The Annals of Statistics
1979, Vol. 7, No. 3, 579–591

APPROXIMATIONS TO BAYESIAN SEQUENTIAL TESTS OF COMPOSITE HYPOTHESES

BY ROBERT FORTUS
Washington State University

This paper deals with approximations to Bayesian sequential tests of composite hypotheses. If the distributions of the data form an exponential or truncation family, then such tests may be described by a continuation region in the space of n, the sample size, and M_n, the sufficient statistics, which are of fixed dimension. In this case Schwarz has been able to describe the asymptotic shape of the continuation region as the sampling cost c approaches zero. We have generalized Schwarz's work by considering more general families of distributions. In this paper the role of M_n is played by the log likelihood function, and we show that the optimal Bayesian stopping rule may be approximated by a stopping rule which depends only on n, c, and two likelihood ratio test statistics.

1. Introduction.

1.1. *Mathematical formulation of the problem.* Let $X_1, X_2, \cdots$ be observations taken one at a time. They are i.i.d., with probability distribution P_θ, where θ is an unknown element of the parameter space Ω, a separable and locally compact metric space. The observations have probability density $f_\theta = dP_\theta/d\lambda$, and λ is assumed to be a σ-finite measure. Both the X_i's and θ could be vectors. We are testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$, where Ω_0 and Ω_1 are disjoint subsets of Ω. The loss function is $L(\theta)$, which is the penalty for a wrong decision when θ is the state of nature. $L(\theta)$ is assumed to be positive on Ω_0 and Ω_1, continuous, bounded above by one, and equal to zero on $(\Omega_0 \cup \Omega_1)'$, where $'$ denotes complement. The existence of an indifference region, a region where the loss due to a wrong decision is zero, is required for our later results. θ has a prior distribution w, and it is assumed that the support of w is all of Ω. The cost of each observation is a constant c. Finally, assume that at least one observation will be taken, which is the case for c small enough. The definitions made in the following discussion will hold throughout the paper.

A discussion concerning the exact solution to the problem can be found in DeGroot (1970, page 300). Let $R(w)$ denote the stopping risk;

$$R(w) = \min\left[\int_{\Omega_0} L(\theta)\,dw(\theta), \int_{\Omega_1} L(\theta)\,dw(\theta)\right]. \tag{1.1}$$

We want to minimize the total risk, the stopping risk plus the cost of sampling, say

$$Y_n = R(w_n) + cn, \tag{1.2}$$

where w_n is the posterior distribution of θ given $X_1, \cdots, X_n$. Let t denote any

---

Received January 1977; revised February 1978.
AMS 1970 *subject classifications.* 62L15, 62F05, 62C10.
Key words and phrases. Optimal Bayes continuation region, optimal stopping rule, stopping risk, convergence of sets.

stopping time. Given θ has prior w, the Bayes risk $\rho_c(w)$ from the optimal sequential procedure is given by

$$\rho_c(w) = \inf_{t \geqslant 1} E_w(Y_t), \tag{1.3}$$

where the inf extends over all stopping times bounded below by one. The optimal stopping rule is

$$t^* = \inf\{n \geqslant 1 : R(w_n) \leqslant \rho_c(w_n)\}. \tag{1.4}$$

1.2. *Review of Schwarz's work.* In most problems finding the exact solution t^* is too difficult, so a logical approach is to try to approximate it. The question to which this paper is addressed is: what stopping rule closely approximates t^* as $c \to 0$?

In this subsection we review the work of Schwarz (1962, 1968), who has considered this problem in the cases that the f_θ form an exponential family, or an exponential truncation family.

For an exponential family, $S_n = X_1 + \cdots + X_n$ is a sufficient statistic of fixed dimensionality, and $R(w_n)$, the stopping risk at time n, can be expressed as $R_w(n, S_n)$, a function of n and S_n. Let w_n denote the distribution of θ given S_n. From (1.4), it follows that the Bayes continuation region $B(c)$ can be defined in the (n, S_n) plane by

$$B(c) = \{(n, S_n) : R_w(n, S_n) > \rho_c(w_n)\}. \tag{1.5}$$

In terms of the function $R_w(n, S_n)$ and $c > 0$, Schwarz defines another family of regions in the (n, S_n) plane by

$$C(c) = \{(n, S_n) : R_w(n, S_n) \geqslant c\}. \tag{1.6}$$

His main result is an explicit expression for the asymptotic representation of $B(c)$ as $c \to 0$. He reaches the result by first getting the asymptotic representation of $C(c)$, and then proving that $C(c)$ and $B(c)$ have the same asymptotic shape as $c \to 0$. What he does is demonstrate pointwise convergence of boundaries. If the support of the prior w is the entire parameter space, as is the case in our formulation of the problem, the asymptotic shape of $B(c)$ does not depend on w.

1.3. *Approach and findings.* Schwarz obtains his asymptotic result for exponential families by fixing a ray in the (n, S_n) plane, and then letting n go to infinity. Our approach is essentially the same; however, in our work the role of S_n is played by the log of the likelihood function, and $B(c)$ and $C(c)$ are defined in an appropriate function space. Our results apply to an extremely wide class of families of distributions, which includes almost any family of mutually absolutely continuous distributions. In this paper we do not consider families of distributions with truncation parameters.

