The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1412-1427

LIKELIHOOD RATIO TESTS FOR SEQUENTIAL k-DECISION PROBLEMS

By Gary Lorden

California Institute of Technology

Sequential tests of separated hypotheses concerning the parameter θ of a Koopman-Darmois family are studied from the point of view of minimizing expected sample sizes pointwise in θ subject to error probability bounds. Sequential versions of the (generalized) likelihood ratio test are shown to exceed the minimum expected sample sizes by at most $M \log \log \underline{\alpha}^{-1}$ uniformly in θ, where $\underline{\alpha}$ is the smallest error probability bound. The proof considers the likelihood ratio tests as ensembles of sequential probability ratio tests and compares them with alternative procedures by constructing alternative ensembles, applying a simple inequality of Wald and a new inequality of similar type. A heuristic approximation is given for the error probabilities of likelihood ratio tests, which provides an upper bound in the case of a normal mean.

1. Introduction. Sequential tests based on the generalized likelihood ratio appeared in a striking way in the (1962) paper of G. Schwarz. He showed in the context of Koopman–Darmois families that Bayes procedures for testing two hypotheses separated by an indifference zone have continuation regions in the plane of n, S_n (the cumulative sum sufficient statistic) which when scaled down by the log of $1/c$, c being the cost per observation, approach a limiting region as $c \to 0$. This limiting region, scaled up by the factor $\log c^{-1}$, is given by the rule: stop when the maximum likelihood in either of the hypotheses is less than c times the unrestricted maximum likelihood. Schwarz also showed how the boundaries of the likelihood ratio test regions can be computed explicitly. These results strongly suggest the use of sequential likelihood ratio tests in applications and raise the question of how close to optimal is their performance. The paper of Wong (1968) demonstrated two forms of asymptotic optimality. The first is the asymptotic Bayes property: the integrated risk of the likelihood ratio procedure is asymptotic to the Bayes risk as $c \to 0$. The second is reminiscent of the optimality property of the SPRT. The expected sample size of the likelihood ratio test at every point in the indifference interval was shown to be asymptotically minimum as $c \to 0$ compared to tests whose error probabilities at the endpoints of the indifference interval are not larger and which also satisfy a restriction on the growth of the expected sample size at the endpoints as $c \to 0$. The present paper is concerned with a strengthening of this result and the extension to k-decision problems.

Specialized to the two-decision problem of Schwarz and Wong, the main theorem together with Remark 3 following it establishes the following result. Among all tests with error probabilities less than a prescribed $\alpha > 0$ at the

Received November 4, 1970; revised January 1972.

endpoints of the indifference interval, the likelihood ratio test with $c = \alpha/D \log \alpha^{-1}$ exceeds the minimum expected sample size by at most $M \log \log \alpha^{-1}$ uniformly over compact subintervals of the natural parameter space (and uniformly over the entire space in the case of normal, binomial, exponential and other families of distributions satisfying assumption III). Since the minimum expected sample sizes are of order $\log \alpha^{-1}$ for small α, the excess sampling required by the likelihood ratio test is at most a constant times the log of the minimum, uniformly over compact subintervals. Explicit values of D and M can be computed from the proof, but are too large to be of practical use. More refined methods of obtaining bounds for specific families are under investigation. Another possibility is the evaluation of the performance of sequential likelihood ratio tests by Monte Carlo methods and comparison with lower bounds such as those of Hoeffding (1953, 1960).

Recent work of Schwarz (1969) has interesting parallels with the present investigation. Considering only the case where the a priori distribution is mutually absolutely continuous with respect to Lebesgue measure and the loss functions are bounded above and below wherever positive, Schwarz's result can be described as follows. A Bayes solution continues sampling if the generalized likelihood ratio is greater than $M_1(\bar{X})c(\log c^{-1})^{\rho}$ but not if it is less than $M_2(\bar{X})c(\log c^{-1})^{\rho-1}$, where M_1, M_2 are functions of the current sample mean, $\bar{X}$. Schwarz determines ρ, but not M_1 and M_2, which may play a greater role than ρ for moderately small c. The closest parallel is with the following version of the present results (see Theorem 3). Stopping when the likelihood ratio is less than $\alpha/D \log \alpha^{-1}$ yields error probabilities less than α, while any procedure with error probabilities less than α has expected sample sizes (over a compact subinterval) at least as large as the test which stops when the likelihood ratio is less than $M^*\alpha$.

A difficulty which immediately confronts the statistician who wishes to apply a sequential likelihood ratio test is how to go about choosing critical values of the likelihood ratio. Obviously it would be highly desirable to have error probability approximations like those of Wald for the SPRT. An heuristic approximation is derived in Section 3 which, based on a result of Lorden (1970), is shown to provide an upper bound on the error probabilities for testing a normal mean. In this derivation, as well as in the main theorem, the formulation admits unequal upper bounds on error probabilities of different decisions.

2. Main results. Independent and identically distributed random variables $X_1, X_2, \cdots$ are observed sequentially, having one of the densities

$$f_\theta(x) = \exp(\theta x - b(\theta)), \qquad \underline{\theta} < \theta < \bar{\theta}$$

with respect to a non-degenerate σ-finite measure. The function $b(\cdot)$ is necessarily convex and infinitely differentiable on $(\underline{\theta}, \bar{\theta})$, which need not be the entire natural parameter space of the family. The first derivative, $b'(\theta)$, equals $E_\theta X$

and the second, $b''(\theta)$, equals $\mathrm{Var}_\theta X$. Let $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \cdots$, and define the log-likelihood function

$$L_n(\theta) = \theta S_n - nb(\theta) .$$

For any n, S_n the log-likelihood is a strictly concave function of θ and hence is either monotonic over $(\underline{\theta}, \bar{\theta})$ or else has a unique maximum at $\hat{\theta}$, the maximum likelihood estimate of θ. In the latter case $b'(\hat{\theta}) = S_n/n$. If the likelihood function is monotonic on $(\underline{\theta}, \bar{\theta})$, set $\hat{\theta} = \bar{\theta}$ or $\underline{\theta}$ according to whether it is increasing or decreasing. In any case denote the supremum of $L_n(\theta)$ for $\underline{\theta} < \theta < \bar{\theta}$ by $L_n(\hat{\theta})$.

The statistical problem is specified by $s + 1$ ($s \geqq 2$) intervals, $H_1 = (\theta_0, \theta_1]$, $H_2 = [\theta_1, \theta_2], \cdots, H_{s+1} = [\theta_s, \theta_{s+1})$, where $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_{s+1} = \bar{\theta}$, together with $k \geqq 2$ decisions and the requirement

$$P_\theta(j\text{th decision}) \leqq \alpha_{ij} \qquad \text{for all} \quad \theta \in H_i \qquad (i = 1, \cdots, s+1; j = 1, \cdots, k) , \tag{1}$$

where the α_{ij}'s belong to $(0, 1]$ and satisfy the following assumptions. (See Remarks 1 and 3 for alternative formulations.)

I. $$\underline{\alpha} = \min \alpha_{ij} \leqq \tfrac{1}{3} ,$$

II. $$\max_j \min(\alpha_{ij}, \alpha_{(i+1)j}) = 1 \qquad \text{for} \quad i = 1, \cdots, s .$$

Assumption I is merely a convenience which makes $\log\log \underline{\alpha}^{-1} \geqq \log\log 3 > 0$. Assumption II is a "separation of hypotheses" type of condition ensuring that adjacent H_i's have at least one "correct" decision in common.

Define

$$\bar{\alpha} = \max\{\alpha_{ij} \mid \alpha_{ij} < 1\} \qquad \text{and} \qquad r = \log \underline{\alpha}/\log \bar{\alpha} \geqq 1 .$$

The bounds obtained in the proof of the theorem can be chosen to depend on the values of the α_{ij}'s < 1 only through $\underline{\alpha}$ and r, or, equivalently, $\bar{\alpha}$ and r. Since $\log\log \bar{\alpha}^{-1} = \log\log \underline{\alpha}^{-1} - \log r$, the bound in (17) of Theorem 1 can be written in the apparently stronger form $M_1(r) \log\log \bar{\alpha}^{-1}$ under the restriction $\bar{\alpha} \leqq \frac{1}{3}$.

