1

Properties of Unified Bayesian-Frequentist Tests[1]

J. O. Berger[†], **B. Boukai**[‡] **and Y. Wang**[‡]

[†]*Purdue University, West-Lafayette, IN*
[‡]*Indiana University - Purdue University, Indianapolis, IN*

Abstract.
The Modified Bayesian-Frequentist test of Berger, Brown and Wolpert (1994), is considered here in the context of normal hypothesis testing. We focus attention on the testing of a precise null hypothesis versus a composite alternative, either the one-sided or the two-sided type. We study the properties of the corresponding modified Bayesian-Frequentist test and in particular the large-sample behavior of its *no decision region* under two different classes of prior p.d.fs; the shifted conjugate class and a domain-restricted noninformative class. It is shown that under these prior classes, the size of the no-decision region of the test is rather small, compared to the relevant sample size. A lower bound on the conditional probability of the type I error is also provided.

Keywords and phrases: Bayes factor, likelihood ratio, composite hypothesis, conditional test, error probabilities.

1.1 Introduction

Over the past twenty years, Shanti S. Gupta has been a major contributor to efforts to integrate Bayesian and Frequentist concepts in statistical inference. An area in which such integration has, until recently, been lacking is hypothesis testing. Indeed, testing statistical hypotheses has been one of the focal points for disagreement between Bayesians and Frequentists.

Given the data, one has to decide whether the underlying statistical model is as described by the null hypothesis H_0 or as described by the alternative

[1]Research was supported, in part, by the National Science Foundation under grant DMS-9303556.

model-hypothesis H_1. The classical Frequentist approach is to construct a *critical rejection* region and report corresponding error probabilities. The incorrect rejection of H_0 (the Type I error), has a designated probability α and the incorrect acceptance of H_0 (the Type II error), has a designated probability β. However, the traditional (α, β)-Frequentist approach has been greatly criticized for reporting these error probabilities independently of the given data. In practice, a common alternative is to compute the test's P-value and report it as a data-dependent measure of the strength of the evidence against the null hypothesis H_0. However, the P-value can be highly misleading as a measure of the evidence provided by the data against the null hypothesis. This was demonstrated by Edwards, Lindman, and Savage (1963), Berger and Sellke (1987), Berger and Delampady (1987) and Delampady and Berger (1990) who have reviewed the practicality of the P-value and explored the dramatic conflict between the P-value and other data dependent measures of evidence.

There have been many attempts to rectify these deficiencies and to modify the classical Frequentist approach by incorporating data-dependent procedures which are based on conditioning. Earlier works in this direction are summarized in Kiefer (1977) and in Berger and Wolpert (1988). In a seminal series of papers, Kiefer (1975, 1976, 1977) and Brownie and Kiefer (1977), the Conditional Frequentist approach was formalized. The basic idea behind this approach is to condition on a statistic measuring the evidential strength of the data, and then to provide error probabilities conditional on the observed value of this statistic. Unfortunately, the approach never achieved substantial popularity, in part because of the difficulty of choosing the statistic upon which to condition (cf., the Discussion of Kiefer, 1977).

The Bayesian approach to testing is based on consideration of the most extreme form of conditioning, namely conditioning on the given data. Suppose we observe the realization, x, of the random variable $X \in \mathcal{X}$ from a density $f(x|\theta)$, with θ being an unknown element of the parameter space Θ. In the sequel, we let $P_\theta(\cdot)$ denote conditional probability given $\theta \in \Theta$. Consider the problem of testing simple versus composite hypotheses as given by

$$H_0: \ \theta = \theta_0 \quad \text{versus} \quad H_1: \ \theta \in \Theta_1, \tag{1.1}$$

where $\theta_0 \notin \Theta_1 \subset \Theta$. We will consider here primarily the two-sided alternative case with $\Theta_1 = \{\theta \in \Theta : \ \theta \neq \theta_0\}$.

Within the Bayesian framework for testing the above hypotheses, one usually specifies the *prior probabilities*, π_0 for H_0 being true and $1 - \pi_0$ for H_1 being true, while assigning to Θ_1 the prior density $(1 - \pi_0)\pi(\theta)$, where $\pi \in \Pi$, some class of proper p.d.fs. over Θ_1. To simplify the presentation, we assume throughout this paper the default prior probability of $\pi_0 = 1/2$ for the simple hypothesis $H_0 : \theta = \theta_0$. Note that H_0 should typically be considered an approximation to the more realistic hypothesis $H_0: \ |\theta - \theta_0| < \epsilon$ for some small ϵ (see Berger and Delampady, 1987).

Let $m_\pi(x)$ and $m_0(x)$ denote the marginal densities of X conditional on H_1 and H_0 being true, respectively; then $m_0(x) = f(x|\theta_0)$ and

$$m_\pi(x) = \int_{\Theta_1} f(x|\theta)\pi(\theta)d\theta. \tag{1.2}$$

For the Classical Frequentist, π might be thought of as a weight function which allows computation of an average likelihood for H_1 (as given by $m_\pi(x)$ in (1.2)). However, for a Bayesian, π is the prior density for θ conditional on H_1 being true. It follows, that for the Bayesian, the test of (1.1) can be reduced to an equivalent test of two *"simple"* hypotheses concerning the marginal distribution of X, namely

$$H_0: \ X \sim m_0(x) \quad \text{versus} \quad H_1: \ X \sim m_\pi(x). \tag{1.3}$$

