A UNIFIED CONDITIONAL FREQUENTIST AND BAYESIAN TEST FOR FIXED AND SEQUENTIAL SIMPLE HYPOTHESIS TESTING*

by

James O. Berger  Lawrence D. Brown
Purdue University  Cornell University

Robert L. Wolpert
Duke University

Technical Report #93-14C

Department of Statistics
Purdue University

February 1993
Revised, February 1994

A UNIFIED CONDITIONAL FREQUENTIST AND BAYESIAN
TEST FOR FIXED AND SEQUENTIAL SIMPLE HYPOTHESIS TESTING*

by
James O. Berger        Lawrence D. Brown
Purdue University       Cornell University
Robert L. Wolpert       Duke University

Summary

Pre-experimental Frequentist error probabilities are arguably inadequate, as summaries of evidence from data, in many hypothesis-testing settings. The Conditional Frequentist may respond to this by identifying certain subsets of the outcome space and reporting a conditional error probability, given the subset of the outcome space in which the observed data lie. Statistical methods consistent with the Likelihood Principle, including Bayesian methods, avoid the problem by a more extreme form of conditioning.

In this paper we prove that the Conditional Frequentist’s method can be made exactly equivalent to the Bayesian’s in simple versus simple hypothesis testing: specifically, we find a conditioning strategy for which the Conditional Frequentist’s reported conditional error probabilities are the same as the Bayesian’s posterior probabilities of error. A Conditional Frequentist who uses such a strategy can exploit other features of the Bayesian approach — for example, the validity of sequential hypothesis tests (including versions of the sequential probability ratio test, or SPRT) even if the stopping rule is incompletely specified.

Short Title: A Unified Frequentist and Bayesian Test
AMS Classification: Primary 62A20, Secondary 62A15.

Keywords: Likelihood Principle, Conditional Frequentist, Bayes Factor, Likelihood Ratio, Significance, Type-I Error, Bayesian Statistics, Stopping Rule Principle.

*Research supported by National Science Foundation Grants DMS–8923071, DMS–8903842, DMS–9303556, and SES–8921227. The project was initiated at the Cornell Workshop on Conditional Inference, sponsored by the Army Mathematical Sciences Institute and the Statistics Center, Cornell University, Ithaca, NY, June 3-14, 1991.
1. INTRODUCTION

1.1 The Basic Conditional Test

We observe the random variable $X \in \mathcal{X}$ and wish to test the simple hypotheses

$$H_0: \text{X has density } f_0(x) \text{ versus } H_1: \text{X has density } f_1(x).$$

Denote the likelihood ratio of $H_0$ to $H_1$ or Bayes factor in favor of $H_0$ by

$$B(x) = \frac{f_0(x)}{f_1(x)}.$$ 

(Note that small values of $B(x)$ correspond to rejection of $H_0$.) Let $F_0$ and $F_1$ be the c.d.f.s of $B(X)$ under $f_0$ and $f_1$, respectively. For ease of exposition, we will assume that $F_0$ and $F_1$ are continuous and invertible, and denote their inverses by $F_0^{-1}$ and $F_1^{-1}$.

The standard equal-tailed (minimax) most powerful test is determined by the critical value $c$, which satisfies

$$F_0(c) = 1 - F_1(c).$$

This test rejects $H_0$ if $B(x) \leq c$, accepts $H_0$ if $B(x) > c$, and has error probabilities of Type I and Type II given by $\alpha = \beta = F_0(c) = 1 - F_1(c)$.

Consider, instead, the testing procedure $T_1$ defined as follows:

if $B(x) \leq c$, reject $H_0$ and report the conditional error probability $\alpha(B) = B/(1 + B)$;
if $B(x) \geq c$, accept $H_0$ and report the conditional error probability $\beta(B) = 1/(1 + B)$.

This simple procedure has a number of attractive properties. First, $\alpha(B)$ and $\beta(B)$ are the posterior probabilities of $H_0$ and $H_1$, respectively (assuming equal prior probabilities), so the reported error probabilities are those that a Bayesian would report. Second, $T_1$ is a test often considered by Likelihoodists. Birnbaum (1961) called $\alpha(B)$ and $\beta(B)$ the "intrinsic significance levels." Third, and quite surprisingly, $T_1$ is also a valid frequentist procedure, more precisely a valid conditional frequentist procedure. (See Section 2 for definitions and justifications.) The situation here thus differs from that in classical testing of simple hypotheses where the unconditional error probabilities $\alpha$ and $\beta$ are typically very different from the Bayesian (or likelihood) $\alpha(B)$ and $\beta(B)$. Indeed, this is one of the
common criticisms of classical testing of simple hypotheses: the reported error probabilities
don't distinguish between data at the boundary of the rejection or acceptance regions and
data deep within the region.

**Example 1.** Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $\mathcal{N}(\theta, 1)$, and that it is desired to test
$H_0: \theta = -1$ versus $H_1: \theta = 1$. Then

$$B = \prod_{i=1}^n \frac{(2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_i + 1)^2\}}{(2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_i - 1)^2\}} = \exp\{-2n \bar{x}\},$$

and $T_1$ becomes:

if $B \leq 1$ (i.e., $\bar{x} \geq 0$), reject $H_0$ and report error probability $\alpha(B) = (1 + \exp\{2n \bar{x}\})^{-1}$;
if $B > 1$ (i.e., $\bar{x} < 0$), accept $H_0$ and report error probability $\beta(B) = (1 + \exp\{-2n \bar{x}\})^{-1}$.

For comparison, the classical Neyman-Pearson test with equal error probabilities utilizes
the same rejection and acceptance regions, but reports the error probabilities (of Type I
and Type II) $\alpha = \beta = 1 - \Phi(\sqrt{n})$, where $\Phi$ is the standard normal c.d.f.. When $n=4$, a
comparison of the reports is given in Table 1. For completeness, we also report the $P$-value
against $H_0$, which here is $1 - \Phi(2(\bar{x} + 1))$.

**Table 1.** Comparison of Error Probabilities When $n = 4$

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>For Test $T_1$</th>
<th>For $N - P$ Test</th>
<th>$P$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha(B)$</td>
<td>$\beta(B)$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.025</td>
</tr>
<tr>
<td>1/4</td>
<td>0.12</td>
<td>0.88</td>
<td>0.025</td>
</tr>
<tr>
<td>1</td>
<td>0.00034</td>
<td>0.99966</td>
<td>0.025</td>
</tr>
</tbody>
</table>

The intuitive attractiveness of $\alpha(B)$ and $\beta(B)$ is clear. If the data is $\bar{x} = 0$, intuition
suggests that the evidence equally supports $H_0: \theta = -1$ and $H_1: \theta = 1$; $\alpha(B)$ and $\beta(B)$ so
indicate, while $\alpha$ and $\beta$ (and the $P$-value) do not. When $\bar{x} = 1$, in contrast, intuition would
suggest overwhelming evidence for $H_1$ (note that $\bar{x} = 1$ is four standard deviations from
$\theta = -1$); again $\alpha(B)$ and $\beta(B)$ reflect this. (We delay further discussion of the $P$-value to
the end of Section 1.4.)
The above succinctly summarizes the major point of the paper: in testing simple hypotheses the counterintuitive \( N - P \) test can be replaced by the much more attractive \( T_1 \), with complete Bayesian, likelihood, and frequentist justification. (Indeed, it will be argued in Section 3 that \( T_1 \) is actually a better frequentist test than the \( N - P \) test.)

