Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem

GEORGE CASELLA and ROGER L. BERGER*

For the one-sided hypothesis testing problem it is shown that it is possible to reconcile Bayesian evidence against H_0, expressed in terms of the posterior probability that H_0 is true, with frequentist evidence against H_0, expressed in terms of the p value. In fact, for many classes of prior distributions it is shown that the infimum of the Bayesian posterior probability of H_0 is equal to the p value; in other cases the infimum is less than the p value. The results are in contrast to recent work of Berger and Sellke (1987) in the two-sided (point null) case, where it was found that the p value is much smaller than the Bayesian infimum. Some comments on the point null problem are also given.

KEY WORDS: Posterior probability; p Value; Prior distribution.

1. INTRODUCTION

In the problem of hypothesis testing, "evidence" can be thought of as a postexperimental (data-based) evaluation of the tenability of the null hypothesis, H_0. To a Bayesian, evidence takes the form of the posterior probability that H_0 is true, while to a frequentist, evidence takes the form of the p value, or the observed level of significance of the result. If the null hypothesis consists of a single point, it has long been known that these two measures of evidence can greatly differ. The famous paper of Lindley (1957) illustrated the possible discrepancy in the normal case.

The question of reconciling these two measures of evidence has been treated in the literature. For the most part, the two-sided (point null) problem has been treated, and the major conclusion has been that the p value tends to overstate the evidence against H_0 (that is, the p value tends to be smaller than a Bayesian posterior probability). Many references can be found in Shafer (1982). Pratt (1965) did state, however, that in the one-sided testing problem the p value can be approximately equal to the posterior probability of H_0.

A slightly different approach to the problem of reconciling evidence was taken by DeGroot (1973). Working in a fairly general setting, DeGroot constructed alternative distributions and found improper priors for which the p value and posterior probability match. DeGroot assumed that the alternative distributions are stochastically ordered, which, although he did not explicitly state it, essentially put him in the one-sided testing problem.

Dickey (1977), in the two-sided problem, considered classes of priors and examined the infimum of the "Bayes factor," which is closely related to the posterior probability of H_0. He also concluded that the p value overstates the evidence against H_0, even when compared with the infimum of Bayesian measures of evidence.

A recent paper by Berger and Sellke (1987) approached the problem of reconciling evidence in a manner similar to Dickey's approach. For the Bayesian measure of evidence they considered the infimum, over a class of priors, of the posterior probability that H_0 is true. For many classes of prior it turns out that this infimum is much greater than the frequentist p value, leading Berger and Sellke to conclude that significance levels "can be highly misleading measures of the evidence provided by the data against the null hypothesis" (p. 112).

Although their arguments are compelling and may lead one to question the worth of p values, their analyses are restricted to the problem of testing a point null hypothesis. Before dismissing p values as measures of evidence, we feel that their behavior should be examined in other hypothesis testing situations.

The testing of a point null hypothesis is one of the most misused statistical procedures. In particular, in the location parameter problem, the point null hypothesis is more the mathematical convenience than the statistical method of choice. Few experimenters, of whom we are aware, want to conclude that "there is a difference." Rather, they are looking to conclude that "the new treatment is better." Thus there is a direction of interest in many experiments, and saddling an experimenter with a two-sided test would not be appropriate.

In this article we consider the problem of reconciling evidence in the one-sided testing problem. We find, in contrast to the results of Berger and Sellke, that evidence can be reconciled. For classes of reasonable, impartial priors, we obtain equality between the infimum of the Bayes posterior probability that H_0 is true and the frequentist p value. In other cases this Bayesian infimum is shown to be a strict lower bound on the p value. Thus the p value may be on the boundary or within the range of Bayesian evidence measures.

In Section 2 we present some necessary preliminaries, including the classes of priors we are considering and how they relate to those considered in the two-sided problem. Section 3 contains the main results concerning the relationship between Bayesian and frequentist evidence, and Section 4 contains comments, in particular about the case of testing a point null hypothesis.

