The Annals of Statistics
1973, Vol. 1, No. 4, 633–643

OPEN-ENDED TESTS FOR KOOPMAN-DARMOIS FAMILIES

BY GARY LORDEN

California Institute of Technology

The generalized likelihood ratio is used to define a stopping rule for rejecting the null hypothesis $\theta = \theta_0$ in favor of $\theta > \theta_0$. Subject to a bound α on the probability of ever stopping in case $\theta = \theta_0$, the expected sample sizes for $\theta > \theta_0$ are minimized within a multiple of $\log \log \alpha^{-1}$, the multiple depending on θ. An heuristic bound on the error probability of a likelihood ratio procedure is derived and verified in the case of a normal mean by consideration of a Wiener process. Useful lower bounds on the small-sample efficiency in the normal case are thereby obtained.

1. Introduction. In his book on sequential analysis (1947), Wald discussed two ways of modifying the sequential probability ratio test (SPRT) to test a simple null hypothesis against a composite alternative. One way is to replace the likelihood ratio used in the SPRT by a weighted likelihood ratio, using a suitably chosen weight function on the alternative hypothesis. The other is to employ the generalized likelihood ratio of classical fixed-sample theory, dividing the maximum likelihood in the alternative by the likelihood for the simple hypothesis. Wald pointed out that the weighted-likelihood-ratio approach has the advantage that an upper bound on the Type I error probability can be obtained in exactly the same way as for the SPRT. Recently Robbins (1970) has exploited this fact by developing elegant methods for obtaining or at least approximating stopping boundaries for "open-ended tests," which, like the one-sided SPRT, continue sampling indefinitely (with prescribed probability) when the null hypothesis is true and stop only if the alternative is to be accepted. His methods are most effective in the case of testing a normal mean, although less sharp estimates are obtained for other cases.

The present paper is an investigation of the generalized likelihood ratio approach to the problem of open-ended tests for Koopman–Darmois families. In this context, the approach leads to easily computed procedures, as was shown in Schwarz (1962). Since the tests are equivalent to simultaneous one-sided SPRT's, it is easy to obtain an upper bound on expected sample sizes. The heart of the investigation concerns the problem of bounding error probabilities, the simple Wald approach not being applicable. For simultaneous one-sided SPRT's with critical value $\alpha < 1$, the estimates in Wong (1968) show that these error probabilities are $o(\alpha \log \alpha^{-1})$ as $\alpha \to 0$, but explicit bounds are not obtained. The bound used in the proof of Theorem 1 is explicit, but of order $\alpha \log \alpha^{-1}$. Nevertheless, it suffices in the proof that the expected sample sizes are minimized

Received November 1970; revised October 1972.

AMS 1970 *subject classifications.* Primary 62L10.

Key words and phrases. Likelihood ratio, sequential probability ratio test, open-ended test, asymptotic efficiency.

within a multiple of $\log \log \alpha^{-1}$, and the latter order of magnitude would not be improved even if the error probabilities were of order $\alpha(\log \alpha^{-1})^{\varepsilon}$ for a small $\varepsilon > 0$. An heuristic approximation to the error probabilities is given in Section 3, indicating that they are of order at most $\alpha(\log \alpha^{-1})^{\frac{1}{2}}$. This approximation is shown to be an upper bound in the case of a normal mean (with known variance) and numerical examples of efficiency in this case are given in Section 4.

Following Remark 2 below there is a brief discussion of the problem of testing $\theta < 0$ vs. $\theta > 0$ (say) with error probabilities bounded by α, β, respectively. Theorem 1 applies to give bounds on the expected sample sizes of the natural procedure based on two open-ended tests performed simultaneously. These bounds give asymptotic efficiency results as $\alpha, \beta \to 0$ for fixed $\theta \neq 0$ (or $\theta \to 0$ slowly). But the procedures and bounds are not of interest for the problem of the optimum behavior of $E_\theta N$ as $\theta \to 0$ for fixed α, β. This much more delicate problem involves iterated logarithm phenomena and was solved in Farrell (1964) using generalized SPRT's. Similar tests were studied in Fabian (1956).