Our main results are given in two theorems. We show that the boundary of $C(c)/\log c^{-1}$ converges to a limiting region as $c \to 0$, and also prove that the boundary of $B(c)/\log c^{-1}$ converges to the same limit. That limit, which does not depend on the prior distribution of θ, can be used as an approximation to the boundary of $B(c)/\log c^{-1}$ for c small. Since we show uniform convergence of boundaries, our results are stronger than Schwarz's pointwise convergence results. Of course, our uniform convergence results can be applied in the exponential family case treated by Schwarz.

An approximation to the optimal stopping rule t^* follows from our convergence results. The approximation holds for any prior on θ and is a function of two likelihood ratio test statistics. For c small, our approximation to the optimal stopping rule is to stop at the first n such that the minimum of the likelihood ratio test statistics for testing Ω_0 vs. Ω_0', and Ω_1 vs. Ω_1', is less than c.

The paper is organized as follows. Some required, preliminary definitions, assumptions and notations are given in Section 2, and the main results appear in Section 3.

2. Preliminary considerations. In this section we present several definitions, assumptions and notations which are needed throughout the paper. These include a definition of the stopping risk as a function of n and a sufficient statistic, and definitions of regions $B(c)$ and $C(c)$ which are analogous to Schwarz's $B(c)$ and $C(c)$ in (1.5) and (1.6).

For $n \geqslant 1$ and fixed $x_1, \cdots, x_n$, we call the function $L_n(\cdot|x_1, \cdots, x_n)$ defined by

$$L_n(\theta|x_1, \cdots, x_n) = \Pi_{i=1}^n f_\theta(x_i),\ \theta \in \Omega,$$

the likelihood function with respect to λ^n, and we call

$$T_n = (\log L_n)/n$$

the log likelihood function with respect to λ^n. Throughout the paper we assume that f_θ is continuous in θ for a.e. x. Let $D(\Omega)$ be the set of all nonnegative, continuous functions f on Ω for which

$$f \not\equiv 0,\ \lim_{\theta \to \infty} f(\theta) = 0,$$

and

$$0 < \min_{i=0,1}\left[\sup_{\Omega_i} f(\theta)\right] < \sup_\Omega f(\theta),$$

and let $E(\Omega)$ be the set of all functions f on Ω for which $e^f \in D(\Omega)$. A standing assumption of the paper is that for all possible realizations $(x_1, \cdots, x_n)$ of $(X_1, \cdots, X_n)$, $L_n(\cdot|x_1, \cdots, x_n) \in D$. Also, a simple application of the factorization theorem gives that the likelihood function is a sufficient statistic, so the log likelihood function is sufficient as well.

Fix $T_n(\cdot|x_1, \cdots, x_n) = f$, and note that, by assumption, f must be in E. The stopping risk can be expressed as a function only of n and f, and is given by

$$R_w(n, f) = \frac{\min_{i=0,1} \int_{\Omega_i} \exp(nf) L(\theta) dw(\theta)}{\int_\Omega \exp(nf) dw(\theta)}, \qquad n \geqslant 1. \tag{2.1}$$

We can compute the stopping risk if we know n and T_n, the sufficient statistic. For

$f \in E$, some important notations are:

$$h_i(f) = \sup_{\Omega_i} e^f, \qquad i = 0, 1,$$

$$h(f) = \sup_{\Omega} e^f,$$

$$\gamma_1(f) = \frac{\min_{i=0,1} h_i(f)}{h(f)},$$

$$\gamma_2(f) = \log(1/\gamma_1(f)),$$

and

$$\gamma_3(f) = 1/\gamma_2(f).$$

Observe that γ_1^n is just the minimum of the likelihood ratio test statistics for testing Ω_0 vs. Ω_0', and Ω_1 vs. Ω_1'. Also, the definitions of D and E imply

$$0 < \min_{i=0,1} h_i(f) < h(f), \tag{2.2}$$

so $\gamma_1(f) < 1$. It is clear from (2.1) and (2.2) that $R_w(n, f) > 0$ for $n \geqslant 1$ and $f \in E$. Of course we are assuming that $w(\Omega_i) > 0$, $i = 0, 1$.

For $c > 0$ and $n \geqslant 1$, the region $C(c)$ is defined in (n, $n \cdot$ log likelihood function) space, and in terms of $R_w(n, f)$, by

$$C(c) = \{(n, nf) : R_w(n, f) \geqslant c\}. \tag{2.3}$$

Noting the definition of the optimal stopping rule t^* in (1.4), we can define the Bayes continuation region $B(c)$ in (n, $n \cdot$ log likelihood function) space by

$$B(c) = \{(n, nf) : R_w(n, f) > \rho_c(w_n^f)\}, \tag{2.4}$$

where w_n^f denotes the distribution of θ given n and the log likelihood function.

Define a set E^* by

$$E^* = \{(n, nf) : n > 0, f \in E\}, \tag{2.5}$$

and let $A(c)$ be any region in E^* which depends on c. A notation which is required for Section 3 is $A(c)/\log c^{-1}$. This notation simply refers to the region in E^* which is determined by the points whose coordinates are the coordinates of the points of $A(c)$, divided by $\log c^{-1}$.