* George Casella is Associate Professor, Biometrics Unit, Cornell University, Ithaca, NY 14853. Roger L. Berger is Associate Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695. This research was done while Professor Casella was Visiting Associate Professor at North Carolina State University. Dr. Casella's work was supported by National Science Foundation Grant MCS85-01973. Dr. Berger's work was supported in part by U.S. Army Research Office Grant DAAG 29-82-K-0168 at Florida State University. The authors thank Raymond J. Carroll for expressing curiosity about the one-sided problem, Mark Berliner for helpful discussions, Thomas Sellke for showing them Lemma 3.1, and Larry Brown for help with Theorem 3.1. They are also grateful to two unknown referees for their insightful comments. This paper is No. 1657 in the Institute of Statistics Mimeo Series, North Carolina State University, and BU-884-M in the Biometrics Unit Series, Cornell University.

Journal of the American Statistical Association
March 1987, Vol. 82, No. 397, Theory and Methods

2. PRELIMINARIES

We consider testing the hypotheses

$$H_0 : \theta \le 0 \qquad \text{versus} \qquad H_1 : \theta > 0 \tag{2.1}$$

based on observing $X = x$, where X has location density $f(x - \theta)$. Throughout this article we will often assume that (a) $f(\cdot)$ is symmetric about zero and (b) $f(x - \theta)$ has monotone likelihood ratio (MLR), but we will explicitly state these assumptions whenever used. Recall that (b) implies that $f(\cdot)$ is unimodal (Barlow and Proschan 1975, p. 76).

If $X = x$ is observed, a frequentist measure of evidence against H_0 is given by the p value

$$p(x) = \Pr(X \ge x \mid \theta = 0) = \int_x^\infty f(t)\, dt. \tag{2.2}$$

A Bayesian measure of evidence, given a prior distribution $\pi(\theta)$, is the probability that H_0 is true given $X = x$,

$$\Pr(H_0 \mid x) = \Pr(\theta \le 0 \mid x) = \frac{\int_{-\infty}^{0} f(x - \theta)\, d\pi(\theta)}{\int_{-\infty}^{\infty} f(x - \theta)\, d\pi(\theta)}. \tag{2.3}$$

Our major point of concern is whether these two measures of evidence can be reconciled, that is, can the p value, in some sense, be regarded as a Bayesian measure of evidence. Since the p value is based on the objective frequentist model, it seems that if reconciliation is possible, we must consider impartial prior distributions. By impartial we mean that the prior distribution gives equal weight to both the null and alternative hypotheses.

Four reasonable classes of distributions are given by

$$\begin{aligned}
\Gamma_A &= \{\text{all distributions giving mass } \tfrac{1}{2} \\
&\qquad \text{to each of } (-\infty, 0] \text{ and } (0, \infty)\} \\
\Gamma_S &= \{\text{all distributions symmetric about zero}\} \\
\Gamma_{US} &= \{\text{all distributions with unimodal densities,} \\
&\qquad \text{symmetric about zero}\} \\
\Gamma_{NOR} &= \{\text{all normal } (0, \tau^2) \text{ distributions, } 0 < \tau^2 < \infty\}.
\end{aligned} \tag{2.4}$$

As our Bayesian measure of evidence we consider inf $\Pr(H_0 \mid x)$, where the infimum is taken over a chosen class of priors. We then examine the relationship between this infimum and $p(x)$ to see if there is agreement. If so, then we have obtained a reconciliation of Bayesian and frequentist measures of evidence.

This development is, of course, similar to that of Berger and Sellke (1987), who considered the two-sided hypothesis test $H_0 : \theta = 0$ versus $H_1 : \theta \ne 0$. They used priors that give probability π_0 and $1 - \pi_0$ to H_0 and H_1, respectively, and spread the mass over H_1 according to a density $g(\theta)$, allowing $g(\cdot)$ to vary within a class of distributions similar to the classes in (2.4). For any numerical calculations they chose $\pi_0 = \frac{1}{2}$, asserting that this provides an impartial prior distribution. We will discuss this choice in Section 4.

For testing $H_0 : \theta \le 0$ versus $H_1 : \theta > 0$, we will mainly be concerned with evidence based on observing $x > 0$. If f is symmetric with MLR, then for $x < 0$, $p(x) > \frac{1}{2}$ and $\inf \Pr(H_0 \mid x) = \frac{1}{2}$, where the infimum is over any class in (2.4) except Γ_A. Thus, for $x < 0$, neither a frequentist nor a Bayesian would consider the data as giving evidence against H_0.