1.2 A Bayesianly Motivated Modification

To a Bayesian, \( T_1 \) will appear to be somewhat unnatural when \( c \neq 1 \). If, say, \( c = 3 \) and \( B(x) = 2 \), then \( T_1 \) would be “reject \( H_0 \), and report conditional error probability \( \alpha(B) = 2/3 \)” The obvious question is “why should \( H_0 \) be rejected if that action has a 2/3 chance of being wrong?”

This might, of course, be a perfectly rational thing to do if the losses in accepting and rejecting are asymmetric. Indeed, in Section 2, \( T_1 \) will be seen to be a true Bayes test for certain losses. For inference without a specified loss, however, this behavior of \( T_1 \) is unappealing and may seem silly to practitioners.

A useful modification of \( T_1 \), that alleviates this difficulty, is to incorporate a “no decision” region. To describe this modification, \( T_1^* \), define

\[
\begin{align*}
  r &= 1 \text{ and } a = F_0^{-1}(1 - F_1(1)) \quad \text{if } F_0(1) \leq 1 - F_1(1); \\
  r &= F_1^{-1}(1 - F_0(1)) \text{ and } a = 1 \quad \text{if } F_0(1) > 1 - F_1(1).
\end{align*}
\]

Then,

- if \( B(x) \leq r \), reject \( H_0 \) and report the conditional error probability \( \alpha(B) = B/(1 + B) \);
- if \( r < B(x) < a \), make no decision;
- if \( B(x) \geq a \), accept \( H_0 \) and report the conditional error probability \( \beta(B) = 1/(1 + B) \).

An example will be given in Section 2.4, which will indicate that the “no decision” region is typically rather small. Indeed, \( T_1^* \) and \( T_1 \) agree (i.e., the “no decision” region disappears) whenever

\[
F_0(1) = 1 - F_1(1),
\]

in which case \( r = a = 1 \) in \( T_1 \). The main situations in which (1.1) is satisfied are situations of “symmetry”; see Section 3.1 and 4.2 for definition and illustration.
1.3 Use of $T_1$ in Sequential Analysis

In sequential testing, use of $T_1$ (or $T_1^*$) can greatly simplify the analysis and has surprising foundational implications. First, an example:

**Example 2.** Consider the scenario of Example 1, but suppose the data is observed sequentially. Letting $N$ denote the (random) stopping time for the experiment, it is still true that, upon stopping at $N = n$, $B_n(x_1,\ldots,x_n) = \exp\{-2n \bar{x}_n\}$. (Bayes factors do not depend on the stopping rule.) Condition (1.1) (which guarantees that $T_1^* = T_1$) can be written (noting that $B_n \leq 1 \iff \bar{x}_n \geq 0$) $P(\bar{X}_N \geq 0|\theta = -1) = P(\bar{X}_N \leq 0|\theta = 1)$. This will be satisfied by stopping rules that are “symmetric,” such as those of the form

"stop" at the first $n$ for which $|\bar{x}_n| \geq g(n), \quad (1.2)$

where $g(n)$ is any nonnegative function such that the stopping rule is proper (i.e., "stops" with probability one). The SPRT with equal error probabilities is of this form, with $g(n) \propto 1/n$.

For this situation, the test $T_1$ becomes

if $\bar{x}_n \geq g(n)$, stop experimentation, reject $H_0$, and report the conditional error probability $\alpha(B_n) = 1/[1 + \exp(2n \bar{x}_n)];$

if $\bar{x}_n < -g(n)$, stop experimentation, accept $H_0$, and report the conditional error probability $\beta(B_n) = 1/[1 + \exp(-2n \bar{x}_n)].$

Observe first that the test $T_1$ is remarkably easy to implement. There is no need for the usually difficult computation of unconditional frequentist error probabilities, as is necessary for the SPRT. Even the choice of $g(n)$ does not necessarily involve such computations; for instance, the choice

$$g(n) = \frac{1}{2n} \log\left(\frac{1}{\alpha^*} - 1\right),$$

is attractive, guaranteeing that the reported conditional error probability will not exceed $\alpha^*$.

The foundational surprise here is that the reported error probabilities arising from $T_1$ do not depend on the specific symmetric stopping rule chosen. This is surprising in light
of the common frequentist opposition to the *Stopping Rule Principle (SRP)*. The SRP states that final inferences should not depend on the stopping rule used to obtain the data (cf., Berger and Berry, 1988, or Berger and Wolpert, 1988). The fact that the "optimal" frequentist procedure, $T_1$, in the above example, seems to ignore the stopping rule in its error report is startling. This issue is discussed more fully in Section 4.

1.4 Background on Conditional Testing and $P$-Values

There is a long history of attempts to modify frequentist theory by utilizing some form of conditioning. Earlier works are summarized in Kiefer (1977) and Berger and Wolpert (1988). Kiefer (1977), together with Kiefer (1975, 1976) and Brownie and Kiefer (1977), formally established the *conditional confidence approach*; a modification of this approach is discussed in Section 2.1, and forms the basis of our analysis. We seek, however (for the problem of testing simple hypotheses), to overcome certain difficulties with the conditional confidence approach, difficulties discussed in the above works, and by the discussants of Kiefer (1977).

Perhaps the main difficulty is that there is a plethora of conditional confidence procedures, and choosing among them is a daunting task. Brown (1978) makes a serious effort in this direction (see Section 3.2), but a general practical prescription seems remote. A second potentially serious difficulty is that conditional confidence procedures that do not have a Bayesian basis can remain incoherent and anti-intuitive. In this regard, Birnbaum (1961) and Barnard (in the discussion of Kiefer, 1977) argue that any effort to develop intuitively sensible conditional answers, for the testing problem we consider, must be based on $B(x) = f_0(x)/f_1(x)$. Even Kiefer (1977) mentioned the appeal of doing so, but felt (as did Birnbaum) that this was incompatible (in general) with a frequentist interpretation. The main thrust of this paper is to show that the two are compatible; that one can retain the coherency and ease of interpretability of $B(x)$, while achieving a conditional frequentist interpretation for its use.

Although this paper is basically a descendant of Kiefer (1977) (and Birnbaum, 1961), there has been a considerable subsequent literature on conditional frequentist testing from the "estimated confidence" perspective. This approach was also proposed in Kiefer (1977) (although it, also, has earlier roots). The focus of the approach (in testing) is to provide
an "estimate" of the indicator function: 1 if $H_0$ is true, 0 if $H_0$ is false. Developments along these lines can be found in Schafsma, Tolboom, and van der Meulen (1989); Hwang, Casella, Robert, Wells, and Farrell (1992); and van der Meulen (1992). Chatterjee and Chattopadhyay (1992) develop a related approach based on a betting interpretation of evidence. Exploration of these new approaches is valuable, but they frequently are susceptible to the same criticisms that we mentioned earlier in relation to the conditional confidence approach.