The final, standing assumption of the paper is that H_0 and H_1 are d-testable hypotheses for some $d \geqslant 1$. Schwarz defines H_0 and H_1 to be d-testable if there exists a fixed sample size test $\Psi_d = \Psi(X_1, \cdots, X_d)$ of H_0 vs. H_1 whose probability of error is bounded by a number less than $\frac{1}{2}$ for any θ in $\Omega_0 \cup \Omega_1$ (Schwarz, 1968). The d-testability assumption is shown there to imply the existence of a $b > 0$ for which

$$C(c) \supset B(c) \supset C(bc \log c^{-1}) \tag{2.6}$$

for all priors w and c sufficiently small. The relation (2.6) will be used in getting the convergence of the boundary of $B(c)/\log c^{-1}$. It should be mentioned that the assumption of d-testability is not very restrictive. Many commonly considered hypothesis testing problems involve d-testable hypotheses. For verification of (2.6), other results concerning d-testability, and examples of d-testable hypotheses, see Fortus (1975).

3. Convergence of $B(c)$.

3.1. *Properties of the stopping risk.* The main result of this subsection is that $R_w(n, f)^{1/n} \to \gamma_1(f)$ as $n \to \infty$ uniformly with respect to f, when f is suitably restricted to a compact set. We also show that for all such f and n sufficiently large, the stopping risk is a decreasing function of n. These results will be applied in subsection 3.2 to get the convergence of $\partial B(c)/\log c^{-1}$ as $c \to 0$.

Fix $T_n(\cdot|x_1, \cdots, x_n) = f$, and consider the integrals in (2.1). $\int_{\Omega_0} \exp(nf)L(\theta)dw(\theta)$ is the nth power of the L_n norm of e^f (restricted to Ω_0) relative to the measure Ldw. Applying a known property of L_n norms, we get that

$$\left(\int_{\Omega_i} \exp(nf)L(\theta)dw(\theta)\right)^{1/n} \to \text{ess sup}_{\Omega_i} e^f, \qquad i = 0, 1,$$

as $n \to \infty$ (Loeve, 1955, page 160). The ess sup here is with respect to the measure Ldw. Note that n is treated as continuous throughout Section 3.

By the continuity of $e^{f(\theta)}$, the above convergence result can be rewritten as

$$\left(\int_{\Omega_i} \exp(nf)L(\theta)dw(\theta)\right)^{1/n} \to h_i(f), \qquad i = 0, 1, \tag{3.1}$$

as $n \to \infty$. Similarly,

$$\left(\int_{\Omega} \exp(nf)dw(\theta)\right)^{1/n} \to h(f) \tag{3.2}$$

as $n \to \infty$. Then it is clear that

$$R_w(n, f)^{1/n} \to \gamma_1(f) < 1 \tag{3.3}$$

as $n \to \infty$.

Before showing that the convergence in (3.3) is uniform, we will define some more necessary notations. Recall that Ω is assumed to be a separable and locally compact metric space, and let σ denote the metric on Ω. Let $\mathcal{F}$ denote the class of Borel subsets of Ω, and assume that the prior distribution w is defined on $\mathcal{F}$. Let Ω^* be the Alexandroff one point compactification of Ω (Royden, 1968, page 168). Then the set E defined in Section 2 can be redefined as the set of all functions f on Ω^* for which e^f is continuous, $e^f \not\equiv 0$, $e^{f(\infty)} = 0$, and (2.2) holds. Note that by the compactness of Ω^*, $e^{f(\theta)}$ achieves $h(f)$ at some $\hat{\theta} = \hat{\theta}(f)$ in Ω.

Now define a metric d_E on E as follows. For f and g in E,

$$d_E(f, g) = \sup_\theta |e^{f(\theta)} - e^{g(\theta)}|. \tag{3.4}$$

Throughout the rest of Section 3, K will denote any compact subset of E; K will consist of functions f on Ω^* such that the functions e^f form an equicontinuous family. In fact, by the compactness of Ω^*, the family of functions e^f, $f \in K$, is uniformly equicontinuous (Dunford and Schwartz, 1964, page 267). That is given $\varepsilon > 0$, there exists a $\delta = \delta(K, \varepsilon)$ such that $|e^{f(\theta_1)} - e^{f(\theta_2)}| < \varepsilon$ whenever $\sigma(\theta_1, \theta_2) < \delta$ for all f in K and θ_1, θ_2 in Ω^*.

For $\alpha \geq 1$, define B_α to be the set of all $f \in E$ for which $h_i(f) \geq 1/\alpha$, $i = 0, 1$, and also $h(f)/h_0(f) \geq 1 + 1/\alpha$ or $h(f)/h_1(f) \geq 1 + 1/\alpha$. We will show that the convergence in (3.3) is uniform with respect to $f \in KB_\alpha$, where KB_α denotes $K \cap B_\alpha$. The sets B_α are defined in order to allow for possible nonuniqueness of maximum likelihood estimates of θ. If we only considered problems in which the likelihood function is strictly positive and has a unique maximum, then we would not need the sets B_α. Indeed, let D_1 denote the set of all $g \in D$ such that $g(\theta) > 0$ for all θ, and let E_1 denote the set of all $f \in E$ for which $e^f \in D_1$. Finally, let E_0 denote the set of all $f \in E$ which assume their maximum only once. If K is compact and contained in $E_0 \cap E_1$, then it is easily shown that $K \subset B_\alpha$ for some α, so $K = KB_\alpha$.