To obtain bounds in the theorem which are uniform on $(\underline{\theta}, \bar{\theta})$ it is necessary that, for example, in a one-sided SPRT of θ_2 versus a θ near $\underline{\theta}$ the extra sampling due to "excess over the boundary" be bounded uniformly in θ. A sufficient condition to meet requirements of this type is the following assumption.

III. $(\theta - \theta_1)^2 b''(\theta)/[(\theta - \theta_1)b'(\theta) - (b(\theta) - b(\theta_1))]^2$ is bounded above as $\theta \to \underline{\theta}, \bar{\theta}$.

It is readily verified that III is equivalent to any condition of the same form with θ_1 replaced by another point in $(\underline{\theta}, \bar{\theta})$.

For $\theta, \theta' \in (\underline{\theta}, \bar{\theta})$ define the information number

$$I(\theta, \theta') = E_\theta \log(f_\theta(X)/f_{\theta'}(X)) = (\theta - \theta')b'(\theta) - (b(\theta) - b(\theta')) \tag{2}$$

and note also that

$$\mathrm{Var}_\theta \log(f_\theta(X)/f_{\theta'}(X)) = (\theta - \theta')^2 b''(\theta) , \tag{3}$$

so that III concerns the ratio of the variance of $\log f_\theta((X)/f_{\theta_1}(X))$ to the square of its mean when θ is true. Since the ratio is continuous, III is implied by

III′. $\underline{\theta}$ and $\bar{\theta}$ belong to the interior of the natural parameter space of the Koopman–Darmois family.

If θ is the mean of normally distributed variables with known variance, then III is satisfied with $\underline{\theta} = -\infty$, $\bar{\theta} = +\infty$. Also, assumption III holds with $(\underline{\theta}, \bar{\theta})$ taken as the full natural parameter space in the case of normal scale parameter, Bernoulli, Poisson, and negative exponential distributions.

Before stating the theorem, it is necessary to define the likelihood ratio tests and examine their structure. For specified γ_{ij}'s $\in (0, 1]$ satisfying assumptions I and II, the likelihood ratio test $(\hat{N}, \hat{d})$ consists of a terminal decision rule $\hat{d}$ and a stopping time $\hat{N} = \min_j \max_i \hat{N}_{ij}$, where $\hat{N}_{ij}$ is the smallest n (or ∞ if there is no n) such that

$$L_n(\hat{\theta}) \geqq \sup_{H_i} L_n(\theta) + \log \gamma_{ij}^{-1}, \tag{4}$$

and $\hat{d}$ is the smallest j such that $\hat{N} = \max_i \hat{N}_{ij}$ (see Remark 2 for a slightly different formulation). Note that for fixed j, values of i such that $\gamma_{ij} = 1$ play no role since (4) always holds. Also, by the concavity of $L_n(\cdot)$, either $\hat{\theta} \in H_i$ and hence $\sup_{H_i} L_n(\theta) = L_n(\hat{\theta})$, or else the supremum of $L_n(\cdot)$ on H_i is attained at the point closest to $\hat{\theta}$, evidently an endpoint.

The structure and performance of likelihood ratio tests is illuminated by investigating the asymptotic behavior of $E_\theta \hat{N}$ as $\underline{\gamma} = \min \gamma_{ij} \to 0$ in the case where the γ_{ij}'s less than one are equal (applicable to the testing problem in the case where $\bar{\alpha} = \underline{\alpha}$). Fix θ belonging to the interior of H_m. It is clear from (4) and the remarks following that in any case $\hat{N}$ stops no later than the first time that for some j with $\gamma_{mj} = 1$

$$L_n(\theta) \geqq \max_{i:\gamma_{ij}<1} [L_n(\varphi_i) + \log \gamma_{ij}^{-1}], \tag{5}$$

where φ_i is the endpoint of H_i closer to θ. Now, in case the γ_{ij}'s appearing in (5) are all equal to $\underline{\gamma}$, Theorem 3.3 of [5] implies that

$$I(\theta)E_\theta \hat{N} \leqq \log \underline{\gamma}^{-1} + O((\log \underline{\gamma}^{-1})^{\frac{1}{2}}) \qquad \text{as} \quad \underline{\gamma} \downarrow 0, \tag{6}$$

where

$$\begin{aligned} I(\theta) &= \max_{j:\gamma_{mj}=1} \min_{i:\gamma_{ij}<1} I(\theta, \varphi_i) \\ &= \max_{j:\gamma_{mj}=1} \min_{i:\gamma_{ij}<1} \sup_{\theta' \in H_i} I(\theta, \theta'). \end{aligned} \tag{7}$$

If θ is the common endpoint of H_m and H_{m+1} the above remarks apply with (5) and (7) restricted to j such that $\gamma_{mj} = \gamma_{(m+1)j} = 1$. Note that by assumption II there is for each θ a j such that the closest to θ of the θ_i's does not occur as a φ_i in (5) or (7). Therefore,

$$\underline{I} = \min_{\underline{\theta}<\theta<\bar{\theta}} I(\theta) > 0.$$

The minimum is attained by virtue of the continuity of $I(\cdot, \theta_i)$ for each $i = 1, \cdots, s$

and the obvious fact that $I(\theta)$ is decreasing on H_1 and increasing on H_{s+1}. The determination of $\underline{I}$ in a specific problem is straightforward since $I(\theta)$ is alternately increasing and decreasing, with local extrema at values of θ such that

$$I(\theta, \varphi_{i_1}) = I(\theta, \varphi_{i_2}) ,$$

where i_1 and i_2 are minimizing choices of i in (7) for different j's. Note that the last relation is equivalent to

$$b'(\theta) = (b(\varphi_{i_1}) - b(\varphi_{i_2}))/(\varphi_{i_1} - \varphi_{i_2}) .$$

As a consequence of the proof of the theorem, obtained by choosing (N, d) to be $(\hat{N}, \hat{d})$, the inequality (6) is actually an equality, in the case of equal error probability bounds. Note, however, that the square root term in (6) is of larger order of magnitude than the difference between $E_\theta \hat{N}$ and the minimum attainable.

It is well known in the case $s = 2$ (e.g. [8], [11]) that sequential likelihood ratio tests are truncated and this result is true in the present context by virtue of the semi-indifference assumption II. In fact,

$$\hat{N} \leqq [\underline{I}^{-1} \log \underline{\gamma}^{-1}] + 1 \quad \text{with probability one,} \tag{8}$$

where $[x]$ denotes the greatest integer $\leqq x$. To verify (8), note that if $\hat{\theta} \in H_m$ for $n = [\underline{I}^{-1} \log \underline{\gamma}^{-1}] + 1$ and j^* is a minimizing j in (7), then $S_n = nb'(\hat{\theta})$ (as shown above) and hence for all $\theta \in \bigcup H_i$ $(i: \gamma_{ij^*} < 1)$

$$\begin{aligned} L_n(\hat{\theta}) - L_n(\theta) &= n[(\hat{\theta} - \theta)b'(\hat{\theta}) - (b(\hat{\theta}) - b(\theta))] \\ &= nI(\hat{\theta}, \theta) \geqq nI(\hat{\theta}) \geqq n\underline{I} > \log \underline{\gamma}^{-1} , \end{aligned} \tag{9}$$

which implies (4) for all i such that $\gamma_{ij^*} < 1$. If $\hat{\theta} = \underline{\theta}$ for $n = [\underline{I}^{-1} \log \underline{\gamma}^{-1}] + 1$, then $S_n \leqq nb'(\theta_1)$, whence $L_n(\theta_1) - L_n(\theta) \geqq nI(\theta_1, \theta) > \log \underline{\gamma}^{-1}$ for $\theta > \theta_1$ as in (9), and (4) is satisfied as above. The case $\hat{\theta} = \bar{\theta}$ is similar.

