A Comparison of Minimal Bayesian Tests of Precise Hypotheses[1]

by

James Berger
Department of Statistics
Purdue University

Technical Report # 90–32C

Department of Statistics
Purdue University

July 1990

[1]Research supported by the National Science Foundation, Grants DMS–8717799 and DMS–8923071.

△

A COMPARISON OF MINIMAL BAYESIAN TESTS OF PRECISE HYPOTHESES

by

James Berger
Department of Statistics
Purdue University

ABSTRACT

The Bayesian viewpoint on precise hypothesis testing is reviewed, with special emphasis on the conflicts with classical P-values that arise. Three methods of actually doing the Bayesian analysis, based on minimal prior input, are presented and compared. These recommended approaches use conjugate or noninformative priors, with the data interpreter required only to specify the prior variance, prior credibility region, or prior density at the null. These methods are illustrated on the problems of testing a normal mean, a binomial proportion, and a multinomial parameter vector. For the latter problem, a particularly efficient scheme for calculation, via Monte Carlo simulation, is developed.

Key words: precise hypotheses, Bayesian testing, minimal prior inputs, Bayes factors, credible sets, conjugate priors, noninformative priors, restricted priors, Monte-Carlo simulation.

1. INTRODUCTION AND BASICS

1.1 Basics and Measures of Evidence

Suppose X having density $f(x|\theta)$ is observed, with θ being an unknown element of the parameter space Θ, and that it is desired to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. (It is shown in Berger and Delampady, 1987, that this is often a good approximation to many realistic scenarios concerning testing of an imprecise hypothesis.) We consider three measures of evidence against H_0, the classical P-value, the weighted likelihood ratio or Bayes factor and the Bayesian posterior probability of H_0.

P-value. Let $T(X)$ be a test statistic, extreme values of which are deemed to be evidence against H_0. If $X = x$ is observed, with corresponding $t = T(x)$, the P-value (or observed significance level) is

$$\alpha = P_{\theta_0}(|T(X)| \geq |t|). \tag{1.1}$$

Bayes factor. Let $g(\theta)$ be a continuous density on $\{\theta \neq \theta_0\}$. Then the Bayes factor, or weighted likelihood ratio of H_0 to H_1, is

$$B = \frac{f(x|\theta_0)}{m_g(x)}, \tag{1.2}$$

where

$$m_g(x) = \int f(x|\theta)g(\theta)d\theta. \tag{1.3}$$

For a Bayesian, g would be the prior density for θ, conditional on H_1 being true. For a likelihoodist, g might be thought of merely as some weight function to allow the computation of an average likelihood for H_1. B might then be called a "weighted likelihood ratio"

for the two hypotheses. Its interpretation is similar to that of a usual likelihood ratio; e.g., a value of $B = 1/10$ means that H_1 is supported ten times as much by the data as is H_0.

Posterior probability. If a Bayesian specifies, in addition to g, the prior probability of H_0, to be denoted by π_0, then the posterior probability of H_0 is

$$P(H_0|x) = \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{m_g(x)}{f(x|\theta_0)}\right]^{-1} = \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{1}{B}\right]^{-1}. \tag{1.4}$$

Example 1. Suppose we observe $\overline{X} \sim N(\theta, \sigma^2/n)$, where σ^2 is known. Then, letting

$$T(\overline{X}) = \sqrt{n}(\overline{X} - \theta_0)/\sigma,$$

one obtains the usual P-value as

$$\alpha = 2[1 - \Phi(|t|)],$$

where Φ is the standard normal cumulative distribution function (cdf).

An easy to analyze density g is the $N(\mu, \tau^2)$ density. Calculation yields that

$$B = \sqrt{1+\rho^{-2}}\exp\left\{-\frac{1}{2}\left[\frac{(t-\rho\eta)^2}{(1+\rho^2)} - \eta^2\right]\right\}, \tag{1.6}$$

where $\rho = \sigma/(\sqrt{n}\tau)$ and $\eta = (\theta_0 - \mu)/\tau$. Note that μ will often by chosen to be θ_0 (so as to have a symmetric "weight function"), in which case

$$B = \sqrt{1+\rho^{-2}}\exp\left\{-\frac{1}{2}\left[\frac{t^2}{(1+\rho^2)}\right]\right\}. \tag{1.7}$$

The posterior probability of H_0 can be found from these formulas and (1.4), provided π_0 is specified.

As a specific example, suppose $\mu = \theta_0, \tau = \sigma$ and $\tau_0 = \frac{1}{2}$. For various t and n, the various measures of evidence are given in Table 1. There P stands for $P(H_0|x)$.

TABLE 1
Measures of evidence, normal example

		n											
t	α	1		5		10		20		50		100	
		B	P	B	P	B	P	B	P	B	P	B	P
1.645	.10	.72	.42	.79	.44	.89	.47	1.27	.56	1.86	.65	2.57	.72
1.960	.05	.54	.35	.49	.33	.59	.37	.72	.42	1.08	.52	1.50	.60
2.576	.01	.27	.21	.15	.13	.16	.14	.19	.16	.28	.22	.37	.27
3.291	.001	.10	.09	.03	.03	.02	.02	.03	.03	.03	.03	.05	.05

1.2 Motivation and History

The main motivation for considering this problem is that the classical P-value and the likelihood or Bayesian answers typically disagree. This is indicated in Table 1 where, for instance, $P(H_0|x)$ is from 5 to 50 times larger than the P-value, α. The Bayesian analysis here is close to that recommended by Jeffreys (1961) as a "standard" Bayesian significance test. (Jeffreys chose a Cauchy form for the prior, but this makes a substantial difference only when $|t|$ is large.) Thus, if $n = 50$ and $t = 1.960$, Jeffreys would conclude that H_0 has probability .52 of being true, although the classical statistician would "reject H_0 at significance level $\alpha = .05$." Admittedly, classical statisticians will warn against interpreting α as the probability that H_0 is true, but surely the classicist feels that $\alpha = .05$ is reasonable cause to doubt H_0, in marked contrast to the Bayesian conclusion. This is perhaps the simplest problem where the Bayesian and classical statistician are in fundamental practical disagreement, and as such, the problem deserves intense study. (Our label "classical

statistician" is admittedly ambiguous; there are statisticians who consider themselves to be "classical," and yet do not view P-values as meaningful measures of evidence, and there are Bayesians who view P-values as useful measures of evidence.)

This phenomenon, of large differences between P-values and likelihood or Bayesian measures in precise hypothesis testing, is very general, as has been documented in Edwards, Lindman, and Savage (1963), Berger and Sellke (1987), Berger and Delampady (1987), Delampady (1989a, b), Delampady and Berger (1990), and the references therein. Indeed, it is shown in these papers that large differences typically exist for *any* reasonable conditional prior g on H_1. When a classical answer disagrees with all sensible Bayesian answers, the classical answer is highly suspect. Indeed, in Berger and Sellke (1987) and Berger and Delampady (1987) it is argued that the P-value is seriously flawed in the sense of conditioning, since it replaces the actual data by the set of all data as or more extreme than that observed. The seemingly inevitable conclusion is that P-values must be replaced by likelihood or Bayesian measures for precise hypothesis testing.

The viewpoint adopted here is that of seeking a "minimal" Bayesian analysis. One of the main attractions of P-values is their simplicity, and so simple Bayesian alternatives are sought. The two facets of Bayesian analysis that we will be concerned with are (i) Simplicity of prior input, and (ii) Calculational simplicity.

The simplest Bayesian analyses are typically "noninformative" prior analyses. Unfortunately, such analyses are usually useless for precise hypothesis testing. In Example 1, for instance, a typical noninformative prior analysis would be done by letting the prior variance, τ^2, go to infinity. From (1.6) or (1.7), however, it can be seen that then $B \to \infty$,

yielding a useless measure of evidence.

There have been efforts at deriving "automatic" priors for use in Bayesian testing of precise hypotheses. These include the original development in Jeffreys (1961), and extensions or alternatives in Zellner and Siow (1980) and Smith and Spiegelhalter (1980). These developments proceed by arguing that "if one must specify a default g for automatic use, then a good such g is ...". In Example 1, Jeffreys argued for a *Cauchy* (θ_0, σ^2) default g, although Smith and Spiegelhalter argued for a constant default g (but a particular constant); these default g actually often give very similar answers.

We agree that, if one must produce an automatic Bayesian significance test, then the Jeffreys, Zellner-Siow or Smith-Spiegelhalter tests are quite satisfactory. Furthermore, we feel that automatic use of such tests is vastly superior to automatic use of P-values, for reasons to be made clear later. Nevertheless, we would argue that either test imposes a particular and highly informative g on the user, and as such cannot claim to be noninformative.

In support of this claim, consider the choice of g in Example 1. It could be argued that $\mu = \theta_0$ is a reasonable "noninformative" choice, but what can be done about τ^2? The Bayes factor, B, depends strongly on τ^2, and we have seen that $\tau^2 = \infty$ yields a ridiculous answer. It seems inescapable that the user must be required to specify τ^2. This would be the "minimal" prior input required in this example. Note that, intuitively, this minimal prior input corresponds to the degree of departure from θ_0 that is anticipated, if H_1 is true.

The "automatic" priors such as that of Jeffreys are, in essence, obtained by rather arbitrarily (but not unreasonably) specifying a value for the "minimal" input. Jeffreys analysis corresponds quite closely to choosing $\tau^2 = \pi\sigma^2/2$ (see the reply to Zellner in Berger and Delampady, 1987). While not unreasonable (it at least puts things on the right "scale") this choice corresponds to a quite specific prior belief. We feel that it is better to acknowledge the necessity of τ^2 as an input, and require its specification. (Note that, in reporting the result of a study, one would typically, say, graph B as a function of τ^2, so that any reader could determine his own B.)