We denote by

$$B_\pi(x) \equiv \frac{m_0(x)}{m_\pi(x)} \tag{1.4}$$

the *likelihood ratio of* H_0 *to* H_1 (or, equivalently, the *Bayes factor in favor of* H_0). With the default choice of prior probability $\pi_0 = 1/2$ of H_0 being true, the posterior probability, given the data, of H_0 being true is

$$Pr(H_0|x) \equiv \alpha^*(B_\pi(x)) = \frac{B_\pi(x)}{1 + B_\pi(x)} \tag{1.5}$$

and the posterior probability, given the data, of H_1 being true is

$$Pr(H_1|x) \equiv \beta^*(B_\pi(x)) = \frac{1}{1 + B_\pi(x)}. \tag{1.6}$$

For a Bayesian, $B_\pi(x)$ in (1.4) is the Bayes factor in favor of H_0; for the default choice of $\pi_0 = 1/2$ it is also the posterior odds in favor of H_0. Clearly, the decision to either *reject* or *accept* H_0 will depend on the observed value of $B_\pi(x)$, where *small* values of $B_\pi(x)$ correspond to rejection of H_0. In this situation, the standard and optimal Bayesian test, for a Bayesian who assumes the loss to be zero for correct decision and ℓ_i for an incorrect decision when H_i is true ($i = 0, 1$), can be written as (defining $\ell = \ell_1/\ell_0$)

$$\mathbf{T}_\ell : \begin{cases} \textit{if } B_\pi(x) \le \ell, & \textit{reject } H_0 \textit{ and report the posterior risk } \ell_0\, \alpha^*(B_\pi(x)); \\[2ex] \textit{if } B_\pi(x) > \ell, & \textit{accept } H_0 \textit{ and report the posterior risk } \ell_1\, \beta^*(B_\pi(x)). \end{cases}$$

For inference without specific losses, one can take in $\mathbf{T}_\ell$ the default values of $\ell_0 = \ell_1 \equiv 1$. With these default losses, the optimal Bayes test becomes

$$\mathbf{T}_1 : \begin{cases} \textit{if } B_\pi(x) \le 1, & \textit{reject } H_0 \textit{ and report the posterior probability } \alpha^*(B_\pi(x)); \\[2ex] \textit{if } B_\pi(x) > 1, & \textit{accept } H_0 \textit{ and report the posterior probability } \beta^*(B_\pi(x)). \end{cases} \tag{1.7}$$

1.2 The New Bayesian-Frequentist Test

The testing of simple versus simple hypotheses was considered by Berger, Brown and Wolpert (1994) who proposed a modification of the Bayesian test above, which includes a structured *no decision region*. The modified Bayesian test was shown to also be a valid Conditional Frequentist test, with error probabilities computed conditionally upon a certain statistic $S(X)$. The surprising aspect of this result is not that both the Bayesian and the Conditional Frequentist might have the same decision rule for rejecting or accepting the null hypothesis (this is not so uncommon), but rather that they will report the same conditional error probabilities (CEP) upon rejecting or accepting. That is, the error probabilities reported by the conditional Frequentist using the proposed conditioning strategy are the same as the posterior probabilities of the relevant errors reported by the Bayesian.

Recently, Berger, Boukai and Wang (1996a) adapted the modified Bayesian test to the composite alternative case as given in (1.1). They have shown that, even in the composite alternative case, the modified test retains its valid Conditional Frequentist interpretations.

Let $\mathcal{B}$ denote the range of $B_\pi(x)$, as x varies over $\mathcal{X}$. We will restrict attention here to the case where $\mathcal{B}$ is an interval that contains 1. Let F_0 and F_1 be the c.d.f.s of $B_\pi(X)$ under m_0 and m_π, respectively. For simplicity, we assume in the following that their inverses, F_0^{-1} and F_1^{-1}, exist over the range $\mathcal{B}$ of $B_\pi(x)$. For any $b \in \mathcal{B}$, define

$$\psi(b) = F_0^{-1}(1 - F_1(b)), \tag{2.1}$$

so that $\psi^{-1}(b) = F_1^{-1}(1 - F_0(b))$, and set

$$r = \min(1, \psi^{-1}(1)) \qquad \text{and} \qquad a = \max(1, \psi(1)). \tag{2.2}$$

The new modified Bayesian test of the hypothesis H_0 versus H_1 in (1.1) (or equivalently (1.3)) is given by

$$\mathbf{T}_1^*: \begin{cases} \textit{if } B_\pi(x) \le r, & \textit{reject } H_0 \textit{ and report the CEP } \alpha^*(B_\pi(x)); \\ \textit{if } r < B_\pi(x) < a, & \textit{make no-decision}; \\ \textit{if } B_\pi(x) > a, & \textit{accept } H_0 \textit{ and report the CEP } \beta^*(B_\pi(x)), \end{cases} \tag{2.3}$$

where $\alpha^*(B_\pi)$ and $\beta^*(B_\pi)$ are as given in (1.5) and (1.6), respectively.

The "surprise" already observed in Berger, Brown and Wolpert (1994) and in Berger, Boukai and Wang (1996a) is that $\mathbf{T}_1^*$ is also a conditional Frequentist test, arising from use of the conditioning statistic

$$S(X) = \min\{B_\pi(X),\ \psi^{-1}(B_\pi(X)\}, \tag{2.4}$$

over the domain $\mathcal{X}^* = \{x \in \mathcal{X} :\ 0 \le S(x) \le r\}$ (whose complement is the no-decision region in the test $\mathbf{T}_1^*$).

Thus, the Conditional Frequentist who uses the acceptance and rejection regions in $\mathbf{T}_1^*$, along with the conditioning statistic in (2.4), will report conditional error probabilities upon accepting or rejecting which are in complete agreement with the Bayesian posterior probabilities.