Finally, use of the $P$-value for testing $H_0$ versus $H_1$ should be mentioned. Because of the intuitive objections to the $N - P$ test that were discussed in Section 1, use of the $P$-value has become quite popular. There is, however, no sound justification for using it in simple versus simple hypothesis testing. It has no Bayesian or likelihood interpretation, typically differing significantly from, say, $T_1$. Thus, in Table 1 for $\bar{x} = 0$, the $P$-value is the anti-intuitive 0.025. And it has no true frequentist justification; for instance, the expected $P$-value under $f_0$, conditional on rejecting, is only one-half the actual probability of rejecting under $f_0$, so that "on average over the rejection region" the $P$-value substantially underestimates the actual Type I error rate. Since $T_1$ is completely justified from all foundational perspectives and is as "data-adaptive" as the $P$-value, $T_1$ is clearly to be preferred.
2. CONDITIONAL FREQUENTIST TESTING

2.1 Basic Elements

The approach we consider here is that formalized by Kiefer (1975, 1976, 1977) and Brownie and Kiefer (1977), called the conditional confidence approach. (See, also, Brown, 1978, and Berger, 1985a,b.) The idea is to partition the sample space, \( \mathcal{X} \), i.e., let

\[
\mathcal{X} = \bigcup_{s \in S} \mathcal{X}_s, \text{ with } \mathcal{X}_s \cap \mathcal{X}_{s'} = \emptyset,
\]

and then develop frequentist measures conditional on \( \mathcal{X}_s \).

For testing, the usual conditional frequentist measures considered are the conditional probabilities of Type I and Type II error,

\[
\alpha(s) = P_0(\text{Type I error} \mid \mathcal{X}_s) = P_0(\text{rejecting} \mid \mathcal{X}_s),
\]

(2.1)

\[
\beta(s) = P_1(\text{Type II error} \mid \mathcal{X}_s) = P_1(\text{accepting} \mid \mathcal{X}_s).
\]

(2.2)

One operates by observing the partition \( \mathcal{X}_s \) in which the data happens to lie, and then reporting the relevant \( \alpha(s) \) or \( \beta(s) \) as the error probability (or, perhaps, reporting both). It is worth noting that, if one treats \( s \) as random under \( f_0 \) and/or \( f_1 \), then

\[
E_0[\alpha(s)] = E_0[P_0(\text{Type I error} \mid \mathcal{X}_s)] = P_0(\text{Type I error}) = \alpha,
\]

(2.3)

\[
E_1[\beta(s)] = P_1(\text{Type II error}) = \beta,
\]

so the conditional tests can be viewed as simply dividing up the overall probabilities of Type I and Type II errors among the various partitions.

It should be remarked that Kiefer worked in terms of "conditional confidence," which is 1-"conditional error." Also, although Kiefer stopped short of explicitly recommending a particular conditional frequentist procedure for simple versus simple hypothesis testing, he (as well as Brown, 1978) seemed to favor the procedure arising from requiring \( \alpha(s) = \beta(s) \) for all \( s \), with \( \{\mathcal{X}_s: s \in S\} \) chosen to be as fine a partition as possible, subject to this constraint. See Kiefer (1977, expressions (3.12)-(3.14)) for development of this procedure. Although this "equal probability continuum" partition has some attractive properties, it can be shown to yield counterintuitive results in many nonsymmetric situations; hence our preference for \( T_1 \) and \( T_1^* \).
2.2 The General Conditional Test for Simple Hypotheses

The tests $T_1$ and $T_1^*$, described in Section 1, can be usefully generalized. Consider the test $T_{t, \rho}$ defined as follows for $0 < t < \infty$ and $0 < \rho < \infty$: define

$$r_{t, \rho} = \ell\rho \quad \text{and} \quad a_{t, \rho} = F_0^{-1}(1 - \rho F_1(\ell\rho)) \quad \text{if} \quad F_0(\ell\rho) \leq 1 - \rho F_1(\ell\rho); \quad (2.4)$$

$$r_{t, \rho} = F_1^{-1}\left(\frac{1}{\rho}[1 - F_0(\ell\rho)]\right) \quad \text{and} \quad a_{t, \rho} = \ell\rho \quad \text{if} \quad F_0(\ell\rho) \geq 1 - \rho F_1(\ell\rho); \quad (2.5)$$

then, the generalization of $T_1^*$ is given by

if $B(x) \leq r_{t, \rho}$, reject $H_0$ and report the conditional error probability $\alpha_{t, \rho}(B) = B/(\rho + B)$;

if $r_{t, \rho} < B(x) < a_{t, \rho}$, make no decision;

if $B(x) \geq a_{t, \rho}$, accept $H_0$ and report the conditional error probability $\beta_{t, \rho}(B) = \rho/(\rho + B)$.

The motivation for considering this test is again Bayesian. Indeed, suppose it is desired to test $H_0: f_0$ versus $H_1: f_1$, with $\pi_0$ and $\pi_1$ being the prior probabilities of $H_0$ and $H_1$, respectively ($\pi_1 = 1 - \pi_0$), and where the loss is assumed to be 0 for a correct decision and $\ell_i$ for an incorrect decision when $H_i$ is true. Then, defining

$$\rho = \pi_1/\pi_0 \quad \text{and} \quad \ell = \ell_1/\ell_0,$$

the optimal Bayes test, $T_{t, \rho}^*$, can be written

if $B \leq \ell\rho$, reject and report posterior risk $\ell\rho \alpha_{t, \rho}(B)$;

if $B > \ell\rho$, accept and report posterior risk $\ell\rho \beta_{t, \rho}(B)$.

This is the test $T_{t, \rho}$, except for the presence in $T_{t, \rho}$ of the "no decision" region $(r_{t, \rho}, a_{t, \rho})$. (Clearly, multiplying $\alpha_{t, \rho}(B)$ by $\ell\rho$ and $\beta_{t, \rho}(B)$ by $\ell_1$ converts the conditional error probabilities in $T_{t, \rho}$ to conditional risks.) The "no decision" region is the price that must be paid to obtain a conditional frequentist interpretation for the optimal Bayes test. This interpretation is developed in the following section.

A strict frequentist can also use $T_{t, \rho}$ effectively. Suppose a frequentist was planning to use the unconditional $N - P$ test with critical value $c$ and error probabilities $\alpha = F_0(c)$ and $\beta = 1 - F_1(c)$. This can be replaced by $T_{t, \rho}^*$, with

$$\rho^* = (1 - \alpha)/(1 - \beta) \quad \text{and} \quad \ell^* = c/\rho^*.$$

8
This test is the generalization of $T_1$. It is straightforward to verify that $T_{\ell^*, \rho^*}$ still has $c$ as a critical value, and it does not have a "no decision" region. But instead of reporting the unconditional $\alpha$ and $\beta$, one now reports the conditional error probabilities $\alpha_{\rho^*}(B)$ and $\beta_{\rho^*}(B)$. Note that the choice of $\ell^*$ and $\rho^*$ is here just viewed as formalism; their interpretation in terms of losses and priors is not necessary.

While we feel that use of $T_{\ell^*, \rho^*}$ is considerably better than use of the $N-P$ test, some of us hesitate to recommend its general use in practice because of the intuitive concerns raised in Section 1.2 and prefer to allow a "no decision" region. This can clearly be left to the personal taste of the practitioner, however.

2.3 Conditional Frequentist Interpretation of $T_{\ell, \rho}$

Define

$$\psi(s) = F_0^{-1}(1 - \rho F_1(s)) \quad \text{for} \quad 0 < s \leq r_{\ell, \rho}. \quad (2.6)$$

Note that $a_{\ell, \rho} = \psi(r_{\ell, \rho})$. Define, for $0 < s \leq r_{\ell, \rho}$,

$$\mathcal{X}_s = \{x \in \mathcal{X}: B(x) = s \text{ or } B(x) = \psi(s)\},$$

and let $\mathcal{X}_0 = \{x \in \mathcal{X}: r_{\ell, \rho} < B(x) < a_{\ell, \rho}\}$. Then $\{\mathcal{X}_s: s \in [0, r_{\ell, \rho}]\}$ is a partition of $\mathcal{X}$.