The proof of the theorem relies on two simple inequalities. The first is a slight modification of Wald's lower bound ([10], page 197) on the expected sample sizes of competitors of the SPRT. A full proof is given so that parts of the argument can be applied conveniently in the proof of Lemma 2.

LEMMA 1. *Suppose N_1 and N_2 are (possibly infinite) stopping times for $X_1, X_2, \cdots$ such that*

$$P_{\theta'}(N_1 < \infty) \leqq \alpha < 1 \qquad \textit{and} \qquad P_\theta(N_2 < \infty) \leqq \beta < 1 .$$

If $I(\theta, \theta') = E_\theta \log (f_\theta(X_1)/f_{\theta'}(X_1)) > 0$, then

$$I(\theta, \theta')E_\theta \min(N_1, N_2) \geqq (1 - \beta) \log \alpha^{-1} - \log 2 . \tag{10}$$

PROOF. Assume $E_\theta \min(N_1, N_2)$ is finite (otherwise (10) is trivial). Wald's equation for the expected value of a stopped sum [1] applies with $S_n = \log (f_\theta(X_1) \cdots f_\theta(X_n)/f_{\theta'}(X_1) \cdots f_{\theta'}(X_n)) = \log(f_{\theta n}/f_{\theta' n})$ and stopping time $\min(N_1, N_2)$ to yield

$$\begin{aligned} &I(\theta, \theta')E_\theta \min(N_1, N_2) \\ &\quad = \int_{\{N_1 < N_2\}} \log(f_{\theta N_1}/f_{\theta' N_1})\, dP_\theta + \int_{\{N_1 \geq N_2 \neq \infty\}} \log(f_{\theta N_2}/f_{\theta' N_2})\, dP_\theta . \end{aligned} \tag{11}$$

Adopting the cnovention that $0 \log a/b = 0$ for all $a, b \geqq 0$, the estimate

$$\int_{\{N_1<N_2\}} \log(f_{\theta N_1}/f_{\theta' N_1})\, dP_\theta \geqq P_\theta(N_1 < N_2) \log \frac{P_\theta(N_1 < N_2)}{P_{\theta'}(N_1 < N_2)} \tag{12}$$

for the first term on the right-hand side of (11) is derived from Jensen's inequality,

$$\int_{\{N_1<N_2\}} (-\log(f_{\theta' N_1}/f_{\theta N_1})) \frac{dP_\theta}{P_\theta(N_1 < N_2)}$$
$$\geqq -\log \int_{\{N_1<N_2\}} (f_{\theta' N_1}/f_{\theta N_1}) \frac{dP_\theta}{P_\theta(N_1 < N_2)},$$

and the relation

$$\int_{\{N_1<N_2\}} (f_{\theta' N_1}/f_{\theta N_1}) \frac{dP_\theta}{P_\theta(N_1 < N_2)} = \sum_{n=1}^\infty \int_{\{N_1=n<N_2\}} (f_{\theta' n}/f_{\theta n}) \frac{dP_\theta}{P_\theta(N_1 < N_2)}$$
$$= \sum_{n=1}^\infty \int_{\{N_1=n<N_2\}} \frac{dP_{\theta'}}{P_\theta(N_1 < N_2)} = \frac{P_{\theta'}(N_1 < N_2)}{P_\theta(N_1 < N_2)}$$

assuming $P_\theta(N_1 < N_2) > 0$ (otherwise (12) holds by the convention stated).

The second term on the right-hand side of (11) is estimated similarly, whence (11) implies

$$\begin{aligned} I(\theta, \theta')E_\theta \min(N_1, N_2) &\geqq P_\theta(N_1 < N_2) \log \frac{P_\theta(N_1 < N_2)}{P_{\theta'}(N_1 < N_2)} \\ &\quad + P_\theta(N_1 \geqq N_2 \neq \infty) \log \frac{P_\theta(N_1 \geqq N_2 \neq \infty)}{P_{\theta'}(N_1 \geqq N_2 \neq \infty)} \\ &\geqq P_\theta(N_1 \leqq N_2) \log(P_{\theta'}(N_1 < N_2))^{-1} - \log 2, \end{aligned} \tag{13}$$

since $\log(P_{\theta'}(N_1 \geqq N_2 \neq \infty))^{-1} \geqq 0$ and $p \log p + (1-p)\log(1-p) \geqq -\log 2$ for all p in $[0, 1]$ and for $p = P_\theta(N_1 < N_2) = 1 - P_\theta(N_1 \geqq N_2 \neq \infty)$ in particular. The bounds $P_\theta(N_1 < N_2) = 1 - P_\theta(N_1 \geqq N_2 \neq \infty) \geqq 1 - P_\theta(N_2 < \infty) \geqq 1 - \beta$ and $P_{\theta'}(N_1 < N_2) \leqq P_{\theta'}(N_1 < \infty) \leqq \alpha$ combined with (13) yield (10).

The second key inequality is needed in cases where $I(\theta, \theta')$ is small.

LEMMA 2. *If $P_{\theta'}(N < \infty) \leqq \alpha < 1$ and $I(\theta, \theta') > 0$, then for all $\varepsilon > 0$*

$$E_\theta(\log \alpha^{-1} - (I(\theta, \theta') + \varepsilon)N)^+ \leqq e^{-1} + (2\varepsilon)^{-1} \operatorname{Var}_\theta \log(f_\theta(X_1)/f_{\theta'}(X_1)),$$

where $e = 2.718 \cdots$.

PROOF. Let $m = (\log \alpha^{-1})/(I(\theta, \theta') + \varepsilon)$. Clearly

$$\begin{aligned} \int_{\{N<m\}} \log(f_{\theta N}/f_{\theta' N})\, dP_\theta &- \int_{\{N<m\}} (I(\theta, \theta') + \varepsilon)N\, dP_\theta \\ &\leqq E_\theta \sup_{n\geqq 0}(\log(f_{\theta n}/f_{\theta' n}) - n(I(\theta, \theta') + \varepsilon)) \\ &\leqq (2\varepsilon)^{-1} \operatorname{Var}_\theta \log(f_\theta(X_1)/f_{\theta'}(X_1)), \end{aligned} \tag{14}$$

the last inequality by Kingman's (1962) result that a random walk with negative mean has expected supremum at most one-half the ratio of its variance to the absolute value of its mean. (Howard Taylor has informed the author that Kingman's proof goes through only if the third moment is finite, but that the general case follows by a truncation argument; in the present context of Koopman–Darmois families the moment generating function exists, hence all moments are finite.)

The estimate

$$\int_{\{N<m\}} \log(f_{\theta N}/f_{\theta' N})\, dP_\theta \geqq P_\theta(N<m) \log \frac{P_\theta(N<m)}{P_{\theta'}(N<m)} \tag{15}$$

for the first term on the left-hand side of (14) is derived exactly like (12). Since $p \log p \geqq -e^{-1}$ for all p in $[0, 1]$ and $P_{\theta'}(N<m) \leqq P_{\theta'}(N<\infty) \leqq \alpha$, (15) implies

$$\begin{aligned}\int_{\{N<m\}} \log(f_{\theta N}/f_{\theta' N})\, dP_\theta &\geqq P_\theta(N<m) \log \alpha^{-1} - e^{-1} \\ &= \int_{\{N<m\}} \log \alpha^{-1}\, dP_\theta - e^{-1},\end{aligned}$$

which combines with (14) to prove the lemma.

THEOREM 1. *There is a $D \geqq 1$ such that the sequential likelihood ratio test $(\hat{N}, d)$ with*

$$\gamma_{ij} = \alpha_{ij}/D \log \underline{\alpha}^{-1} \quad \text{if} \quad \alpha_{ij} < 1, \ = 1 \quad \text{if} \quad \alpha_{ij} = 1 \tag{16}$$

satisfies (1) *whenever the α_{ij}'s satisfy* I *and* II. *Denote by $n(\theta)$ the infimum of $E_\theta N$ over all tests (N, d) for which* (1) *holds. Under assumptions* I—III *there is an increasing function $M(\cdot)$ on $[1, \infty)$ such that*

$$E_\theta \hat{N} - n(\theta) \leqq M(r) \log\log \underline{\alpha}^{-1} \qquad \text{for all} \quad \theta \in (\underline{\theta}, \bar{\theta}), \tag{17}$$

where $r = \log \underline{\alpha}/\log \bar{\alpha}$.