2. Asymptotic efficiency results. Independent and identically distributed random variables $X_1, X_2, \cdots$ are observed sequentially with density

$$f_\theta(x) = \exp(\theta x - b(\theta)) \qquad \text{for some } \theta \in (\underline{\theta}, \bar{\theta})$$

with respect to a non-degenerate σ-finite measure, μ. Stopping times N (possibly randomized) will be required to satisfy

$$P_0(N < \infty) \leqq \alpha \tag{1}$$

for prescribed $\alpha \in (0, \frac{1}{3})$. (Reparameterize if necessary to shift the boundary point between null and alternative hypotheses to $\theta = 0$.) Also assume without loss of generality that $b(0) = 0$. Let $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \cdots$ and note that

$$\log \frac{f_\theta(X_1) \cdots f_\theta(X_n)}{f_0(X_1) \cdots f_0(X_n)} = \theta S_n - nb(\theta)\,,$$

so that one-sided SPRT's of f_0 against f_θ, $\theta > 0$, stop as soon as

$$S_n > \frac{\log \gamma^{-1}}{\theta} + n\frac{b(\theta)}{\theta} \tag{2}$$

for prescribed $\gamma \in (0, 1)$. 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 Koopman–Darmois family. The first derivative, $b'(\theta)$, equals $E_\theta X$ and the second, $b''(\theta)$, equals $\operatorname{Var}_\theta X$. An easy calculation shows that the information number, $E_\theta \log(f_\theta(X)/f_0(X))$, equals

$$\theta b'(\theta) - b(\theta) = I(\theta)\,, \tag{3}$$

while the variance of $\log(f_\theta(X)/f_0(X))$ under θ equals $\theta^2 b''(\theta)$.

Define a likelihood ratio open-ended test of $\theta = 0$ vs. $\bar{\theta} > \theta \geqq \theta_1 > 0$ as a stopping time, $N(\theta_1, \gamma)$: the smallest $n \geqq 1$ (or ∞ if there is no n) such that

$$S_n > \inf_{\theta_1 \leq \theta \leq \bar{\theta}} \left[\frac{\log \gamma^{-1}}{\theta} + n \frac{b(\theta)}{\theta} \right]. \tag{4}$$

For the alternative $\theta \leqq \theta_2 < 0$, define $N(\theta_2, \gamma)$ similarly.

Although the infimum in (4) is readily computed in some cases, e.g. that of a normal mean, it is simpler in many cases such as the normal scale parameter and negative exponential distribution to formulate the critical inequality in terms of $\bar{X}_n \doteq S_n/n$ and n, as in Schwarz (1962). First note that (4) is equivalent to

$$\sup_{\theta_1 \leq \theta < \bar{\theta}} [\theta S_n - nb(\theta)] > \log \gamma^{-1}. \tag{5}$$

In case $b'(\theta_1) \leqq \bar{X}_n < b'(\bar{\theta})$, the supremum is attained at $q(\bar{X}_n)$, where q is the inverse of the increasing function b'. In this case (5) is equivalent to

$$\bar{X}_n q(\bar{X}_n) - b(q(\bar{X}_n)) > \frac{\log \gamma^{-1}}{n}.$$

In case $\bar{X}_n < b'(\theta_1)$, the supremum is attained at θ_1; and if $\bar{X}_n \geqq b'(\bar{\theta})$, the supremum is approached as $\theta \to \bar{\theta}$ (and attained at $\bar{\theta}$ if the latter belongs to the natural parameter space).

THEOREM 1. *If $\hat{N} = N(\theta_1, \gamma)$ with $\gamma = \alpha/3(I(\theta_1)^{-1} + 1)^2 \log \alpha^{-1}$, then $\hat{N}$ satisfies* (1) *and*

$$E_\theta \hat{N} \leqq \frac{\log \alpha^{-1} + \log \log \alpha^{-1}}{I(\theta)} + 2 \frac{\log (3^{\frac{1}{2}}(I(\theta_1)^{-1} + 1))}{I(\theta)} + \frac{\theta^2 b''(\theta)}{(I(\theta))^2} + 1 \tag{6}$$

for all $\theta \in [\theta_1, \bar{\theta})$. If N satisfies (1), *then*

$$E_\theta N \geqq \frac{\log \alpha^{-1}}{I(\theta)} \qquad \textit{for } \theta \neq 0. \tag{7}$$

PROOF. The lower bound (7) is similar to Wald's lower bound on average sample numbers ((1947) page 197). Using Wald's equation for a randomly stopped sum (see Farrell (1964) for a proof) and Jensen's inequality,

$$\begin{aligned} I(\theta)E_\theta N &= E_\theta \left(-\log \frac{f_0(X_1) \cdots f_0(X_N)}{f_\theta(X_1) \cdots f_\theta(X_N)} \right) \\ &\geqq -\log E_\theta \frac{f_0(X_1) \cdots f_0(X_N)}{f_\theta(X_1) \cdots f_\theta(X_N)} = -\log P_0(N < \infty) \geqq \log \alpha^{-1}. \end{aligned}$$

Inequality (6) follows at once from the fact that

$$E_\theta N(\theta_1, \gamma) \leqq \frac{\log \gamma^{-1}}{I(\theta)} + \frac{\theta^2 b''(\theta)}{(I(\theta))^2} + 1 \qquad \text{for } \theta \in [\theta_1, \bar{\theta}), \tag{8}$$

by virtue of the choice of γ in the theorem. Relation (8) holds because $E_\theta N(\theta_1, \gamma)$ is no larger than the expected time until (2) holds, which is bounded by the right-hand side of (8), using Wald's equation and Theorem 1 of Lorden (1970). (The latter states that the expected "excess over the boundary" is at most the second moment of $\log (f_\theta(X)/f_0(X))$ divided by the first moment.)

