The Annals of Statistics
1975, Vol. 3, No. 6, 1267–1282

APPROXIMATIONS TO THE EXPECTED SAMPLE SIZE OF CERTAIN SEQUENTIAL TESTS

BY M. POLLAK[1] AND D. SIEGMUND

Columbia University

This paper presents asymptotic formulae, lower and upper bounds for the expected sample size of certain sequential tests of the parameter of an exponential family of distributions. The tests involved are tests of power one based on mixture-type stopping rules and tests for the detecting of change in the underlying distribution. Analysis for incorrect assumptions of the underlying distribution yields asymptotic formulae for such cases, showing robustness of the original formulae. Monte Carlo results indicate the validity of asymptotic formulae for sample sizes one would expect in practical applications.

1. Introduction and summary. Let J denote an open interval of real numbers. Assume that for each $\theta \in J$ P_θ is a probability measure under which $x_1, x_2, \cdots$ are independent random variables with probability density function $h_\theta(x) = \exp(\theta x - \Psi(\theta))$ with respect to some σ-finite measure ν. Let $s_n = \sum_1^n x_k$ $(n = 0, 1, \cdots)$. For a given $\theta_0 \in J$ and F a probability distribution on J define

$$f(x, t) = \int_J \exp[(y - \theta_0)x - t(\Psi(y) - \Psi(\theta_0))]\, dF(y)$$

and

$$T = \inf\{n : f(s_n, n) \geqq a\} \qquad (a > 1)\,. \tag{1}$$

Statistical applications of the stopping rule (1) have been discussed, for example, in [9] (see also Section 5), where as a consequence of a simple argument due to Wald it was shown that (cf. Lemma 6)

$$P_{\theta_0}\{T < \infty\} \leqq 1/a \qquad (a > 1)\,. \tag{2}$$

For these applications it is desirable to have approximations to $E_\theta T$ for $\theta \neq \theta_0$. Previous investigations have focused for the most part on the asymptotic behavior of $E_\theta T$ as $\theta \to \theta_0$ (cf. [3], [11], and [12]). However, it was pointed out in [12] that there is reason to think that an asymptotic analysis of $E_\theta T$ as $a \to \infty$ will provide better numerical approximations for a wide range of values of θ and a. In this paper we begin with such an analysis and indicate the accuracy of our results by comparing them with some Monte Carlo simulations.

Putting $I(\theta) = (\theta - \theta_0)\Psi'(\theta) - (\Psi(\theta) - \Psi(\theta_0))$, we state our first result as

THEOREM 1. *Let $\theta \neq \theta_0$ be such that F' exists in a neighborhood of θ and is positive and continuous at θ. Then as $a \to \infty$*

$$E_\theta T \cong 2 \log a + \log ((\log a)/I(\theta)) - \log (2\pi(F'(\theta))^2/\Psi''(\theta)) - 1]/2I(\theta) + o(1)\,. \tag{3}$$

Received September 1973; revised December 1974.

[1] Now at Hebrew University.

AMS 1970 *subject classifications.* Primary 62L10, 60G40; Secondary 62L05.

Key words and phrases. Sequential tests, tests of power one, expected sample size, stopping rules, asymptotic formulae, detecting a change, Monte Carlo.

REMARK. The approximate equality (denoted by $\cong$) appears in Theorem 1 because, following a tradition going back to Wald [13], we have neglected the "excess over the boundary" $\log f(S_T, T) - \log a$. A more detailed analysis shows, at least whenever the distribution of X_1 is nonlattice, that the difference between the left- and right-hand sides of (3) converges to a positive finite limit which in principle can be evaluated in terms of ladder variable distributions. We have not included this refinement because (a) the required analysis is quite complicated, (b) the additional term obtained requires a computer for evaluation in those cases in which evaluation is possible, and (c) the improvement in accuracy is rather small, particularly in light of the investment of effort required. Details of such an analysis with different applications will appear in [4]. Simple (distribution dependent) upper bounds for this additional term are discussed in Section 7.