There is a substantial literature on the subject of Bayesian testing of a precise hypothesis. Among the many references to analyses with particular priors, as in Example 1, are Jeffreys (1957, 1961), Good (1958, 1965, 1967, 1983 1985, 1986), Lindley (1957, 1961, 1965, 1977), Raiffa and Schlaiffer (1961), Edwards, Lindman and Savage (1963), Smith (1965), Zellner (1971, 1984), Dickey (1971, 1973, 1974, 1980), Lempers (1971), Leamer (1978), Smith and Spiegelhalter (1980), Zellner and Siow (1980), Diamond and Forrester (1983) and Gómez and de la Horra Navarro (1984). Many of these works specifically discuss the relationship of $P(H_0|x)$ to significance levels; other papers in which such comparisons are made include Hildreth (1963), Pratt (1965), DeGroot (1973), Dempster (1973), Dickey (1977), Bernardo (1980), Hill (1982), Shafer (1982), Good (1984), Berger (1986), Berger and Sellke (1987), Casella and Berger (1987), Delampady (1989a, 1989b), Berger and Delampady (1987), and Delampady and Berger (1990).

1.3 Overview

In this paper, we consider three "minimal input" Bayesian analyses for precise hypoth-

esis testing. These correspond to three possible choices for classes of conditional priors, g, on H_1. The classes considered are a conjugate class, a domain restricted noninformative class, and a multiplicative noninformative class. These classes are introduced in Section 2.1. Section 2.2 considers application of these classes to Example 1; Section 2.3 deals with their application to the binomial problem; and Section 2.4 considers the multinomial situation. For the latter problem, a new Monte-Carlo simulation technique is developed to perform needed numerical integrations. Section 3 presents conclusions and additional comments.

From now on, we will focus on the Bayes factor B, rather than $P(H_0|x)$. The reason is primarily our goal of "minimality"; $P(H_0|x)$ also requires specification of π_0. While specification of π_0 may be necessary for an individual trying to make a decision about H_0, the data influences the decision only through B; hence it suffices to present B as the output of the data.

2. THE MINIMAL BAYESIAN ANALYSES

2.1 Three Classes of Priors

One can, of course, specify g in (1.2) in a completely subjective fashion, and we certainly recommend that such be done when possible. When this cannot be done, g must be selected from a simple class of priors. Here are three classes that have been considered:

$$\mathcal{G}_{\mathcal{C}} = \{g_\tau,\ 0 < \tau < \overline{\tau} :\ g_\tau \text{ is a conjugate prior with mean } \theta_0 \text{ and standard deviation } \tau\};$$

$$\mathcal{G}_{\mathcal{RN}} = \{c_r \pi^*(\theta) 1_{\Theta_r}(\theta),\ 0 < r < \infty :\ \pi^* \text{ is a noninformative prior, } 1_{\Theta_r} \text{ is the indi-}$$

cator function on a compact prior credible set Θ_r "centered" at θ_0, and c_r is the normalization constant} ;

$\mathcal{G}_{\mathcal{N}} = \{c\pi^*(\theta)/\pi^*(\theta_0),\ 0 < c < \infty :\ \pi^*$ is an improper noninformative prior, and c is the desired density of g at $\theta_0\}$.

A few words about the rationales for these three classes are in order. (Additional discussion can be found in Section 3.) The classes $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{RN}}$ are based on the idea that the "minimal" input is a specification, in some sense, of the believed possible "distance" of θ from θ_0 under H_1. For $\mathcal{G}_{\mathcal{C}}$, "distance from θ_0" is measured by standard deviation τ; here $\overline{\tau}$ is the largest (often ∞) standard deviation that can be attained by a conjugate prior with mean θ_0. When θ is a vector, as in Section 2.4, it will be more convenient to let τ stand for a covariance matrix multiplier that is related to the standard deviations.

For $\mathcal{G}_{\mathcal{RN}}$, "distance from θ_0" is measured by choice of a credible set Θ_r in which it is thought that θ is "likely" to lie (under H_1). For instance, in Example 1, a natural choice for Θ_r is $\Theta_r = \{\theta : |\theta - \theta_0| \leq r\}$. It is convenient to define "likely" by specifying that r should be chosen so that Θ_r is an 80% prior credible set. This choice is clearly arbitrary, but has the advantage of roughly calibrating $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{RN}}$. Once r is chosen, it is reasonable to select the prior to be noninformative on Θ_r.

The motivation for $\mathcal{G}_{\mathcal{N}}$ is quite different, since it consists only of improper priors. The class arises out of the observation that, for a fixed smooth density g on H_1 and a fixed P-value, the Bayes factor can be approximated for moderate to large sample sizes by

$$B \cong \frac{f(x|\theta_0)}{g(\theta_0)\int f(x|\theta)d\theta}.$$

In this situation (which is not uncommon), the only feature of g that is required is $g(\theta_0)$. This suggests using $g(\theta_0)$ as the "minimal" Bayesian input; for calculational convenience, it is often appealing to choose the prior corresponding to this input to be a noninformative prior with the given density at θ_0. See Smith and Spiegelhalter (1980) for an alternative motivation.

Although apparently the simplest of the classes of priors, subjectively choosing $g(\theta_0)$ is not easy. Indeed, the only technique for subjectively choosing $g(\theta_0)$ that we are aware of is to develop a proper subjective prior and let $g(\theta_0)$ be its density at θ_0. In the following we will assume that one either specifies τ from $\mathcal{G}_{\mathcal{C}}$ or r from $\mathcal{G}_{\mathcal{RN}}$, in which case $g(\theta_0)$ is either $g_\tau(\theta_0)$ or $c_r\pi^*(\theta_0)$. (The main reason for using $\mathcal{G}_{\mathcal{N}}$ after specifying τ or r would be calculational simplicity.)

2.2 Application to Testing a Normal Mean

Consider the scenario of Example 1. The natural choices for the three classes of priors are (since $\pi^*(\theta) = 1$ is the usual noninformative prior)

$$\mathcal{G}_{\mathcal{C}} = \{g_\tau = \text{ the } \mathcal{N}(\theta_0, \tau^2) \text{ density}, \ 0 < \tau < \infty\},$$

$$\mathcal{G}_{\mathcal{RN}} = \{g_r = \text{ the Uniform } (\theta_0 - r, \theta_0 + r) \text{ density}, \ 0 < r < \infty\};$$

$$\mathcal{G}_{\mathcal{N}} = \{g_c = \text{ the constant (improper) density } c \text{ on } (-\infty, \infty), \ 0 < c < \infty\}.$$

For priors in these classes, the Bayes factors are given, respectively, by

$$B_{\mathcal{C}}(\tau) = \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2}t^2/(1 + \frac{\sigma^2}{n\tau^2})\right\},$$

$$B_{\mathcal{RN}}(r) = \frac{2\sqrt{n}}{\sigma} r\phi(t) / \left[\Phi\left(\frac{\sqrt{n}r}{\sigma} - |t|\right) - \Phi\left(-\frac{\sqrt{n}r}{\sigma} - |t|\right)\right],$$

$$B_{\mathcal{N}}(c) = \frac{\sqrt{n}}{\sigma} c^{-1}\phi(t);$$

recall that $t = \sqrt{n}(\overline{x} - \theta_0)/\sigma$, and ϕ and Φ stand for the standard normal density and c.d.f., respectively.

To compare these Bayes factors, we relate τ, r, and c as follows. First, to relate τ and r recall the suggestion that r be chosen so that $(\theta_0 - r, \theta_0 + r)$ is an 80% credible set for θ under H_1. If g were $g_\tau = \mathcal{N}(\theta_0, \tau^2)$, the 80% credible set would be $(\theta_0 - (1.28)\tau, \theta_0 + (1.28)\tau)$, so the natural relationship is $r = (1.28)\tau$. The two suggestions for c given in Section 2.1 were $g_\tau(\theta_0)$ and $g_r(\theta_0)$, which for $r = (1.28)\tau$ can both be shown to approximately equal $c = 1/[(2.5)\tau]$; hence we use this choice in the following.

With the above choices of r and c, the Bayes factors $B_{\mathcal{C}}, B_{\mathcal{RN}}$, and $B_{\mathcal{N}}$ can be written as the following functions of $\tau^* = \sqrt{n}\tau/\sigma$:

$$B_{\mathcal{C}}(\tau) = (1 + \tau^{*^2})^{1/2} \exp\{-\frac{1}{2}t^2/(1 + \tau^{*-2})\},$$

$$B_{\mathcal{RN}}(r) = (2.56)\tau^*\phi(t)/[\Phi((1.28)\tau^* - |t|) - \Phi(-(1.28)\tau^* - |t|)],$$

$$B_{\mathcal{N}}(c) = (2.5)\tau^*\phi(t).$$

Figures 1, 2, and 3 graph these three Bayes factors, as functions of τ^*, for $t = 1.645, t = 1.96$, and $t = 2.576$, respectively (corresponding to P-values of 0.10, 0.05, and 0.01.) Note that $B_{\mathcal{C}}$ and $B_{\mathcal{RN}}$ are very similar, and both agree with $B_{\mathcal{N}}$ for large $\tau^* = \sqrt{n}\tau/\sigma$. For small τ^*, $B_{\mathcal{RN}}$ behaves quite differently, and should not be used; recall that the motivation for $B_{\mathcal{RN}}$ was asymptotic, as $n \to \infty$.

Note also the fact that the Bayes factors tend to be much higher than the corresponding P-values (except for the inappropriate $B_{\mathcal{N}}$ for small τ^*). This, again, is one reason that it is very important to develop Bayes factors in precise hypothesis testing.

Note, finally, that, for this problem, Jeffreys (1961) proposed use of Cauchy priors. As

mentioned earlier, his results are essentially the same (except for quite large $|t|$) as those obtained by use of a $\mathcal{N}(\theta_0, \pi\sigma^2/2)$ prior. We do not consider the class of Cauchy priors because computation of B then requires numerical integration. Also, the Cauchy class has no significant advantages over the classes considered here.