PROOF. A crude estimate of the error probabilities of $(\hat{N}, \hat{d})$ will suffice. Wald's upper bound [10] for SPRT error probabilities yields for any fixed n

$$P_{\theta_i}(f_{\theta_i n} \leqq \gamma f_{\theta n}) \leqq P_{\theta_i}(f_{\theta_i m} \leqq \gamma f_{\theta m} \text{ for some } m \geqq 1) \leqq \gamma \qquad \text{for } 0<\gamma<1. \tag{18}$$

Therefore, for $\theta > \theta_i$ $(i = 1, \cdots, s)$

$$P_{\theta_i}\left(S_n \geqq \frac{\log \gamma^{-1}}{\theta - \theta_i} + n\frac{b(\theta) - b(\theta_i)}{\theta - \theta_i}\right) \leqq \gamma. \tag{19}$$

Now

$$P_{\theta_i}\left(S_n > \inf_{\theta > \theta_i}\left[\frac{\log \gamma^{-1}}{\theta - \theta_i} + n\frac{b(\theta) - b(\theta_i)}{\theta - \theta_i}\right]\right) \leqq \gamma, \tag{20}$$

since a sequence of θ's approximating the infimum can be chosen so as to express the left-hand side as the limit of probabilities in (19). Rewriting the left-hand side of (20),

$$P_{\theta_i}(L_n(\hat{\theta}) > L_n(\theta_i) + \log \gamma^{-1} \text{ and } \hat{\theta} > \theta_i) \leqq \gamma, \tag{21}$$

and, letting $\gamma \downarrow \gamma_{ij}$,

$$P_{\theta_i}(L_n(\hat{\theta}) \geqq L_n(\theta_i) + \log \gamma_{ij}^{-1} \text{ and } \hat{\theta} > \theta_i) \leqq \gamma_{ij}. \tag{22}$$

Since every S_n is stochastically smaller for $\theta < \theta_i$, evidently

$$(23) \qquad P_\theta(L_n(\hat\theta) \geqq L_n(\theta_i) + \log \gamma_{ij}^{-1} \quad \text{and} \quad \hat\theta > \theta_i) \leqq \gamma_{ij} \qquad \text{for} \quad \theta \leqq \theta_i$$

and, similarly, for $i \geqq 1$

$$(24) \qquad P_\theta(L_n(\hat\theta) \geqq L_n(\theta_{i-1}) + \log \gamma_{ij}^{-1} \quad \text{and} \quad \hat\theta < \theta_{i-1}) \leqq \gamma_{ij} \qquad \text{for} \quad \theta \geqq \theta_{i-1}\,.$$

Fix $i \neq 1,\ s+1$ and consider j such that $\gamma_{ij} < 1$. Evidently (4) cannot be satisfied unless one of the events in (23) and (24) occurs. Hence, for $\theta \in H_i = [\theta_{i-1}, \theta_i]$, $i \neq 1,\ s+1$,

$$(25) \qquad P_\theta(\hat N_{ij} = n) \leqq 2\gamma_{ij}\,.$$

Since the choices of γ_{ij} in (16) result in $\underline{\gamma} = \min \gamma_{ij} < \underline{\alpha} \leqq \frac{1}{3}$, $\log \underline{\gamma}^{-1}$ is greater than 1 and, by (8),

$$(26) \qquad \hat N \leqq (\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1}\,.$$

Therefore $\hat d = j$ only if $\hat N_{ij} < (\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1}$, whence by (25)

$$(27) \qquad P_\theta(\hat d = j) \leqq 2(\underline{I}^{-1} + 1)\gamma_{ij} \log \underline{\gamma}^{-1} \qquad \text{for} \quad \theta \in H_i\,,$$

$i \neq 1,\ s+1$. In case $i = 1$ or $s+1$, (27) holds without the factor of 2, since only one side need be considered in the above argument. (Wong ([11]) has shown that in the case where all γ_{ij}'s < 1 are equal the error probabilities are of some smaller order of magnitude than the bounds (27) as $\underline{\gamma} \to 0$. Also, the heuristics and the results for the normal case in Section 3 suggest bounds of order $\underline{\gamma}(\log \underline{\gamma}^{-1})^{\frac{1}{2}}$, at least when $(\underline{\theta}, \bar\theta)$ is a compact subinterval of the natural parameter space. However, the order of magnitude of the bound $\log\log \underline{\alpha}^{-1}$ in (17) is unaffected if the factor $\log \underline{\gamma}^{-1}$ in (27) is replaced by any positive root of $\log \underline{\gamma}^{-1}$.)

By (27), the choices of γ_{ij}'s in (16) are sufficient for (1) if D is chosen to satisfy

$$2(\underline{I}^{-1} + 1)D^{-1} \log(D\underline{\alpha}^{-1} \log \underline{\alpha}^{-1}) \leqq \log \alpha^{-1}\,.$$

Routine calculation shows that, for example, $D \geqq e$ satisfying $D/\log D \geqq 6(\underline{I}^{-1} + 1)$ is sufficient.

Fix $\theta \in (\underline{\theta}, \bar\theta)$ and let $J = \{j \mid \alpha_{ij} = 1$ for all i such that $\theta \in H_i\}$ (θ may be an endpoint of two H_i's). For $j \in J$, define $N_{ij} = 1$ if $\alpha_{ij} = 1$, and if $\alpha_{ij} < 1$ let N_{ij} be the smallest n (or ∞ if there is no n) such that

$$(28) \qquad L_n(\theta) \geqq L_n(\varphi_i) + \log \gamma_{ij}^{-1}\,,$$

where φ_i is the endpoint of H_i closer to θ. A comparison of definitions using (28) and (4), establishes that $N_{ij} \geqq \hat N_{ij}$ for all $j \in J$, $i = 1, \cdots, s+1$, so that

$$(29) \qquad \hat N = \min_j \max_i \hat N_{ij} \leqq \max_i N_{ij} \qquad \text{for all} \quad j \in J\,.$$

Given a fixed procedure (N, d) satisfying (1), the choice of which may depend on θ, define

$$(30) \qquad \tilde N_{ij} = \min(N_{ij}, N\{d = j \ \text{ or } \ d \notin J\})$$

for $j \in J$, $i = 1, \cdots, s+1$, where the second term of the minimum equals N if

$d = j$ or $d \notin J$ and equals $+\infty$ otherwise. Now, for some $j \in J$ the event $\{d = j$ or $d \notin J\}$ occurs, whence $\tilde{N}_{ij} \leqq N$ for all i by (30). For the same $j \in J$, (29) yields $\hat{N} \leqq N_{ij}$ for some i; hence,

$$\hat{N} - N \leqq N_{ij} - \tilde{N}_{ij} \quad \text{for some} \quad i\,, \tag{31}$$

$\hat{N}$ and N_{ij} being finite with probability one for all θ. Therefore,

$$\hat{N} - N \leqq \textstyle\sum_{j \in J, i} (N_{ij} - \tilde{N}_{ij}) \text{ a.s.} \quad \text{for all} \quad \theta\,, \tag{32}$$

since the summands are nonnegative. Taking expectations,

$$E_\theta \hat{N} - E_\theta N \leqq \textstyle\sum_{j \in J, i} (E_\theta N_{ij} - E_\theta \tilde{N}_{ij})\,. \tag{33}$$

To estimate the summands in (33), first fix $j \in J$ and note that for i such that $\alpha_{ij} = 1$ both expectations are zero by definition. Thus, consider only i such that $\alpha_{ij} < 1$. Apply Wald's lower bound (10) to $E_\theta \tilde{N}_{ij}$ with $\theta' = \varphi_i$, $N_1 = \min(N_{ij}, N\{d = j\})$, $N_2 = N\{d \notin J\}$. Now

$$P_{\varphi_i}(N_1 < \infty) \leqq P_{\varphi_i}(N_{ij} < \infty) + P_{\varphi_i}(d = j) \leqq \gamma_{ij} + \alpha_{ij} \leqq 2\alpha_{ij}\,,$$

since (N, d) satisfies (1) and the usual error probability bound applies to the one-sided SPRT N_{ij}.