Theorem 1 is proved in Section 2, and in Section 3 is given a slight variation of Theorem 1 in which the random variables $x_1, x_2, \cdots$ may have a different distribution than that defined by the density $\exp(\theta x - \Psi(\theta))$. Monte Carlo approximations to $E_\theta T$ are presented in Section 4 and compared with the analytic approximations of Theorems 1 and 2. In Sections 5 and 6 we discuss the problem of detecting a change in the parameter of a distribution and obtain results closely related to Theorem 1. Miscellaneous remarks appear in Section 7.

2. Proof of Theorem 1. Before beginning the proof of Theorem 1 we make several notational simplifications which we use throughout the rest of this paper. Let $\mu = E_\theta x_1$ and $\sigma^2 = \mathrm{Var}_\theta x_1$. It is easy to see that $\mu = \Psi'(\theta)$ and $\sigma^2 = \Psi''(\theta)$. By translating the x's we may assume that $\Psi'(\theta_0) = E_{\theta_0} x_1 = 0$. By putting $\nu_0(dx) = h_{\theta_0}(x)\nu(dx)$ so that under P_θ the random variables $x_1, x_2, \cdots$ have relative to ν_0 the probability density function $h_\theta(x)/h_{\theta_0}(x) = \exp[(\theta - \theta_0)x - (\Psi(\theta) - \Psi(\theta_0))]$, we may by relabelling the parameter space J assume that $\theta_0 = 0$ and $\Psi(\theta_0) = 0$. With these conventions we have $f(x, t) = \int \exp[yx - t\Psi(y)]\, dF(y)$ and $I(\theta) = \theta\Psi'(\theta) - \Psi(\theta)$.

We defer until later several lemmas referred to in the proof of Theorem 1.

Without loss of generality assume that $\theta > 0$ and hence, by the strict convexity of Ψ, $\mu = \Psi'(\theta) > 0$.

It is shown in Lemma 1 below that $E_\theta T < \infty$. By the definition of T

$$\log a \leqq \theta s_T - T\Psi(\theta) - 1/2 \log T + \log[T^{\frac{1}{2}} \textstyle\int_J \exp[(y - \theta)s_T - T(\Psi(y) - \Psi(\theta))]\, dF(y)]\,. \tag{4}$$

We proceed to analyse the different expressions in (4).

Let $0 < \varepsilon < 1$ be arbitrary and let $0 < \delta_1 < \delta_2$ be positive numbers to be specified later. Let $n_1 = (1 - \varepsilon)(\log a)/I(\theta)$ and $A = \{T > n_1, \max_{n \geqq n_1} |n^{-1}s_n - \mu| < \delta_1\}$. Obviously $P_\theta(A^c) \leqq P_\theta\{T \leqq n_1\} + P_\theta\{\max_{n \geqq n_1} |n^{-1}s_n - \mu| \geqq \delta_1\}$ and hence by Lemmas 2 and 3 below, for some $\lambda > 0$

$$P_\theta(A^c) = O(a^{-\lambda}) \qquad (a \to \infty)\,. \tag{5}$$

By Wald's lemma

$$\begin{aligned}
&\int_A (\theta s_T - T\Psi(\theta))\, dP_\theta \\
&\quad = \int_A [\theta(s_T - \mu T) + TI(\theta)]\, dP_\theta \\
&\quad = I(\theta)E_\theta T - \theta \int_{A^c} (s_T - \mu T)\, dP_\theta - I(\theta) \int_{A^c} T\, dP_\theta\,.
\end{aligned} \tag{6}$$

By (5), the Schwarz inequality, Wald's lemma for squared sums (cf. [2], page 23) and Lemma 5

$$0 < \int_{A^c} T\, dP_\theta \leqq (E_\theta T^2 P_\theta(A^c))^{\frac{1}{2}} = o(1)\,, \tag{7}$$

and

$$\int_{A^c} (s_T - \mu T)\, dP_\theta \leqq (E_\theta(s_T - \mu T)^2 P_\theta(A^c))^{\frac{1}{2}} = (E_\theta T\sigma^2 P_\theta A^c)^{\frac{1}{2}} = o(1) \tag{8}$$

as $a \to \infty$, which together with (6) show that

$$\int_A (\theta s_T - T\Psi(\theta))\, dP_\theta = I(\theta)E_\theta T + o(1) \qquad (a \to \infty)\,. \tag{9}$$