LEMMA 3.1. *Let K be any compact subset of E, and let $\alpha \geq 1$. Then*

$$R_w(n, f)^{1/n} \to \gamma_1(f)$$

as $n \to \infty$ uniformly with respect to $f \in KB_\alpha$.

PROOF. A simple application of Hölder's inequality and Fatou's lemma gives that $[\int_{\Omega_i} e^{nf} L(\theta) dw(\theta)]^{1/n}$, $i = 0, 1$, and $[\int_\Omega e^{nf} dw(\theta)]^{1/n}$ are nondecreasing in n, and lower semicontinuous with respect to f. Thus, by Dini's theorem, the convergence in (3.1) and (3.2) is uniform with respect to $f \in KB_\alpha$, so the convergence in (3.3) is uniform as well (Royden, 1968, page 162).

Lemma 3.2, which follows, is needed only for the proof of Lemma 3.3.

LEMMA 3.2. *Let K be any compact subset of E, let $\alpha \geq 1$, let $0 < \varepsilon < 1/\alpha$, and let $K_\varepsilon^f = \{\theta : e^{f(\theta)} > h(f) - \varepsilon\}$. Then*

$$\inf_{f \in KB_\alpha} w(K_\varepsilon^f) > 0.$$

PROOF. Let $G_\varepsilon(f) = w(K_\varepsilon^f)$, and consider the following three statements.

(a) $G_\varepsilon(f) > 0$ for all $f \in KB_\alpha$;

(b) $\liminf_{f_n \to f} G_\varepsilon(f_n) \geq G_{\varepsilon/2}(f)$, where $f \in KB_\alpha$ and $f_n \in KB_\alpha$, $n \geq 1$;

(c) $\inf_{f \in KB_\alpha} G_\varepsilon(f) > 0$.

The truth of (a) follows directly from the continuity of e^f, and the assumption that the support of w is all of Ω.

For f_n, $n \geq 1$, in KB_α, let ν_n be the distribution of e^{f_n} with respect to the probability space $(\Omega, \mathcal{F}, w)$. Also, for $f \in KB_\alpha$, let ν be the distribution of e^f with respect to $(\Omega, \mathcal{F}, w)$. If $f_n \to f$, then $\nu_n \Rightarrow \nu$ and $h(f_n) \to h(f)$, where $\Rightarrow$ denotes weak convergence. Thus, given ε, there exists an n' such that $|h(f_n) - h(f)| < \varepsilon/2$ for all $n \geq n'$, and it can be shown easily, by the extended Helly-Bray lemma, that

$$\liminf_{f_n \to f} w\{\theta : e^{f_n(\theta)} > h(f_n) - \varepsilon\} \geq w\{\theta : e^{f(\theta)} > h(f) - \varepsilon/2\}$$

(Billingsley, 1968, page 12). Therefore, (b) holds.

Now we will show that (a) and (b) imply (c), and the proof will be done. There exists a sequence f_n, $n \geq 1$, in KB_α such that

$$G_\varepsilon(f_n) \to \inf_{f \in KB_\alpha} G_\varepsilon(f) = I$$

as $n \to \infty$. By the compactness of K, there exists a subsequence $f_{n_k} \to f$, and

$$I = \liminf_{k\to\infty} G_\varepsilon(f_{n_k}) \geqslant G_{\varepsilon/2}(f) > 0.$$

LEMMA 3.3. *For any $f \in E$, $R_w(n, f) \to 0$ as $n \to \infty$. Moreover, if K is any compact subset of E and $\alpha \geqslant 1$, then $R_w(n, f) \to 0$ as $n \to \infty$ uniformly with respect to $f \in KB_\alpha$, and there exists an $n_0 = n_0(K, \alpha)$ such that $R_w(n, f)$ is a decreasing function of n for all $f \in KB_\alpha$ and $n \geqslant n_0$.*

PROOF. The first assertion follows directly from (3.3). The assertion that $R_w(n, f) \to 0$ as $n \to \infty$ uniformly with respect to $f \in KB_\alpha$ follows from Lemma 3.1, and the fact that $\gamma_1(f) \leqslant 1/(1 + 1/\alpha)$ for $f \in KB_\alpha$.

The final assertion of the lemma remains to be proved. Pick any $f \in KB_\alpha$. By the definition of B_α, without loss of generality, we can assume that $h(f)/h_0(f) \geqslant 1 + 1/\alpha$. Let $h^*(f) = (h_0(f) + h(f))/2$, and let $g_f(\theta) = e^{f(\theta)}/h^*(f)$; then the stopping risk can be expressed by

$$R_w(n, f) = \frac{\min_{i=0,1} \int_{\Omega_i} g_f^n(\theta) L(\theta)\, dw(\theta)}{\int_\Omega g_f^n(\theta)\, dw(\theta)}. \tag{3.5}$$

Now consider the denominator in (3.5). Observe that

$$\frac{d}{dn}\left[\int_\Omega g_f^n dw\right] = \int_\Omega g_f^n \log g_f dw = D.$$