To be more precise, let

$$\alpha(s) \equiv P_{\theta_0}(\textit{rejecting } H_0 \mid S(X) = s) \tag{2.5}$$

and

$$\beta(\theta|s) \equiv P_{\theta}(\textit{accepting } H_0 \mid S(X) = s) \tag{2.6}$$

denote the conditional probabilities of Type I and Type II error as reported by the Conditional Frequentist who utilizes $\mathbf{T}_1^*$ along with the conditioning statistic (2.4) in order to test $H_0:\ \theta = \theta_0$ versus $H_1:\ \theta \in \Theta_1$, as in (1.1).

Theorem 1 [BBW (1996a)] *For the test $\mathbf{T}_1^*$ of H_0 versus H_1 above, and the the conditioning statistic $S(X)$ as given in (2.4), we have: $\alpha(s) \equiv \alpha^*(B_\pi)$ and*

$$E^{\pi_1(\theta|s)}[\beta(\theta|s)] \equiv \beta^*(B_\pi), \tag{2.7}$$

where $\pi_1(\theta|s)$ denotes the posterior p.d.f. of θ conditional on H_1 being true and on the observed value s of $S(X)$

For a further discussion of this result and the Conditional Frequentist interpretation of this modified Bayesian test, see Berger, Boukai and Wang (1996a).

It is clear from the definitions of a and r that the *no decision region* in $\mathbf{T}_\pi^*$ disappears whenever $F_0(1) = 1 - F_1(1)$, in which case $r = a = 1$. This can happen in cases with *Likelihood Ratio Symmetry* (for a definition see Berger, Brown, and Wolpert (1994)). Without the presence of the *no decision region* in $\mathbf{T}_1^*$, it would be the optimal Bayes test $\mathbf{T}_1$ (as presented in (1.7)) for a Bayesian who assumes equal prior probabilities of the hypotheses as well as equal losses for the two incorrect decisions (as we do in this paper). For such a Bayesian, the report of the conditional error probabilities $\alpha^*(B_\pi)$ and $\beta^*(B_\pi)$ (being the posterior probabilities of H_0 and H_1, respectively) has valid Bayesian interpretations. The inclusion of the *no decision region* in the test is primarily intended to eliminate possible conflict between the intuitive interpretation of the conditional error probabilities and the respective decision to either reject or accept the null hypothesis (see Berger, Brown, and Wolpert (1994) for discussion). It should be viewed as the *price* that must be paid to obtain a conditional Frequentist interpretation for the optimal Bayes test.

The notation in (1.2) and (1.4) emphasizes the dependence of the testing procedure $\mathbf{T}_1^*$, and especially of the constants r and a defined in (2.2) (through the marginal p.d.f. m_π), on the particular choice of a prior p.d.f. $\pi \in \Pi$ for θ over Θ_1. Moreover, these constants are also greatly affected by the sample size n. These effects, combined, will determine the "size" of the *no decision region* on which the Conditional Frequentist and the Bayesian might report differing error probabilities. The test would be impractical if the size of the "no decision region" is too large.

This Chapter has three goals. One is to study the size of the *no-decision region*, both numerically (for small sample sizes) and asymptotically. It will be shown that the *no decision region* does appear to be largely ignorable. A second goal is to show that, in the studied problems, $r = 1$. Then $\mathbf{T}_1^*$ simplifies considerably in that the *no decision region* is never involved in rejection of H_0 (assuming that one would not want to consider rejection when $B_\pi > 1$). Since practitioners are (historically) mainly interested in rejection of H_0, this means that the *no decision region* can typically be completely ignored. The final goal of the paper is to show how results from robust Bayesian testing can be employed to yield lower bounds on conditional Frequentist error probabilities over wide classes of π (or, equivalently, over wide classes of conditioning statistics $S(X)$). This can greatly ease concerns about appropriateness of any particular π (or $S(X)$). For reasons of space limitations, proofs of the main results presented here are omitted. They can be found in Berger, Boukai and Wang (1996b).

1.3 Normal Testing

Suppose that $X_1, X_2, \ldots, X_n$ are n i.i.d. random variables from a normal distribution having unknown mean θ and known variance σ^2, that is, the $\mathcal{N}(\theta, \sigma^2)$ distribution. We denote by $\bar{X}_n = \sum X_i/n$ their usual average and recall that $\bar{X}_n \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$. We denote by $\phi(z)$ and $\Phi(z)$ the p.d.f. and the c.d.f., respectively, of the standard normal distribution $\mathcal{N}(0,1)$, so that

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z^2}{2}\} \qquad \text{and} \qquad \Phi(z) = \int_{-\infty}^{z} \phi(x)\, dx.$$

Given the observed value $\bar{x}_n$ of $\bar{X}_n$, we wish to test, as in (1.1), the hypotheses

$$H_0 : \ \theta = \theta_0 \ \text{ versus } \ H_1 : \ \theta \in \Theta_1, \tag{3.1}$$

where $\theta_0 \notin \Theta_1 \subset \Theta$. We assume that, conditional on H_1 being true, the prior distribution of θ has a continuous p.d.f. $\pi(\theta)$ over the set Θ_1 ($\pi \in \Pi$).