Under a "symmetry" condition to be defined in Section 3.1, this partition turns out to be a "maximal ancillary" partition. An ancillary partition is a partition in which, for each $s$, $\mathcal{X}_s$ has the same probability (or density) under $f_0$ as under $f_1$. The term "maximal" means that no finer partition, among partitions based on the sufficient statistic $B(X)$, can be ancillary. In the nonsymmetric case, the above partition is not ancillary and hence has no clear intrinsic justification; rather, it is extrinsically justified by yielding error probabilities that are equal to the intuitively attractive and foundationally secure posterior probabilities.

**Theorem 1.** For the test $T_{\ell, \rho}$ and the above partition, if $s > 0$ then

$$\alpha(s) = P_0(\text{rejecting } H_0 | \mathcal{X}_s) = \alpha_{\rho}(B) = B/(\rho + B);$$

$$\beta(s) = P_1(\text{accepting } H_0 | \mathcal{X}_s) = \beta_{\rho}(B) = \rho/(\rho + B).$$

*Proof.* Let $f_i^*(\cdot)$ denote the density of the (sufficient) statistic $B(X)$ under $f_i$, $i = 1, 2$. We first show that

$$f_0^*(b) = b f_1^*(b). \quad (2.7)$$
To see this, observe that

\[
\int_0^b f_0^*(y)dy = P_0(B(X) \leq b)
\]

\[
= \int_{\{x : B(x) \leq b\}} f_0(x)dx
\]

\[
= \int_{\{x : B(x) \leq b\}} B(x)f_1(x)dx
\]

\[
= \int_0^b yf_1^*(y)dy,
\]

the last step following from the change of variables \( y = B(x) \). Differentiating both sides with respect to \( b \) establishes (2.7). Calculus also yields

\[
\frac{d}{ds} \psi(s) = -\rho f_1^*(s)/f_0^*(\psi(s)). \tag{2.8}
\]

Applying first (2.8) and then (2.7) thus yields

\[
\alpha(s) = P_0(\text{rejecting} | \mathcal{X}_s)
\]

\[
= f_0^*(s)/[f_0^*(s) + f_0^*(\psi(s))|\frac{d}{ds} \psi(s)|]
\]

\[
= f_0^*(s)/[f_0^*(s) + \rho f_1^*(s)]
\]

\[
= B/[B + \rho],
\]

concluding the proof for \( \alpha(s) \). The proof for \( \beta(s) \) is similar. \( \square \)

2.4 Discussion of \( T_{\ell, \rho} \)

As promised, the test \( T_{\ell, \rho} \) simultaneously has a Bayesian and a frequentist justification, with the decision and the reported risk (loss \( \times \) error probability) being the same under either paradigm. Of course, the interpretation of the risk (or error probability) will differ for Bayesians and frequentists, but this will be irrelevant to practice. Indeed, those who see merit in both the Bayesian and frequentist philosophies might be especially pleased in the dual interpretation of the reported risk.

Strict frequentists might complain that, in operation, \( T_{\ell, \rho} \) does not provide all needed information. Frequentist dogma asserts that one cannot, say, only look at the Type I error probability when rejecting, but must also look at Type II error. In the conditional frequentist setting, this dogma would imply that both \( \alpha(s) \) and \( \beta(s) \) must be reported, whether
one rejects or accepts. The given procedure, $T_{t,\rho}$, only provides $\alpha(s)$ upon rejection, and $\beta(s)$ upon acceptance.

There are several possible replies to this concern. First of all, the map (2.6) is available and could be used to compute the other error probability if really desired. In practice, however, we suspect that this will not be done, primarily because it is almost impossible to see how one should, say, use $\beta(s)$ upon rejecting. Second, it can be argued that the source of the frequentist dogma concerning the reporting of both error probabilities lies in an attempt to intuitively compensate for not conditioning; such compensation is clearly not needed for $T_{t,\rho}$. Finally, in the very important "symmetric" situation of Sections 3.1 and 4.2, we will see that $\alpha(s) = \beta(s)$ for all $s$, so that the issue does not then arise.

The major potential thorn in the use of $T_{t,\rho}$ is the presence of the "no decision" region $(r_{t,\rho}, a_{t,\rho})$. In a sense, this is the region over which one cannot force agreement between Bayesians and frequentists. If this region is too large, then the utility of the test $T_{t,\rho}$ is reduced (although, recall that a frequentist who is less concerned about agreement with Bayesians can just use $T_1$ or $T_{t^*,\rho^*}$, neither of which has a "no decision" region).

An encouraging fact is that it can be shown, in general, that, for any $\ell$ or any $\rho$, a solution in the other variable can be found to

$$r_{t,\rho} = a_{t,\rho} = \ell \rho,$$  \hspace{1cm} (2.9)

in which case the "no decision" region is empty. It can also be shown that the solutions to (2.9) in $\ell$ and $\rho$ are inversely related. To investigate how large the "no decision" region can be, consider the following example.

**Example 3.** Suppose $X_1, \ldots, X_n$ are i.i.d. exponential, with density $f(x_i|\theta) = \theta^{-1} \exp\{-x_i/\theta\}$ for $x_i > 0$, and that it is desired to test $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$, where $\theta_0 < \theta_1$. Defining $\gamma = \theta_0/\theta_1$, it is immediate that

$$B = \gamma^{-n} \exp\{-n(1 - \gamma)\bar{x}/\theta_0\},$$

which is monotonically decreasing in $\bar{x}$. Note that $0 \leq B \leq \gamma^{-n}$. Also, computation yields, for the CDF of $B$ under $H_i$,

$$F_i(b) = 1 - \Gamma(n, \frac{-\gamma i \log(b \gamma^n)}{(1 - \gamma)}),$$  \hspace{1cm} (2.10)
where $\Gamma(n,t) = \left(1/\Gamma(n)\right) \int_0^t y^{n-1} \exp\{-y\} dy$ is the incomplete Gamma function.

Consider, first, the effect of $\ell$, the loss ratio $\ell_1/\ell_0$. For the special case $n = 1$,

$$F_0(b) = (b\gamma)^{1/(1-\gamma)} \quad \text{and} \quad F_1(b) = (b\gamma)^{\gamma/(1-\gamma)}.$$

From these and (2.4) and (2.5), $r_{\ell,\rho}$ and $a_{\ell,\rho}$ can be explicitly obtained. When $\rho = 1$ and $\gamma = \frac{1}{2}$ for instance (say, when testing $H_0$: $\theta = 1$ versus $H_1$: $\theta = 2$ with equal prior probabilities),

$$r_{\ell,1} = \min\{\ell, 2(1-\ell/2)^2\} \quad \text{and} \quad a_{\ell,1} = \max\{\ell, 2(1-\ell/2)^{1/2}\}$$

if $0 \leq \ell \leq 2$; when $\ell > 2$, $r_{\ell,1} = 0$ and $a_{\ell,1} = 2$. (Note that the range of $B$ here is $0 \leq B \leq 2$.) These are plotted in Figure 1(a).

![Graph](image)

(a) $n = 1$ \hspace{2cm} (b) $n = \infty$

**Figure 1.** The “no decision” region, as a function of $\ell$, for testing $H_0$: $\theta = 1$ versus $H_1$: $\theta = 2$ under equal prior probabilities.