Let $0 < \eta < 1$, and let $\delta_2 < \min(\mu, I(\theta)/\theta)$ be so small that for all $|y - \theta| < \delta_2$, $F'(y)$ exists,

$$F'(\theta)(1 - \eta) \leqq F'(y) \leqq F'(\theta)(1 + \eta)\,, \tag{10}$$

and

$$\tfrac{1}{2}(y - \theta)^2\sigma^2(1 - \eta) \leqq \Psi(y) - \Psi(\theta) - (y - \theta)\mu \leqq \tfrac{1}{2}(y - \theta)^2\sigma^2(1 + \eta)\,. \tag{11}$$

Considering first that part of the integral over J in (4) which comes from $|y - \theta| \leqq \delta_2$, and using (10) and (11), we obtain (letting $\varphi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$)

$$\begin{aligned}
&T^{\frac{1}{2}} \int_{|y-\theta|<\delta_2} \exp[(y - \theta)s_T - T(\Psi(y) - \Psi(\theta))]\, dF(y) \\
&\quad \leqq (2\pi)^{\frac{1}{2}} \exp\left(\frac{(s_T - \mu T)^2}{2\sigma^2(1 - \eta)T}\right) T^{\frac{1}{2}} \\
&\qquad \times \int_{|y-\theta|<\delta_2} \varphi\left((\sigma^2(1 - \eta)T)^{\frac{1}{2}}\left[y - \theta - \left(\frac{s_T - \mu T}{\sigma^2(1 - \eta)T}\right)\right]\right) dF(y) \\
&\quad \leqq (2\pi/\sigma^2(1 - \eta))^{\frac{1}{2}} \exp\left(\frac{(s_T - \mu T)^2}{2\sigma^2(1 - \eta)T}\right) F'(\theta)(1 + \eta)\,.
\end{aligned} \tag{12}$$

It is shown in Lemma 4 that for sufficiently small δ_1, δ_2, on A

$$T^{\frac{1}{2}} \int_{|y-\theta|>\delta_2} \exp[(y - \theta)s_T - T(\Psi(y) - \Psi(\theta))]\, dF(y)$$

is majorized by a nonrandom quantity which converges to 0 as $a \to \infty$. Hence by (12)

$$\begin{aligned}
&\int_A (\log T^{\frac{1}{2}} \int_J \exp[(y - \theta)s_T - T(\Psi(y) - \Psi(\theta))]\, dF(y))\, dP_\theta \\
&\quad \leqq 1/2 \log (2\pi/\sigma^2(1 - \eta)) + \log F'(\theta)(1 + \eta) \\
&\qquad + [2\sigma^2(1 - \eta)]^{-1} \int_A (s_T - \mu T)^2/T\, dP_\theta + o(1)\,.
\end{aligned} \tag{13}$$

Moreover, by the definition of A, Wald's lemma for squared sums, and Lemma 5

$$\begin{aligned}
\int_A (s_T - \mu T)^2/T\, dP_\theta &\leqq n_1^{-1}E_\theta(s_T - \mu T)^2 \\
&= \sigma^2 n_1^{-1}E_\theta T = \sigma^2(1 - \varepsilon)^{-1} + o(1)\,.
\end{aligned} \tag{14}$$

By integrating (4) with respect to P_θ over A and using (5), (9), (13) and (14) we obtain as $a \to \infty$

$$I(\theta)E_\theta T \geqq \log a + 1/2 \log n_1 - 1/2 \log [2\pi(F'(\theta)(1+\eta))^2/\sigma^2(1-\eta)] \\ - [2(1-\eta)(1-\varepsilon)]^{-1} + o(1) .$$

Since ε and η are arbitrary, this proves that the right-hand side of (3) is asymptotically a lower bound for $E_\theta T$. If we neglect the "excess" $E_\theta \log f(s_T, T) - \log a$, a similar argument completes the proof of Theorem 1.

LEMMA 1. $E_\theta T < \infty$.