The remainder of the proof is concerned with showing that the choice of γ is sufficient for $\hat{N}$ to satisfy (1). To determine the infimum in (4), note that the derivative with respect to θ of the bracketed quantity is

$$-\frac{\log \gamma^{-1}}{\theta^2} + n\frac{\theta b'(\theta) - b(\theta)}{\theta^2} = \theta^{-2}(nI(\theta) - \log \gamma^{-1}) , \tag{9}$$

which is nonnegative on $[\theta_1, \bar{\theta})$ if $nI(\theta_1) \geqq \log \gamma^{-1}$. Thus for $n \geqq \log \gamma^{-1}/I(\theta_1)$ the infimum is attained at θ_1 and hence

$$\begin{aligned} &P_0(\infty > N \geqq \log \gamma^{-1}/I(\theta)) \\ &\quad \leqq P_0\left(S_n > \frac{\log \gamma^{-1}}{\theta_1} + n\frac{b(\theta_1)}{\theta_1} \text{ for some } n \geqq 1\right) \leqq \gamma , \end{aligned} \tag{10}$$

the last inequality coming from Wald's upper bound (1947) on SPRT error probabilities.

For fixed $n < \log \gamma^{-1}/(\theta_1)$, a sequence $\{\theta_k\}$ can be chosen along which the infimum in (4) is approached, and the probability that strict inequality holds in (4) is the limit of the probabilities that (2) holds for $\theta = \theta_k$. The latter probabilities are at most γ, again by Wald's upper bound, and thus

$$P_0\left(S_n > \inf_{\theta_1 \leq \theta < \bar{\theta}}\left[\frac{\log \gamma^{-1}}{\theta} + n\frac{b(\theta)}{\theta}\right]\right) \leqq \gamma , \tag{11}$$

or, equivalently,

$$P_0(\hat{N} = n) \leqq \gamma . \tag{12}$$

By (10) and (12),

$$P_0(\hat{N} < \infty) < \gamma\left(\frac{\log \gamma^{-1}}{I(\theta_1)} + 1\right) < \gamma(\log \gamma^{-1})(I(\theta_1)^{-1} + 1) , \tag{13}$$

since $\log \gamma^{-1} > \log \alpha^{-1} > 1$.

The choice $\gamma = \alpha/3(I(\theta_1)^{-1} + 1)^2 \log \alpha^{-1}$ suffices to make the extreme right member of (13) less than α, since

$$\begin{aligned} &\frac{\alpha \log (3(I(\theta_1)^{-1} + 1)^2\alpha^{-1} \log \alpha^{-1})}{3(I(\theta_1)^{-1} + 1) \log \alpha^{-1}} \\ &\quad \leqq \frac{2\alpha \log (3^{\frac{1}{2}}(I(\theta_1)^{-1} + 1))}{3(I(\theta_1)^{-1} + 1)} + \frac{\frac{3}{2}\alpha \log \alpha^{-1}}{3(I(\theta_1)^{-1} + 1) \log \alpha^{-1}} < \tfrac{1}{2}\alpha + \tfrac{1}{2}\alpha = \alpha , \end{aligned}$$

using the estimates $\log x < x^{\frac{1}{2}}$ and $\log x/x \leqq e^{-1}$ for $x \geqq 1$.

REMARK 1. By letting $\theta_1 \to 0$ as $\alpha \to 0$ (e.g. set $I(\theta_1)^{-1} = \log \alpha^{-1}$) one can obtain a class of procedures $\hat{N}(\alpha) = N(\theta_1(\alpha), \gamma(\alpha))$ such that for all $\theta > 0$ there is an $M(\theta)$ such that

$$E_\theta \hat{N}(\alpha) - \frac{\log \alpha^{-1}}{I(\theta)} \leqq 3\frac{\log \log \alpha^{-1}}{I(\theta)} + M(\theta) \tag{14}$$

for all $\alpha < \exp(-1/I(\theta))$. Since $0 = b(0) = b(\theta) - \theta b'(\theta) + \frac{1}{2}\theta^2 b''(\xi)$, for some

$0 < \xi < \theta$, and b'' is continuous, $I(\theta) = \theta b'(\theta) - b(\theta) \sim \frac{1}{2}b''(0)\theta^2$ as $\theta \to 0$, so that the choice of θ_1 just made results in $\theta_1 \sim (\log \alpha^{-1})^{-\frac{1}{2}} \cdot$ (const.) as $\alpha \to 0$. Other negative powers of $\log \alpha^{-1}$ yield similar results, and lead to bounds of the form (14) with 3 replaced by any number larger than one.