We will show the existence of an $n_1 = n_1(K, \alpha)$ such that $D > 0$ (that is, the denominator in (3.5) is an increasing function of n) for all $n \geqslant n_1$.

$$\begin{aligned} D &= \int_{0<g_f<1} g_f^n \log g_f dw + \int_{g_f>1} g_f^n \log g_f dw \\ &\geqslant \qquad -e^{-1} \qquad\quad + \qquad\quad I \end{aligned} \tag{3.6}$$

Let $0 < \varepsilon < 1/\alpha$. By Lemma 3.2 and the definition of g_f, there exists an $l = l(K, \alpha, \varepsilon)$ such that

$$w\{\theta : g_f(\theta) > \sup_\Omega g_f(\theta) - \varepsilon/h^*(f)\} \geqslant l. \tag{3.7}$$

Since $f \in KB_\alpha$, $h_0(f) \geqslant 1/\alpha$, so $h^*(f) \geqslant 1/\alpha$. Also, it can be shown easily that

$$\sup_\Omega g_f(\theta) \geqslant \frac{2 + 2/\alpha}{2 + 1/\alpha} = b(\alpha).$$

Thus, from (3.7), it follows that

$$w\{\theta : g_f(\theta) > b(\alpha) - \alpha\varepsilon\} \geqslant l. \tag{3.8}$$

Set $\varepsilon = (b(\alpha) - 1)/2\alpha$ in (3.8), and consider I in (3.6). It is clear from (3.8) that

$$I \geqslant (b(\alpha)/2 + 1/2)^n \log(b(\alpha)/2 + 1/2) l,$$

where $b(\alpha)/2 + 1/2 > 1$. By the above inequality, there exists an $n_1 = n_1(K, \alpha)$ such that $I > e^{-1}$ (and thus $D > 0$) for all $n \geqslant n_1$.

Observe that if $h_1(f) < h^*(f)$, then $\sup_{\Omega_i} g_f(\theta) \leqslant 1$, $i = 0, 1$, and the numerator in (3.5) is a nonincreasing function of n. Now consider the case that $h_1(f) \geqslant h^*(f)$.

To finish this proof, it is enough to show the existence of an $n_2 = n_2(K, \alpha)$ such that the numerator in (3.5) is a nonincreasing function of n for $n \geqslant n_2$. If $h_1(f) > h^*(f)$, it follows easily that

$$h_1(f) \geqslant h_0(f) + 1/2\alpha^2. \tag{3.9}$$

Let $\varepsilon = 1/4\alpha^2$. Since the convergence in (3.1) is uniform with respect to $f \in KB_\alpha$, there exists an $n_2 = n_2(K, \alpha)$ such that

$$\left\{\int_{\Omega_1} e^{nf(\theta)}L(\theta)dw(\theta)\right\}^{1/n} \geqslant h_1(f) - \varepsilon \tag{3.10}$$

for $n \geqslant n_2$. Clearly,

$$\left\{\int_{\Omega_0} e^{nf(\theta)}L(\theta)dw(\theta)\right\}^{1/n} \leqslant h_0(f) \tag{3.11}$$

for all n. The inequalities (3.9), (3.10) and (3.11) imply that

$$\min_{i=0,\,1}\int_{\Omega_i} e^{nf(\theta)}L(\theta)dw(\theta) = \int_{\Omega_0} e^{nf(\theta)}L(\theta)dw(\theta)$$

for $n \geqslant n_2$. Thus, for $n \geqslant n_2$, the numerator in (3.5) is $\int_{\Omega_0} g_f^n(\theta)L(\theta)dw(\theta)$, which is a nonincreasing function of n.

The n_0 in the statement of the lemma is the maximum of $n_1(K, \alpha)$ and $n_2(K, \alpha)$.

LEMMA 3.4. *Let d_E be the metric on E which is defined in* (3.4), *and let τ be the metric on* $[0, \infty)$ *which is defined by* $\tau(x, y) = |x - y|$, $x \geqslant 0 \leqslant y$. *Finally, let d be the metric on* $[0, \infty) \times E$ *which is defined by*

$$d[(x, f), (y, g)]^2 = \left[d_E(f, g)^2 + \tau(x, y)^2\right] \tag{3.12}$$

for f and g in E and $x \geqslant 0 \leqslant y$. Then $R_w(n, f)$ is jointly continuous in (n, f).

PROOF. It is sufficient to show that if $(n_k, f_k) \to_d (n, f)$ as $k \to \infty$, then $R_w(n_k, f_k) \to R_w(n, f)$. The result follows from (3.4), (3.12), and the dominated convergence theorem.

3.2. *Convergence.* In this subsection we define what is meant by convergence of sets, and then show that $\partial C(c)/\log c^{-1}$ and $\partial B(c)/\log c^{-1}$ converge to the same limiting set as $c \to 0$. The convergence result leads to a reasonable approximation to the optimal stopping rule t^*.