PROOF. Let $\tau = \min(T, n) - 1$, so that $\tau + 1$ is a stopping time. From the definition of T

$$\log a \geqq \theta s_\tau - \Psi(\theta)\tau + \log \textstyle\int_{|y-\theta|<\delta} \exp[(y-\theta)s_\tau - \tau(\Psi(y) - \Psi(\theta))]\, dF(y) .$$

Using the expansion $\Psi(y) = \Psi(\theta) + (y-\theta)\Psi'(\theta) + (\frac{1}{2}(y-\theta)^2)\Psi''(\xi)$, and assuming that δ is taken so small that $|y-\theta|\Psi''(\xi) \leqq 1$ for all $|y-\theta| \leqq \delta$, we obtain

$$\log a > \theta(s_\tau - \mu\tau) + I(\theta)\tau + \log \textstyle\int_{|y-\theta|<\delta} \exp[(y-\theta)(s_\tau - \mu\tau)]\, dF(y) - \delta_\tau ,$$

and hence by Jensen's inequality

$$\log a \geqq \theta(s_\tau - \mu\tau) + (I(\theta) - \delta)\tau + \log F(\theta-\delta, \theta+\delta) \\ + (c-\theta)(s_\tau - \mu\tau) , \tag{15}$$

where $c = \int_{|y-\theta|<\delta} y\, dF(y)/F(\theta-\delta, \theta+\delta)$. Hence

$$(I(\theta) - \delta)\tau \leqq -c(s_{\tau+1} - \mu(\tau+1)) + c|x_{\tau+1} - \mu| \\ + \log a - \log F(\theta-\delta, \theta+\delta) . \tag{16}$$

By Wald's lemma and the Schwarz inequality $E_\theta(s_{\tau+1} - \mu(\tau+1)) = 0$ and $E_\theta(|x_{\tau+1} - \mu|) \leqq [E_\theta(\sum_1^{\tau+1}(x_k - \mu)^2)]^{\frac{1}{2}} = \sigma(E_\theta(\tau+1))^{\frac{1}{2}}$. Hence if δ is so small that $I(\theta) - \delta > 0$, we see from (16) that $E_\theta \tau$ remains bounded as $n \to \infty$, and hence $E_\theta T < \infty$.

LEMMA 2. *For any $\delta > 0$ there exist $0 < \lambda < \infty$, $0 < \alpha < \infty$ such that*

$$P_\theta\left\{\max_{n \geqq r}\left|\frac{s_n}{n} - \mu\right| \geqq \delta\right\} \leqq \alpha \exp(-\lambda r) .$$

PROOF. Obviously

$$P_\theta\left\{\max_{n \geqq r}\left|\frac{s_n}{n} - \mu\right| \geqq \delta\right\} \leqq \textstyle\sum_{n \geqq r} P_\theta\left\{\left|\frac{s_n}{n} - \mu\right| \geqq \delta\right\} .$$

For all $\xi > \theta > 0$

$$\begin{aligned} &P_\theta\{s_n - n\mu \geqq n\delta\} \\ &\quad = \textstyle\int_{\{s_n - n\mu \geqq n\delta\}} \exp[(\theta-\xi)s_n - n(\Psi(\theta) - \Psi(\xi))]\, dP_\xi \\ &\quad \leqq \exp[-n((\mu+\delta)(\xi-\theta) - (\Psi(\xi) - \Psi(\theta))]P_\xi\{s_n - n\mu \geqq n\delta\} , \end{aligned}$$

and hence since $\Psi(\xi) - \Psi(\theta) \sim \mu(\xi - \theta)$ as $\xi \downarrow \theta$, by choosing ξ sufficiently close to θ, for some $\lambda' = \lambda'(\delta, \theta) > 0$, $P_\theta\{s_n - n\mu \geqq n\delta\} \leqq \exp(-\lambda' n)$. Analogously one gets that $P_\theta\{s_n - n\mu \leqq -n\delta\} \leqq \exp(-\lambda'' n)$ for some λ'', so that for some $\lambda = \lambda(\delta, \theta) > 0$

$$P_\theta\{|s_n - n\mu| \geqq n\delta\} \leqq \exp(-\lambda n) . \tag{17}$$

The lemma follows easily from (17) by summing over $n \geqq r$.