REMARK 2. The open-ended test $N(\theta_1, \gamma)$ has $E_\theta N = \infty$ for $\theta > 0$ closer (in terms of information numbers) to 0 than to θ_1. In some applications this may be quite acceptable, and the interval of positive θ's where this happens can be regarded as a kind of "indifference zone." However, if it is desired to consider only procedures which satisfy $E_\theta N < \infty$ for all $\theta > 0$, then one can simply use combinations of the form $N^* = \min_k N(\theta_k, \gamma_k)$ where $\theta_k \to 0$ and (say) $\gamma_k = \alpha 2^{-k}/3(I(\theta_k)^{-1} + 1)^2 \log(\alpha^{-1}2^k)$, so that $P_0(N^* < \infty) \leqq \alpha(2^{-1} + 2^{-2} + 2^{-3} + \cdots) = \alpha$. Then (6) holds for N^* if α is replaced by $\alpha/2$ and similar inequalities hold for $\theta \in (\theta_{k+1}, \theta_k)$, $k = 1, 2, \cdots$. Note that bounds of the form (14) hold for N^* also. The stopping boundaries for N^* can be calculated "piecemeal" as n increases, passing from $N(\theta_k, \gamma_k)$ to $N(\theta_{k+1}, \gamma_{k+1})$ as soon as

$$\inf_{\theta_{k+1} \leqq \theta < \theta_k} \left[\frac{\log \gamma_{k+1}^{-1}}{\theta} + n\,\frac{b(\theta)}{\theta}\right] < \frac{\log \gamma_k^{-1}}{\theta_k} + n\,\frac{b(\theta_k)}{\theta_k}\,.$$

Suppose it is desired to test $\theta < 0$ vs. $\theta > 0$ with error probabilities less than α, β, respectively. An obvious procedure is to stop at $\tilde{N} = \min(N_1^*, N_2^*)$ where $N_1^* = \min_k N(\theta_k, \gamma_k)$ with $\theta_k \downarrow 0$ and $N_2^* = \min_k N(-\theta_k, \gamma_k')$, where γ_k' is defined using β in place of α. If $\theta < 0$ or $\theta > 0$ is chosen according as N_2^*, (resp.) N_1^* stops first, then clearly

$$P_\theta(\text{stop and decide } \theta > 0) \leqq \alpha \qquad \text{for all } \theta \leqq 0 \tag{15}$$

and

$$P_\theta(\text{stop and decide } \theta < 0) \leqq \beta \qquad \text{for all } \theta \geqq 0\,. \tag{16}$$

Upper bounds of the form (14) clearly hold for $E_\theta \tilde{N}$ for $\theta \neq 0$ and are to be compared with the lower bound (Lemma 1 of Lorden (1972))

$$I(\theta)E_\theta N \geqq (1-\alpha)\log \alpha^{-1} - \log 2 \geqq \log \alpha^{-1} - \log 2 - e^{-1} \tag{17}$$

for $\theta > 0$ (and a similar bound for $\theta < 0$) which is satisfied by the stopping time N of any test satisfying (15) and (16). Omitting $\theta = 0$ in (15) and (16) is only an apparent weakening of these requirements, because it is easy to show by the dominated convergence theorem that $P_\theta(N \leqq n)$ is continuous in θ for each n, and, letting $n \to \infty$, that $P_\theta(N < \infty)$ is lower-semi-continuous.

3. Error probability approximations. Assume that $\bar{\theta}$ belongs to the natural parameter space and $I(\bar{\theta}) < \log \gamma^{-1}$. Let $m \geqq 1$ be the largest integer such that $mI(\bar{\theta}) \leqq \log \gamma^{-1}$, and let M be the smallest integer such that $MI(\theta_1) \geqq \log \gamma^{-1}$. For $\theta > 0$, let $N(\theta)$ be the smallest n (or ∞ if there is no n) such that (2) holds, i.e. the stopping time of the one-sided SPRT of f_0 against f_θ specified by γ. In the definition of $N(\theta_1, \gamma)$, (4), the infimum is evidently attained when $m < n < M$

at the solution, θ_n (say), of $nI(\theta) = \log \gamma^{-1}$, since the derivative, (9), changes from negative to positive at θ_n. Hence, if $N(\theta_1, \gamma) = n > m$, then $N(\theta_n) \leqq n$ and since $N(\theta_1, \gamma) \leqq N(\theta_n)$ in any case ($\theta_n \in (\theta_1, \bar{\theta})$), evidently $N(\theta_n) = n$. Thus

$$P_0(N(\theta_1, \gamma) = n) \leqq P_0(N(\theta_n) = n) \qquad \text{for } M > n > m . \tag{18}$$