Also, since decisions outside J are "incorrect" for θ,

$$P_\theta(N_2 < \infty) \leqq k\bar{\alpha} = k\underline{\alpha}^{1/r} \leqq k\alpha_{ij}^{1/r}\,,$$

where $r = \log \underline{\alpha}/\log \bar{\alpha}$, so that by Lemma 1

$$\begin{aligned} I(\theta, \varphi_i) E_\theta \tilde{N}_{ij} &\geqq (1 - k\alpha_{ij}^{1/r}) \log(2\alpha_{ij})^{-1} - \log 2 \\ &\geqq (1 - k\alpha_{ij}^{1/r}) r \log \alpha_{ij}^{-1/r} - 2 \log 2 \\ &\geqq \log \alpha_{ij}^{-1} - kre^{-1} - \log 4\,, \end{aligned} \tag{34}$$

since $x \log x^{-1} \leqq e^{-1}$.

Applying Wald's equation to obtain the expected sample size of a one-sided SPRT,

$$I(\theta, \varphi_1) E_\theta N_{ij} = \log \gamma_{ij}^{-1} + \Delta\,,$$

where Δ is expected excess over the boundary, which by Theorem 1 of [6] is at most $E_\theta Z^2/E_\theta Z$, where $Z = \log(f_\theta(X)/f_{\varphi_i}(X))$. Therefore,

$$E_\theta N_{ij} \leqq (I(\theta, \varphi_i))^{-1}(\log \alpha_{ij}^{-1} + \log\log \underline{\alpha}^{-1} + \log D) + 1 + \rho(\theta, \varphi_i)\,, \tag{35}$$

where

$$\rho(\theta, \varphi_i) = (\theta - \varphi_i)^2 b''(\theta)/((\theta - \varphi_i) b'(\theta) - (b(\theta) - b(\varphi_i)))^2\,.$$

Combining (34) and (35),

$$\begin{aligned} E_\theta N_{ij} - E_\theta \tilde{N}_{ij} \leqq\; & (I(\theta, \varphi_i))^{-1}(kre^{-1} + \log 4D + \log\log \underline{\alpha}^{-1}) \\ & + 1 + \rho(\theta, \varphi_i)\,. \end{aligned} \tag{36}$$

Adding estimates of the form (36) for $j \in J$ and i such that $\alpha_{ij} < 1$ and using the fact that $\log\log \underline{\alpha}^{-1} \geqq \log\log 3$ it follows from (33) that

$$E_\theta \hat{N} - E_\theta N \leqq M_\theta(r) \log\log \underline{\alpha}^{-1}\,. \tag{37}$$

Now, r is fixed and $M_\theta(r)$ is bounded above uniformly on any subset of $(\underline{\theta}, \bar{\theta})$ where $\max_i \rho(\theta, \varphi_i)$ and $\max_i (I(\theta, \varphi_i))^{-1}$ are bounded, the maximum being taken over i such that $\alpha_{ij} < 1$ for some $j \in J$. Let $V_1, \cdots, V_s$, denote disjoint deleted neighborhoods of $\theta_1, \cdots, \theta_s$, respectively (i.e., $\theta_i \notin V_i$), to be chosen later. If the endpoints of V_1 and V_s are in the natural parameter space, then continuity considerations and assumption III imply that $\rho(\theta, \varphi_i)$ and $\max_i (I(\theta, \varphi_i))^{-1}$ are bounded above uniformly on the complement of $\bigcup_i V_i$. Therefore, there is an $M(r)$ such that

$$E_\theta \hat{N} - E_\theta N \leqq M(r) \log\log \underline{\alpha}^{-1} \qquad \text{for} \quad \theta \in (\underline{\theta}, \bar{\theta}) - \textstyle\bigcup_i V_i\,. \tag{38}$$

Now consider a fixed θ belonging to V_m, $m = 1, \cdots, s$. Let $J = \{j \mid \alpha_{mj} = \alpha_{(m+1)j} = 1\}$. Define N_{ij} as before, but change the definition of $\tilde{N}_{ij}$, (30), to

$$\tilde{N}_{ij} = \min(N_{ij}, N\{d = j\})\,. \tag{39}$$

In case $d \in J$, (32) holds by an argument similar to the one above and, hence,

$$\hat{N} - N \leqq \textstyle\sum_{j \in J, i} (N_{ij} - \tilde{N}_{ij}) + ((\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1} - N\{d \notin J\})^+ \tag{40}$$

since all terms are nonnegative and since the final term suffices in case $d \notin J$ by virtue of (8).

To estimate $E_\theta N_{ij} - E_\theta \tilde{N}_{ij}$, note that Lemma 1 applies with $N_2 \equiv \infty$, $\beta = 0$, and $N_1 = \tilde{N}_{ij}$ to yield

$$I(\theta, \varphi_i) E_\theta \tilde{N}_{ij} \geqq \log(4\alpha_{ij})^{-1}\,, \tag{41}$$

by virtue of $P_{\varphi_i}(N_1 < \infty) \leqq 2\alpha_{ij}$ as before. Using the upper bound (35) on $E_\theta N_{ij}$,

$$E_\theta N_{ij} - E_\theta \tilde{N}_{ij} \leqq (I(\theta, \varphi_i))^{-1}(\log 4D + \log\log \underline{\alpha}^{-1}) + 1 + \rho(\theta, \varphi_i)\,. \tag{42}$$

Now, $\theta \in V_m$ and, by the definition of J, θ_m does not occur as a φ_i for any $\alpha_{ij} < 1$. Hence, $\max_i (I(\theta, \varphi_i))^{-1}$ and $\max_i \rho(\theta, \varphi_i)$ occurring in (42) are bounded above uniformly on $\bigcup_i V_i$, and by (40) there is an $M(r)$ such that for $\theta \in \bigcup V_i$

$$E_\theta \hat{N} - E_\theta N \leqq M(r) \log\log \underline{\alpha}^{-1} + E_\theta((\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1} - N\{d \notin J\})^+\,. \tag{43}$$

An upper bound on the last term in (43) is obtained by applying Lemma 2 with $\theta' = \theta_m$, $N = N_2$, and $\alpha = k\underline{\alpha}^{1/r} \geqq k\bar{\alpha} \geqq P_{\theta_m}(N\{d \notin J\} < \infty)$. Noting that

$$\log \underline{\gamma}^{-1} = \log(\underline{\alpha}^{-1} \log \underline{\alpha}^{-1}) + \log D \leqq (1 + e^{-1}) \log \underline{\alpha}^{-1} + \log D$$

since $\log x \leqq x^{1/e}$ for $x \geqq 1$, the above choice of α leads to

$$(\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1} \leqq A \log \alpha^{-1} + B\,, \qquad \text{where}$$

$$A = r(1 + e^{-1})(\underline{I}^{-1} + 1) \qquad \text{and} \qquad B = (\underline{I}^{-1} + 1)(\log D + r(1 + e^{-1}) \log k)\,.$$

Thus

$$\begin{aligned} E_\theta((\underline{I}^{-1} + 1) \log \underline{\gamma}^{-1} - N\{d \notin J\})^+ &\leqq E_\theta(A \log \alpha^{-1} + B - N\{d \notin J\})^+ \\ &\leqq B + A E_\theta(\log \alpha^{-1} - A^{-1} N\{d \notin J\})^+\,. \end{aligned} \tag{44}$$