LEMMA 3. *Let* $0 < \varepsilon < 1$ *and* $n_1 = (1 - \varepsilon) \log a / I(\theta)$. *There exists* $\lambda > 0$ *such that*

$$P_\theta\{T \leqq n_1\} = O(a^{-\lambda}) \qquad as \quad a \to \infty .$$

PROOF (cf. Lemmas 6 and 13 of [12]). For any $x > 0$

$$\begin{aligned} P_\theta\{T \leqq n_1\} &\leqq P_\theta\{s_{n_1} - \mu n_1 \geqq x\sigma n_1^{\frac{1}{2}}\} \\ &\quad + \int_{\{T \leqq n_1, s_{n_1} - \mu n_1 < x\sigma n_1^{\frac{1}{2}}\}} \exp[\theta s_{n_1} - n_1 \Psi(\theta)]\, dP_0 \\ &\leqq P_\theta\{s_{n_1} - \mu n_1 \geqq x\sigma n_1^{\frac{1}{2}}\} \\ &\quad + \exp[I(\theta) n_1 + \theta x \sigma n_1^{\frac{1}{2}}] P_0\{T \leqq n_1\} . \end{aligned} \tag{18}$$

Since by (2) $P_0\{T < \infty\} \leqq a^{-1}$, for $x = \varepsilon(I(\theta) \log a)^{\frac{1}{2}} / 2\theta\sigma$, the second term on the right-hand side of (18) is $O(a^{-\varepsilon/2})$. The required estimate for this choice of x for the first term on the right-hand side of (18) follows easily from (17).

LEMMA 4. *Given* $\delta_2 < \min(\mu, I(\theta)/\theta)$, *for all* δ_1 *so small that the inequalities* (20) *and* (21) *below are satisfied, on* A

$$\int_{|y-\theta| > \delta_2} T^{\frac{1}{2}} \exp[(y - \theta)s_T - T(\Psi(y) - \Psi(\theta))]\, dF(y) \leqq \varepsilon(a) , \tag{19}$$

where $\varepsilon(a)$ *is nonrandom and* $\to 0$ *as* $a \to \infty$.

PROOF. By the mean value theorem $\Psi(y) = \Psi(\theta) + (y - \theta)\Psi'(\xi)$ for some ξ between θ and y. Since Ψ is strictly convex, Ψ' is strictly increasing and ξ is an increasing function of y for $y > \theta$. Hence we may take δ_1 so small that for all $y \geqq \theta + \delta_2$

$$\Psi(y) \geqq \Psi(\theta) + (y - \theta)(\Psi'(\theta) + 2\delta_1) \tag{20}$$

and also

$$4\delta_1 < \delta_2 \inf_{0 \leqq x \leqq \theta} \Psi''(x) . \tag{21}$$

Now split the range of integration in (19) into three pieces according as $y \leqq 0$, $0 < y < \theta - \delta_2$, and $y > \theta + \delta_2$. Denote the resulting integrals by I_1, I_2, and I_3 respectively. Then on A for sufficiently large a

$$I_1 \leqq T^{\frac{1}{2}} \exp[-(\theta(\mu - \delta_1) - \Psi(\theta))T] \leqq n_1^{\frac{1}{2}} \exp[-n_1(I(\theta) - \delta_1\theta)] . \tag{22}$$

Also, by (20) on A for large a

$$\begin{aligned} I_3 &\leqq T^{\frac{1}{2}} \int_{y \geqq \theta + \delta_2} \exp[(y - \theta)(\mu + \delta_1)T - T(y - \theta)(\mu + 2\delta_1)]\, dF(y) \\ &\leqq n_1^{\frac{1}{2}} \exp(-\delta_1\delta_2 n_1) . \end{aligned} \tag{23}$$

Using a two-term Taylor series expansion and (21), we find that on A for large a

$$\begin{aligned} I_2 &\leqq T^{\frac{1}{2}} \textstyle\int_{0<y<\theta-\delta_2} \exp[T|y-\theta|(\delta_1 - \frac{1}{2}|y-\theta|\Psi''(\xi))]\, dF(y) \\ &\leqq T^{\frac{1}{2}} \textstyle\int_{0<y<\theta-\delta_2} \exp[T|y-\theta|(\delta_1 - 2\delta_1)]\, dF(y) \\ &\leqq n_1^{\frac{1}{2}} \exp(-\delta_1\delta_2 n_1)\,. \end{aligned} \tag{24}$$

(22), (23), and (24) complete the proof.