2.3 Application to Testing a Binomial Proportion

Suppose $X \sim$ Binomial (n, θ), so that $0 \leq \theta \leq 1$ and

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \ x = 0, 1, \ldots, n.$$

Then the Bayes factor (1.2) is given by

$$B = \frac{\theta_0^x(1-\theta_0)^{n-x}}{\int_0^1 \theta^x(1-\theta)^{n-x}g(\theta)d\theta}. \tag{2.1}$$

Natural conjugate densities are Beta densities; the Beta density with mean θ_0 and standard deviation τ is (on (0, 1))

$$g_\tau(\theta) = \frac{\Gamma(K_\tau)}{\Gamma(\theta_0 K_\tau)\Gamma((1-\theta_0)K_\tau)}\theta^{[\theta_0 K_\tau - 1]}(1-\theta)^{[(1-\theta_0)K_\tau - 1]}, \tag{2.2}$$

where $K_\tau = [\tau^{-2}\theta_0(1-\theta_0) - 1]$ and Γ is the Gamma function. Only standard deviations $0 < \tau < \sqrt{\theta_0(1-\theta_0)}$ yield proper Beta densities, so the conjugate class becomes

$$\mathcal{G}_\mathcal{C} = \{g_\tau \text{ in (2.2)}, \ 0 < \tau < \sqrt{\theta_0(1-\theta_0)}\}. \tag{2.3}$$

The Bayes factor for g_τ can easily be shown to be (defining $\prod_{i=0}^{-1} a_i \equiv 1$)

$$\begin{aligned} B_\mathcal{C}(\tau) &= \frac{\Gamma(\theta_0 K_\tau)\Gamma((1-\theta_0)K_\tau)\Gamma(n+K_\tau)\theta_0^x(1-\theta_0)^{n-x}}{\Gamma(K_\tau)\Gamma(x+\theta_0 K_\tau)\Gamma(n-x+(1-\theta_0)K_\tau)} \\ &= \frac{\prod_{i=0}^{n-1}(K_\tau + i)}{\left[\prod_{i=0}^{x-1}\left(K_\tau + \frac{i}{\theta_0}\right)\right]\left[\prod_{i=0}^{n-x-1}\left(K_\tau + \frac{i}{(1-\theta_0)}\right)\right]}. \end{aligned} \tag{2.4}$$

To develop $\mathcal{G}_{\mathcal{RN}}$ and $\mathcal{G}_{\mathcal{N}}$ for the binomial problem, it is necessary to select a noninformative prior for θ. The three common choices are $\pi(\theta) = 1$, $\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, and

$$\pi^*(\theta) = \theta^{-1}(1-\theta)^{-1};$$

the last is improper and works best for our purposes, because it allows one to choose compact sets Θ_r for which any θ_0 is the mean or median for priors in $\mathcal{G}_{\mathcal{RN}}$. The main disadvantage of π^* is that its use in $\mathcal{G}_{\mathcal{N}}$ requires that $0 < x < n$.

One could choose a variety of different Θ_r to define $\mathcal{G}_{\mathcal{RN}}$. Here we consider

$$\Theta_r = ([1 + e^r(\theta_0^{-1} - 1)]^{-1}, [1 + e^{-r}(\theta_0^{-1} - 1)]^{-1}), \tag{2.5}$$

so that

$$c_r = 1/\int_{\Theta_r} \pi^*(\theta)d\theta = 1/(2r); \tag{2.6}$$

this follows immediately from the observation that the density of

$$\xi = \log\frac{\theta}{(1-\theta)} - \log\frac{\theta_0}{(1-\theta_0)} \tag{2.7}$$

is $\pi(\xi) = 1$ (on $(-\infty, \infty)$), if θ has density $\pi^*(\theta)$, and that Θ_r transforms into $\xi \in (-r, r)$. In fact, this was the motivation for consideration of Θ_r, being as the above log odds transformation "centers" the problem at θ_0 and is often thought to be a "normalizing" transformation. (A somewhat better normalizing transformation was used to define analogous regions in Berger and Delampady (1987), but the choice here is computationally more convenient.)

The Bayes factors resulting from use of $g_c \in \mathcal{G}_{\mathcal{N}}$ and $g_r \in \mathcal{G}_{\mathcal{RN}}$ are, respectively (and recalling that we assume $0 < x < n$ for $\mathcal{G}_{\mathcal{N}}$),

$$B_{\mathcal{N}}(c) = \frac{\theta_0^x(1-\theta_0)^{n-x}}{c\theta_0(1-\theta_0)\int_0^1 \theta^{(x-1)}(1-\theta)^{(n-x-1)}d\theta} = \binom{n-2}{x-1}(n-1)c^{-1}\theta_0^{(x-1)}(1-\theta_0)^{(n-x-1)}, \tag{2.8}$$

$$B_{\mathcal{RN}}(r) = 2r\theta_0^x(1-\theta_0)^{(n-x)}\Big/\int_{\Theta_r}\theta^{(x-1)}(1-\theta)^{(n-x-1)}d\theta = \frac{2r\theta_0^x(1-\theta_0)^{(n-x)}}{\sum_{i=0}^{(n-x-1)}\binom{n-x-1}{i}\frac{(-1)^i}{(x+i)}[\mu^{(x+i)}-\ell^{(x+i)}]}, \tag{2.9}$$

where (ℓ, μ) is the interval in (2.5) and $\binom{m}{q} = m!/[q!(m-q)!]$ is the usual binomial coefficient (with $\binom{m}{0} \equiv 1$).

In Section 2.1, it was suggested that r be chosen to yield an 80% prior credible set for θ, while one suggestion for c was to choose τ for $g_\tau \in \mathcal{G}_{\mathcal{C}}$, and then set $c = g_\tau(\theta_0)$. This could again result in putting the "minimal inputs" on a similar scale for comparison of the corresponding Bayes factors, but there are technical difficulties in doing so. Hence we, instead, present separate graphs of $B_{\mathcal{C}}(\tau)$ and $B_{\mathcal{RN}}(r)$ for three different situations, each corresponding to three different P-values, α. The three situations are:

(i) $\alpha = 0.096$, with $n = 15, x = 6$, and $\theta_0 = 0.20$;

(ii) $\alpha = 0.049$, with $n = 50, x = 16$, and $\theta_0 = 0.20$;

(iii) $\alpha = 0.010$, with $n = 20, x = 9$, and $\theta_0 = 0.20$.

Figures 4 and 5 graph $B_{\mathcal{C}}(\tau)$ and $B_{\mathcal{RN}}(r)$, respectively, for each of these three cases. Note that the range of possible standard deviations, τ, is $0 < \tau < \sqrt{\theta_0(1-\theta_0)} = 0.4$. We have not graphed $B_{\mathcal{N}}(c)$, since it is of the simple form K/c; for cases (i), (ii), and (iii),

K is, respectively, 0.967, 1.11, and 0.229. To interpret these, note that $c \approx 5$ would be a reasonable value.

Figures 4 and 5 show that the classes $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{RN}}$ behave very similarly. The class $\mathcal{G}_{\mathcal{RN}}$ is quite a bit larger, so the minimum of $B_{\mathcal{RN}}(r)$ is smaller than that of $B_{\mathcal{C}}(\tau)$, but not dramatically so. As expected, all minimum Bayes factors are dramatically larger than the corresponding P-values. Finally, it may seem surprising that the Bayes factor corresponding to $\alpha = 0.049$ is *larger* than that corresponding to $\alpha = 0.096$, for large τ or r. The difference is that $n = 50$ in the case where $\alpha = 0.096$, while $n = 15$ in the case where $\alpha = 0.049$; sample size does not have much effect on the minimum Bayes factor, but has a large effect on how fast the Bayes factor increases as one moves away from this minimum.

2.4 Application to Multinomial Testing

Suppose $X = (X_1, \ldots, X_{p+1}) \sim$ Multinomial (n, θ), where the x_i are nonnegative integers, $\sum_{i=1}^{p+1} x_i = n$, and

$$\theta \in \Theta = \{(\theta_1, \ldots, \theta_p) : 0 < \theta_i < 1 \text{ for } i = 1, \ldots, p, \text{ and } \sum_{i=1}^{p} \theta_i < 1\}.$$

Defining $\theta_{p+1} \equiv 1 - \sum_{i=1}^{p} \theta_i$, the resulting likelihood for θ given x is then proportional to

$$\ell(\theta) = \prod_{i=1}^{p+1} \theta_i^{x_i}, \tag{2.10}$$

and the Bayes factor (1.2) for testing

$$H_0 : \theta = \theta^0 = (\theta_1^0, \ldots, \theta_0^0) \text{ vs. } H_1 : \theta \neq \theta^0$$

is

$$B = \ell(\theta^0) / \int_{\Theta} \ell(\theta) g(\theta) d\theta. \tag{2.11}$$

The natural conjugate densities with mean θ^0 are Dirichlet densities given by

$$g_\tau(\theta) = \Gamma(\tau) \prod_{i=1}^{p+1} [\theta_i^{(\tau\theta_i^0 - 1)} / \Gamma(\tau\theta_i^0)]. \tag{2.12}$$

Here τ is not the standard deviation, but is a covariance matrix multiplier in the sense that

$$\text{Covariance } (\theta) = \frac{1}{(\tau+1)} \left(\text{diag } \{\theta^0\} - \theta^0 (\theta^0)^t \right), \tag{2.13}$$

where diag $\{\theta^0\}$ is the diagonal matrix with diagonal given by θ^0. To elicit τ, one could specify the variance of a "typical" θ_i, and solve for τ. Or one could elicit the "average" variance, $\overline{v}$, of the θ_i, and solve for τ via

$$\tau = (p\overline{v})^{-1} \left[1 - \sum_{i=1}^{p+1} (\theta_i^0)^2 \right] - 1. \tag{2.14}$$