If $n \leqq m$, the derivative, (9), is nonpositive on $[\theta_1, \theta)$ since $nI(\bar{\theta}) \leqq \log \gamma^{-1}$, and hence the infimum is attained at $\bar{\theta}$. Therefore,

$$P_0(N(\theta_1, \gamma) \leqq m) = P_0(N(\bar{\theta}) \leqq m) . \tag{19}$$

If $n \geqq M$, the derivative, (9), is nonnegative on $[\theta_1, \theta)$, the infimum is attained at θ_1 and, reasoning as in (18),

$$P_0(M \leqq N(\theta_1, \gamma) < \infty) \leqq P_0(M \leqq N(\theta_1) < \infty) . \tag{20}$$

For all n and $\theta > 0$

$$\begin{aligned} P_0(N(\theta) = n) &= \int_{\{N(\theta)=n\}} f_0(x_1) \cdots f_0(x_n)\, d\mu^n \\ &\leqq \int_{\{N(\theta)=n\}} \gamma f_\theta(x_1) \cdots f_\theta(x_n)\, d\mu^n = \gamma P_\theta(N(\theta) = n) \end{aligned} \tag{21}$$

since $N(\theta) = n$ only if $f_0(x_1) \cdots f_0(x_n) \leqq \gamma f_\theta(x_1) \cdots f_\theta(x_n)$.

Combining (18)—(21),

$$\begin{aligned} P_0(N(\theta_1, \gamma) < \infty) \leqq \gamma P_{\bar{\theta}}(N(\bar{\theta}) \leqq m) &+ \gamma P_{\theta_1}(M \leqq N(\theta_1) < \infty) \\ &+ \gamma \textstyle\sum_{n=m+1}^{M-1} P_{\theta_n}(N(\theta_n) = n) . \end{aligned} \tag{22}$$

As shown in Wald (1947), when θ_n is true and γ is small, $N(\theta_n)$ is approximately normally distributed with mean $\log \gamma^{-1}/I(\theta_n) = n$ and variance

$$\frac{(\log \gamma^{-1}) \operatorname{Var}_{\theta_n}(\log(f_{\theta_n}(X)/f_0(X)))}{(E_{\theta_n} \log(f_{\theta_n}(X)/f_0(X)))^3} = \frac{\log \gamma^{-1} \cdot \theta_n{}^2 b''(\theta_n)}{I(\theta_n)^3} .$$

This suggests the approximations

$$P_{\theta_n}(N(\theta_n) = n) \approx \frac{I(\theta_n)^{\frac{3}{2}}}{\theta_n(2\pi b''(\theta_n) \log \gamma^{-1})^{\frac{1}{2}}} \tag{23}$$

and, similarly,

$$P_{\bar{\theta}}(N(\bar{\theta}) \leqq m) \approx \tfrac{1}{2} \qquad \text{and} \qquad P_{\theta_1}(M \leqq N(\theta_1) < \infty) \approx \tfrac{1}{2} . \tag{24}$$

Regarding n as a continuous variable on the interval from $(\log \gamma^{-1})/I(\bar{\theta}) \approx m$, to $(\log \gamma^{-1})/I(\theta_1) \approx M$, with $nI(\theta_n) \equiv \log \gamma^{-1}$, (23) and (24) yield the following modification of (22).

$$P_0(N < \infty) \lessapprox \gamma + \gamma \int_{\log \gamma^{-1}/I(\bar{\theta})}^{\log \gamma^{-1}/I(\theta_1)} \frac{I(\theta_n)^{\frac{3}{2}}}{\theta_n(2\pi b''(\theta_n) \log \gamma^{-1})^{\frac{1}{2}}}\, dn . \tag{25}$$

Change variables by letting $\theta = \theta_n$ and obtain from the relation $I(\theta_n) = (\log \gamma^{-1})/n$

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

so that

$$P_0(N < \infty) \lessapprox \gamma + \gamma(\log \gamma^{-1})^{\frac{1}{2}} \int_{\theta_1}^{\bar\theta} \left(\frac{b''(\theta)}{2\pi I(\theta)}\right)^{\frac{1}{2}} d\theta \,. \tag{26}$$

In case θ is the mean of a normal distribution with variance one, (26) becomes

$$P_0(N < \infty) \lessapprox \gamma\left[1 + \left(\frac{\log \gamma^{-1}}{\pi}\right)^{\frac{1}{2}} \log\left(\frac{\bar\theta}{\theta_1}\right)\right]. \tag{27}$$

It will be verified in the next section that the "approximate inequality" (27) is an actual inequality in this case.