Recall that $R_w(n, f) > 0$ for $n \geqslant 1$ and $f \in E$. By Lemmas 3.3 and 3.4, $R_w(n, f) \to 0$ as $n \to \infty$, and $R_w(n, f)$ is continuous in (n, f). Thus, for fixed f and c small enough, there exists an $n \geqslant 1$ such that $R_w(n, f) = c$. For $f \in E$ and sufficiently small $c > 0$, define $n^* = n^*(c, f)$ by

$$n^* = \inf\{n \geqslant 1 : R_w(n, f) = c\}. \tag{3.13}$$

LEMMA 3.5. *Let K be any compact subset of E, and let $\alpha \geqslant 1$. For $c > 0$ and $f \in KB_\alpha$, let $n^* = n^*(c, f)$ defined in* (3.13). *Then $n^* \to \infty$ as $c \to 0$ uniformly with respect to $f \in KB_\alpha$.*

PROOF. Pick any $f \in KB_\alpha$, and suppose that $n^*(c, f) \leqslant n_1$. Then

$$\min_{f \in KB_\alpha,\ 1 \leqslant n \leqslant n_1} R_w(n, f) \leqslant c. \tag{3.14}$$

By the continuity of $R_w(n, f)$ and the compactness of $KB_\alpha \times [1, n_1]$, $R_w(n, f)$ assumes its minimum over $KB_\alpha \times [1, n_1]$. That minimum clearly is bigger than 0, because $R_w(n, f) > 0$ for all $f \in KB_\alpha$ and $n \geqslant 1$. Therefore, (3.14) is not true for c sufficiently small. There exists a $c_0 > 0$ such that our original supposition is false for $c \leqslant c_0$.

Let d be the metric on $[0, \infty) \times E$ which is defined in (3.12), and let E^* be as in (2.5). It is clear that $E^* \subset [0, \infty) \times E$, so (E^*, d^*) is a metric space, where d^* is the metric d restricted to E^*. Now define a metric ρ on all compact subsets of E^* as follows. If A and B are compact subsets of E^*,

$$\rho(A, B) = \sup \inf_{a \in A;\ b \in B} d^*(a, b) + \sup \inf_{b \in B;\ a \in A} d^*(a, b). \tag{3.15}$$

For compact $K \subset E$, $\alpha \geqslant 1$, and sufficiently small $c > 0$, let

$$C^{\#}_{K,\alpha}(c) = \{(n^*, n^* \cdot f) : f \in KB_\alpha\}, \tag{3.16}$$

where n^* is the $n^*(c, f)$ in (3.13). Also, let

$$H_{K,\alpha} = \{(n, nf) : f \in KB_\alpha, n > 0\}. \tag{3.17}$$

LEMMA 3.6. *For any compact $K \subset E$, $\alpha \geqslant 1$, and c sufficiently small,*

$$\partial C(c) \cap H_{K,\alpha} = C^{\#}_{K,\alpha}(c).$$

PROOF. Let n_0 be the $n_0(K, \alpha)$ in Lemma 3.3, so $R_w(n, f)$ is a decreasing function of n for all $f \in KB_\alpha$ and $n \geqslant n_0$. By Lemma 3.5 there exists a $c_0 > 0$ such that if $c \leqslant c_0$, then $n^*(c, f) > n_0$ for all $f \in KB_\alpha$. Assume that $c \leqslant c_0$, and recall (2.3).

Because $n^* > n_0$ for all $f \in KB_\alpha$, every neighborhood about a point $(n^*, n^* \cdot f)$ in $C^{\#}_{K,\alpha}(c)$ contains a point (n, nf) for which $R_w(n, f) < c$; that is, every neighborhood about a point in $C^{\#}_{K,\alpha}(c)$ contains a point which is not in $C(c)$. It is clear that every neighborhood about a point in $C^{\#}_{K,\alpha}(c)$ contains a point in $C(c)$. Thus, $C^{\#}_{K,\alpha}(c) \subset \partial C(c) \cap H_{K,\alpha}$. To show that $\partial C(c) \cap H_{K,\alpha} \subset C^{\#}_{K,\alpha}(c)$ for $c \leqslant c_0$, appeal to the continuity of $R_w(n, f)$ in (n, f) and the definition of $C^{\#}_{K,\alpha}(c)$.

LEMMA 3.7. *For any compact $K \subset E$, $\alpha \geqslant 1$, and sufficiently small $c > 0$, $C^{\#}_{K,\alpha}(c)$ is a compact subset of E^*.*

PROOF. Pick any $c > 0$, and suppose it is small enough that the conclusion of the previous lemma holds. Let V_n, $n \geqslant 1$, be any sequence in $C^{\#}_{K,\alpha}(c)$. Then

$$V_n = (n^*(c, f_n), n^*(c, f_n) \cdot f_n),$$

for f_n, $n \geqslant 1$, in KB_α. We need to show that V_n has a subsequence which converges to a point in $C^{\#}_{K,\alpha}(c)$.

By the compactness of KB_α, f_n has a subsequence f_{n_k} which converges to some $f \in KB_\alpha$. By Lemma 3.3, $R_w(n, f) \to 0$ as $n \to \infty$ uniformly with respect to $f \in$

KB_α, so it is clear from the definition of n^* that

$$\sup_{f \in KB_\alpha} n^*(c, f) < \infty$$

for fixed $c > 0$. Therefore, $n^*(c, f_{n_k})$ is a bounded sequence of positive numbers, and has a convergent subsequence $n^*(c, f_{n_{k_j}})$. Now we can conclude that

$$V_{n_{k_j}} = \left(n^*(c, f_{n_{k_j}}), n^*(c, f_{n_{k_j}}) \cdot f_{n_{k_j}}\right) \to_{d^*} (t, tf) = V,$$

where $f \in KB_\alpha$ and $t \geq 0$. By Lemma 3.6, $C^{\#}_{K,\alpha}(c) = \partial C(c) \cap H_{K,\alpha}$, which is a closed set, so $V \in C^{\#}_{K,\alpha}(c)$.

