$$

Therefore the extreme points, (u^*, u_1^*, u_2^*), of g satisfy

$$\frac{u_2^*F(u_2^*)}{(u_2^*-2a+b)} = \frac{u_1^*F(u_1^*)}{(u_1^*-2a+b)}. \tag{18}$$

They also satisfy $\frac{\partial g}{\partial u_1} = 0$. Since,

$$\frac{\partial g}{\partial u_1} = \frac{(u_2-u_1)\left[(u_2-u)\{\frac{d}{du_1}u_1F(u_1)\} - u_2F(u_2)\right]}{2(u-2a+b)(u_2-u_1)^2} + \frac{(u_2-u)u_1F(u_1) + (u-u_1)u_2F(u_2)}{2(u-2a+b)(u_2-u_1)^2},$$

and $u_1 < u < u_2$, we get

$$(u_2-u_1)\frac{d}{du_1}u_1F(u_1) + u_1F(u_1) - u_2F(u_2) = 0. \tag{19}$$

But, from (17),

$$u_2^*F(u_2^*) - u_1^*F(u_1^*) = u_1^*F(u_1^*)\left(\frac{u_2^*-u_1^*}{u_1^*-2a+b}\right).$$

Therefore (18) reduces to

$$\frac{f(x|u_1^*) + f(x|-u_1^*)}{2} - \frac{1}{2(u_1^*-2a+b)}\left(\int_{-u_1^*}^{u_1^*} f(x|v)\,dv - 2bf(x|0)\right)$$

$$= \quad 0, \tag{20}$$

since

$$\begin{aligned}\frac{d}{dr} rF(r) &= \frac{d}{dr}\int_{-r}^{r} f(x|v)\,dv - 2bf(x|0) \\ &= f(x|r) + f(x|-r).\end{aligned}$$

Now observe that

$$\begin{aligned}&\frac{d}{dr}\Big(\frac{1}{2(r-2a+b)}\big(\int_{-r}^{r} f(x|v)\,dv - 2bf(x|0)\big)\Big)|_{r=u_1^*} \\ &= \frac{f(x|u_1^*) + f(x|-u_1^*)}{2} - \frac{1}{2(u_1^*-2a+b)}\left(\int_{-u_1^*}^{u_1^*} f(x|v)\,dv - 2bf(x|0)\right).\end{aligned}$$

Therefore (19) implies that, $u_1^* = a$ or $u_1^* = r^*$, where r^* is the unique maximum of $\frac{1}{2(r-2a+b)}\left(\int_{-r}^{r} f(x|v)\,dv - 2bf(x|0)\right)$. But if $u_1^* = r^*$, then $u_2^* = r^*$ from (17) and the fact that r^* is unique. This is a contradiction. On the other hand, if $u_1^* = a$, then since

$$aF(a) = \int_{-a}^{a} f(x|v)\,dv - \int_{-a}^{a} f(x|v)\,dv = 0,$$

it follows that

$$\begin{aligned}g(u^*, a, u_2^*) &= \frac{1}{2(u^*-2a+b)}\frac{(u^*-a)}{(u_2^*-a)}u_2^*F(u_2^*) \\ &= \frac{1}{2(u_2^*-2a+b)}\left(\int_{-u_2^*}^{u_2^*} f(x|v)\,dv - 2bf(x|0)\right) \\ &\quad \times\frac{(u^*-a)(u_2^*-2a+b)}{(u_2^*-a)(u^*-2a+b)}u_2^*F(u_2^*) \\ &< \frac{1}{2(u_2^*-2a+b)}\left(\int_{-u_2^*}^{u_2^*} f(x|v)\,dv - 2bf(x|0)\right),\end{aligned}$$

the last step following from the fact that,

$$\begin{aligned}\frac{(u^*-a)(u_2^*-2a+b)}{(u_2^*-a)(u^*-2a+b)} &= \left(1+\frac{(b-a)}{(u_2^*-a)}\right) / \left(1+\frac{(b-a)}{(u^*-a)}\right) \\ &< 1.\end{aligned}$$

This proves that (a, u^*, u_2^*) can not be a maximum for g. Therefore we conclude that

$$u_1^* = u^* = u_2^*,$$

which proves step 3 and hence the Theorem. □

3.2 Verification for the Normal Example

Now it will be shown that for the normal example discussed earlier all the three necessary conditions are satisfied and that Theorem 1 is thus applicable.