Let

$$t \equiv \sqrt{n}(\bar{x}_n - \theta_0)/\sigma$$

denote the "standard" statistic for many of the normal testing problems. It is easy to verify that the marginal p.d.fs. of t, as calculated under H_0 and H_1, are given (respectively) by

$$m_0(t) = \phi(t) \equiv \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} \tag{3.2}$$

and

$$m_\pi(t) = \int_{\Theta_1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(t - \sqrt{n}(\theta - \theta_0)/\sigma)^2\right\} \pi(\theta) d\theta. \tag{3.3}$$

It follows from (3.2) and (3.3) that the problem of testing the hypotheses (3.1) is reduced to that of testing the hypotheses

$$H_0 : \ t \sim m_0(t) \quad \text{versus} \quad H_1 : \ t \sim m_\pi(t), \tag{3.4}$$

which, according to (2.3), is based on the Bayes factor

$$B_\pi(t) = \frac{\phi(t)}{\int_{\Theta_1} \phi(t - \sqrt{n}(\theta - \theta_0)/\sigma)\pi(\theta)d\theta}. \tag{3.5}$$

The corresponding modified Bayesian testing procedure $\mathbf{T}_1^*$ is given in (2.3) with critical testing values $r, \ a$ obtained from (2.2).

Ideally, the prior π in (3.2) should be specified, whenever possible, in a completely subjective fashion. However, when this cannot be it is useful to consider classes of priors that depend on what could be termed *minimal Bayesian input.* Here, we consider (for the normal testing problem at hand) two possible choices for the class of conditional priors π. These classes are the (shifted) conjugate normal class and a uniform class.

To be specific, when $\Theta_1 = \{\theta \neq \theta_0\}$ (as corresponds to the "two-sided" normal test), these classes are

$$\begin{aligned} \Pi_c &= \{\pi : \pi(\theta) = \frac{1}{\tau}\phi(\frac{\theta - \theta_1}{\tau}) \text{ for } \theta \neq \theta_0, -\infty < \theta_1 < \infty, \ 0 < \tau < \infty\}, \\ \Pi_u &= \{\pi : \pi(\theta) = \mathcal{U}(\theta_0 - \tau, \ \theta_0 + \tau), \ 0 < \tau < \infty\}. \end{aligned}$$

The class Π_c of normal priors allows a mean other than θ_0, although the choice $\theta_1 = \theta_0$ is the common choice. The class Π_u of symmetric (about θ_0) uniform priors is of interest not only because they are appealing in their simplicity, but also because they form the extreme points of the class of all symmetric (about θ_0) unimodal distributions. When $\Theta_1 = \{\theta > \theta_0\}$ (as corresponds to the "one-sided" normal test), we will consider the class of one-sided normal priors given by

$$\Pi_+ = \{\pi : \pi(\theta) = \frac{2}{\tau}\phi(\frac{\theta - \theta_0}{\tau}) \text{ for } \theta > \theta_0, \ 0 < \tau < \infty\}.$$

In all these cases, the prior p.d.f. π, depends (heavily) on the value of τ which has to be specified; this is the required *minimal prior input.* Intuitively, this minimal prior input corresponds to the degree of departure from θ_0 that is anticipated, conditional on H_1 being true. (For Π_c one must also specify θ_1 but, since the default choice of $\theta_1 = \theta_0$ can always be made, specification of θ_1 is not absolutely necessary.)

To clarify (and thus also to emphasize) the choice of the relevant class of prior p.d.f.'s, we will write throughout this chapter, $m_c, B_c, \ F_c$ and $m_u, B_u, \ F_u$

for the marginal p.d.f. m_π, the Bayes factor B_π and its c.d.f. F_1 (under H_1), corresponding to the classes Π_c and Π_u, respectively. Similar notation is used for the case of Π_+.

1.4 The "Two-Sided" Normal Test with Shifted Conjugate Prior

We consider the problem of testing the hypotheses (3.1) assuming the shifted conjugate prior for θ. That is $\pi \in \Pi_c$ over $\Theta_1 \equiv \{\theta \neq \theta_0\}$. Thus, conditional on H_1 being true, θ has the $\mathcal{N}(\theta_1,\ \tau^2)$ prior distribution. Without loss of generality, we assume throughout that $\theta_0 > \theta_1$.

Since $\pi(\theta)$ is now specified, $m_\pi(t)$ in (3.3) can be computed as

$$m_c(t) = \frac{1}{\sqrt{2\pi}\sqrt{1+\lambda}} \exp\left\{ \frac{-(t+\sqrt{\lambda}\Delta)^2}{2(1+\lambda)} \right\}, \tag{4.1}$$

where $\Delta = (\theta_0 - \theta_1)/\tau$ and $\lambda = n\tau^2/\sigma^2$. Thus, with $\pi \in \Pi_c$, $t \sim \mathcal{N}(-\sqrt{\lambda}\Delta,\ 1+\lambda)$, under H_1. The parameter θ_1 is the conditional prior mean of θ, given H_1 is true. This allows, under H_1, a measurable *shift* of the conditional prior p.d.f. of θ away from H_0. When $\Delta = 0$, the prior p.d.f. is symmetric about θ_0. This choice of Δ is often considered as the default choice for applications, and is used in Example 1 below. Also in Example 1, the default choice of $\tau^2 = 2\sigma^2$ was made; the resulting $\mathcal{N}(0, 2\sigma^2)$ prior is similar to the Cauchy $(0, \sigma^2)$ default prior recommended by Jeffreys (1961).