3. COMPARING MEASURES OF EVIDENCE

In this section we consider prior distributions contained in the classes given in (2.4) and various types of sampling densities. We compare $\inf \Pr(H_0 \mid x)$ with $p(x)$ under different assumptions and find many situations in which $\inf \Pr(H_0 \mid x) \le p(x)$. For the classes Γ_{US} and Γ_{NOR}, as well as some others, we show that $\inf \Pr(H_0 \mid x) = p(x)$ if f is symmetric and has MLR.

We begin with a computational lemma that will facilitate many subsequent calculations. The essence of the lemma is that $\inf \Pr(H_0 \mid x)$ is the same whether we take the infimum over a given class of priors or over the class of all mixtures of members of the class. Since many interesting classes can be expressed as mixtures of simpler distributions, this lemma will prove to be extremely helpful.

Lemma 3.1. Let $\Gamma = \{\pi_\alpha : \alpha \in \mathcal{A}\}$ be a class of prior distributions on the real line indexed by the set $\mathcal{A}$. Let Γ_M be the set of all mixtures of elements of Γ, that is,

$$\pi \in \Gamma_M \leftrightarrow \pi(B) = \int_{\mathcal{A}} \pi_\alpha(B)\, dP(\alpha)$$

for some probability measure P on $\mathcal{A}$ and all measurable B. Then

$$\inf_{\pi \in \Gamma_M} \Pr(H_0 \mid x) = \inf_{\pi_\alpha \in \Gamma} \Pr(H_0 \mid x). \tag{3.1}$$

Proof. We use the notation $\Pr_\pi(H_0 \mid x)$ to indicate that π is the prior used in calculating a posterior probability.

Consider the random triple (A, θ, x) with joint distribution defined by the following. The distribution of $X \mid \Theta = \theta$ has density $f(x - \theta)$, the distribution of $\Theta \mid A = \alpha$ is π_α, and the distribution of A is P. Then for any $\pi \in \Gamma_M$,

$$\begin{aligned}
\Pr_\pi(H_0 \mid x) &= \Pr_\pi(\Theta \le 0 \mid X = x) \\
&= E_A[\Pr(\Theta \le 0 \mid A = \alpha, X = x) \mid X = x] \\
&= E_A[\Pr_{\pi_\alpha}(\Theta \le 0 \mid X = x) \mid X = x] \\
&\ge E_A[\inf_{\alpha \in \mathcal{A}} \Pr_{\pi_\alpha}(\Theta \le 0 \mid X = x) \mid X = x] \\
&= \inf_{\alpha \in \mathcal{A}} \Pr_{\pi_\alpha}(\Theta \le 0 \mid x = x) \\
&= \inf_{\pi_\alpha \in \Gamma} \Pr(H_0 \mid x).
\end{aligned}$$

The opposite inequality is true since $\Gamma \subset \Gamma_M$, and (3.1) is established.

We note that this theorem can be proved in greater generality than is done here, but as stated it will serve our purposes.

By using Lemma 3.1 we can obtain conditions under which $p(x)$ is an upper bound on $\inf \Pr(H_0 \mid x)$ for the class Γ_S through consideration of a smaller class contained in Γ_S, $\Gamma_{2PS} = \{$all two-point distributions symmetric about 0$\}$, since Γ_S is the class of all mixtures of distributions in Γ_{2PS}.