If $V_1, \cdots, V_s$ are chosen so that, say,

$$I(\theta, \theta_m) \leqq \tfrac{1}{2} A^{-1} \qquad \text{for} \quad \theta \in V_m\,, \qquad m = 1, \cdots, s\,, \tag{45}$$

then applying Lemma 2 with $\varepsilon = A^{-1} - I(\theta, \theta_m)$ to the last term in (44) and using (45) yields

$$E_\theta((\underline{I}^{-1} + 1)\log \underline{\gamma}^{-1} - N\{d \notin J\})^+ \leqq B + Ae^{-1} + A^2(\theta - \theta_m)^2 b''(\theta) , \tag{46}$$

using (3) for the variance term in Lemma 2. If V_1 and V_s are chosen so that their endpoints are interior to the natural parameter space, then the last term in (46) is uniformly bounded by virtue of the continuity of b''. Choose $V_1, \cdots, V_s$ to satisfy this last requirement and (45). By (43) and (46), and the fact that $\log\log \underline{\alpha}^{-1} > \log\log 3$, there is an $M'(r)$ such that

$$E_\theta \hat{N} - E_\theta N \leqq M'(r) \log\log \underline{\alpha}^{-1} \qquad \text{for} \quad \theta \in \textstyle\bigcup_i V_i . \tag{47}$$

Since the larger of $M'(r)$ in (47) and $M(r)$ in (38) suffices for all $\theta \in (\underline{\theta}, \bar{\theta})$ and (N, d) may be chosen differently for different θ's, (17) is proved.

The following remarks deal with modifications of the above formulations and other refinements.

REMARK 1. Note that the proof goes through (with slightly larger bounds) if the restriction on (N, d) is weakened by multiplying all α_{ij}'s less than one by k, with the choice of $(\hat{N}, \hat{d})$ unchanged. This observation suffices to show that a similar theorem holds if the testing problem is reformulated with (1) replaced by

$$\textstyle\sum_{j \in J_i} P_\theta(j\text{th decision}) \leqq \alpha_i \qquad \text{for all} \quad \theta \in H_i , \qquad i = 1, \cdots, s+1 , \tag{1$'$}$$

where J_i is the subset of decisions "incorrect" for H_i (corresponding to the set of j's with $\alpha_{ij} < 1$ in the original formulation), because (1)$'$ is stronger than (1) with $\alpha_{ij} = \alpha_i$ for $j \in J_i$, $= 1$ for $j \notin J_i$, but is weaker than (1) with $\alpha_{ij} = \alpha_i/k$ for $j \in J_i$, $= 1$ for $j \notin J_i$. Of course, in the case $s = 2$ considered by Schwarz and Wong the two formulations are identical.

REMARK 2. A natural alternative definition of $\hat{N}$ is the smallest n such that, for some j, (4) holds for all i. This "memoryless" version of $\hat{N}$ requires (possibly) more sampling, since (4) may not hold for all i simultaneously at time $n = \max_i \hat{N}_{ij}$; however, unlike the original $\hat{N}$, it makes the decision to stop on the basis of the sufficient statistic (n, S_n) only. The theorem holds for this modification of $\hat{N}$, but an additional estimate is needed to bound the expected time required, after N_θ stops, for the "memoryless" version of N_θ (defined by (5)) to stop. Evidently this is no larger than the time required for $L_n(\theta)$ to increase by at least Y relative to every $L_n(\varphi_i)$ such that $\alpha_{ij} < 1$ for some $j \in J$, where

$$Y = \textstyle\sum_i [L_{\hat{N}_{ij}}(\theta) - L_{\hat{N}_{ij}}(\varphi_i) - (L_{N_\theta}(\theta) - L_{N_\theta}(\varphi_i))]^+ .$$

Given Y, the (conditional) expected time required for this increase is bounded by $2Y \min_i (I(\theta, \varphi_i))^{-1} + B$ (where B can be calculated by applying Theorem 1 of [6] for $\theta < \theta_1$, $\theta > \theta_s$, and Theorem 3.3 of [5] for $\theta_1 \leqq \theta \leqq \theta_s$). Clearly, each term in Y is stochastically smaller than $-\inf_{n \geqq 0}[L_n(\theta) - L_n(\varphi_i)]$, whose expectation is bounded above by $b''(\theta)(\theta - \varphi_i)^2/2I(\theta, \varphi_i)$ by applying Kingman's bound, used for (12). Thus the expectation of the extra time is bounded above by s times $b''(\theta)\max_i(\theta-\varphi_i)^2/(\min_i I(\theta,\varphi_i))^2+B$, which is easily seen to be bounded uniformly in θ by virtue of assumption III. Adding this bound in the derivations of (38) and (47) is the only change in the proof required to handle the "memoryless" $\hat N$. Once again, the discussion does not apply to the case $s=2$ where the two versions of $\hat N$ are identical.

REMARK 3. Another alternative to the error probability formulation (1) is

$$(1)'' \qquad P_{\theta_i}(j\text{th decision}) \leqq \alpha_{ij} \qquad (i=1,\cdots,s;\quad i=1,\cdots,k),$$

which does not restrict the probabilities of the decisions when $\theta\neq\theta_1,\cdots,\theta_s$. The theorem holds provided $\hat N$ is changed by defining $\hat N_{ij}$ as the smallest n such that

$$L_n(\hat\theta)\geqq L_n(\theta_i)+\log\gamma_{ij}^{-1}.$$

The proof goes through with obvious modifications if $J=\{1,\cdots,k\}$ for $\theta\in(\underline{\theta},\bar\theta)-\bigcup_i\bar V_i$ ($\bar V_i$ contains θ_i) and $J=\{j\,|\,\alpha_{mj}=1\}$ for $\theta\in\bar V_m$ (the chief modifications being the use of (11) rather than (10) in the first case and the obvious redefining of $I(\theta)$, I, and D). Such an $\hat N$ may differ markedly from those defined above, or may be identical, *depending on the* α_{ij}*'s*. One can, for example, make a decision incorrect for θ_1,θ_2 when $\hat\theta\in(\theta_1,\theta_2)$. In the case $s=2$, one can "reject the hypothesis $\theta\leqq\theta_1$" in favor of "$\theta\geqq\theta_2$" when $\hat\theta<\theta_1$! However, it is easily seen that the two versions of $\hat N$ are identical if $\bar\alpha$ and $\underline{\alpha}$ are equal (or sufficiently close) and if (1)$''$ is equivalent to the restriction of (1) to $\theta=\theta_1,\cdots,\theta_s$. An example of this is Wong's formulation of the $s=2$ case, wherein $\bar\alpha=\underline{\alpha}$. In this case, or anytime the new $\hat N$ is identical with the original formulation, one obtains a "bonus": the theorem holds with the class of competitors, (N,d), enlarged to include all those tests satisfying the restriction of (1) to $\theta=\theta_1,\cdots,\theta_s$.

REMARK 4. Several refinements in the derivation of bounds can be made in case $\bar\alpha=\underline{\alpha}$. The easiest one is based on the observation that the concavity of $L_n(\cdot)$ implies that for fixed θ and j only two values of i are needed to determine $\max_i N_{ij}$: those corresponding to the closest φ_i to the right of θ and the closest φ_i to the left. Thus in relations like (32) and (33), and throughout the proof, one need consider only (one or) two values of i for each $j\in J$.

Another refinement of the proof of the theorem leads to the result

THEOREM 2. *Under assumptions* I *and* II, *for every* $\theta\in(\underline{\theta},\bar\theta)$ *there is an* M_θ *such that*

$$(48)\qquad I(\theta)(E_\theta\hat N-n(\theta))\leqq\log\log\underline{\alpha}^{-1}+M_\theta\cdot(\log\log\underline{\alpha}^{-1})^{\frac12},$$

provided $\bar\alpha=\underline{\alpha}$.