From Figure 1(a), it is apparent that $\ell$ can have a dramatic effect. First of all, unless $0 < \ell < 2$, the “no decision” region is the entire space. And for $\ell$ small or near 2, the “no decision” region is most of the space. Note that, at $\ell = \sqrt{5} - 1 \approx 1.24$, the “no decision” region is empty. The resulting test, $T_{1.24,1}$ is thus the test $T_1$ defined in Section 1.1, and could beneficially be used by a frequentist instead of the minimax test (which has critical value $B = 1.24$).
Turning to large $n$, three cases can be distinguished. Technical details are given in the Appendix.

**Case 1.** $\rho > 1$: Then, as $n \to \infty$, $a_{\ell, \rho} = \ell \rho$ and

$$ r_{\ell, \rho} = \exp\{(1 - \gamma^{-1} - \log \gamma)n + (1 - \gamma^{-1})[z_{(1-\rho^{-1})}\sqrt{n} + \frac{1}{3}(z_{\rho}^2 - 1) - 1] + o(1)\}, \quad (2.11) $$

where $z_{\alpha}$ denotes the $\alpha^{th}$ quantile of the standard normal distribution. Note that $(1 - \gamma^{-1} - \log \gamma) < 0$ (recall that $0 < \gamma < 1$), so that $r_{\ell, \rho}$ decreases to zero exponentially fast as $n \to \infty$.

**Case 2.** $\rho < 1$: Then, as $n \to \infty$, $r_{\ell, \rho} = \ell \rho$ and

$$ a_{\ell, \rho} = \exp\{(\gamma - 1 - \log \gamma)n + (\gamma - 1)[z_{\rho}\sqrt{n} + \frac{1}{3}(z_{\rho}^2 - 1)] + o(1)\}. \quad (2.12) $$

Note that $(\gamma - 1 - \log \gamma) > 0$ (for $0 < \gamma < 1$), so that $a_{\ell, \rho}$ increases to $\infty$ exponentially fast as $n \to \infty$.

**Case 3A.** $\rho = 1$, $\ell > -(1 + g(\gamma))$: Here

$$ g(\gamma) \equiv (1 - \gamma)(\log \gamma)/(1 - \gamma + \gamma \log \gamma). \quad (2.13) $$

Then, as $n \to \infty$, $a_{\ell, 1} = \ell$, and

$$ r_{\ell, 1} = \exp[-\ell/(1 + g(\gamma))]^{g(\gamma)} + o(1). \quad (2.14) $$

If $\gamma = 1/2$, then the condition on $\ell$ is $\ell > 1.258$, and $r_{\ell, 1} \approx (1.679)\ell^{(-1.258)}$. Note that, if $\ell = 1.258$, then the "no decision" region is (asymptotically) empty.

**Case 3B.** $\rho = 1$, $\ell < -(1 + g(\gamma))$: Then $r_{\ell, 1} = \ell$ and, as $n \to \infty$,

$$ a_{\ell, 1} = \ell[-\ell/(1 + g(\gamma))]^{-g(\gamma)/(1+g(\gamma))} + o(1). \quad (2.15) $$

If $\gamma = 1/2$, then $a_{\ell, 1} \approx (1.510)\ell^{(-0.795)}$. For the "objective" choice $\ell = 1$, note that this asymptotic "no decision" region of $(1,1.510)$ is quite close to the corresponding region $(1, \sqrt{2})$ when $n = 1$, seemingly indicating a stability of the "no decision" region with respect to sample size (in this case).
Discussion of Example 2: When $\rho \neq 1$, (2.11) and (2.12) show that the “no decision” region grows enormously as $n \to \infty$. Furthermore, it can be shown in Cases 1 and 2, respectively, that, as $n \to \infty$,

$$P_{\theta_1}(\text{“no decision”}) \to 1 - \rho^{-1}, \quad P_{\theta_0}(\text{“no decision”}) \to 1 - \rho.$$

Hence the “no decision” region is even nonnegligible probabilistically.

The story is very different when $\rho = 1$. Then (2.14) and (2.15) show that $r_{\ell,1}$ and $a_{\ell,1}$ stay bounded. Indeed, as shown in Figure 1(b) for the case $\gamma = 1/2$, the “no decision” region for $n = \infty$ is remarkably similar to that for $n = 1$, unless $\ell$ is extreme. And the probabilities of the “no decision” region under $\theta_0$ and $\theta_1$ go to zero exponentially fast as $n \to \infty$.

While generalization from a single example is hazardous, we expect the above pattern to hold for other distributions. Thus, for large $n$ and $\rho \neq 1$, the “no decision” region might be too large to make $T_{\ell,\rho}$ attractive. But for $\rho = 1$ (which, of course, is the natural “balanced” or “noninformative” assumption) we expect the “no decision” region to be small, even for large sample sizes.

From a theoretical perspective, there are various oddities here. Perhaps the most interesting is that the loss ratio, $\ell$, and prior odds ratio, $\rho$, have very different effects, in contrast to usual Bayesian reasoning. Selecting $\rho = 1$ seems to be important, while $\ell = 1$ has no special effect.
3. SYMMETRY AND OPTIMALITY

3.1 Likelihood Ratio Symmetry

A particularly attractive situation arises when the following is satisfied:

LRS Property: The testing problem is said to possess likelihood ratio symmetry if \( f_0(X)/f_1(X) \) has the same distribution under \( f_0 \) as \( f_1(X)/f_0(X) \) has under \( f_1 \).

Example 4. If \( X = (X_1, \ldots, X_n) \) arises from a coordinatewise symmetric location density \( f(x|\theta) = g(|x_1 - \theta|, \ldots, |x_n - \theta|) \), and it is desired to test \( H_0: \theta = \theta_0 \) versus \( H_1: \theta = \theta_1 \), then it is straightforward to show that the LRS Property holds.

One benefit from the LRS Property is indicated in the following lemma.

Lemma 1. If the LRS Property holds, then \( F_0(1) = 1 - F_1(1) \); hence the “no-decision” region for the test \( T_1^* \) in Section 1.2 is empty and \( a = r = 1 \). (Clearly, \( T_1^* \) is then equivalent to \( T_1 \).)

Proof. Clearly

\[
F_0(b) = P_0(B \leq b) = P_1(1/B \leq b) = P_1(1/b \leq B) = 1 - F_1(1/b).
\]

Setting \( b = 1 \) yields the result. \( \square \)

Thus, under the LRS Property, the Bayesian (with \( \rho = 1 \) and \( \ell = 1 \)) and the conditional frequentist using \( T_1 \) always report identical numbers. This was noticed by Kiefer (1977), who seemed quite happy with \( T_1 \) in this case, in part because it also then corresponds with his “equal probability continuum” procedure (see Section 2.1).

3.2 Optimality

Consider the general conditional frequentist testing scenario defined in Section 2.1. Any partition \( \{X_s: s \in S\} \) of \( X \) corresponds to a possible conditional frequentist test, \( T \). A natural question to ask (from a frequentist perspective) is whether an optimal \( T \) (i.e., optimal partition) exists.