LEMMA 5. *$T \sim \log a/I(\theta)$ in probability and for $i = 1$ or 2 $E_\theta T^i \sim (\log a/I(\theta))^i$ as $a \to \infty$.*

PROOF. That T is asymptotically at least as large as $\log a/I(\theta)$ in probability follows easily from (4), (5), (12), and Lemma 4. Hence

$$\liminf_{a\to\infty} E_\theta T/(\log a)^i \geqq (I(\theta))^{-i} \qquad (i = 1 \text{ or } 2)\,,$$

and to complete the proof it suffices to show that

$$\limsup E_\theta T^i/(\log a)^i \leqq (I(\theta))^{-i} \qquad (i = 1, 2)\,.$$

Reasoning as in the proof of Lemma 1 yields (16) with τ replaced by $T - 1$. The remainder of the argument is a straight-forward application of Wald's lemmas for the first and second moments.

For further reference we also record (cf. [2], page 40)

LEMMA 6. *Let $z_1, z_2, \cdots$ be a nonnegative supermartingale sequence. Then*

$$P\{z_n \geqq b \text{ for some } n \geqq 1\} \leqq Ez_1/b \qquad (b > Ez_1)\,.$$

REMARK. Since $z_n = f(s_n, n)$, $n = 1, 2, \cdots$ is a nonnegative martingale under P_0 with $E_0 z_1 = 1$, the inequality (2) is a special case of Lemma 6.

3. Case of incorrect distribution. It may occur that the random variables $x_1, x_2, \cdots$ do not have the assumed probability density function $\exp(\theta x - \Psi(\theta))$ used in the definition (1) of the stopping rule T. For example, it may be convenient to assume a normal model ($\Psi(\theta) = \theta^2/2$) even though the x's are not normal. If the x's are *subnormal* in the sense that their probability density function (relative to $P_0 \circ x_1^{-1}$) is

$$\exp(\theta x - g(\theta))\,, \tag{25}$$

where $g(\theta) \leqq \theta^2/2$ for all θ, then (2) continues to hold for T defined by (1) with $\Psi(\theta) = \theta^2/2$. (An example of the subnormal case occurs if the x's assume the values $+1$ and -1 with probabilities p and $q = 1 - p$ respectively and $\theta = \frac{1}{2}\log p/q$, so $g(\theta) = \log[(e^\theta + e^{-\theta})/2] \leqq \theta^2/2$.) Alternatively, according to Theorem 2 of [10] if we use the normal model $\Psi(\theta) = \theta^2/2$, then for random variables having an arbitrary distribution with mean 0 and variance 1 the inequality (2) holds approximately for measures F which are suitably concentrated about 0.

In this section we assume that under some probability measure P, $x_1, x_2, \cdots$ are independent and identically distributed. We continue to assume that T is defined in terms of a mixture of probability density functions of the form $\exp(\theta x - \Psi(\theta))$, as in (1). We do not assume, however, that this is the correct density function so that (2) may or may not be true. Let μ and σ^2 denote Ex_1 and Var x_1 respectively. We shall assume that $Ee^{\lambda x_1} < \infty$ for λ in some neighborhood of 0, although this assumption is much stronger than necessary.

THEOREM 2. *Assume that* $\mu \neq 0$ *and that there exists* $\theta \neq 0$ *such that*

$$\mu = \Psi'(\theta) . \tag{26}$$

Assume also that F' *exists in some neighborhood of* θ *and is positive and continuous at* θ. *If for some* $\underline{\sigma}^2 > 0$

$$\Psi''(y) \geqq \underline{\sigma}^2 \tag{27}$$

for all y *in an interval of* F*-measure* 1, *then as* $a \to \infty$

$$E(T) \cong [2 \log a + \log((\log a)/I(\theta)) - \log(2\pi(F'(\theta))^2/\Psi''(\theta)) - \sigma^2/\Psi''(\theta)]/2I(\theta) + o(1) . \tag{28}$$

REMARK. The condition (27) is always satisfied for the normal model $\Psi(y) = y^2/2$, for in that case $\Psi''(y) \equiv 1$. It is possible to give examples to show that condition (27) cannot be completely eliminated.