Combining (4.1) and (3.1), in (3.5), we obtain a simple expression for the Bayes factor $B_c(t)$, as

$$B_c(t) = \sqrt{1+\lambda}\exp\left\{ -\frac{\lambda}{2(1+\lambda)}(t - \frac{\Delta}{\sqrt{\lambda}})^2 + \frac{\Delta^2}{2} \right\}. \tag{4.2}$$

Observe that, for fixed $\lambda > 0$ and $\Delta \in \mathbb{R}$, $B_c(t)$ attains its maximum value at $t^* = \Delta/\sqrt{\lambda}$, at which point

$$B_c(t^*) \equiv B_c^* \equiv \sqrt{1+\lambda}\exp\{\frac{\Delta^2}{2}\} > 1.$$

For any $b \in \mathcal{B} = (0,\ B_c^*)$, let $t_b^\pm$ be the two solutions of the equation $B_c(t) = b$. Then it follows from expression (4.2) for $B_c(t)$ that

$$t_b^\pm = \frac{\Delta}{\sqrt{\lambda}} \pm \sqrt{\frac{1+\lambda}{\lambda}}\sqrt{\Delta^2 + \log(\frac{1+\lambda}{b^2})}. \tag{4.3}$$

It is straightforward to verify that, for all $b \in \mathcal{B}$, the marginal c.d.f.'s of B_c under m_0 and m_c are given (respectively) by

$$F_0(b) = \Phi(t_b^-) + \Phi(-t_b^+) \tag{4.4}$$

and

$$F_c(b) = \Phi(\frac{t_b^- + \sqrt{\lambda}\Delta}{\sqrt{1+\lambda}}) + \Phi(\frac{-t_b^+ - \sqrt{\lambda}\Delta}{\sqrt{1+\lambda}}). \tag{4.5}$$

Lemma 1 . *For any $\lambda > 0$, $\Delta \in \mathbb{R}$ and $\pi \in \Pi_c$, we have $F_0(1) \leq 1 - F_c(1)$, so that $\psi(1) > 1$ in (2.2).*

The proof of this result can be found in Berger, Boukai and Wang (1996b); we omit the details here for reasons of space.

By Lemma 1, $\psi(1) > 1$ so that $r = 1$ and $a = \psi(1)$ in (2.3). Accordingly, when $\pi \in \Pi_c$, the modified Bayesian test of the hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ becomes,

$$\mathbf{T}_1^* : \begin{cases} \text{if } B_c(t) \leq 1, & \text{reject } H_0 \text{ and report the CEP } \alpha^*(B_c(t)); \\ \text{if } 1 < B_c(t) < a, & \text{make no decision;} \\ \text{if } B_c(t) \geq a, & \text{accept } H_0 \text{ and report the CEP } \beta^*(B_c(t)). \end{cases} \tag{4.6}$$

This is an important simplification because it is natural to even consider rejection only when $B_c \leq 1$ (at least for a default Bayesian), and then the *no decision region* never comes into play.

Unfortunately, there is no explicit expression for a in this case. But it can be computed numerically as the solution of the equation $F_0(a) = 1 - F_c(1)$, which, in view of (4.4)-(4.5), may be written as

$$\Phi(t_a^-) + \Phi(-t_a^+) = \Phi(\frac{t_1^+ + \sqrt{\lambda}\Delta}{\sqrt{1+\lambda}}) - \Phi(\frac{t_1^- + \sqrt{\lambda}\Delta}{\sqrt{1+\lambda}}), \tag{4.7}$$

where $t_a^\pm$ and $t_1^\pm$ are as defined by (4.3) with $b \equiv a$ and $b \equiv 1$, respectively. It is clear that this solution $a \equiv a(\lambda,\ \Delta)$ depends on both Δ and on τ and the sample size n, through λ. Table 1 below lists several values of a evaluated numerically for different choices of λ and Δ.

Example 1. Fisher and Belle (1993) provide the birthweights in grams of some $n = 15$ cases of SIDS (Sudden Infant Death Syndrome) born in King County in 1977:

2013	3827	3090	3260	4309
3374	3544	2835	3487	3289
3714	2240	2041	3629	3345

With the standing assumption of normality and a supposed known standard deviation of $\sigma = 800$ g, we consider the test of

$$H_0 :\ \theta = 3300 \quad \text{versus} \quad H_1 :\ \theta \neq 3300;$$

here 3300 g is the overall average birthweight in King County in 1977 (which can effectively be considered to be known), so that H_0 would correspond to the (believable) hypothesis that SIDS is not related to birthweight. We apply the test (4.6) with $\theta_1 = \theta_0$ (i.e. $\Delta = 0$) and the default choice of $\tau^2 = 2\sigma^2$. From Table 1, we find that $a \equiv a(30, 0) = 3.051$, and simple calculations yield $t = 0.485$ and $B_c(t) = 4.968$, so that $B_c(t) > a$. Thus, according to $\mathbf{T}_1^*$, we accept H_0 and report the CEP $\beta^*(B_c) = 1/(1 + 4.968) = 0.168$.

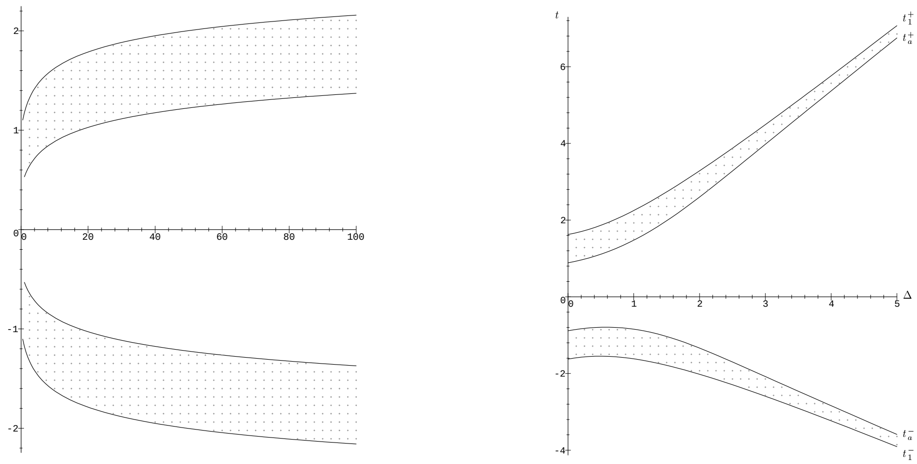
Table 1. Selected Values of $a(\lambda, \Delta)$ for the Two-sided Test with the Conjugate Prior