Theorem 3.1. For the hypotheses in (2.1), if f is symmetric and has MLR and if $x > 0$, then

$$\inf_{\pi \in \Gamma_S} \Pr(H_0 \mid x) = \inf_{\pi \in \Gamma_{2PS}} \Pr(H_0 \mid x) \le p(x). \tag{3.2}$$

Proof. The equality in (3.2) follows from Lemma 3.1. For the $\pi \in \Gamma_{2PS}$ that gives probability $\frac{1}{2}$ to the two points $\theta = \pm k$ we have

$$\Pr(H_0 \mid x) = \frac{f(x+k)}{f(x-k) + f(x+k)} .$$

The assumptions on f imply that, for $x > 0$, $\Pr(H_0 \mid x)$ is decreasing in k and hence

$$\inf_{\pi \in \Gamma_{2PS}} \Pr(H_0 \mid x) = \lim_{k \to \infty} \frac{f(x+k)}{f(x-k) + f(x+k)}$$

$$= \lim_{k \to \infty} \frac{1}{1 + (f(k-x)/f(k+x))} ,$$

where we have used the symmetry of f in the second equality. For the remainder of the proof assume that $f'(t)$ exists for all t and the support of f is the entire real line. If either of these conditions fail to hold, the proof can be suitably modified.

Since f has MLR we can write $f(t) = \exp[-g(t)]$, where g is convex, that is, f is log-concave. Now

$$\begin{aligned} f(k-x)/f(k+x) &= \exp\{g(k+x) - g(k-x)\} \\ &\ge \exp\{2xg'(k-x)\}, \end{aligned} \tag{3.3}$$

by the convexity of g. Define $l = \lim_{t \to \infty} g'(t)$, which must exist since $g'(t)$ is increasing. If $l = \infty$ the theorem is trivally true, so assume that $l < \infty$. Substituting l for $g'(k - x)$ in (3.3) gives a lower bound on the ratio $f(k - x)/f(k + x)$, and it then follows that

$$\inf_{\pi \in \Gamma_{2PS}} \Pr(H_0 \mid x) \le \frac{1}{1 + e^{2lx}} .$$

Next note that for $t > 0$, the ratio

$$f(t)/e^{-lt} = e^{lt - g(t)}$$

is increasing in t, since $l \ge g'(t)$. This implies that

$$p(x) = \frac{\int_x^\infty f(t)\, dt}{2\int_0^\infty f(t)\, dt} \ge \frac{\int_x^\infty e^{-lt}\, dt}{2\int_0^\infty e^{-lt}\, dt} = \frac{1}{2} e^{-lx}$$

by an application of the Neyman–Pearson lemma together with a corollary relating power to size (Lehmann 1959, corollary 1, p. 67).

Combining this inequality with that for $\inf \Pr(H_0 \mid x)$, it is straightforward to verify that

$$p(x) \ge \frac{1}{2} e^{-lx} > \frac{1}{1 + e^{2lx}} \ge \inf_{\pi \in \Gamma_{2PS}} \Pr(H_0 \mid x),$$

proving the theorem.

For densities f whose support is the entire real line, it must be the case that $l \ne 0$, so the inequality between $\inf_{\pi \in \Gamma_{2PS}} \Pr(H_0 \mid x)$ and $p(x)$ is strict. If f has bounded support, then equality may be attained.

Table 1 gives explicit expressions for some common distributions, the first three satisfying the conditions of Theorem 3.1. Note in particular that the values calculated for the double exponential distribution are equal to the bounds obtained in the previous proof, suggesting that this distribution plays some role as a "boundary" distribution. The Cauchy distribution, which is symmetric but does not have MLR, does not attain its infimum at $k = \infty$ but rather at $k = (x^2 + 1)^{1/2}$. The exponential distribution, which has MLR but is asymmetric, attains its infimum at $k = x$. For both of these distributions $p(x)$ is greater than $\inf \Pr(H_0 \mid x)$.

We now turn to the class of distributions Γ_{US}, all priors with symmetric unimodal densities. We can, in fact, demonstrate equality between $p(x)$ and $\inf \Pr(H_0 \mid x)$ for this class. We will again use Lemma 3.1 and the fact that Γ_{US} is the set of all mixtures of $U_S = \{$all symmetric uniform distributions$\}$.