An examination of the proof of Theorem 1 shows that only Lemma 3 required that $\exp(\theta x - \Psi(\theta))$ be the correct probability density function. (On the other hand, Lemma 3 applies not only to stopping rules defined by (1) but to any stopping rule satisfying (2).) The following result may be used instead of Lemma 3 to complete the proof of Theorem 2.

LEMMA 7. *Under the conditions of Theorem* 2, *for any* $0 < \varepsilon < 1$, *there exists* $\delta > 0$ *such that*

$$P\{T \leqq (1 - \varepsilon)(\log a)/I(\theta)\} \leqq 2 \exp(-\delta(\log a)^{\frac{1}{2}}) . \tag{29}$$

PROOF. Using a two-term Taylor expansion, (26), and (27) we obtain the following chain of inequalities:

$$\begin{aligned}
&\log \textstyle\int_J \exp(ys_n - n\Psi(y))\, dF(y) \\
&\quad = \theta s_n - n\Psi(\theta) + \log \textstyle\int_J \exp[(y - \theta)s_n - n(\Psi(y) - \Psi(\theta))]\, dF(y) \\
&\quad \leqq \theta(s_n - n\mu) + nI(\theta) \\
&\qquad\quad + \log \textstyle\int_J \exp[(y - \theta)(s_n - n\mu) - n\underline{\sigma}^2(y - \theta)^2/2]\, dF(y) \\
&\quad = \theta(s_n - n\mu) + nI(\theta) + (s_n - n\mu)^2/2\underline{\sigma}^2 n \\
&\qquad\quad + \log \textstyle\int_J \exp[-1/2n\underline{\sigma}^2(y - \theta - (s_n - n\mu)/n\underline{\sigma}^2)^2]\, dF(y) \\
&\quad \leqq \theta(s_n - n\mu) + nI(\theta) + (s_n - n\mu)^2/2\underline{\sigma}^2 n .
\end{aligned}$$

Hence, letting $c = \log a$, $n_1 = (1 - \varepsilon)c/I(\theta)$, and $B = \{|s_n - n\mu| \leqq \delta_1 c^{\frac{1}{2}} + \delta_2 n$ for all $n\}$, for all δ_1 and δ_2 sufficiently small and $n \leqq n_1$, on B the preceding quantity is

$$\begin{aligned}&\leqq |\theta|(\delta_1 c^{\frac{1}{2}} + \delta_2 n) + nI(\theta) + (1/2n\underline{\sigma}^2)(\delta_1^2 c + 2\delta_1\delta_2 c^{\frac{1}{2}} n + \delta_2^2 n^2)\\ &\leqq |\theta|(\delta_1 c^{\frac{1}{2}} + \delta_2 n_1) + n_1 I(\theta) + (1/2\underline{\sigma}^2)(\delta_1^2 c + 2\delta_1\delta_2 c^{\frac{1}{2}} + \delta_2^2 n_1)\\ &< (1 - \varepsilon/2)c < c ,\end{aligned}$$

so that $B \cap \{T < n_1\} = 0$. Hence $P\{T < n_1\} \doteq P(\{T < n_1\} \cap B^c) \leqq P(B^c)$, and to complete the proof it suffices to show that for some $\delta > 0$

$$P(B^c) \leqq 2\exp(-\delta c^{\frac{1}{2}}) . \tag{30}$$

Let $g(\lambda) = \log E[\exp(\lambda(x_1 - \mu))]$ and $b = \delta_1 c^{\frac{1}{2}}$. Obviously for $\lambda > 0$

$$\begin{aligned}&P\{s_n - n\mu \geqq b + \delta_2 n \quad \text{for some} \quad n\}\\ &\qquad = P\{\exp[\lambda(s_n - n\mu) - ng(\lambda)]\\ &\qquad\qquad \geqq \exp[\lambda b + n(\lambda\delta_2 - g(\lambda))] \quad \text{for some} \quad n\} .\end{aligned} \tag{31}$$

Now $g(\lambda) = O(\lambda^2)$ as $\lambda \to 0$ and hence for small positive λ $\lambda\delta_2 - g(\lambda) > 0$. For any such λ the right-hand side of (31) is majorized by

$$P\{\exp[\lambda(s_n - n\mu) - ng(\lambda)] \geqq \exp(\lambda b) \quad \text{for some} \quad n\} \leqq \exp(-\lambda b) , \tag{32}$$

where the inequality (32) follows from Lemma 6. Putting $b = \delta_1 c^{\frac{1}{2}}$ and $\delta = \lambda\delta_1$ in (31) and (32) yields (30), which completes the proof.