λ	$\|\Delta\|$=0	1	2	3	4	5
1	1.317	1.655	1.777	1.793	1.780	1.802
5	1.932	2.506	3.306	3.449	3.483	3.500
10	2.321	2.966	4.206	4.617	4.683	4.710
15	2.576	3.256	4.744	5.442	5.559	5.593
20	2.768	3.471	5.121	6.085	6.272	6.314
25	2.922	3.642	5.407	6.608	6.882	6.936
30	3.051	3.783	5.637	7.046	7.421	7.490
40	3.260	4.010	5.990	7.749	8.343	8.455
50	3.425	4.188	6.257	8.293	9.116	9.287
60	3.563	4.336	6.470	8.732	9.781	10.026
70	3.681	4.462	6.647	9.096	10.362	10.694
80	3.784	4.571	6.798	9.404	10.878	11.305
90	3.876	4.668	6.929	9.671	11.338	11.868
100	3.958	4.756	7.045	9.903	11.754	12.390

One can, alternatively, write the test $\mathbf{T}_1^*$ in terms of the standard statistic, t, as follows:

$$\mathbf{T}_1^* : \begin{cases} \text{if } \ t \le t_1^- \text{ or } t \ge t_1^+, & \text{reject } H_0 \text{ and report the CEP } \alpha^*(B_c(t)); \\ \text{if } \ t_1^- < t < t_a^- \text{ or } t_a^+ < t < t_1^+, & \text{make no decision;} \\ \text{if } \ t_a^- \le t \le t_a^+, & \text{accept } H_0 \text{ and report the CEP } \beta^*(B_c(t)). \end{cases}$$

Figure 1 below illustrates the effects of the *"shift"* parameter Δ and of the sample size n (through λ) on the *no decision region* corresponding to the test $\mathbf{T}_1^*$ as expressed in terms of the standard statistic t. Note the symmetry of the regions when $\Delta = 0$ and that the size of the *no decision region* decreases as Δ increases.



(a) (b)

Figure 1. Illustration of the no-decision region
(a) with $\Delta = 0$ and (b) with $\lambda = 10$ (fixed)

As we can see in (4.6) and in these figures, the magnitude of $a \equiv \psi(1)$ determines the "size" of the *no decision region.* If this 'size' is too large, the provided procedure will not be useful in practice. We first show that, for the testing problem at hand, a grows as $\log n$ for fixed Δ, and is bounded for fixed λ. Recall that $\lambda = n\tau^2/\sigma^2$. Here and henceforth, we will use the notation $f(y) \sim g(y)$, as $y \to \infty$ to mean $\lim_{y\to\infty} f(y)/g(y) = 1$.

Theorem 2 *Assume $a(\lambda, \Delta)$ satisfies the equation $F_0(a) = 1 - F_c(1)$. Then*
(a) for each fixed Δ, $\quad a(\lambda, \Delta) \sim \log \lambda$*, as* $\lambda \to \infty$*;*
(b) for each fixed λ, $\quad \lim\limits_{\Delta\to\infty} a(\lambda, \Delta) = (1+\lambda)^{\frac{1}{2}\frac{\lambda}{1+\lambda-\sqrt{1+\lambda}}}$.

See Berger, Boukai and Wang (1996b) for a detailed proof. A similar result on the asymptotic behavior of $t_a^{\pm}$ and $t_1^{\pm}$ can be obtained directly from Theorem 2.

Corollary 1 *For $\pi \in \Pi_c$ and with fixed Δ, $|t_1|^{\pm} \sim \sqrt{\log(n)}$ and $|t_a|^{\pm} \sim \sqrt{\log(n)}$ as $n \to \infty$.*

While the value of a is growing with the sample size n, so that the no-decision region is growing in size, a much more relevant concern in practice is the probability (under H_0 and H_1) of falling in the no-decision region. The following immediate corollary shows that this probability is vanishingly small asymptotically.

Corollary 2 *For $\pi \in \Pi_c$ and with fixed Δ,*

$$P_i(\textit{no-decision}) = O(n^{-\frac{1}{2}}(\log(n))^{\frac{1}{2}}), \quad i = 0, 1, \quad \textit{as} \;\; n \to \infty.$$

1.5 The "Two-Sided" Test with Uniform Prior

We consider here the same two-sided normal testing problem as in Section 4 (i.e. with $\Theta_1 = \{\theta \neq \theta_0\}$ in (3.1)), but now with the uniform prior over Θ_1, so that $\pi \in \Pi_u$. It is straightforward to verify that, with $\pi \in \Pi_u$, the marginal p.d.f., $m_\pi(t)$ in (3.3), can be written as

$$m_\pi(t) = \frac{1}{\lambda}[\Phi(t+\lambda) - \Phi(t-\lambda)], \tag{5.1}$$

where we have set $\lambda = \sqrt{n}\tau/\sigma$. From (3.2) and (4.1), the corresponding expression for the Bayes factor B_u, from (3.5), is

$$B_u(t) = 2\lambda\,\phi(t)/[\Phi(t+\lambda) - \Phi(t-\lambda)], \tag{5.2}$$

which has a maximum value $B_u^* \equiv B_u(0) = \sqrt{2/\pi}\lambda\,/[2\Phi(\lambda) - 1)]$. The situation here is somewhat more complicated than that of the previous section, since the Bayes factor $B_u(t)$ in (5.2) involves the normal c.d.f., $\Phi(\cdot)$. To arrive at similar results to those of Section 4, we present the following series of Lemmas regarding the Bayes factor B_u. Their proofs are provided in Berger, Boukai and Wang (1996b).