4. Testing the mean drift of a Wiener process. In this section it is verified that the approximation (26) gives an upper bound on the error probability in the case of normal distributions with known variance. Variance one is assumed for convenience. This result is obtained as a consequence of a similar result for the continuous-time analogue, a Wiener process. Specifically, we have to bound the probability that a standard Wiener process, $X(t)$, on $[0, \infty)$ crosses a continuous boundary, $h(t)$. Theorem 2 gives such a bound for the class of concave and continuously differentiable h, which includes the likelihood ratio test boundaries. The author is grateful to the referee for pointing out that the methods and results of Itô and McKean (1965) and Strassen (1967) are similar. The argument on page 34 of Itô and McKean (1965) differs from the proof of Theorem 2 primarily in the use of step function approximation and first passage time distribution rather than polygonal approximation and the distribution (29) below. This difference accounts for the improvement upon the inequality 5) on page 34 of their book, which amounts to a factor of two in our application. The deduction of (39) below from Strassen's (1967) work can be carried out with a bit of care (see Theorem 1.2 and the strengthening remarked upon following (22) of his paper), but would be inappropriate here because of the much greater refinement and complexity of his arguments, which are aimed at considerably more delicate results.

Given a standard Wiener process $X(t)$ on $[0, \infty)$ and a function $h(t)$ on $(0, \infty)$ define

$$T(h) = \inf\{t \mid X(t) > h(t)\} \,. \tag{28}$$

For the linear case, $r(t) = a + bt \, (a \geqq 0)$, $T(r)$ has a known distribution (defective if $b > 0$)

$$P(0 < T(r) < d) = \int_0^d \frac{a}{(2\pi t^3)^{\frac{1}{2}}} \exp\left(-\frac{(a + bt)^2}{2t}\right) dt \,, \tag{29}$$

which is readily derived from Doob's (1949) result

$$P(T(r) < \infty) = \exp(-2ab^+) \tag{30}$$

by the method Anderson (1960) used to solve a more complicated problem involving multiple crossings of linear boundaries. In fact, (29) can be derived as a limiting case of (5.5) in Anderson's paper. It is more instructive, however, to use Anderson's method directly. Given $X(d) = y$, the process

$$\frac{d+u}{d}\left(X\left(\frac{du}{d+u}\right) - \frac{uy}{d+u}\right), \qquad u \geqq 0,$$

is a standard Wiener process. Applying Doob's result, (30), one obtains

$$P(0 \leqq T(r) < d \mid X(d) = y) = \exp\left(-2a\left(\frac{a-y}{d} + b\right)^{+}\right). \tag{31}$$

Integrating (31) with respect to the normal distribution of $X(d)$ and simplifying leads to (29).

The following theorem gives a generalization of (29) in the form of a bound on the distribution function of $T(h)$ for concave h.

THEOREM 2. *If h is concave and piecewise continuously differentiable on* $(0, \infty)$ *and* $h(0^+) \geqq 0$, *then*

$$P(c < T(h) < d) \leqq \int_c^d \frac{h(t) - th'(t)}{(2\pi t^3)^{\frac{1}{2}}} \exp\left(-\frac{h^2(t)}{2t}\right) dt \tag{32}$$

for $0 \leqq c < d \leqq \infty$.

PROOF. Assume h is continuously differentiable, since the modification needed for the piecewise case will be clear. Consider first the case where $0 < c < d < \infty$. Given a partition $c = a_0 < a_1 < \cdots < a_n = d$ of (c, d) into n subintervals, let r_k be the linear function determined by the points $(a_{k-1}, h(a_{k-1}))$ and $(a_k, h(a_k))$, $k = 1, \cdots, n$. It is clear from the concavity of h that $h \leqq r_k$ on $(0, a_{k-1})$ and $h \geqq r_k$ on $[a_{k-1}, a_k)$. Therefore,

$$\begin{aligned} P(c \leqq T(h) < d) &= \textstyle\sum_{k=1}^n P(a_{k-1} \leqq T(h) < a_k) \\ &\leqq \textstyle\sum_{k=1}^n P(a_{k-1} \leqq T(r_k) < a_k) \\ &= \textstyle\sum_{k=1}^n \int_{a_{k-1}}^{a_k} \frac{r_k(t) - tr_k'(t)}{(2\pi t^3)^{\frac{1}{2}}} \exp\left(-\frac{r_k^2(t)}{2t}\right) dt \end{aligned} \tag{33}$$

by (29). The last summation can be written as

$$\int_c^d \frac{g_n(t) - tg_n'(t)}{(2\pi t^3)^{\frac{1}{2}}} \exp\left(-\frac{g_n^2(t)}{2t}\right) dt, \tag{34}$$

where $g_n(t) = \min(r_1(t), \cdots, r_n(t))$ and g_n' exists except at the a_k's. For each $n = 1, 2, \cdots$, choose a partition of (c, d) into n subintervals, the maximum width tending to zero as $n \to \infty$. The continuity of h and h' insure that the integrand in (34) approaches the integrand in (32) for almost every t in (c, d). Note that

$$h'(a_k) \leqq g_n'(t) \leqq h'(a_{k-1}) \qquad \text{for all } t \text{ in } (a_k, a_{k-1}). \tag{35}$$

A routine application of the bounded convergence theorem shows that the integral in (34) approaches the integral in (32), proving (32) for $0 < c < d < \infty$, and the latter restriction is easily removed using monotone convergence.