Theorem 3.2. For the hypotheses in (2.1), if f is symmetric and has MLR and if $x > 0$, then

$$\inf_{\pi \in \Gamma_{US}} \Pr(H_0 \mid x) = \inf_{\pi \in U_s} \Pr(H_0 \mid x) = p(x). \tag{3.4}$$

Proof. The first equality in (3.4) follows from Lemma

Table 1. P Values and inf $\Pr(H_0 \mid x)$ for the Classes of Symmetric Two-Point Distributions and All Symmetric Distributions ($x > 0$)

Distribution	$p(x)$	$\inf \Pr(H_0 \mid x)$
Normal	$1 - \Phi(x)$	0
Double exponential	$\frac{1}{2}e^{-x}$	$(1 + e^{2x})^{-1}$
Logistic	$(1 + e^{x})^{-1}$	$(1 + e^{2x})^{-1}$
Cauchy	$\frac{1}{2} - \dfrac{\tan^{-1}x}{\pi}$	$\dfrac{1 + [x - (x^2 + 1)^{1/2}]^2}{2 + [x - (x^2 + 1)^{1/2}]^2 + [x + (x^2 + 1)^{1/2}]^2}$
Exponential	e^{-x}	$(1 + e^{2x})^{-1}$

3.1. To prove the second equality let $\pi(\theta)$ be uniform $(-k, k)$. Then

$$\Pr(H_0 \mid x) = \frac{\int_{-k}^{0} f(x - \theta)\, d\theta}{\int_{-k}^{k} f(x - \theta)\, d\theta} \tag{3.5}$$

and

$$\frac{d}{dk} \Pr(H_0 \mid x) = \left(\frac{f(x - k) + f(x + k)}{\int_{-k}^{k} f(x - \theta)\, d\theta} \right) \times \left[\frac{f(x + k)}{f(x - k) + f(x + k)} - \Pr(H_0 \mid x) \right].$$

We will now establish that $\Pr(H_0 \mid x)$, as a function of k, has no minimum on the interior of $(0, \infty)$. Suppose that $k = k_0$ satisfies

$$\frac{d}{dk} \Pr(H_0 \mid x) \big|_{k=k_0} = 0.$$

It is straightforward to establish that the sign of the second derivative, evaluated at $k = k_0$, is given by

$$\operatorname{sgn} \frac{d^2}{dk^2} \Pr(H_0 \mid x) \big|_{k=k_0} = \operatorname{sgn} \frac{d}{dk} \frac{f(x + k)}{f(x - k) + f(x + k)} \Bigg|_{k=k_0}. \tag{3.6}$$

Since f is symmetric and has MLR, the ratio $f(x + k)/f(x - k)$ is decreasing in k for fixed $x > 0$. Therefore, the sign of (3.6) is always negative, so any interior extremum can only be a maximum. The minimum is, therefore, attained on the boundary, and it is straightforward to check from (3.5) that

$$\inf_{\pi \in U_S} \Pr(H_0 \mid x) = \lim_{k \to \infty} \frac{\int_{-k}^{0} f(x - \theta)\, d\theta}{\int_{-k}^{k} f(x - \theta)\, d\theta} = \int_{-\infty}^{0} f(x - \theta)\, d\theta = p(x).$$

In Theorem 3.2, as well as Theorem 3.3, the infimum equals the value of $\Pr(H_0 \mid x)$ associated with the improper prior, Lebesgue measure on $(-\infty, \infty)$. Indeed, the theorems are proved by considering a sequence of priors converging to this "uniform $(-\infty, \infty)$" prior. In other examples, however, such as the Cauchy and exponential examples following Theorem 3.4, the infimum is less than the value for this limiting uniform prior.

Certain subclasses of Γ_{US} might also be of interest, for example, Γ_{NOR}, the class of all normal priors with mean zero. Theorem 3.3 shows that any class, like Γ_{NOR}, that consists of all scale transformations of a bounded, symmetric, and unimodal density will have $\inf \Pr(H_0 \mid x) = p(x)$ if f is symmetric with MLR. Furthermore, by using Lemma 3.1, this equality will hold for mixtures over these classes. For example, by considering scale mixtures of normal distributions in Γ_{NOR}, we could obtain a class that included all t distributions.