Lemma 2 *Let $\pi \in \Pi_u$, and $0 < b < B_u^*$ be fixed. Then there is only one positive solution t_b to the equation $B_u(t_b) = b$, and furthermore, this positive solution t_b satisfies $t_b \sim \sqrt{2\log\lambda}$ as $\lambda \to \infty$.*

In practice, for all $b > 0$, the value of t_b can be determined numerically as the solution of $B_u(t_b) = b$, which by (5.2) can be written as

$$2\lambda\exp\{-\frac{t_b^2}{2}\} = b\int_{t_b-\lambda}^{t_b+\lambda}\exp\{-\frac{s^2}{2}\}ds. \tag{5.3}$$

This identity is used repeatedly in the sequel. It is also used to obtain expressions for $F_0(1)$ and $F_u(1)$ as are given by:

$$F_0(1) = 2\Phi(-t_1) \tag{5.4}$$

and

$$F_u(1) = \frac{1}{\lambda}\int_{t_1}^{\infty}[\Phi(t+\lambda) - \Phi(t-\lambda)]dt. \tag{5.5}$$

With the result of Lemma 2, the two c.d.f.'s, (5.4)-(5.5), can be evaluated asymptotically, as summarized in Lemma 3 below.

Lemma 3 *If $\pi \in \Pi_u$, then $F_0(1) \sim \frac{1}{\lambda\sqrt{2\log\lambda}}$ and $1 - F_u(1) \sim \frac{\sqrt{2\log\lambda}}{\lambda}$ as $\lambda \to \infty$. Hence, for sufficiently large λ, $F_0(1) < 1 - F_u(1)$, so that $\psi(1) > 1$ in (2.2).*

Consequently, the form of the modified Bayesian test $\mathbf{T}_1^*$, in this case, is exactly as given in (4.6), with the Bayes factor $B_c(t)$ replaced by $B_u(t)$ from (5.2), and with $a \equiv \psi(1)$, as determined by the equation $F_0(a) = 1 - F_u(1)$. In fact, it follows from (5.3)-(5.5) that the numerical values of t_1, t_a and of a (see Table 2 below), for each fixed λ, must satisfy the following set of equations:

$$\begin{cases} 2\lambda \exp\{-\frac{t_1^2}{2}\} = \int_{t_1-\lambda}^{t_1+\lambda} \exp\{-\frac{s^2}{2}\}ds; \\ 2\lambda\Phi(t_a) = \int_{t_1}^{\infty}[\Phi(t+\lambda) - \Phi(t-\lambda)]dt; \\ a = \frac{2\lambda\exp\{-\frac{t_a^2}{2}\}}{\int_{t_a-\lambda}^{t_a+\lambda}\exp\{-\frac{s^2}{2}\}ds}. \end{cases}$$

In similarity to the conjugate prior case discussed in Section 4, a still has a nice asymptotic behavior, as presented in Theorem 3 (see Berger, Boukai and Wang, 1996b, for a detailed proof).

Theorem 3 *For $\pi \in \Pi_u$, we have $a \sim 2\log\lambda$ and*

$$P_i(\textit{no-decision}) = O(\lambda^{-1}(2\log(\lambda))^{\frac{1}{2}}), \quad i = 0, 1, \quad \textit{as } \lambda \to \infty.$$

With $\lambda = \sqrt{n}\tau/\sigma$, the results of this theorem should be compared to those of Theorem 2 (a) and Corollary 2.

Table 2. Selected Values of a, t_a and t_1 for the Two-sided Test with a Uniform Prior

λ	a	t_a	t_1
1	1.131	0.478	1.049
2	1.442	0.623	1.184
3	1.814	0.762	1.360
4	2.178	0.875	1.528
5	2.496	0.969	1.664
10	3.559	1.271	2.038

λ	a	t_a	t_1
20	4.693	1.565	2.354
30	5.379	1.728	2.520
40	5.874	1.840	2.632
50	6.262	1.924	2.715
70	6.853	2.048	2.836
100	7.486	2.175	2.960

1.6 Lower Bound on α^*

A lower bound on the conditional error probability, $\alpha^*(B_\pi)$, over a class Π of priors is clearly of great interest here. It provides a bound on the amount of evidence in support of the null hypothesis (through the posterior probability of H_0) obtained in the modified Bayesian test $\mathbf{T}_1^*$. While the conditional Frequentist might be reluctant to employ a particular prior $\pi \in \Pi$, the lower bound on the error probability is quite compelling.

Theorem 4 *Let $\pi \in \Pi_c$ and let $B_c(t)$ be as defined by (4.2). Then*

$$\alpha^*(B_c(t)) \equiv \frac{B_c(t)}{1+B_c(t)} \geq [1+e^{\frac{t^2}{2}}]^{-1}, \quad \forall t \in \mathbb{R}.$$

Unfortunately, there is no closed form expression for the lower bound in the uniform prior case. However, it can be computed numerically. In Table 3 we provide selected numerical values of the lower bound of α^* over the two classes of priors considered here, for selected values of t along with the corresponding $P-$values. The implications are quite surprising. For instance, a $P-$value of 0.05 corresponds to a conditional Frequentist error-probability of *at least* 0.290 for any $\pi \in \Pi_u$ (and indeed any unimodal prior symmetric about θ_0). Thus a $P-$value of 0.05 does not seem to be substantial evidence against H_0 from almost any reasonable conditional Frequentist measure.