The likelihood ratio test for normally distributed observations with mean θ and variance one stops at the first n such that

$$S_n > \inf_{\theta_1 \leqq \theta < \bar{\theta}} \left[\frac{\log \gamma^{-1}}{\theta} + \tfrac{1}{2} n\theta \right], \tag{36}$$

or, equivalently,

$$\begin{aligned} S_n &> \frac{\log \gamma^{-1}}{\bar{\theta}} + \tfrac{1}{2}\bar{\theta} n && \text{for } n < 2\bar{\theta}^{-2} \log \gamma^{-1} \\ &> (2n \log \gamma^{-1})^{\frac{1}{2}} && \text{for } 2\bar{\theta}^{-2} \log \gamma^{-1} \leqq n \leqq 2\theta_1^{-2} \log \gamma^{-1} \\ &> \frac{\log \gamma^{-1}}{\theta_1} + \tfrac{1}{2}\theta_1 n && \text{for } n \geqq 2\theta_1^{-2} \log \gamma^{-1}. \end{aligned} \tag{37}$$

The probability that (37) holds for some $n = 1, 2, \cdots$ is not greater than the probability of the continuous-time analogue of (37), obtained by putting a standard Brownian motion, $X(t)$ $(t \geqq 0)$, in place of S_n and t in place of n on the right-hand side. An upper bound on the latter probability is available from Theorem 2 since the function h defined by the right-hand side of (37) (for the continuous variable t) is concave and continuously differentiable on $(0, \infty)$. Relation (31) yields the upper bound

$$\begin{aligned} &\int_0^{t_0} \frac{\log \gamma^{-1}}{(2\pi\bar{\theta}^2 t^3)^{\frac{1}{2}}} \exp\left(-\frac{(2 \log \gamma^{-1} + \bar{\theta}^2 t)^2}{8\bar{\theta}^2 t} \right) dt + \int_{t_0}^{t_1} \frac{\gamma (\log \gamma^{-1})^{\frac{1}{2}}}{2\pi^{\frac{1}{2}} t} dt \\ &\qquad + \int_{t_1}^{\infty} \frac{\log \gamma^{-1}}{(2\pi\theta_1^2 t^3)^{\frac{1}{2}}} \exp\left(-\frac{(2 \log \gamma^{-1} + \theta_1^2 t)^2}{8\theta_1^2 t} \right) dt, \end{aligned} \tag{38}$$

where $t_0 = 2\bar{\theta}^{-2} \log \gamma^{-1}$ and $t_1 = 2\theta_1^{-2} \log \gamma^{-1}$. The change of variables $u = (\theta_1/\bar{\theta})^2 t$ transforms the last term into an integral on (t_0, ∞) and the integrand is the same as in the first term. Combining these two, we obtain an integral of the type in (29), expressing the probability of ever reaching the line $\bar{\theta}^{-1} \log \gamma^{-1} + \frac{1}{2}\bar{\theta} t$. By Doob's result, this probability is γ. Using this and evaluating the second integral in (38), the bound simplifies to

$$\gamma \left[1 + \frac{(\log \gamma^{-1})^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}} \log \frac{t_1}{t_0} \right] = \gamma \left[1 + \left(\frac{\log \gamma^{-1}}{\pi} \right)^{\frac{1}{2}} \log \left(\frac{\bar{\theta}}{\theta_1} \right) \right], \tag{39}$$

which agrees with (27), the normal case version of the heuristic bound in (26).

In applying this bound to the normal case it should be noted that the upper line on the right-hand side of (37) plays no role if $\bar{\theta} \geqq (2 \log \gamma^{-1})^{\frac{1}{2}}$. Thus, all choices of $\bar{\theta} \geqq (2 \log \gamma^{-1})^{\frac{1}{2}}$ yield equivalent tests (including $\bar{\theta} = \infty$) and the bound (39) holds in this case with $\bar{\theta}$ replaced by $(2 \log \gamma^{-1})^{\frac{1}{2}}$.

For the analogous testing problem for Wiener processes, explicit lower bounds on efficiency can be derived as follows.