1. Notice that

$$\begin{aligned} H(r) &= \frac{1}{2r}\left[\Phi(r-t)-\Phi(-r-t)\right], \\ \frac{dH(r)}{dr} &= \frac{1}{2r^2}\left[r\left(\phi(r-t)+\phi(-r-t)\right)-\left(\Phi(r-t)-\Phi(-r-t)\right)\right] \text{ and} \\ \frac{d^2H(r)}{dr^2} &= \frac{1}{r}\left[\frac{\phi'(r-t)-\phi'(-r-t)}{2}-2\frac{dH(r)}{dr}\right]. \end{aligned}$$

It can be shown that $H(r)$ has its unique maximum at the first non-zero solution of $\frac{dH(r)}{dr}=0$. (Result 2 in the Appendix of Delampady (1986a) has a detailed proof.)

2. Here observe that

$$\begin{aligned} H_a(r) &= \frac{1}{2(r-2a+b)}\left[\Phi(r-t)-\Phi(-r-t)-\Phi(a-t)+\Phi(-a-t)\right], \\ \frac{dH_a(r)}{dr} &= \frac{1}{2(r-2a+b)^2}\{(r-2a+b)\left(\phi(r-t)+\phi(-r-t)\right) \\ &\quad -\left(\Phi(r-t)-\Phi(-r-t)-\Phi(a-t)+\Phi(-a-t)\right)\} \qquad (21) \end{aligned}$$

and

$$\frac{d^2H_a(r)}{dr^2} = \frac{1}{(r-2a+b)}\left[\frac{\phi'(r-t)-\phi'(-r-t)}{2}-2\frac{dH_a(r)}{dr}\right].$$

Therefore condition 2. follows exactly as condition 1.

3. As explained earlier in the example, in Theorem 2, we are only interested in situations where $t > a+1$. □

Let G^*_{US} be the class of all unimodal symmetric distributions which assign 1/2 mass to 0. Then the following result is immediate.

Note. *Under conditions 1–3 of Theorem 1,*

$$\lim_{a\to 0}\inf_{\pi\in G_{US}} P^\pi(H_0|x) = \inf_{\pi\in G^*_{US}} P^\pi(H_0|x).$$

Table 1: Summary

a^*	$\underline{P}(H_0 \mid t)$		
	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
0	.1092	.2900	.3890
.01	.1092	.2900	.3890
.02	.1089	.2894	.3886
.30	.0818	.2499	.3559

4 Conclusions

First of all, we have seen in this chapter that, it is possible to derive the lower bounds on the posterior probabilities over "objective" classes of priors for interval hypotheses. Further, these lower bounds for small interval hypotheses do approximate the corresponding lower bounds for point null hypotheses as Result 1 shows. Summary of the lower bounds graphed in Figure 1 makes the latter point even clearer. This is where the testing of interval nulls were studied in the normal case. Recall that a^* here stands for the length of the half interval in multiples of standard error.

In fact, if the standardized half length, $a^* \leq .02$, these lower bounds remain approximately the same as those for testing the point null. Even for a 3/5 standard error long interval there isn't much of a reduction in these values. This reinforces our belief that the large values of the lower bounds on the posterior probabilities of the null hypothesis are not due to assigning mass to the point null hypothesis, and that P–values are instead the suspect measures.

It was also observed in the Normal example that, as the length of the interval of the null hypothesis increases, the lower bound on the posterior probability of the null hypothesis approaches the P–value. This is interesting and is expected in many situations (cf. Casella and Berger(1987)).

Acknowledgements. The author is grateful to his doctoral thesis advisor, Professor James Berger, for all his valuable suggestions, attention and help in this work. Thanks are also due to Professors H. Rubin and T. Sellke for their help.

References

[1] Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York, Springer-Verlag.

[2] Berger, J. and Berliner, L. M. (1986), "Robust Bayes and Empirical Bayes Analysis with ϵ–Contaminated Priors", *Annals of Statistics*, **14**, 461–486.

[3] Berger, J. and Delampady, M. (1986), "Bayesian Testing of Precise Hypotheses", Tech. Report 86-38, Department of Statistics, Purdue University.