Theorem 3.3. Let $g(\theta)$ be any bounded, symmetric, and unimodal prior density, and consider the class

$$\Gamma^\sigma(g) = \{\pi_\sigma : \pi_\sigma(\theta) = g(\theta/\sigma)/\sigma,\ \sigma > 0\}. \tag{3.7}$$

For the hypotheses in (2.1), if f is symmetric and has MLR and if $x > 0$, then

$$\inf_{\pi_\sigma \in \Gamma^\sigma(g)} \Pr(H_0 \mid x) = p(x).$$

Proof. Since $\Gamma^\sigma(g) \subset \Gamma_{US}$, by Theorem 3.2

$$\inf_{\pi_\sigma \in \Gamma^\sigma(g)} \Pr(H_0 \mid x) \geq p(x). \tag{3.8}$$

To establish the opposite inequality, write

$$\inf_{\pi_\sigma \in \Gamma^\sigma(g)} \Pr(H_0 \mid x) \leq \lim_{\sigma \to \infty} \Pr_{\pi_\sigma}(H_0 \mid x) = \lim_{\sigma \to \infty} \frac{\int_{-\infty}^{0} f(x - \theta) g(\theta/\sigma)\, d\theta}{\int_{-\infty}^{\infty} f(x - \theta) g(\theta/\sigma)\, d\theta}.$$

The boundedness of g allows us to apply the dominated convergence theorem to bring the limit inside the integral. Furthermore, since g is symmetric and unimodal, $\lim_{\sigma \to \infty} g(\theta/\sigma) = g_0$ (say) exists and is positive. Thus

$$\lim_{\sigma \to \infty} \Pr_{\pi_\sigma}(H_0 \mid x) = \frac{\int_{-\infty}^{0} f(x - \theta) g_0\, d\theta}{\int_{-\infty}^{\infty} f(x - \theta) g_0\, d\theta} = p(x),$$

establishing that $\inf_{\pi_\sigma \in \Gamma^\sigma(g)} \Pr(H_0 \mid x) \leq p(x)$, which together with (3.8) proves the theorem.

The conditions on g and f may be relaxed and a similar theorem can be proved. Since the proof of Theorem 3.4 is similar to that of Theorem 3.3, we omit it.

Theorem 3.4. Let f be any density, and let g be any prior that is bounded and left- and right-continuous at zero. Denote $\lim_{\theta \uparrow 0} g(\theta) = g(0^-)$ and $\lim_{\theta \downarrow 0} g(\theta) = g(0^+)$, and define the class $\Gamma^\sigma(g)$ as in (3.7). Then for the hypotheses in (2.1), if x is such that

$$\max\{g(0^-)p(x), \quad g(0^+)[1 - p(x)]\} > 0,$$

$$\inf_{\pi_\sigma \in \Gamma^\sigma(g)} \Pr(H_0 \mid x) \leq \lim_{\sigma \to \infty} \Pr_{\pi_\sigma}(H_0 \mid x) = \frac{g(0^-)p(x)}{g(0^-)p(x) + g(0^+)[1 - p(x)]}. \tag{3.9}$$

Note in particular that in Theorem 3.4, if $g(0^+) = g(0^-)$, then the right-most expression in (3.9) is $p(x)$. This shows that for any location sampling density the infimum over such classes of scale transformations is bounded above by the p value. If f is not symmetric or does not have MLR, then strict inequality may obtain in (3.9). We will mention two examples. For both, Theorem 3.4 implies that $\inf \Pr(H_0 \mid x) \le p(x)$, but, in fact, the inequality is strict. For each example we let g be the uniform$(-1, 1)$ density so that $\Gamma^\sigma(g) = U_S$. Let $\pi_k \in U_S$ denote the uniform $(-k, k)$ density.

Let f be a Cauchy density, which is symmetric but does not have MLR. For π_k it is straightforward to calculate

$$\Pr_{\pi_k}(H_0 \mid x) = \frac{\tan^{-1}(x + k) - \tan^{-1}(x)}{\tan^{-1}(x + k) - \tan^{-1}(x - k)}.$$

For fixed $x > 0$, $\Pr_{\pi_k}(H_0 \mid x)$ is not monotone in k, but rather attains a unique minimum at a finite value of k. Table 2 lists the minimizing values of k, $\inf \Pr(H_0 \mid x)$, and the p value for selected values of x. Examination of Table 2 shows that $\inf \Pr(H_0 \mid x) < p(x)$; this observation held true for more extensive calculations that are not reported here.