Now we will give a definition of convergence of sets in E^*. For $c > 0$, let $A(c)$ be a region in E^* which depends on c. Then we will say that $A(c)$ converges to a limit $L \in E^*$ as $c \to 0$, and we will write that $A(c) \to L$, if and only if

$$A(c) \cap H_{K,\alpha} \to_\rho L \cap H_{K,\alpha} \tag{3.18}$$

as $c \to 0$ for any compact $K \subset E$ and $\alpha \geq 1$. Because ρ was defined on compact subsets of E^*, (3.18) makes sense only if $A(c) \cap H_{K,\alpha}$ is compact for c sufficiently small, and $L \cap H_{K,\alpha}$ is compact.

THEOREM 3.1. *Let $C(c)$ be as in* (2.3), *and let $\gamma_3(f)$ be as defined in Section* 2. *Then,*

$$\partial C(c)/\log c^{-1} \to \{(\gamma_3(f), \gamma_3(f) \cdot f) : f \in E\} \tag{3.19}$$

as $c \to 0$.

PROOF. Let L denote $\{(\gamma_3(f), \gamma_3(f) \cdot f) : f \in E\}$, and observe that $\partial C(c)/\log c^{-1} \cap H_{K,\alpha} = (\partial C(c) \cap H_{K,\alpha})/\log c^{-1}$. Then by our definition of convergence, we need to show that $(\partial C(c) \cap H_{K,\alpha})/\log c^{-1} \to_\rho L \cap H_{K,\alpha}$ as $c \to 0$, for any compact $K \subset E$ and $\alpha \geq 1$. It is clear from the compactness of KB_α and the definition of $\gamma_3(f)$ that every sequence in $L \cap H_{K,\alpha}$ has a convergent subsequence, so $L \cap H_{K,\alpha}$ is compact. By Lemmas 3.6 and 3.7, for c sufficiently small, $\partial C(c) \cap H_{K,\alpha} = C^{\#}_{K,\alpha}(c)$, which is a compact subset of E^*. Thus, we need to show that

$$C^{\#}_{K,\alpha}(c)/\log c^{-1} \to_\rho L \cap H_{K,\alpha} \tag{3.20}$$

as $c \to 0$.

Let n^* be the $n^*(c, f)$ in (3.13). By Lemmas 3.1 and 3.5,

$$R_w(n^*, f)^{1/n^*} \to \gamma_1(f)$$

as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. From the definition of n^*, it follows that $c^{1/n^*} \to \gamma_1(f)$ as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. Then

$$n^*/\log c^{-1} \to \frac{1}{\log(1/\gamma_1(f))} = 1/\gamma_2(f) = \gamma_3(f)$$

as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. Therefore,

$$d^*\left[(n^*, n^* \cdot f)/\log c^{-1}, (\gamma_3(f), \gamma_3(f) \cdot f)\right] \to 0 \tag{3.21}$$

as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. The convergence in (3.20) follows from (3.15), (3.16) and (3.21).

LEMMA 3.8. *Suppose K is any compact subset of E, $\alpha \geqslant 1$, n_0 is the $n_0(K, \alpha)$ in Lemma 3.3, $B(c)$ is the Bayes continuation region defined in (2.4), and $(n', n'f)$ is any point in $\partial B(c) \cap H_{K,\alpha}$. Then there exists a $c_0 = c_0(K, \alpha)$ such that $n' > n_0$ when $c \leqslant c_0$.*

PROOF. Recall the assumption that H_0 and H_1 are d-testable, and let c be small enough so that (2.6) holds for some $b > 0$. Clearly, there exists a sequence $(n_k, n_k f_k)$, $k \geqslant 1$, such that $(n_k, n_k f_k) \notin B(c)$, and $(n_k, n_k f_k) \to (n', n'f)$. By (2.6), $(n_k, n_k f_k) \notin C(bc \log c^{-1})$, so $R_w(n_k, f_k) < bc \log c^{-1}$ for all k. Thus, by the continuity of $R_w(n, f)$,

$$R_w(n', f) \leqslant bc \log c^{-1}.$$

The rest of the argument is very similar to the proof of Lemma 3.5. Suppose $n' \leqslant n_0$. Then

$$\min_{f \in KB_\alpha,\, 1 \leqslant n \leqslant n_0} R_w(n, f) \leqslant bc \log c^{-1}.$$

By the compactness of $KB_\alpha \times [1, n_0]$, the continuity of $R_w(n, f)$, and the fact that $R_w(n, f) > 0$ for all $f \in KB_\alpha$ and $n \geqslant 1$, the above inequality does not hold for c sufficiently small.

THEOREM 3.2. *Let $B(c)$ be as in* (2.4). *Then*

$$\partial B(c)/\log c^{-1} \to \{(\gamma_3(f), \gamma_3(f) \cdot f) : f \in E\} \tag{3.22}$$

as $c \to 0$.