Brown (1978) studied this problem, and proposed the following approach for simple versus simple hypothesis testing. Let \( h(\cdot) \) be a nondecreasing, convex function (e.g., \( h(v) = \ldots \)
\( u^2 \), and define the \( B\)-utility of a conditional frequentist test \( T \) to be the pair \((U_T(0), U_T(1))\), where

\[
U_T(i) = E_{f_i}[h(1 - \max\{\alpha_T(s), \beta_T(s)\})],
\]

with \( \alpha_T(s) \) and \( \beta_T(s) \) being the conditional error probabilities defined in (2.1) and (2.2). It is desirable to have the \( U_T(i) \) large, because of the following two points:

(i) Since \( h \) is nondecreasing, reducing both conditional error probabilities (clearly desirable) will cause the \( U_T(i) \) to increase;

(ii) Since \( h \) is convex, making the conditional error probabilities more variable in \( s \) will cause the \( U_T(i) \) to increase; variability in \( \alpha_T(s) \) and \( \beta_T(s) \) is desirable, because it allows for reflection of varying evidentiary strength for different data.

As an illustration of this last point, the following lemma shows that, under the LRS Property, the conditional frequentist test \( T_1 \) is superior to the classical Neyman-Pearson test with equal error probabilities.

**Lemma 2.** Suppose the LRS Property holds, and let \( T_0 \) denote the classical Neyman-Pearson test with rejection region \( \{B \leq 1\} \). Then

\[
U_{T_1}(i) \geq U_{T_0}(i), \quad i = 0, 1,
\]

with strict inequality if \( h \) is strictly convex.

**Proof.** The LRS Property implies that \( \alpha_{T_1}(B) = \beta_{T_1}(B) \), and \( \alpha_{T_0} = \beta_{T_0} \). Thus Jensen's inequality yields

\[
U_{T_1}(i) = E_{f_i}^B[h(1 - \alpha_{T_1}(B))] \\
\geq h(1 - E_{f_i}^B[\alpha_{T_1}(B)]) \\
= h(1 - \alpha_{T_0}) \quad \text{(using (2.3))} \\
= U_{T_0}(i),
\]

with strict inequality if \( h \) is strictly convex. \( \square \)

It is, furthermore, clear that the same reasoning applies to any other symmetric conditional frequentist test, i.e., any test for which each \( \mathcal{X}_s \) is defined as

\[
\mathcal{X}_s = \{x \in \mathcal{X}: B(x) \in A_s \text{ or } 1/B(x) \in A_s\},
\]

16
for some set $A_s$. If $\alpha_T(\cdot)$ for such a test differs from $\alpha_{T_1}(\cdot)$ with positive probability and $h$ is strictly convex, then $U_{T_1}(i) > U_T(i)$.

Being "best" among all symmetric conditional tests is quite compelling, but Brown (1978) also establishes two optimality properties of $T_1$ (under the LRS Property) among all conditional frequentist tests. First, he shows that $T_1$ is the unique test that is $U$-admissible for all $h$; thus, for any other test, $T^*$, one can find a nondecreasing convex $h$ and another test $T^{**}$ such that $U_{T^{**}}(i) \geq U_{T^*}(i)$, with strict inequality for $i = 0$ or 1.

The second global property that Brown establishes is that $T_1$ is the unique test that is totally maximin, i.e., for which

$$\min\{U_{T_1}(0), U_{T_1}(1)\} = \sup_T \min\{U_T(0), U_T(1)\}$$

holds for all nondecreasing convex $h$. (Brown’s uniqueness result applies because, here, $B(X)$ is assumed to have a nonatomic distribution.)

Brown further suggests that, among monotone procedures, $T_1$ is probably strictly optimal (i.e., minimizes both $U_T(0)$ and $U_T(1)$). These optimality results are all particularly compelling because of the great generality allowed in choice of $h$. The bottom line is that, using reasonable frequentist criteria alone, $T_1$ appears to be best, in a variety of ways, among all valid (conditional) frequentist tests, at least when the LRS Property holds. It does not seem to be possible to establish such strong optimality of $T_1$ if the LRS Property does not hold, but $T_1$ undoubtedly remains admissible and reasonable even then.
4. SEQUENTIAL TESTS

4.1 The General Conditional Sequential Test

Suppose $X = (X_1, X_2, \ldots)$ is a sequential sample and that it is desired to test $H_0$: $X \sim f_0$ versus $H_1$: $X \sim f_1$. By this we mean that, for $i = 0, 1$, $f_i = \{f_{i,1}, f_{i,2}, \ldots\}$ with $X^{(n)} = (X_1, \ldots, X_n)$ having density $f_{i,n}(x^{(n)})$ for $n \geq 1$. Define

$$B_n = f_{0,n}(x^{(n)})/f_{1,n}(x^{(n)});$$

let $N$ denote the stopping time of the sequential experiment (a proper stopping rule being assumed); and let $F_i(\cdot)$ be the c.d.f. of $B_N$ under $f_i$, $i = 0, 1$.

The situation is slightly more complicated than that discussed previously in the paper, because $F_i$ is typically not invertible. Usually, however, the following is satisfied:

**Condition S:** The stopping rule is such that $B_N \not\in (R, A)$, where $R \leq 1 \leq A$; the rejection and acceptance regions are $\{B_N \leq R\}$, $\{B_N \geq A\}$, respectively; and, for $i = 0, 1$, $F_i(b)$ is invertible for $b \not\in (R, A)$, with $F_i(R) = F_i(A)$.

**Example 5. The Sequential Probability Ratio Test:** The famous Wald SPRT is defined by

if $B_n \leq R$, stop sampling, reject $H_0$, and report error probability $\alpha = P_0(B_N \leq R)$;

if $B_n \geq A$, stop sampling, accept $H_1$, and report error probability $\beta = P_1(B_N \geq A)$.

It is immediate that $B_N \not\in (R, A)$, and that the rejection and acceptance regions are as stated in Condition S. (Note that stopping and rejecting or accepting when $\{B_n \leq R\}$ or $\{B_n \geq A\}$ is considerably stronger than the Condition S assumption $\{B_N \leq R\}$ and $\{B_N \geq A\}$; the former actually states that one stops and rejects or accepts upon crossing $R$ or $A$, while the latter only states that, if one happens to stop and has crossed $R$ or $A$, then one must reject or accept, respectively). It is virtually always the case that $R \leq 1 \leq A$ for the SPRT.

Under Condition S, the appropriate generalization of $T_1^*$ in Section 1.2 is defined as $T_1^*$ but with $B(x)$ replaced by $B_N$ and $r$ and $a$ defined by

$$r = R \quad \text{and} \quad a = F_0^{-1}(1 - F_1(A)) \quad \text{if } F_0(R) \leq 1 - F_1(A);$$

$$r = F_1^{-1}(1 - F_0(R)) \quad \text{and} \quad a = A \quad \text{if } F_0(R) > 1 - F_1(A).$$
It is straightforward to show that the conditional frequentist interpretation of this test is still valid. Note, however, that this generalized $T_1^*$ does not specify the stopping rule; it merely gives the conclusion to report, upon stopping.

There are two difficulties with $T_1^*$. First, computing $F_0^{-1}$ or $F_1^{-1}$ can be quite challenging in sequential settings. Second, it is somewhat troubling if the “no decision” region $(r, a)$ is larger than the initial $(R, A)$, since the latter is often, in a sense, constructed to be the desired “no decision” region.

These difficulties disappear and, indeed, turn into a delightful advantage if the stopping rule is chosen so that

$$F_0(R) = 1 - F_1(A). \tag{4.2}$$

This is equivalent to the condition that the classical sequential test is constructed to have equal error probabilities, $\alpha = \beta$. Then the conditional frequentist test can be written:

- if $B_N \leq R$, reject $H_0$ and report the conditional error probability $\alpha(B_N) = B_N/(1 + B_N)$;
- if $B_N \geq A$, accept $H_0$ and report the conditional error probability $\beta(B_N) = 1/(1 + B_N)$.

$$\tag{4.3}$$

This is the analogue of the test $T_1$ defined in Section 1.1. Again note, however, that this does not specify when to stop, just what to do upon stopping.