Any $\tau > 0$ is possible, so

$$\mathcal{G}_\mathcal{C} = \{g_\tau \text{ as in } (2.12) : \tau > 0\}.$$

For $g_\tau(\theta)$ in $\mathcal{G}_\mathcal{C}$, the Bayes factor can be written in the alternative forms

$$\begin{aligned} B_\mathcal{C}(\tau) &= \frac{\Gamma(\tau+n)}{\Gamma(\tau)} \prod_{i=1}^{p+1} \frac{(\theta_i^0)^{x_i} \Gamma(\tau\theta_i^0)}{\Gamma(\tau\theta_i^0 + x_i)} \\ &= \left[\prod_{j=0}^{(n-1)} (\tau + j) \right] \Big/ \prod_{i=1}^{(p+1)} \left\{ \prod_{j=0}^{(x_i-1)} (\tau + \frac{j}{\theta_i^0}) \right\}, \end{aligned} \tag{2.15}$$

again adopting the convention that $\prod_{i=0}^{-1} a_i = 1$. Note that, for the common choice $\theta^0 = (p+1)^{-1}(1, \ldots, 1)$, the Bayes factor becomes

$$B_\mathcal{C}(\tau) = \prod_{j=0}^{n-1} \frac{(\tau + j)}{[\tau + j(p+1)]^{m_j}}, \text{ where } m_j = \#\{x_i > j\}. \tag{2.16}$$

Analogously to the binomial situation, we define $\mathcal{G}_{\mathcal{RN}}$ and $\mathcal{G}_{\mathcal{N}}$, using formulae in Section 2.1, with the noninformative prior

$$\pi^*(\theta) = \prod_{i=1}^{p+1} \theta_i^{-1}.$$

It will be necessary, for use of this prior with $\mathcal{G}_{\mathcal{N}}$, to assume that all $x_i > 0$.

For $g_c \in \mathcal{G}_{\mathcal{N}}$, i.e.,

$$g_c(\theta) = c \prod_{i=1}^{p+1} (\theta_i^0/\theta_i)^{-1}, \tag{2.17}$$

the Bayes factor becomes

$$B_{\mathcal{N}}(c) = \frac{\Gamma(n)}{c} \prod_{i=1}^{(p+1)} \left[(\theta_i^0)^{(x_i-1)}/\Gamma(x_i) \right]. \tag{2.18}$$

Again, the choice $c = g_\tau(\theta_0)$ (g_τ from (2.12)) is recommended and will be used in our example.

It is not easy to define a sensible $\mathcal{G}_{\mathcal{RN}}$, here, for which computation of the Bayes factor is relatively easy. One possibility is to define

$$\Theta_r = \{\theta : \xi_i(\theta) \in (-r, r) \text{ for } i = 1, \ldots, p\},$$

where $\xi_i(\theta)$ is the "centered" log odds

$$\xi_i(\theta) = \log \frac{\theta_i}{(1-\theta_i)} - \log \frac{\theta_i^0}{(1-\theta_i^0)}. \tag{2.19}$$

For $p = 1$, this reduces to the set (2.5) for the binomial situation. This would suggest determining r by "assuming" that the ξ_i are i.i.d. so that, if Θ_r is to have probability 0.8, say, then $P(\xi_i \in (-r, r)) = (0.8)^{1/p}$. One could then choose r by considering a "typical"

ξ_i and subjectively choosing an r to achieve this probability for $(-r, r)$. Note that one could have different intervals $(-r_i, r_i)$ corresponding to each ξ_i; the following analysis is essentially identical for this case.

The normalization constant corresponding to $\pi^*(\theta)1_{\Theta_r}(\theta)$ is

$$c_r = 1/\int_{\Theta_r} \pi^*(\theta)d\theta = 1/(2r)^p;$$

this follows, as in the binomial case, from the observation that, under the transformation $\theta \to \xi(\theta)$, the prior transforms into the constant (improper) prior 1 on R^p. This completes the specification of $\mathcal{G}_{\mathcal{RN}}$.

For $\pi_r \in \mathcal{G}_{\mathcal{RN}}$, the Bayes factor can be shown to equal

$$B_{\mathcal{RN}}(r) = \frac{(2r)^p \prod_{i=1}^{p+1} (\theta_i^0)^{x_i}}{\int_{\Theta_r} \left[\prod_{i=1}^{p+1} \theta_i^{(x_i-1)} d\theta_i\right]} = (2r)^p / \int_{(-r,r)^p} \ell(\underset{\sim}{\xi}) d\underset{\sim}{\xi}, \tag{2.20}$$

where $\underset{\sim}{\xi} = (\xi_1, \ldots, \xi_p), (-r, r)^p = (-r, r) \times (-r, r) \times \ldots (-r, r)$, and

$$\ell(\underset{\sim}{\xi}) = \left[\theta_{p+1}^0 + \sum_{i=1}^{p} \theta_i^0 e^{\xi_i}\right]^{-n} e^{\sum_{i=1}^{p} \xi_i x_i}. \tag{2.21}$$

A Monte-Carlo approximation to the integral in (2.20) (see the Appendix for discussion) is, for $r > 0$,

$$\hat{B}_{\mathcal{RN}}(r) = \frac{(2r)^p m \left(\prod_{i=1}^{p} \tau_i d_i(r)\right)^{-1}}{\sum_{j=1}^{m} \left[\ell(\underset{\sim}{\xi}^{(i)}(r)) \prod_{i=1}^{p} (1 + |\xi_i^{(j)}(r) - \hat{\xi}_i|/\tau_i)^2\right]}, \tag{2.22}$$

where, for $i = 1, \ldots, p$,

$$\hat{\xi}_i = \log\left(\frac{x_i \theta^0_{p+1}}{x_{p+1}\theta^0_i}\right), \qquad \tau_i = (.675)\left(\frac{1}{x_i} + \frac{1}{x_{p+1}}\right)^{1/2}, \tag{2.23}$$

$$c_i(r) = -\frac{(r + \hat{\xi}_i)}{(\tau_i + |r + \hat{\xi}_i|)}, \qquad d_i(r) = \frac{(r - \hat{\xi}_i)}{(\tau_i + |r - \hat{\xi}_i|)} - c_i(r),$$

$$\xi_i^{(j)}(r) = \hat{\xi}_i + \tau_i W_{i,j}(1 - |W_{i,j}|)^{-1}, \; W_{i,j} = c_i(r) + d_i(r)U_{i,j}; \tag{2.24}$$

here the $\{U_{i,j}\}$ are i.i.d. $\mathcal{U}(0,1)$ random variables.

For examples, we consider the three situations

(i) $\alpha = 0.100, p = 2, n = 13, x_1 = 7, x_2 = 5, x_3 = 1, \theta^0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$;

(ii) $\alpha = 0.053, p = 3, n = 11, x_1 = 7, x_2 = x_3 = x_4 = 1, \theta^0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$;

(iii) $\alpha = 0.008, p = 3, n = 14, x_1 = 9, x_2 = 3, x_3 = x_4 = 1, \theta^0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

Figures 6 and 7 graph $B_{\mathcal{C}}$ from (2.15) (or (2.16)) and $\hat{B}_{\mathcal{RN}}(r)$ from (2.22), respectively, for each of these three cases. We graph $B_{\mathcal{C}}$ versus τ^{-1} instead of $B_{\mathcal{C}}(\tau)$ versus τ so that the shape is similar to other graphs. We have not graphed $B_{\mathcal{N}}(c)$ since it is of the simple form K/c; for cases (i), (ii), and (iii), K is, respectively, 0.469, 0.3076, and 0.0736.

3. COMMENTS AND CONCLUSIONS

It should be stressed that a "minimal" Bayesian analysis will not always be appropriate. For instance, it may be the case that, under H_1, there is a value θ_1 (or nearby region) that is particularly likely apriori. Then it would be appropriate to use a density g (or class $\mathcal{G}$) on $\theta \neq \theta_0$ that is "centered" at θ_1. Such situations tend to be situations requiring completely subjective Bayesian analysis, however, as opposed to the "as-objective-as-possible" type of analysis we are considering.

The main purpose of this paper was to illustrate and compare three approaches to minimal Bayesian testing of precise hypotheses. Table 2 summarizes the comparison.

TABLE 2
Comparison of prior classes

Class	Calculational Simplicity	Applicability	Difficulty of Input Specification	Size of Class
$\mathcal{G}_{\mathcal{C}}$	easy	exponential families	easy	moderate
$\mathcal{G}_{\mathcal{RN}}$	moderate	general	easiest	large
$\mathcal{G}_{\mathcal{N}}$	easy	general	difficult	moderate

The key calculational question is whether or not it is necessary to resort to numerical integration, as with $\mathcal{G}_{\mathcal{RN}}$ in the multinomial problem. For exponential families, where conjugate priors typically exist, $B_{\mathcal{C}}$ and $B_{\mathcal{N}}$ can usually be evaluated in closed form, while $B_{\mathcal{RN}}$ often requires numerical integration. Use of $\mathcal{G}_{\mathcal{RN}}$, on the other hand, is not limited to exponential families, as is use of $\mathcal{G}_{\mathcal{C}}$.

The "minimal" inputs needed to drive each of the three analyses have quite different interpretations. The input for $\mathcal{G}_{\mathcal{RN}}$ is probably the easiest, since one needs only specify an 80% prior credible set under H_1. Specifying a prior variance, for use of $\mathcal{G}_{\mathcal{C}}$, is also quite easy, though not all practitioners will be comfortable with variance. Most difficult is choice of the constant for $\mathcal{G}_{\mathcal{N}}$. Indeed, only indirect methods of assessment seem to be available.

The column "size of class" reflects, to some extent, how convincing the class will be to skeptics. The idea is that, in specifying $B_{\mathcal{C}}(\tau)$ or $B_{\mathcal{RN}}(r)$ or $B_{\mathcal{N}}(c)$, one is really specifying an "envelope" of Bayes factors that covers also all priors that are mixtures of priors in the

class. In particular, the indicated lower bound for the class can also be shown to be the lower bound over all associated mixture priors.