[4] Berger, J. and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of Significance Levels and Evidence", *Journal of the American Statistical Association*, **82**.

[5] Berkson, J. (1942), "Tests of Significance Considered as Evidence", *Journal of the American Statistical Association*, **37**, 325–335.

[6] Casella, G. and Berger, R. L. (1987), "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem", *Journal of the American Statistical Association*, **82**.

[7] Delampady, M. (1986a), "Testing a Precise Hypothesis: Interpreting P–values from a Robust Bayesian Viewpoint", Ph. D Thesis, Purdue University.

[8] Delampady, M. (1986c), " Lower Bounds on Bayes Factors for Invariant Testing Situations", Tech. Report 86-36, Department of Statistics, Purdue University.

[9] Dickey, J. M. (1971), "The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters", *Annals of Mathematical Statistics*, **42**, 204–223.

[10] Dickey, J. M. (1973), "Scientific Reporting", *Journal of the Royal Statistical Society B*, **35**, 285–305.

[11] Dickey, J. M. (1974), "Bayesian Alternatives to the F-test and Least Squares Estimate in the Linear Model", *Studies in Bayesian Econometrics and Statistics*, S. E. Fienberg and A. Zellner (Eds.), Amsterdam: North-Holland.

[12] Dickey, J. M. (1977), "Is the Tail Area Useful as an Approximate Bayes Factor", *Journal of the American Statistical Association*, **72**, 138–142.

[13] Dickey, J. M. and Lientz, B. P. (1970), "The Weighted Likelihood Ratio, Sharp Hypotheses About Chances, The Order of a Markov Chain", *Annals of Mathematical Statistics*, **41**, 214–226.

[14] Edwards, W., Lindman, H., and Savage, L. J. (1963), "Bayesian Statistical Inference for Psychological Research", *Psychological Review*, **70**, 193–242. (Reprinted in *Robustness of Bayesian Analyses*, J. Kadane (Eds.), Amsterdam: North-Holland.)

[15] Freedman, D. (1971), *Brownian Motion and Diffusion*, San Francisco: Holden-Day.

[16] Good, I. J. (1950), *Probability and the Weighing of Evidence*, London: Charles Griffin.

[17] Good, I. J. (1958), "Significance Tests in Parallel and in Series", *Journal of the American Statistical Association*, **53**, 799–813.

[18] Good, I. J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, Cambridge: M. I. T. Press.

[19] Good, I. J. (1967), "A Bayesian Significance Test for the Multinomial Distribution", *Journal of the Royal Statistical Society B*, **29**, 399–431.

[20] Good, I. J. (1983), *Good Thinking: The Foundations of Probability and Its Applications*, Minneapolis: University of Minnesota Press.

[21] Hildreth, C. (1963), "Bayesian Statisticians and Remote Clients", *Econometrica*, **31**, 422–438.

[22] Jeffreys, H. (1961), *Theory of Probability (3rd Edition)*, Oxford: Oxford University Press.

[23] Leamer, E. E. (1978), *Specification Searches: Ad Hoc Inference With Nonexperimental Data*, New York: Wiley.

[24] Lempers, F. B. (1971), *Posterior Probabilities of Alternative Linear Models*, Rotterdam: University of Rotterdam Press.

[25] Lindley, D. V. (1957), "A Statistical Paradox", *Biometrika*, **44**, 187–192.

[26] Lindley, D. V. (1961), "The Use of Prior Probability Distributions in Statistical Inference and Decision", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press.

[27] Lindley, D. V. (1977), "A Problem in Forensic Science", *Biometrika*, **64**, 207–213.

[28] Pratt, J. W. (1965), "Bayesian Interpretation of Standard Inference Statements (with Discussion)", *Journal of the Royal Statistical Society B*, **27**, 169–203.

[29] Smith, A. F. M. and Spiegelhalter, D. J. (1980), "Bayes Factors and Choice Criteria for Linear Models", *Journal of the Royal Statistical Society B*, **42**, 213–220.

[30] Zellner, A. (1984), "Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results", *Basic Issues in Econometrics*, Chicago: University of Chicago Press.

[31] Zellner, A. and Siow, A. (1980), "Posterior Odds Ratios for Selected Regression Hypotheses", *Bayesian Statistics*, J. M. Bernardo et. al. (Eds.), Velencia: University Press.