PROOF. The basic idea, which can be applied also in the general case where $\bar\alpha\neq\underline{\alpha}$, is to consider an N_θ' defined like N_θ, but with α_{ij}'s in place of γ_{ij}'s. The derivation of (37), applied with N_θ' in place of N_θ, leads to a bound M_θ (r equals one), omitting the factor $\log\log\underline{\alpha}^{-1}$. The same result holds for the "memoryless" version of N_θ' if the M_θ's are increased by a suitable constant M, by virtue of the argument in Remark 2. Hence, (48) follows from

$$I(\theta)E_\theta(N_\theta - N_\theta') \leqq \log\log \underline{\alpha}^{-1} + M_\theta \cdot (\log\log \underline{\alpha}^{-1})^{\frac{1}{2}}\,, \tag{49}$$

where N_θ, N_θ' denote memoryless versions. To derive (49), first consider the expected wait after N_θ' stops for (5) to hold *for some j attaining the maximum in the definition of* $I(\theta)$, (7). Either the wait is zero or else, for some θ_m with $I(\theta, \theta_m) < I(\theta)$,

$$L_{N_\theta'}(\theta) - L_{N_\theta'}(\theta_m) \geqq \log \underline{\alpha}^{-1} \tag{50}$$

in order that (5) holds for a non-maximizing j'. Choose a maximizing j and note that (5) holds for this j as soon as $L_n(\theta)$ increases relative to each $L_n(\varphi_i)$ for this j by Y, where

$$\begin{aligned} Y = \textstyle\sum_i [\log \underline{\alpha}^{-1} - (L_{N_\theta'}(\theta) - L_{N_\theta'}(\varphi_i))]^+ &\leqq \textstyle\sum_{i,m} [L_{N_\theta'}(\varphi_i) - L_{N_\theta'}(\theta_m)]^+ \\ &\leqq \textstyle\sum_{i,m} \sup_{n \geqq 1} (L_n(\varphi_i) - L_n(\theta_m))^+ \end{aligned}$$

using (50), with m ranging over those values $1, \cdots, s$ such that $I(\theta, \theta_m) < I(\theta)$. Since $E_\theta(L_1(\varphi_i) - L_n(\theta_m))$ equals $I(\theta, \theta_m) - I(\theta, \varphi_i)$, which is negative, Kingman's bound and an argument like the one in Remark 2 show that the expected wait for the required increase in bounded by M_θ' (say) independent of $\underline{\alpha}$. To prove (49), then, it evidently suffices to show that after (5) holds for a maximizing j the expected wait for N_θ satisfies a bound like (49). Theorem 3.3 of [5] suffices for this task, since the minimum value of $I(\theta, \varphi_i)$ under consideration is $I(\theta)$ and $\log(D \log \underline{\alpha}^{-1})$ is the required increase in $L_n(\theta)$ relative to each $L_n(\varphi_i)$. In fact the square-root term in (48) and (49) can be replaced by a constant in case $I(\theta)$ is attained by a unique j.

The following result generalizes a claim made in the introduction.

THEOREM 3. *Under assumptions* I, II, III′, *there is an $M^*(r)$ such that any procedure satisfying* (1) *has expected sample sizes for all $\theta \in (\underline{\theta}, \bar{\theta})$ which are larger than the likelihood ratio test with $\gamma_{ij} = M^*(r)\alpha_{ij}$ when $\alpha_{ij} < 1$, $= 1$ when $\alpha_{ij} = 1$.*

PROOF. Define N_θ' like N_θ in the proof of Theorem 1, but with α_{ij}'s in place of γ_{ij}'s. It will first be shown that there is an $M(r)$ such that

$$E_\theta N_\theta' - n(\theta) \leqq M(r) \tag{51}$$

for all $\theta \in (\underline{\theta}, \bar{\theta})$. Let $V_1, \cdots, V_s$ be the deleted neighborhoods of $\theta_1, \cdots, \theta_s$ chosen in the proof of Theorem 1. The derivation of (38), with N_θ' in place of N_θ, leads to (51) for $\theta \in (\underline{\theta}, \bar{\theta}) - \bigcup_i V_i$. For $\theta \in \bigcup_i V_i$, the argument used to derive (47) applies with $\log \underline{\gamma}^{-1}$ replaced by $\log \underline{\alpha}^{-1}$, $\hat{N}$ replaced by N_θ', and other obvious changes to yield a bound of the form (51) for $\theta \in \bigcup_i V_i$, hence for all $\theta \in (\underline{\theta}, \bar{\theta})$.

Now let N_θ'' denote the modification of N_θ' obtained by replacing the α_{ij}'s < 1 by $M^*(r)\alpha_{ij}$. The theorem follows from (51) if $M^*(r)$ can be chosen so that

$$E_\theta(N_\theta' - N_\theta'') > M(r) \qquad \text{for all} \quad \theta \in (\underline{\theta}, \bar{\theta})\,. \tag{52}$$

After N_θ'' stops, the expected wait for N_θ' to stop is at least the time for $L_n(\theta)$ to increase relative to a certain θ_m (say, the smallest θ_m involved in an N_{ij}'' determining N_θ'') by $\log M^*(r) - W$, where W is the excess over the boundary for $L_n(\theta) - L_n(\theta_m)$. Given W, the conditional expectation of the wait is by Wald's equation at least $(\log M^*(r) - W)/\max_m I(\theta, \theta_m)$. Hence

$$E_\theta(N_\theta' - N_\theta'') \geqq (\log M^*(r) - E_\theta W)/\max_m I(\theta, \theta_m) . \tag{53}$$

Since $I(\cdot, \theta_m)$ is continuous on the natural parameter space, $\max_m I(\cdot, \theta_m)$ is bounded above on $(\underline{\theta}, \bar{\theta})$ under assumption III′. Therefore, the proof is complete once it is shown that $E_\theta W$ is bounded above uniformly in θ, since $M^*(r)$ can then be chosen large enough to make the right-hand side of (53) larger than $M(r)$. Applying Theorem 1 of [6],

$$E_\theta W \leqq k \sum_m \frac{(\theta - \theta_m)^2 b''(\theta)}{I(\theta, \theta_m)} , \tag{54}$$

where only θ_m's occurring as φ_i's in the definition of N_θ'' are included in the sum. This proviso insures that $I(\theta, \theta_m)$ is bounded below uniformly since $\theta \in V_m$ implies that θ_m is excluded from the sum. The factor of k in (54) takes into account the possibility that a θ_m may occur as a φ_i for more than one value of j, with different α_{ij}'s. Since the numerators $(\theta - \theta_m)^2 b''(\theta)$ in (54) are bounded above by virtue of assumption III′, $E_\theta W$ is bounded above uniformly and the proof is complete.

3. Open-ended tests and error probability approximations. The sequential likelihood ratio test for the case $s = 2$ can be described as follows. Stop and reject $\theta \leqq \theta_1$ the first time

$$L_n(\hat{\theta}) \geqq L_n(\theta_1) + \log \gamma_1^{-1} \qquad \text{and} \qquad \hat{\theta} > \theta_1 . \tag{55}$$

and stop and reject $\theta \geqq \theta_2$ the first time

$$L_n(\hat{\theta}) \geqq L_n(\theta_2) + \log \gamma_2^{-1} \qquad \text{and} \qquad \hat{\theta} < \theta_2 , \tag{56}$$

where γ_1, γ_2 are chosen to yield error probabilities less than prescribed α, β. The criterion (55) for rejecting $\theta \leqq \theta_1$ exemplifies what H. Robbins has called an "open-ended test" of the hypothesis $\theta \leqq \theta_1$ against the alternative $\theta > \theta_1$. The prototype for open-ended tests is the one-sided SPRT "reject f_0 when $f_0/f_1 \leqq A$". Open-ended sequential likelihood ratio tests are studied in [7]. It should be noted that, just as an ordinary SPRT is obtained by performing two one-sided SPRT's simultaneously, so also the sequential likelihood ratio test for $s = 2$ is obtained by performing two open-ended tests; namely, the ones defined by (55) and (56). In fact, the general k-decision likelihood ratio tests of Section 2 are equivalent to performing simultaneously k open-ended tests, $\max_i \hat{N}_{ij}$ for $j = 1, \cdots, k$, and stopping the first time any one of them stops.