For our second example, let f be an exponential location density that has MLR but is asymmetric. For $x > 0$ and $\pi_k \in U_S$ we have

$$\Pr_{\pi_k}(H_0 \mid x) = [\exp(k) - 1]/\{\exp[k + \min(k, x)] - 1\},$$

which is minimized (in k) at $k = x$, with minimum

$$\inf \Pr(H_0 \mid x) = (e^x - 1)/(e^{2x} - 1) < e^{-x} = p(x).$$

So again, strict inequality obtains in (3.9).

In fact, for small values of x, the p value can be regarded as a conservative Bayesian measure in this example. It is straightforward to calculate

$$\sup \Pr(H_0 \mid x) = \max\{\tfrac{1}{2}, e^{-x}\} = \max\{\tfrac{1}{2}, p(x)\},$$

so, in particular, if $x \le \log 2$, then $p(x)$ is larger than $\Pr(H_0 \mid x)$ for every prior in the class.

Finally, we turn to the class Γ_A, which contains all distributions giving mass $\frac{1}{2}$ to each of H_0 and H_1 and might be considered the broadest class of impartial priors. This class, however, is really too broad to be of any practical interest, since, for any density f, $\inf \Pr(H_0 \mid x) = 0$. To verify this, let g be any bounded density in Γ_A with $g(0^-) = 0$ and $g(0^+) > 0$. Then if $p(x) < 1$, Theorem 3.4 shows that $\inf_{\pi\in\Gamma_A} \Pr(H_0 \mid x) = \inf_{\pi\in\Gamma^\sigma(g)} \Pr(H_0 \mid x) = 0$.

The restriction that the priors give equal probability to H_0 and H_1, however, has little weight in the previous argument. A prior, g, could assign probability arbitrarily near one to H_0 and still we would have $\inf_{\Gamma^\sigma(g)} \Pr(H_0 \mid x) = 0$ if $g(0^-) = 0$ and $g(0^+) > 0$. It is important to note that, for any class of priors Γ possessing densities, if the class is closed under scale transformations, then Theorem 3.4 gives an upper bound on $\inf \Pr(H_0 \mid x)$ that depends only on the local behavior of g, the density of any element of Γ, at 0.

Table 2. P Values and inf $\Pr(H_0 \mid x)$ for X ~ Cauchy, Infimum Over U_s

x	k_{min}	$p(x)$	inf $\Pr(H_0 \mid x)$
.2	2.363	.437	.429
.4	2.444	.379	.363
.6	2.570	.328	.306
.8	2.727	.285	.260
1.0	2.913	.250	.222
1.2	3.112	.221	.192
1.4	3.323	.197	.168
1.6	3.541	.178	.148
1.8	3.768	.161	.132
2.0	3.994	.148	.119
2.5	4.572	.121	.094
3.0	5.158	.102	.077
3.5	5.746	.089	.065
4.0	6.326	.078	.056
5.0	7.492	.063	.044
10.0	13.175	.032	.020
25.0	29.610	.013	.007
50.0	56.260	.006	.004
75.0	82.429	.004	.002
100.0	108.599	.003	.002

4. COMMENTS

For the problem of testing a one-sided hypothesis in a location-parameter family, it is possible to reconcile measures of evidence between the Bayesian and frequentist approaches. The phrase "the probability that H_0 is true" has no meaning within frequency theory, but it has been argued that practitioners sometimes attach such a meaning to the p value. Since the p value, in the cases considered, is an upper bound on the infimum of $\Pr(H_0 \mid x)$ it lies within or at the boundary of a range of Bayesian measures of evidence demonstrating the extent to which the Bayesian terminology can be attached. In particular, for the Cauchy (non-MLR) and exponential (asymmetric) sampling densities we found that, for various classes of priors, $\inf \Pr(H_0 \mid x) < p(x)$ so that $p(x)$ is, in fact, equal to $\Pr(H_0 \mid x)$ for some prior in the class (the prior depending on x).