PROOF. For any compact $K \subset E$ and $\alpha \geqslant 1$, let $(n', n'f)$ be any point in $\partial B(c) \cap H_{K,\alpha}$, and let $L = \{(\gamma_3(f), \gamma_3(f) \cdot f) : f \in E\}$. We will show that

$$d^*[(n', n'f)/\log c^{-1}, (\gamma_3(f), \gamma_3(f) \cdot f)] \to 0 \tag{3.23}$$

as $c \to 0$ uniformly with respect to f. Then by (3.15) and (3.23), we will have that

$$\partial B(c)/\log c^{-1} \cap H_{K,\alpha} \to_\rho L \cap H_{K,\alpha}$$

as $c \to 0$, and the conclusion of the theorem will follow from our definition of convergence of sets.

By Lemmas 3.5 and 3.8, there exists a $c_1 = c_1(K, \alpha)$ such that $n^*(c, f) > n_0$, $n^*(bc \log c^{-1}, f) > n_0$, and $n' > n_0$ for $c \leqslant c_1$, where n_0 is the $n_0(K, \alpha)$ in Lemma 3.3 and n^* is defined in (3.13). In the proof of Lemma 3.8, we showed that for any $(n', n'f) \in \partial B(c) \cap H_{K,\alpha}$ and c sufficiently small, $R_w(n', f) \leqslant bc \log c^{-1}$ for some $b > 0$. Then it follows from the definitions of n^* and n_0 that

$$n' \geqslant n^*(bc \log c^{-1}, f) \tag{3.24}$$

for $c \leqslant c_1$. Similarly, it can be shown that for $c \leqslant c_1$,

$$n' \leqslant n^*(c, f). \tag{3.25}$$

In the proof of Theorem 3.1, we showed that

$$n^*(c, f)/\log c^{-1} \to \gamma_3(f) \tag{3.26}$$

as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. Since $\log(bc \log c^{-1})^{-1} = \log c^{-1} + o(\log c^{-1})$,

$$n^*(bc \log c^{-1}, f)/\log c^{-1} \to \gamma_3(f) \tag{3.27}$$

as $c \to 0$ uniformly with respect to $f \in KB_\alpha$. Thus by (3.24) through (3.27),

$$n'/\log c^{-1} \to \gamma_3(f) \tag{3.28}$$

as $c \to 0$ uniformly with respect to f. From (3.28), we immediately can conclude that (3.23) is satisfied as $c \to 0$ uniformly with respect to f, so the proof is done.

Certainly, the limit in (3.22) is a reasonable approximation to the boundary of $B(c)/\log c^{-1}$ for c small. Because $\partial C(c)/\log c^{-1}$ also converges to that limit, it makes sense to use the stopping rule t_c, which is defined by

$$t_c = \inf\{n \geq 1 : (n, nT_n) \notin C(c)\} = \inf\{n \geq 1 : R_w(n, T_n) < c\},$$

as an approximation to the optimal rule t^*. A simple heuristic argument shows that $C(c)$ looks like $\{(n, nT_n) : \gamma_1(T_n)^n \geq c\}$ for c small. Thus, a reasonable approximation to t^* for c small is

$$t_c' = \inf\{n : \gamma_1(T_n)^n < c\}.$$

By the definition of γ_1, t_c' is the first n such that the minimum of the likelihood ratio test statistics for testing Ω_0 vs. Ω_0', and Ω_1 vs. Ω_1', is less than c.

In the exponential family case, Schwarz showed that the second-order approximations to the continuation regions depend on the prior distribution of θ, through its atoms and zeros of its density (Schwarz, 1969). Fushimi (1967) found that without a second-order correction, the approximations remain bad for c as small as .00000001. Observe that the limit in (3.22) and t_c' are independent of the prior w. However, second-order approximations to $B(c)$ and $C(c)$ could depend on w.

Acknowledgment. I would like to thank Dr. Michael Woodroofe for his outstanding advice, assistance and encouragement. Also, I would like to thank the referee for his helpful comments and suggestions.

REFERENCES

(1) Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley and Sons, New York.

(2) DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

(3) Dunford, N. and Schwartz, J. T. (1964). *Linear Operators, Part* 1. Inter-Science Publishers, New York.

(4) Fortus, R. (1975). Approximations to Bayesian sequential tests of composite hypotheses. Unpublished thesis, Univ. Michigan.

(5) Fushimi, M. (1967). On the rate of convergence of asymptotically optimal Bayes tests. *Rep. Statist. Appl. Res. Un. Japan Sci. Engrs.* **14** 1–7.

(6) Loeve, M. (1955). *Probability Theory*. D. Van Nostrand, New York.

(7) ROYDEN, H. L. (1968). *Real Analysis 2nd ed.* The Macmillan Company, New York.

(8) SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

(9) SCHWARZ, G. (1968). Asymptotic shapes for sequential test of truncation parameters. *Ann. Math. Statist.* **39** 2038–2043.

(10) SCHWARZ, G. (1969). A second order approximation to optimal sampling regions. *Ann. Math. Statist.* **70** 313–315.

DEPARTMENT OF PURE AND APPLIED MATHEMATICS
WASHINGTON STATE UNIVERSITY
PULLMAN, WASHINGTON 99164