Of the three classes considered here, mixtures of the $g_r \in \mathcal{G}_{\mathcal{RN}}$ are by far the most general, typically consisting of any density that is nonincreasing, in some sense, as one moves away from θ_0. Thus $B_{\mathcal{RN}}(r)$ will typically be smaller (as a function) than, say, $B_{\mathcal{C}}(\tau)$ (once r and τ are appropriately scaled). While $B_{\mathcal{RN}}(r)$ may be more convincing to the skeptic (who views with suspicion the fact that the Bayes factors are much larger than P-values), it may seem unreasonable to a Bayesian, since the priors $g_r \in \mathcal{G}_{\mathcal{RN}}$ are less smooth than, say, the $g_\tau \in \mathcal{G}_{\mathcal{C}}$.

Our overall recommendation is to use $\mathcal{G}_{\mathcal{C}}$ and $B_{\mathcal{C}}$ if working within the exponential family *and* if the notion of variance is a comfortable one for those who will be interpreting $B_{\mathcal{C}}$. Otherwise, use of $\mathcal{G}_{\mathcal{RN}}$ and $B_{\mathcal{RN}}$ is indicated. We recommend against use of $\mathcal{G}_{\mathcal{N}}$ and $B_{\mathcal{N}}$, because of difficulty of interpretation and inaccuracies when only small departures from θ_0 are expected.

Finally, it should be mentioned that a Bayes factor is seldom sufficient as an inference. Typically, one must also produce a "confidence" or "credible" set for θ, given that H_1 is true. (The Bayes factor provides the evidence against H_0, and the credible set indicates how large a departure from θ_0 to expect, if indeed H_1 is true.) There is much to be said for abandoning $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{RN}}$ at this point, and simply using a noninformative prior, $\pi^*(\theta)$, to derive the credible set under H_1. Such would be the "minimal input" Bayesian credible set.

It might seem contradictory to be recommending different priors for Bayes factors and credible sets, but recall that we are excluding the possibility of a completely subjective analysis. The "minimal-input" needed for Bayes factors is not the same as that needed for credible sets, and may in fact be misleading for the latter. For instance, specification of $g_\tau \in \mathcal{G}_{\mathcal{RN}}$ will typically give much the same Bayes factor as specification of $g_\tau \in \mathcal{G}_{\mathcal{C}}$, but the two can easily result in very different credible sets. Rather than trying to assess which of these is more reasonable for credible sets, one can often do as well or better by simply using a credible set for a noninformative prior.

REFERENCES

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

Berger, J. (1986). Are P-values reasonable measures of accuracy. In *Pacific Statistical Congress* (I.S. Francis et al., eds). North-Holland, Amsterdam.

Berger, J. and Delampady, M. (1987). Testing precise hypotheses (with discussion). *Statist. Science*, **2**, 317–352.

Berger, J. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of P-values and evidence (with discussion). *J. Amer. Statist. Assoc.*, **82**, 112–139.

Berger, J. and Wolpert, R.L. (1984). *The Likelihood Principle.* IMS, Hayward, California.

Bernardo, J.M. (1980). A Bayesian analysis of classical hypothesis testing. In *Bayesian Statistics* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds), 605–618. University Press, Valencia.

Casella, G. and Berger, R.L. (1987). Reconciling Bayesian and frequentist evidence in the

one-sided testing problem. *J. Amer. Statist. Assoc.*, **82**, 106–111.

DeGroot, M.H. (1973). Doing what comes naturally: Interpreting a tail area as a posterior probability or as a likelihood ratio. *J. Amer. Statist. Assoc.*, **68**, 966–969.

Delampady, M. (1989a). Lower bounds on Bayes factors for interval hypotheses. *J. Amer. Statist. Assoc.*, **84**, 120–124.

Delampady, M. (1989b). Lower bounds on Bayes factors for invariant testing situations. *J. Multivariate Anal.*, **28**, 227–246.

Delampady, M. and Berger, J. (1990). Lower bounds on posterior probabilities for multinomial and chi-squared tests. *Ann. Statist.*, **18**.

Dempster, A.P. (1973). The direct use of likelihood for significance testing. In *Proc. of the Conference on Foundational Questions in Statistical Inference* (O. Barndorff-Nielsen et al., eds.) 335–352. Dept. Theoretical Statistics, Univ. Aarhus.

Diamond, G.A. and Forrester, J.S. (1983). Clinical trials and statistical verdicts: Probable grounds for appeal. *Ann. Internal Med.*, **98**, 385–394.

Dickey, J.M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Statist.*, **42**, 204–223.

Dickey, J.M. (1973). Scientific reporting. *J. Roy. Statist. Soc. Ser. B*, **35**, 285–305.

Dickey, J.M. (1974). Bayesian alternatives to the F test and least squares estimate in the linear model. In *Studies in Bayesian Econometrics and Statistics* (S.E. Feinberg and A. Zellner, eds.), 515–554. North-Holland, Amsterdam.

Dickey, J.M. (1976). Approximate posterior distributions. *J. Amer. Statist. Assoc.*, **71**, 680–689.

Dickey, J.M. (1977). Is the tail area useful as an approximate Bayes factor? *J. Amer. Statist. Assoc.*, **72**, 138–142.

Dickey, J.M. (1980). Approximate coherence for regression model inference – with a new analysis of Fisher's Broadback Wheatfield example. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (A. Zellner, ed.), 333–354. North-Holland, Amsterdam.

Edwards, W., Lindman, H. and Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.*, **70**, 193–242. Reprinted in *Robustness of Bayesian Analysis* (J.B. Kadane, ed.). North-Holland, Amsterdam, 1984.

Gómez Villegas, M.A. and De La Horra Navarro, J. (1984). Aproximacion de factores Bayes. *Cuad. Bioestadist.*, **2**, 355–361.

Good, I.J. (1958). Significance tests in parallel and in series. *J. Amer. Statist. Assoc.*, **53**, 799–813.

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

Good, I.J. (1967). A Bayesian significance test for the multinomial distribution. *J. Roy. Statist. Soc. Ser. B*, **29**, 399–431.

Good, I.J. (1983). *Good Thinking: The Foundations of Probability and Its Applications.*

Univ. Minnesota Press, Minneapolis.

Good, I.J. (1984). Notes C140, C144, C199, C200 and C201. *J. Statist. Comput. Simulation*, **19**.

Good, I.J. (1985). Weight of evidence: A brief survey. In *Bayesian Statistics 2* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), 249–270. North-Holland, Amsterdam.

Good, I.J. (1986). A flexible Bayesian model for comparing two treatments. *J. Statist. Comput. Simulation*, **26**, 301–305.

Hildreth, C. (1963). Bayesian statisticians and remote clients. *Econometrica*, **31**, 422–438.

Hill, B. (1982). Comment on "Lindley's paradox," by G. Shafer. *J. Amer. Statist. Assoc.*, **77**, 344–347.

Jeffreys, H. (1957). *Scientific Inference.* Cambridge Univ. Press, Cambridge.

Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford Univ. Press, London.

Leamer, E.E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* Wiley, New York.

Lempers, F.B. (1971). *Posterior Probabilities of Alternative Linear Models.* Univ. Rotterdam Press, Rotterdam.

Lindley, D.V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Lindley, D.V. (1961). The use of prior probability distributions in statistical inference and decisions. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 453–468. Univ.

California Press.

Lindley, D.V. (1965). *An Introduction to Probability and Statistics from a Bayesian Viewpoint*, **1, 2**. Cambridge Univ. Press, Cambridge.

Lindley, D.V. (1977). A problem in forensic science. *Biometrika*, **64**, 207–213.

Pratt, J.W. (1965). Bayesian interpretation of standard inference statements (with discussion). *J. Roy. Statist. Soc. Ser. B*, **27**, 169–203.

Raiffa, H. and Schlaiffer, R. (1961). *Applied Statistical Decision Theory.* Division of Research, Graduate School of Business Administration, Harvard Univ.

Rubin, H. (1971). A decision-theoretic approach to the problem of testing a null hypothesis. In *Statistical Decision Theory and Related Topics* (S.S. Gupta and J. Yackel, eds.). Academic, New York.

Shafer, G. (1982). Lindley's paradox (with discussion). *J. Amer. Statist. Assoc.*, **77**, 325–351.

Smith, A.F.M. and Spiegelhalter, D.J. (1980). Bayes factors and choice criteria for linear models. *J. Roy. Statist. Soc. Ser. B*, **42**, 213–220.

Smith, C.A.B. (1965). Personal probability and statistical analysis. *J. Roy. Statist. Soc. Ser. A*, **128**, 469–499.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Economics.* Wiley, New York.

Zellner, A. (1984). Posterior odds ratios for regression hypotheses: General considerations and some specific results. In *Basic Issues in Econometrics* (A. Zellner, ed.), 275–305.

Univ. Chicago Press, Chicago.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In *Bayesian Statistics* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), 585–603. University Press, Valencia.

APPENDIX

That $\hat{B}_{\mathcal{RN}}(r)$ in (2.22) is a Monte-Carlo approximation to $B_{\mathcal{RN}}(r)$ follows directly from the observation that the $\xi_i^{(j)}(r)$ in (2.24) are independent random variables with density

$$(A.1) \qquad g_i(\xi_i^{(j)}(r)) = 1/\left[\tau_i d_i(r)(1 + |\xi_i^{(j)}(r) - \hat{\xi}_i|/\tau_i)^2\right],$$

so that the expression in (2.22) is the usual Monte-Carlo approximation with importance function $\prod_{i=1}^{p} g_i(\cdot)$ (cf. Berger, 1985, for discussion). The reason for choosing this importance function is that it is easily computable, easy to generate random variables from, has fat tails, and mimics the likelihood function on the domain of integration.