For a fixed value, R, of the ratio t_1/t_0, let

$$k(\gamma) = 1 + \frac{(\log \gamma^{-1})^{\frac{1}{2}} \log R}{2\pi^{\frac{1}{2}}}. \tag{40}$$

To guarantee a prescribed error probability bound, α, one should choose γ so that $\gamma k(\gamma) = \alpha$. The expected time to stop when $EX(t) = \theta t$ is at most the expected time for the standard $X(t)$ to reach $h(t) = \log \gamma^{-1}/\theta - \frac{1}{2}\theta t$,

$$ET(h) = \frac{\log \gamma^{-1}}{\frac{1}{2}\theta^2} \quad \text{for} \quad \left(\frac{2 \log \gamma^{-1}}{t_1}\right)^{\frac{1}{2}} \leqq \theta \leqq R^{\frac{1}{2}} \left(\frac{2 \log \gamma^{-1}}{t_1}\right)^{\frac{1}{2}}, \tag{41}$$

while the minimum possible expected time is

$$\frac{\log \alpha^{-1}}{\frac{1}{2}\theta^2} = \frac{\log \gamma^{-1} - \log k(\gamma)}{\frac{1}{2}\theta^2},$$

so that the efficiency, $e(\alpha)$, of the chosen procedure is at least

$$1 - \frac{\log k(\gamma)}{\log \gamma^{-1}}$$

for θ in the prescribed range.

By the choice of γ

$$\log \alpha^{-1} = \log \gamma^{-1} - \log k(\gamma) = e(\alpha) \log \gamma^{-1}.$$

Thus

$$\log\left[1 + \frac{(\log \alpha^{-1})^{\frac{1}{2}} \log R}{2\pi^{\frac{1}{2}}(e(\alpha))^{\frac{1}{2}}}\right] = \log k(\gamma) = \left(\frac{1}{e(\alpha)} - 1\right) \log \alpha^{-1}. \tag{42}$$

From (42) one can determine for a prescribed level, α, and desired efficiency, $e(\alpha)$, the time span R over which these two can be guaranteed by the appropriate maximum likelihood procedure. Results for illustrative cases are given in the following table.

TABLE 1

Values of R for prescribed α and $e(\alpha)$

$e(\alpha)$	$\alpha = .05$	$\alpha = .01$
.80	7.7	24.4
.75	20.9	183
.70	87.7	5250
.65	760	2.13×10^6
.60	24,300	2.60×10^{11}

As in Theorem 1, the maximum likelihood procedures are here compared with the optimum procedures chosen separately for each parameter value.

To illustrate the performance of a particular procedure over a broad range of θ values, the choices $\theta_1 = .1$, $\bar{\theta} = 1.423$, and $\gamma = 8/7000$ were made, which yield an error probability $\alpha \leqq .0056$. This choice was made to facilitate comparison with the example in Section 4 of Robbins (1970). The minimum $E_\theta N$ subject to this error probability bound at $\theta = 0$ is $10.4/\theta^2$ for $\theta > 0$. The following upper bounds on $E_\theta \hat{N}$ for the likelihood ratio procedure were obtained in the Wiener process and (discrete) normal cases, the latter using the usual bound on excess over the boundary (Wald (1947)).

		Likelihood ratio test	
θ	$\min E_\theta N$	Wiener process	normal variables
.1	1037	1355	1363
.3	115	151	154
1	10.4	13.5	14.8
2	2.6	3.7	4.7

Reducing θ_1 to .01, while reducing γ to keep $\alpha \leqq .0056$, results in only an 8% increase in the bounds on $E_\theta \hat{N}$ for the Wiener process and the differences between these and the normal case bounds are unchanged. The resulting bound in the normal case for $\theta = .01$ is 147,500, compared with a min $E_\theta N$ of 103,700.

Using h^* to denote the continuous version of the test boundary in (37), with t_0 and t_1 as in (38), Theorem 2 can be used to show that

$$P(X(t) > h^*(t) \text{ for some } t \leqq t_1) \leqq \gamma\left[\frac{1}{2} + \frac{(\log \gamma^{-1})^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}} \log \frac{t_1}{t_0}\right].$$

This estimate is useful when open-ended tests are truncated, as in Lorden (1972), where k-decision procedures are presented as simultaneous open-ended tests.

REFERENCES

ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

FABIAN, V. (1956). A decision function. *Czechoslovak Math. J.* **6** 31–45.

FARRELL, R. H. (1964). Asymptotic behavior of expected sample size in certain one sided tests. *Ann. Math. Statist* **35** 36–72.

ITÔ, K. and MCKEAN, H. P., JR. (1965). *Diffusion Processes and their Sample Paths.* Academic Press, New York.

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

LORDEN, G. (1972). Likelihood ratio tests for sequential k-decision problems. *Ann. Math. Statist.* **43** 1412–1427.

ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397–1409.

ROBBINS, H. and SIEGMUND, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.* **41** 1410–1429.

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

STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** 315–343.

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

WONG, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *Ann. Math. Statist.* **39** 1244–1236.

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