It is of considerable interest that the conditional error probabilities are available explicitly here, while classical (unconditional) error probabilities are typically very hard to compute. Even for the SPRT, computation of $\alpha$ and $\beta$ usually requires difficult analysis of the “overshoot,” the amount by which $B_N$ overshoots $R$ or $A$ (cf., Siegmund, 1985). The conditional error probabilities are not only trivially computable, but, interestingly, incorporate the overshoot into the error statement; the more the overshoot, the less the stated error.

4.2 Symmetric Sequential Tests

In Section 4.1, it was assumed that stopping is governed by a stopping rule separate from $T_1^*$. In this section we explore the extent to which that assumption can be relaxed.
We will consider the situation in which $X_1, X_2, \ldots$ arise as i.i.d. observations from $f_0(x_i)$ or $f_1(x_i)$, with $Y_i = f_0(X_i)/f_1(X_i)$ having the LRS property (i.e., the distribution of $Y_i$ under $f_0$ is the same as that of $1/Y_i$ under $f_1$). We also assume that the stopping rule $\tau = \{\tau_1, \tau_2, \ldots\}$ depends only on $Y_1, Y_2, \ldots$ and is symmetric in the sense that

$$\tau_i(Y_1, \ldots, Y_i) = \tau_i(1/Y_1, \ldots, 1/Y_i)$$

(4.4)

for all $i \geq 1$. (As usual, $\tau_i$ gives the probability, typically zero or one, of stopping sampling upon observing $Y_1, \ldots, Y_i$. More general stopping rules involving other chance mechanisms could be allowed, so long as they are “noninformative” — see Berger and Wolpert, 1988, for definition — but the generality here suffices to make the basic point.)

**Example 6.** A common class of symmetric stopping rules is given by

$$\tau_n(Y_1, \ldots, Y_n) = \begin{cases} 1 \text{ (i.e., stop)} & \text{if } |\sum_{i=1}^n \log Y_i| \geq g(n) \\ 0 \text{ (i.e., continue)} & \text{if } |\sum_{i=1}^n \log Y_i| < g(n), \end{cases}$$

where $g(n)$ is any arbitrary function for which $\tau$ is a proper stopping rule (i.e., is guaranteed to eventually result in “stop”). The SPRT with symmetric boundaries, $A = R^{-1}$, corresponds to $g(n) = \log A$.

For this symmetric situation, it is straightforward to verify that (4.2) holds, so that (4.3) defines $T_1^*$; note that it is not even necessary here to explicitly calculate the quantities in (4.2). That (4.3) defines a valid conditional frequentist test, in this situation, was already recognized by Kiefer (1977). Finally, note that if $\tau$ is restricted to be as above (i.e., depends only on the Bayes factors $B_n = \prod_{i=1}^n Y_i$), then Brown (1978) applies and shows that $T_1^*$ in (4.3) defines the “optimal” conditional frequentist test. To understand the practical ramifications of this situation, consider the following example.

**Example 7.** A sequential experiment is conducted involving i.i.d. $N(\theta, 1)$ data for testing $H_0$: $\theta = 0$ versus $H_1$: $\theta = 1$ under a symmetric stopping rule (or at least a rule for which $\alpha = \beta$). Suppose the report states that sampling stopped after 20 observations, with $\bar{x}_{20} = 0.7$. One can then “replace” whatever sequential test was used by $T_1^*$ in (4.3). Computing

$$B_{20} = \prod_{i=1}^{20} \left[ f(x_i|0)/f(x_i|1) \right] = \exp\{-20(\bar{x}_{20} - \frac{1}{2})\} = 0.018,$$

20
it follows that your conclusion should be to reject $H_0$, with associated conditional error probability $\alpha(B_{20}) = B_{20}/(1 + B_{20}) = 0.018$. This will be a "better" conclusion than that reached in the study (unless they also used $T_1^*$). Note that you do not need to explicitly know the stopping rule used in order to perform the optimal analysis. This is quite attractive in practice because, all too often, the exact stopping rule is incompletely specified and, perhaps, even incompletely known to the experimenters! For instance, if the experimenters had not prespecified the stopping rule, but simply monitored the data stream and stopped when they wished, analysis with $T_1^*$ would still be possible (under the weak assumption of symmetry).

The ramifications for sequential testing of simple hypotheses are profound. First, it appears to be possible to avoid the typically difficult classical computations involving the stopping rule. Second, and more importantly, it seems that pre-experimental analysis can be avoided; one can simply start collecting data and stop whenever one desires, as long as $T_1^*$ is then employed. This last fact strongly refutes the usual frequentist argument against the Stopping Rule Principle, the argument which asserts that allowing one to monitor the data stream and arbitrarily stop allows a "biasing" of the result.

It is interesting to consider why the stopping rule "disappears" here. If we were to actually compute the conditioning partitions, $\mathcal{X}_s$, corresponding to $T_1^*$, we would indeed need to know the stopping rule. But since we will choose the partitions which guarantee $\alpha(B_n)$ and $\beta(B_n)$ as the conditional error probabilities, there is no need to actually compute the $\mathcal{X}_s$.

6. CONCLUSIONS AND GENERALIZATION

For the simple versus simple testing problem, we feel that the procedures proposed in this paper should become standard statistical practice. They are easy to understand, interpret, and use (the sequential versions, for instance, often being much easier than, say, the SPRT); they are correct from Bayesian and likelihood perspectives; and they are valid (and often optimal) frequentist procedures.

A second conclusion from the paper is foundational: frequentist theory, itself, seems to suggest that optimal conditional frequentist procedures will ignore the stopping rule in sequential experimentation. At the very least, the classical argument against the Stopping
Rule Principle is dramatically weakened by the results here.

While the testing of simple hypotheses is often considered as a "practical" approximation in sequential settings, it is admittedly very specialized. From a Bayesian perspective, however, any problem of testing $H_0: X$ has density $f_0(x|\theta_0)$ versus $H_1: X$ has density $f_1(x|\theta_1)$, where $\theta_0$ and/or $\theta_1$ are unknown, can be reduced to simple versus simple testing: the Bayes factor of $H_0$ to $H_1$ is

$$B(x) = m_0(x)/m_1(x)$$

$$\equiv \frac{\int f_0(x|\theta_0)\pi_0(d\theta_0)}{\int f_1(x|\theta_1)\pi_1(d\theta_1)},$$

where $\pi_i$ is the prior distribution of $\theta_i$, $i = 0,1$, so that a Bayesian is implicitly testing $H_0: X$ has density $m_0$ versus $H_1: X$ has density $m_1$. As the latter test is simple versus simple, $T_1$ and/or $T_1^*$ can be applied. The key question, however, is whether or not these tests can be given a satisfactory (conditional) frequentist interpretation in the original problem.

If $H_0$ is simple (i.e., $\theta_0$ is absent or, equivalently, assumes a specific value), then the answer is — yes! The conditional type I error probability is precisely the posterior probability of $H_0$ (assuming equal prior probabilities of the hypotheses), while the conditional type II error probability (which for $T_1$ or $T_1^*$ is, of course, the posterior probability of $H_1$), has an interpretation as a (posterior) expected frequentist type II error. Clarification and discussion of this idea will be presented elsewhere, but the preliminary indication is that Bayesian and frequentist testing may be generally compatible; the severe conflicts that have, in the past, been observed between the two are, perhaps, simply due to use of an inferior frequentist test, namely the unconditional test.