In elaboration of this last point, note that the usual "observed likelihood" approximation to $\ell(\xi)$ is proportional to a $\mathcal{N}_p(\hat{\xi}, \mathbb{I})$ density, where $\hat{\xi} = (\hat{\xi}_1, \ldots, \hat{\xi}_p)$ with the $\hat{\xi}_i$ defined in (2.23), and

$$\mathbb{I} = \text{ diag } \left\{\frac{1}{x_1}, \ldots, \frac{1}{x_p}\right\} + \frac{1}{x_{p+1}} \ (1),$$

with diag { } denoting a diagonal matrix with the given diagonal entries, and (1) denoting the $p \times p$ matrix of all ones. Here $\hat{\xi}$ is the m.l.e. of $\ell(\xi)$, and $\mathbb{I}$ is the inverse of the observed information corresponding to $\ell(\xi)$ (cf. Berger, 1985). Because fatter tails than normal are

desirable for an importance function, we consider, for ξ_i, the importance function

$$(A.2) \qquad g_i^*(\xi_i) = \frac{1}{2\tau_i(1 + |\xi_i - \hat{\xi}_i|/\tau_i)^2},$$

with quartiles chosen to match the normal density (hence the choice of τ_i in (2.23)). For simplicity, we consider the ξ_i to be independent in this new importance function.

The final alteration needed arises because the domain of integration is $\xi_i \in (-r, r), i = 1, \ldots, p$. One of the pleasant features of g_i^* in (A.2) is that the conditional density obtained by conditioning on $\xi_i \in (-r, r)$ becomes the also simple (A.1). Hence the densities in (A.1) define the actual importance function used. Note that, if one attempted to incorporate into the importance function the dependence among the original ξ_i (or that in the $N_p(\hat{\xi}, \hat{\Sigma})$ approximation), the ability to transform the importance function to have range precisely equal to the domain of integration would be lost.

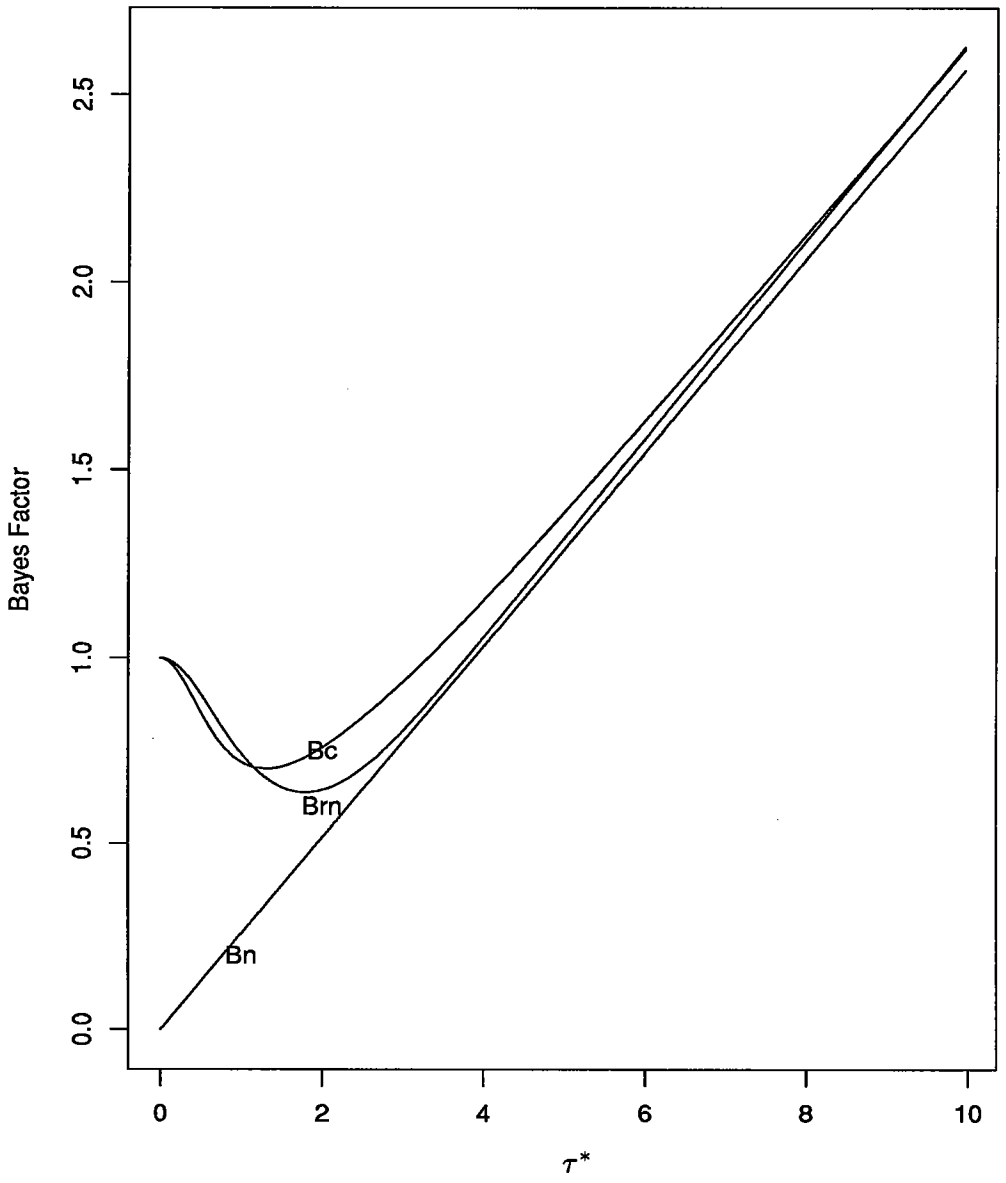

Figure 1. Bayes factors as a function of $\tau^* = \sqrt{n}\tau/\sigma$ when $t = 1.645$ (P-value=0.10).

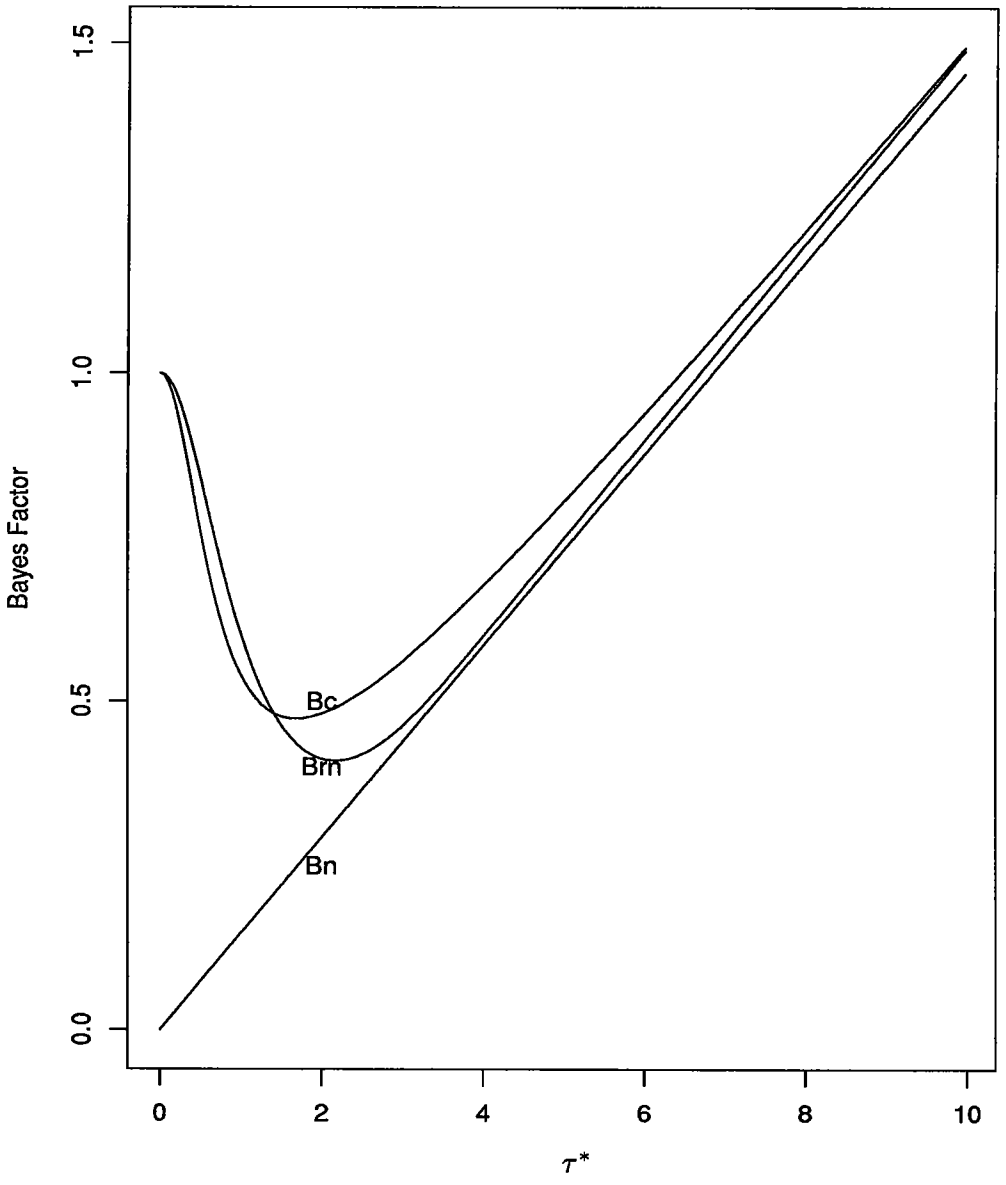

Figure 2. Bayes factors as a function of $\tau^* = \sqrt{n}\tau/\sigma$ when $t = 1.96$ (P-value=0.05).