The discrepancies observed by Berger and Sellke (1987) in the two-sided (point null) case do not carry over to the problems considered here. This leads to the question of determining what factors are crucial in differentiating the two problems. It seems that if some prior mass is concentrated at a point (or in a small interval) and the remainder is allowed to vary over H_1, then discrepancies between Bayesian and frequentist measures will obtain. In fact, Berger and Sellke note that for testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, the p value and the Bayesian infimum are quite different. [For example, for $X \sim n(\theta, 1)$, an observed $x = 1.645$ will give a p value of .05, while over all priors for which mass $\frac{1}{2}$ is concentrated at zero, $\inf \Pr(H_0 \mid x = 1.645) = .21$.]

Seen in another light, however, placing a point mass of $\frac{1}{2}$ at H_0 may not be representative of an impartial prior distribution. For the problem of testing $H_0 : \theta \le 0$ versus $H_1 : \theta > 0$, consider priors of the form

$$\pi(\theta) = \pi_0 h(\theta) + (1 - \pi_0)g(\theta), \tag{4.1}$$

where π_0 is a fixed number and $h(\theta)$ and $g(\theta)$ are proper prior densities on $(-\infty, 0]$ and $(0, \infty)$, respectively. It then follows that, if f is unimodal with mode 0 and $x > 0$,

$$\sup_h \Pr(H_0 \mid x) = \sup_h \times \frac{\pi_0 \int_{-\infty}^{0} f(x-\theta)h(\theta)\,d\theta}{\left[\pi_0 \int_{-\infty}^{0} f(x-\theta)h(\theta)\,d\theta + (1-\pi_0)\int_{0}^{\infty} f(x-\theta)g(\theta)\,d\theta\right]}$$

$$= \frac{\pi_0 f(x)}{\pi_0 f(x) + (1-\pi_0)\int_0^{\infty} f(x-\theta)g(\theta)\,d\theta}, \quad (4.2)$$

and the last expression is equal to $\Pr(H_0 \mid x)$ for the hypotheses $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ with prior giving mass π_0 to $\theta = 0$ and having density $(1 - \pi_0)g(\theta)$ if $\theta > 0$. Thus concentrating mass on the point null hypothesis is biasing the prior in favor of H_0 as much as possible (for fixed g) in this one-sided testing problem.

The calculation in (4.2) casts doubt on the reasonableness of regarding $\pi_0 = \frac{1}{2}$ as impartial. In fact, it is not clear to us if any prior that concentrates mass at a point can be viewed as an impartial prior. Therefore, it is not surprising that the p value and Bayesian evidence differ in the normal example given previously. Setting $\pi_0 = \frac{1}{2}$ actually reflects a bias toward H_0, which is reflected in the Bayesian measure of evidence.

Indeed, any class of priors that fixes the probability distribution on one hypothesis and allows the probability distribution on the other hypothesis to vary might lead to extreme posterior probabilities. For example, consider prior densities of the form

$$\pi(\theta) = \pi_0 h(\theta/\sigma_1)/\sigma_1 + (1 - \pi_0)g(\theta/\sigma_2)/\sigma_2,$$

where h and g are as defined previously. Then under conditions similar to those of Theorem 3.3, if σ_2 is fixed,

$$\lim_{\sigma_2\to\infty} \Pr(H_0 \mid x) = 1,$$

but if σ_2 is fixed, then

$$\lim_{\sigma_1\to\infty} \Pr(H_0 \mid x) = 0.$$

Clearly, there are classes of priors for which there are large discrepancies between $\inf \Pr(H_0 \mid x)$ and $p(x)$; the fact remains, however, that reconciliation of measures of evidence is possible between the Bayesian and frequentist approaches. Since these measures can overlap one another, interpretations of one school of thought can have meaning within the other and, contrary to the message of Berger and Sellke, p values may not always overstate evidence against H_0 in that $\Pr(H_0 \mid x) < p(x)$ for some priors under consideration.

[Received November 1985. Revised January 1986.]

REFERENCES

Barlow, R., and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart & Winston.

Berger, J. O., and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence," *Journal of the American Statistical Association*, 82, 112–122.

DeGroot, M. H. (1973), "Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or Likelihood Ratio," *Journal of the American Statistical Association*, 68, 966–969.

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

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley.

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

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

Shafer, G. (1982), "Lindley's Paradox," *Journal of the American Statistical Association*, 77, 325–351.