This reduction to open-ended tests is important for two reasons. First, it is a useful way to set up the application of a sequential likelihood ratio test and to compute stopping boundaries. Second, it facilitates the approximation of error probabilities. In the case of an SPRT, Wald's approximations of the error probabilities differ only by factors of $(1-\alpha)$ and $(1-\beta)$ from the error probability approximations for each of the one-sided SPRT's involved. This suggests bounding the error probabilities of sequential likelihood ratio tests in terms of open-ended tests. From (8), there is an upper bound, M, on the number of observations. Thus, letting N be the extended stopping variable defined by (55), the error probability under θ_1 of the test defined by (55) and (56) is at most $P_{\theta_1}(N \leqq M)$. An approach to the problem of approximating probabilities of this type is the following.

Assume that $\bar{\theta}$ belongs to the natural parameter space and $I(\bar{\theta}, \theta_1) < \log \gamma_1^{-1}$. Let $m \geqq 1$ be the largest integer such that $mI(\bar{\theta}, \theta_1) \leqq \log \gamma_1^{-1}$. For $n > m$, define $\theta_n{}^*$ as the solution in $\theta \in (\theta_1, \bar{\theta})$ of $nI(\theta, \theta_1) = \log \gamma_1^{-1}$. Let $N(\theta)$ be the smallest n (or ∞ if there is no n) such that

$$L_n(\theta) \geqq L_n(\theta_1) + \log \gamma_1^{-1}\,. \tag{57}$$

Note that (55) is equivalent to

$$S_n \geqq \inf_{\theta_1 < \theta \leqq \bar{\theta}} \left[\frac{\log \gamma_1^{-1}}{\theta - \theta_1} + n\,\frac{b(\theta) - b(\theta_1)}{\theta - \theta_1}\right], \tag{58}$$

and for $n > m$ the infimum is attained at $\theta_n{}^*$. From this and the fact that $N \leqq N(\theta_n{}^*)$ by definition, it follows that $N = n > m$ only if $N(\theta_n{}^*) = n$. Therefore,

$$P_{\theta_1}(N = n) \leqq P_{\theta_1}(N(\theta_n{}^*) = n) \qquad \text{for} \quad n > m\,. \tag{59}$$

If $n \leqq m$, the infimum in (58) is attained at $\bar{\theta}$, so that

$$P_{\theta_1}(N \leqq m) = P_{\theta_1}(N(\bar{\theta}) \leqq m)\,. \tag{60}$$

By the cancellation of densities argument used for Wald's upper bound on SPRT error probabilities [10]

$$P_{\theta_1}(N(\theta) = n) \leqq \gamma_1 P_\theta(N(\theta) = n) \qquad \text{for all} \quad n\,. \tag{61}$$

By (59)—(61)

$$P_{\theta_1}(N \leqq M) \leqq \gamma_1(P_{\bar{\theta}}(N(\bar{\theta}) \leqq m) + \textstyle\sum_{n=m+1}^{M} P_{\theta_n{}^*}(N(\theta_n{}^*) = n)\,. \tag{62}$$

As shown in [10], when $\theta_n{}^*$ is true and γ_1 is small, $N(\theta_n{}^*)$ is approximately normally distributed with mean $\log \gamma_1^{-1}/I(\theta_n{}^*, \theta_1) = n$ and variance

$$\frac{(\log \gamma_1^{-1}) \operatorname{Var}_{\theta_n{}^*}(\log (f_{\theta_n{}^*}(X)/f_{\theta_1}(X)))}{(E_{\theta_n{}^*} \log (f_{\theta_n{}^*}(X)/f_{\theta_1}(X)))^3} = \frac{\log \gamma_1^{-1} \cdot (\theta_n{}^* - \theta_1)^2 b''(\theta_n{}^*)}{(I(\theta_n{}^*, \theta_1))^3}\,.$$

This suggests the approximations

$$P_{\theta_n{}^*}(N(\theta_n{}^*) = n) \approx \frac{(I(\theta_n{}^*, \theta_1))^{\frac{3}{2}}}{(\theta_n{}^* - \theta_1)(2\pi b''(\theta_n{}^*) \log \gamma_1^{-1})^{\frac{1}{2}}} \tag{63}$$

and

$$P_{\bar{\theta}}(N(\bar{\theta}) \leqq m) \approx \tfrac{1}{2}\gamma_1\,. \tag{64}$$

Regarding n as a continuous variable on the interval from $(\log \gamma_1^{-1})/I(\bar{\theta}, \theta_1)$ to M, with $nI(\theta_n^*, \theta_1) \equiv \log \gamma_1^{-1}$, one obtains from (63) and (64) the following approximation to the upper bound, (62),

$$(65) \qquad P_{\theta_1}(N \leqq M) \lesssim \tfrac{1}{2}\gamma_1 + \gamma_1 \int_{(\log \gamma_1^{-1})/I(\bar{\theta}, \theta_1)}^{M} \frac{(I(\theta_n^*, \theta_1))^{\frac{3}{2}}}{(\theta_n^* - \theta_1)(2\pi b''(\theta_n^*) \log \gamma_1^{-1})^{\frac{1}{2}}} \, dn \, .$$

Making the change of variable $\theta = \theta_n^*$ and differentiating the relation defining θ_n^*,

$$(\theta_n^* - \theta_1) b''(\theta_n^*) \frac{d\theta_n^*}{dn} = -\frac{\log \gamma_1^{-1}}{n^2} = -\frac{(I(\theta_n^*, \theta_1))^2}{\log \gamma_1^{-1}}$$

and hence

$$(66) \qquad P_{\theta_1}(N \leqq M) \lesssim \tfrac{1}{2}\gamma_1 + \gamma_1 (\log \gamma_1^{-1})^{\frac{1}{2}} \int_{\theta_M^*}^{\bar{\theta}} \left(\frac{b''(\theta)}{2\pi I(\theta, \theta_1)}\right)^{\frac{1}{2}} d\theta \, .$$

In the case of a normal mean θ (variance one), (66) becomes

$$(67) \qquad P_{\theta_1}(\text{reject } \theta \leqq \theta_1) \lesssim \gamma_1 \left[\tfrac{1}{2} + \pi^{-1} (\log \gamma_1^{-1})^{\frac{1}{2}} \log \frac{\bar{\theta} - \theta_1}{\theta_M^* - \theta_1}\right] .$$

It is shown in [7] that inequality does hold in (67). The numerical examples in that paper indicate that choosing $\bar{\theta}$ equal to or not much larger than θ_2 significantly reduces the error probability. In case $s > 2$, approximations similar to (66) can be derived by considering, for example, the probability under θ_1 that (55) holds for some n in a restricted range $[M_1, M_2]$ and transforming sums like the one in (62) into integrals like the one in (66).

Acknowledgment. The author wishes to thank the referee for helpful suggestions.

REFERENCES

[1] FARRELL, R. H. (1964). Limit theorems for stopped random walks. *Ann. Math. Statist.* **35** 1332–1343.

[2] HOEFFDING, W. (1953). A lower bound for the average sample number of a sequential test. *Ann. Math. Statist.* **24** 127–130.

[3] HOEFFDING, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.

[4] KINGMAN, J. F. C. (1962). Some inequalities for the queue GI/G/1. *Biometrika* **49** 315–324.

[5] LORDEN, G. (1967). Integrated risk of asymptotically Bayes sequential tests. *Ann. Math. Statist.* **38** 1399–1422.

[6] LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527.

[7] LORDEN, G. (1970). Open-ended tests for Koopman–Darmois families. Submitted to *Ann. Math. Statist.*

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

[9] SCHWARZ, G. (1969) A second-order approximation to optimal sampling regions. *Ann. Math. Statist.* **40** 313–314.

[10] WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

[11] WONG, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *Ann. Math. Statist.* **39** 1244–1263.

ALFRED P. SLOAN LABORATORY OF
MATHEMATICS AND PHYSICS
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASEDENA, CALIFORNIA 91109