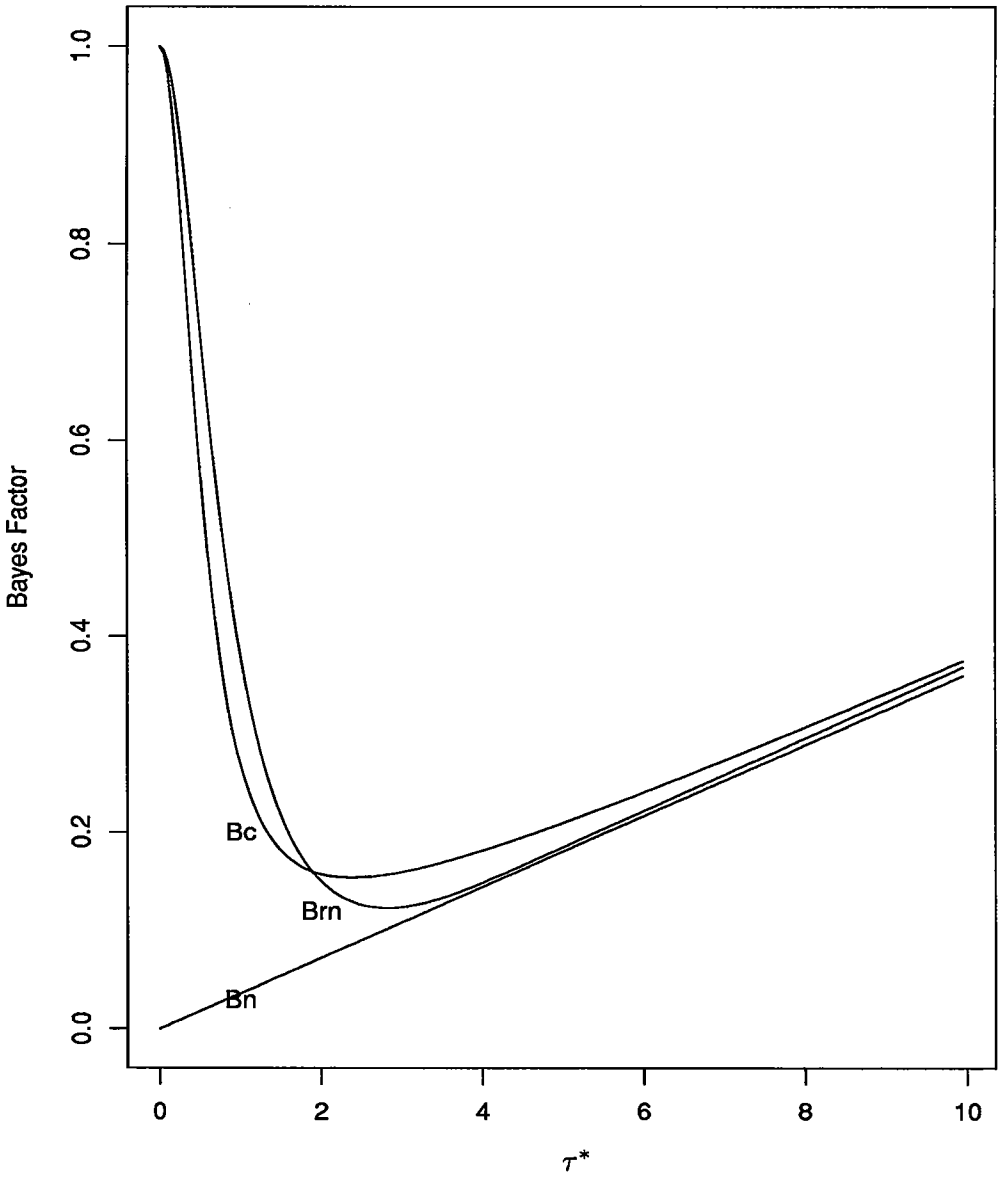

Figure 3. Bayes factors as a function of $\tau^* = \sqrt{n}\tau/\sigma$ when $t = 2.576$ (P-value=0.01).

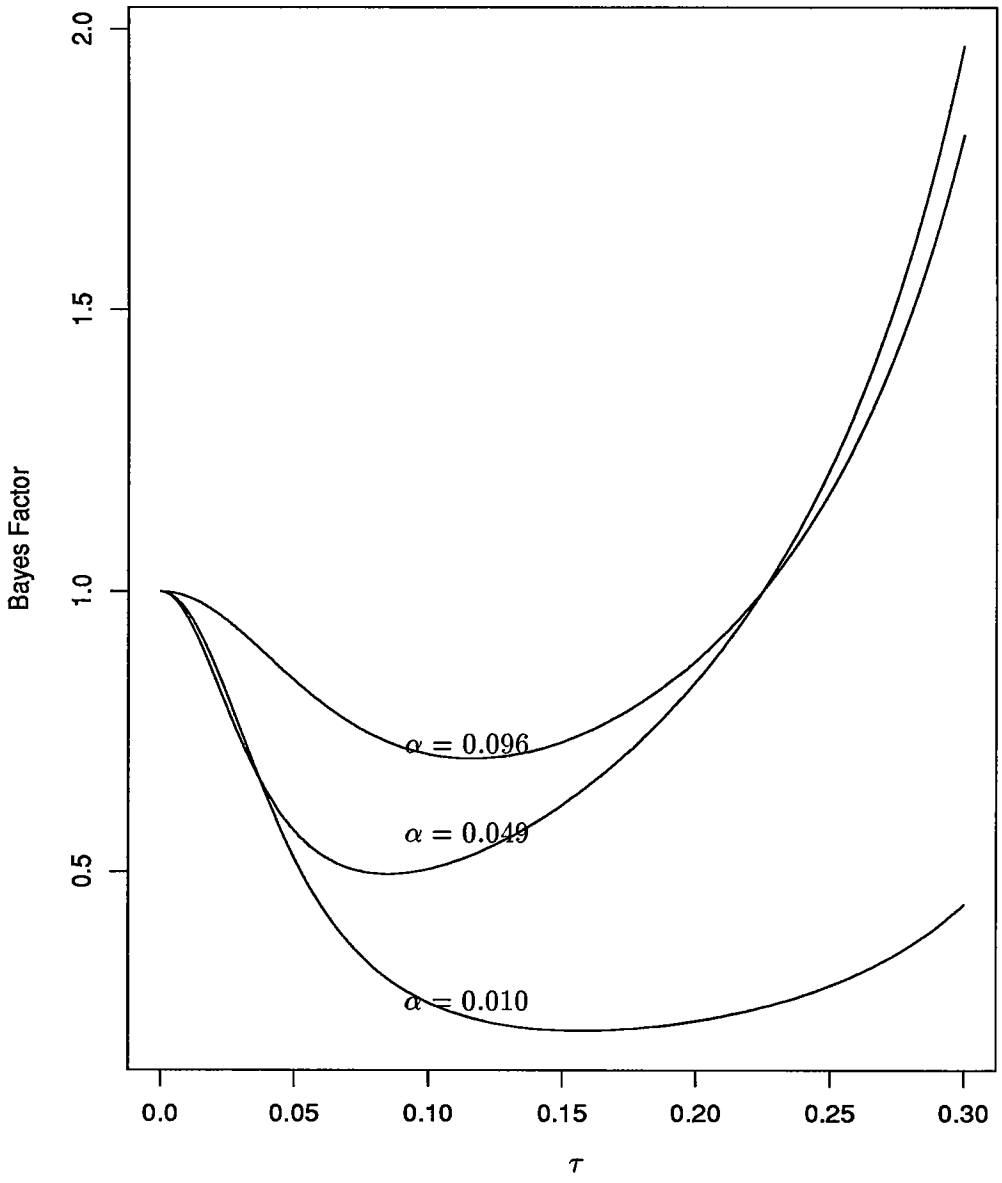

Figure 4. Bayes factors $B_c(\tau)$ corresponding to the P-values $\alpha = 0.096, \alpha = 0.049$, and $\alpha = 0.010$.