It is of interest that the "better" conditional frequentist tests will (in the composite hypothesis case) depend on the prior distributions assigned to the unknown parameters of the composite hypotheses. Robust Bayesian theory can perhaps provide conditional frequentist tests that are guaranteed to be better than the unconditional tests, but the preliminary indication for composite hypotheses is that some utilization of prior information will be necessary in defining good (conditional) frequentist tests. Note, however, that this use of prior information will probably be no more severe than is the customary use of prior information in selecting power levels for unconditional frequentist testing.
APPENDIX: TECHNICAL DETAILS FOR EXAMPLE 3

Lemma 3. For the incomplete Gamma function $\Gamma(n, t)$, the following hold:

(i) If $c \neq 1$, then as $n \to \infty$

$$\Gamma(n, cn + d + o(1)) = 1_{(1, \infty)}(c) + (ce^{1-c})^n \frac{\exp\{d(c^{-1} - 1)\}}{\sqrt{2\pi n(1-c)}} (1 + o(1)). \quad (A1)$$

(ii) If, as $n \to \infty$,

$$\Gamma(n, t_n) = \xi + o\left(\frac{1}{\sqrt{n}}\right) \quad (A2)$$

for $0 < \xi < 1$, then

$$t_n = n + z_\xi \sqrt{n} + \frac{1}{3}(z_\xi^2 - 1) + o(1). \quad (A3)$$

Proof. First consider $c > 1$ in part (i). A valid expansion of $\Gamma(n, \cdot)$ is

$$\Gamma(n, t) = 1 - \frac{1}{\Gamma(n)}t^{(n-1)}e^{-t[1 + \frac{(n-1)}{t} + \frac{(n-1)(n-2)}{t^2} + \cdots]}.$$ 

Setting $t_n = cn + d + o(1)$, it is straightforward to show that

$$[1 + \frac{(n-1)}{t_n} + \frac{(n-1)(n-2)}{t_n^2} + \cdots] \longrightarrow [1 + \frac{1}{c} + \frac{1}{c^2} + \cdots] = \frac{c}{c-1}.$$ 

Approximating $\Gamma(n)$ by Stirling’s formula then yields

$$\Gamma(n, t_n) = 1 - \frac{(cn + d + o(1))^{n-1}e^{-(cn+d+o(1))c}}{e^{-n}n^{(n-1)/2}\sqrt{2\pi(c-1)}} (1 + o(1)).$$

Clearly

$$\frac{(cn + d + o(1))^{n-1}}{n^{n-1}} = c^{n-1}(1 + \frac{d + o(1)}{cn})^{n-1} = c^{n-1}e^{d/c}(1 + o(1)),$$

from which the result follows. The proof of part (i) for $c < 1$ is similar, but now uses the expansion

$$\Gamma(n, t) = e^{-t} \sum_{k=n}^{\infty} \frac{t^k}{k!}.$$ 

Details are omitted.
To prove part (ii), we use the fact that \( \Gamma(n,t) = F(2t|2n) \), where \( F(\cdot|\nu) \) is the c.d.f. of the chi-squared distribution with \( \nu \) degrees of freedom. Thus (A2) can be rewritten

\[
F(2t_n|2n) = \xi + o\left(\frac{1}{\sqrt{n}}\right) \equiv \xi_n.
\]

But, as \( n \to \infty \), the \( \xi_n^{th} \) quantile of the chi-squared \((2n)\) distribution is

\[
\chi^2_{\xi_n} = 2n + z_{\xi_n} \sqrt{4n} + \frac{2}{3}(z^2_{\xi_n} - 1) + o(1)
\]

\[
= 2n + z\sqrt{4n} + \frac{2}{3}(z^2 - 1) + o(1).
\]

This directly yields (A3). \( \square \)

From (2.10), observe that

\[
F_0(\ell_\rho) = 1 - \Gamma(n, c_0 n + d_0), \quad F_1(\ell_\rho) = 1 - \Gamma(n, c_1 n + d_1), \quad (A4)
\]

where

\[
c_0 = -\log \gamma \quad (1 - \gamma), \quad d_0 = -\log(\ell_\rho) \quad (1 - \gamma), \quad c_1 = \gamma c_0, \quad d_1 = \gamma d_0. \quad (A5)
\]

Since \( 0 < \gamma < 1 \), it can be shown that \( c_0 > 1 \) and \( 0 < c_1 < 1 \). It is then immediate, from (A1) and the fact that \( c \exp\{1 - c\} < 1 \) for \( c \neq 1 \), that

\[
F_0(\ell_\rho) = o\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad F_1(\ell_\rho) = 1 - o\left(\frac{1}{\sqrt{n}}\right). \quad (A6)
\]

Verification of (2.11) and (2.12):

Consider, first, the case \( \rho > 1 \). It is clear from (A6) that \( F_0(\ell_\rho) > 1 - \rho F_1(\ell_\rho) \) for large \( n \). Hence (see (2.5)), \( a_{\ell_\rho} = \ell_\rho \) and

\[
F_1(r_{\ell_\rho}) = \frac{1}{\rho} \left[1 - F_0(\ell_\rho)\right] = \frac{1}{\rho} + o\left(\frac{1}{\sqrt{n}}\right).
\]

This can be rewritten as

\[
\Gamma(n, c_1 n - \frac{\gamma \log(r_{\ell_\rho})}{(1 - \gamma)}) = 1 - \frac{1}{\rho} + o\left(\frac{1}{\sqrt{n}}\right).
\]

Hence, (A3) yields

\[
c_1 n - \frac{\gamma \log(r_{\ell_\rho})}{(1 - \gamma)} = n + z_{(1-1/\rho)\sqrt{n}} + \frac{1}{3}(z^2_{(1-1/\rho)} - 1) + o(1).
\]
Solving for \( r_{\ell,\rho} \) yields (2.11). The derivation of (2.12) for \( \rho < 1 \) is similar.

\[ \square \]

**Verification of (2.14) and (2.15):**

Since \( \rho = 1 \), the inequality defining applicability of (2.4) or (2.5) is \( F_0(\ell) \geq 1 - F_1(\ell) \). Using (A1), (A4), and (A5), this condition will be satisfied as \( n \to \infty \) only if \( \ell > -(1+g(\gamma)) \) (see (2.13)). From (2.5), we then know that \( a_{\ell,1} = \ell \) and \( F_1(r_{\ell,1}) = 1 - F_0(\ell) \), which can be rewritten, using (2.10) and (A4), as

\[
\Gamma(n, c_1 n - \frac{\gamma \log(r_{\ell,1})}{1 - \gamma}) = \Gamma(n, c_0 n + d_0) . \tag{A7}
\]

The solution to this equation is given in (2.14). To show this, note that then

\[
\log(r_{\ell,1}) = \log(\ell[-\ell/(1 + g(\gamma))]^{g(\gamma)}) + o(1),
\]

so that (A1) can be applied to both sides of (A7); algebra then verifies their equality.

If \( \ell < -(1 + g(\gamma)) \), (2.4) applies. Thus \( r_{\ell,1} = \ell \), and an argument similar to that above verifies (2.15).

**REFERENCES**