Table 3. Lower Bound on $\alpha^*(B_c)$ and $\alpha^*(B_u)$

$P-$Value	t	$\pi \in \Pi_c$	$\pi \in \Pi_u$
0.10	1.645	0.205	0.389
0.05	1.960	0.128	0.290
0.01	2.576	0.035	0.109
0.001	3.291	0.004	0.018

These results on lower bound are simply "borrowed" from the robust Bayesian literature (cf., Edwards, Lindman and Savage (1963) and Berger and Sellke (1987)). One of the strengths of the dual Bayesian-Frequentist interpretation of $\mathbf{T}_1^*$ is that known Bayesian results can directly be carried over and given a conditional Frequentist interpretation.

1.7 The "One-Sided" Test

We continue with the same testing problem as in Section 4, but now with $\Theta_1 = \{\theta > \theta_0\}$ and a prior p.d.f $\pi \in \Pi_+$. With this prior p.d.f., the marginal p.d.f. (3.3) of t, conditional on H_1, becomes

$$m_+(t) = \frac{2}{\sqrt{1+\lambda}}\phi(\frac{t}{\sqrt{(1+\lambda)}})\Phi(\frac{\sqrt{\lambda}t}{\sqrt{1+\lambda}}),$$

where, as in Section 4, $\lambda = n\tau^2/\sigma^2$. Hence, the corresponding Bayes factor may be written as

$$B_+(t) = \frac{\sqrt{1+\lambda}}{2}\exp\{\frac{-\lambda t^2}{2(1+\lambda)}\}\left(\Phi(\frac{\sqrt{\lambda}t}{\sqrt{1+\lambda}})\right)^{-1}.$$

It can be verified that, in this case also, the *no decision region* is of the form $(1, a)$, where a is determined numerically by the following set of equations:

$$\begin{cases} B_+(t_1) = 1; \\ 1 - \Phi(t_a) = 2\int_{-\infty}^{t_1/\sqrt{1+\lambda}} \Phi(\lambda t)\phi(t)dt; \\ a = B_+(t_a). \end{cases}$$

Thus, the corresponding test $\mathbf{T}_1^*$ (as presented in terms of the standard test statistic t) is

$$\mathbf{T}_1^* : \begin{cases} \text{if } t \ge t_1, & \text{reject } H_0 \text{ and report the CEP } \alpha^*(B_+(t)); \\ \text{if } t_a < t < t_1\,, & \text{make no decision;} \\ \text{if } t \le t_a\,, & \text{accept } H_0 \text{ and report the CEP } \beta^*(B_+(t)). \end{cases}$$

Table 4 below presents values of a, t_a and t_1 as were evaluated numerically for selected choices of λ. Note that the *no decision region* is somewhat smaller than that in the two-sided test, unless λ is very small.

Table 4. Selected Values of a, t_a and t_1 for the One-sided Test

λ	a	t_a	t_1
1	1.271	0.183	0.560
2	1.448	0.262	0.731
3	1.580	0.320	0.841
4	1.858	0.367	0.923
5	1.774	0.406	0.987
10	2.084	0.541	1.190

λ	a	t_a	t_1
20	2.436	0.690	1.390
30	2.659	0.781	1.505
40	2.825	0.847	1.584
50	2.956	0.898	1.645
70	3.161	0.976	1.734
100	3.385	1.057	1.825

References

Berger, J. O., Boukai, B. and Wang, Y. (1996a). Unified frequentist and Bayesian testing of a precise hypothesis. To appear in *Statistical Science*.

Berger, J. O., Boukai, B. and Wang, Y. (1996b). Asymptotic properties of modified Bayesian-Frequentist tests. *Preprint Series in Mathematics and Statistics, # PR-96-17*, Department of Mathematical Sciences, Indiana University Purdue University, Indianapolis, IN.

Berger, J. O., Brown, L. D. and Wolpert, R. L. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing. *Annals of Statistics* **22**, 1787–1807.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statist. Sci.*, **3**, 317-352.

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of P-values and evidence. *J. Amer. Statist. Assoc.*, **82**, 112–122.

Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle*, 2nd Ed. Institute of Mathematical Statistics, Hayward, CA.

Brownie, C. and Kiefer, J. (1977). The ideas of conditional confidence in the simplest setting. *Commun. Statist.- Theory Meth.* **A 6(8)**, 691–751.

Delampady, M. and Berger, J. O. (1990). Lower bounds on Bayes factors for the multinomial distribution, with application to chi-squared tests of fit. *The Annals of Statistics*, **18**, 1295–1316.

Edwards, W., Lindman, H. and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psych. Rev.*, **70**, 193–242.

Fisher, L. D. and Van Belle, G. (1993). *Biostatistics: A Methodology for the Health Sciences.* New York: John Wiley and Son.

Jeffreys, H. (1961). *Theory of Probability*. London: Oxford University Press.

Kiefer, J. (1975). Conditional confidence approach in multi decision problems. In: P. R. Krishnaiah, Ed. *Multivariate Analysis IV.* New-York: Academic Press.

Kiefer, J. (1976). Admissibility of conditional confidence procedures. *Ann. Math. Statist.* **4**, 836–865.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators. (with discussion). *J. Amer. Statist. Assoc.* **72**, 789–827.

Wolpert, R. L. (1995). Testing simple hypotheses. In *Studies in Classification, Data Analysis, and Knowledge Organization*, Vol. 7. (H. H. Bock and W. Polasek, Eds.). Springer-Verlag, Heidelberg, 289–297.