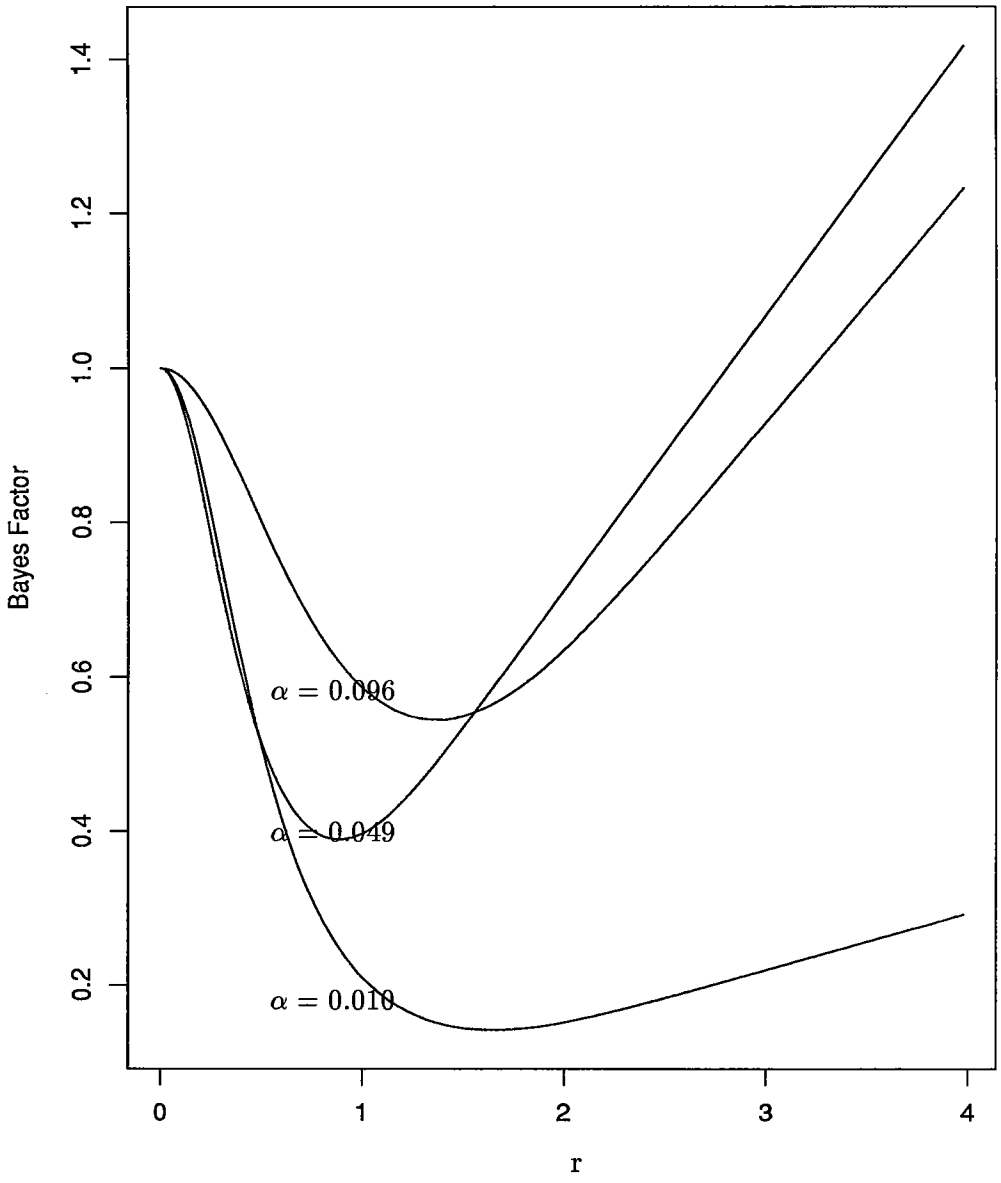

Figure 5. Bayes factors $B_{rn}(r)$ corresponding to the P-values $\alpha = 0.096, \alpha = 0.049$, and $\alpha = 0.010$.

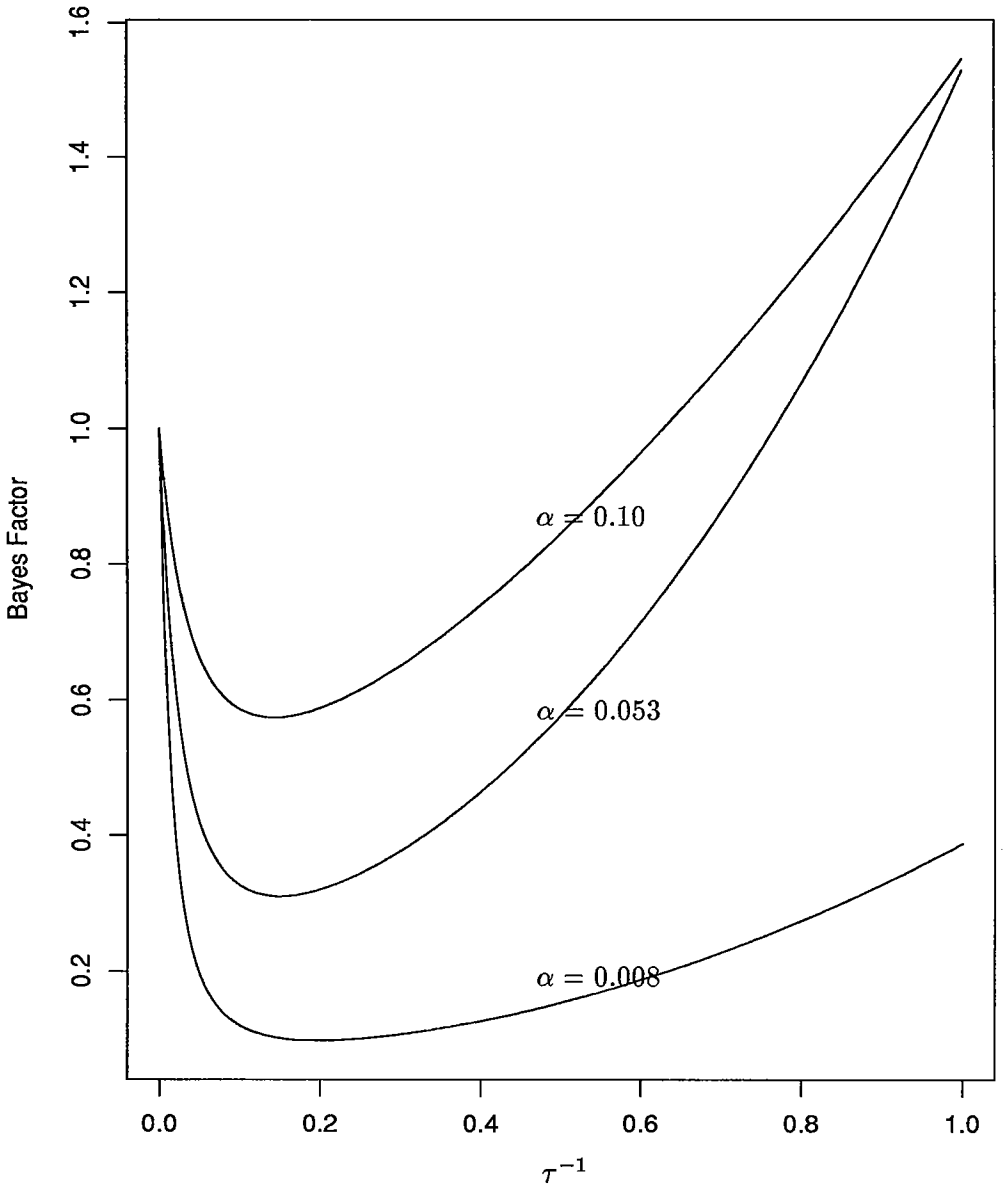

Figure 6. Bayes factors B_c, graphed versus τ^{-1}, corresponding to the P-values $\alpha = 0.10, \alpha = 0.053$, and $\alpha = 0.008$.

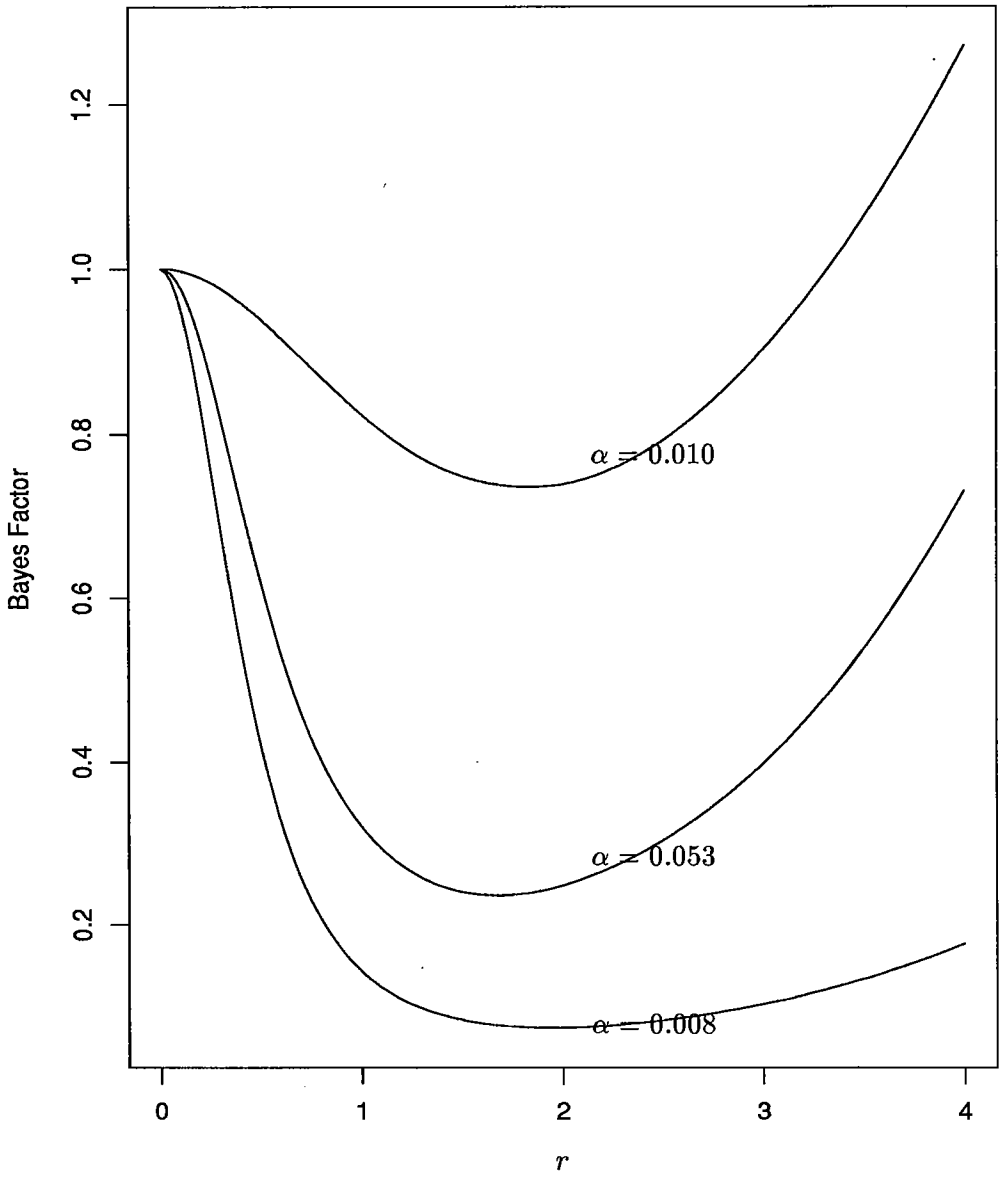

Figure 7. Approximate Bayes factors $\hat{B}_{rn}(r)$ corresponding to the P-values $\alpha = 0.010, \alpha = 0.053$, and $\alpha = 0.008$.