4. Monte Carlo approximations. The following tables give 1000 run Monte Carlo approximations to $E_\theta T$ together with the appropriate theoretical asymptotic

TABLE 1

x_1 *is* $N(\mu, 1)$, $\Psi(y) = y^2/2$, $dF(y) = 2\varphi(y)\,dy \quad (y > 0)$
$\qquad = 0 \qquad (y < 0)$

	μ					
a	.30		.50		.75	
	Theory	Monte Carlo	Theory	Monte Carlo	Theory	Monte Carlo
20	87.7	94.1	28.1	30.9	11.6	14.7
100	128.3	133.8	42.7	44.6	18.1	20.7
1000	183.9	183.4	62.8	64.1	27.0	30.3

TABLE 2

x_1 *is* $N(\mu, 1)$, $\Psi(y) = y^2/2$, $dF(y) = \varphi(y)dy \quad (-\infty < y < \infty)$

	μ					
a	.30		.50		.75	
	Theory	Monte Carlo	Theory	Monte Carlo	Theory	Monte Carlo
20	103.1	109.3	33.7	37.3	14.1	16.5
100	143.7	153.3	48.3	51.6	20.6	22.7
1000	199.3	207.5	68.3	72.0	29.5	30.9

approximations from Theorem 1 or Theorem 2. In all cases Theoretical and Monte Carlo results are in close agreement.

Tables 1 and 2 give results for a normal model and normally distributed x's. Let $\varphi(y) = (2\pi)^{-\frac{1}{2}} \exp(-y^2/2)$.

It turns out that a slightly better agreement with the Monte Carlo results is given by the asymptotic upper bound for $E_\theta T$ of Remark (a) in Section 7. For example for the row $a = 100$ of Table 1 the asymptotic upper bounds are 136.8, 45.7, and 19.4 for $\mu = .3$, .5, and .75 respectively.

Table 3 gives results for a normal model and Bernoulli x's. The measure F is a $N(0, 1)$ distribution as in Table 2. Similar results have been obtained for F a half-normal distribution as in Table 1, but have not been included.

Table 4 deals with a Bernoulli model and Bernoulli x's. The measure F is a logistic distribution on θ-space which arises from a uniform distribution on $p = \exp(2\theta)/(\exp(2\theta) + 1)$.

TABLE 3

$P\{x_1 = 1\} = p = 1 - P\{x_1 = -1\}$, $\Psi(y) = y^2/2$, $dF(y) = \varphi(y)\,dy$

$(\mu = 2p - 1)$

a	μ = .30 Theory	.30 Monte Carlo	.50 Theory	.50 Monte Carlo
20	104.1	110.4	34.7	37.6
100	144.7	146.5	49.3	53.3
1000	200.3	205.0	69.3	73.3

TABLE 4

$P_\theta\{x_1 = 1\} = p = 1 - P_\theta\{x_1 = -1\}$, $\Psi(y) = \log\cosh y$,

$dF(y) = 2e^{2y}\,dy/(e^{2y} + 1)^2$

$(\theta = 1/2 \log(p/(1-p)), \mu = 2p - 1)$

a	μ = .30 Theory	.30 Monte Carlo	.50 Theory	.50 Monte Carlo
20	96.5	102.3	30.4	34.1
100	136.4	140.6	44.4	46.2
1000	191.2	189.9	63.5	66.0

5. Detecting a change in distribution. Let ν be an arbitrary nonnegative integer or $+\infty$, and let $\omega \leqq 0 < \theta$. Let $P_{\omega,\theta}^{(\nu)}$ denote a probability under which $x_1, \cdots$ are independent, $x_1, \cdots, x_\nu$ with probability density function $\exp(\omega x - \Psi(\omega))$ and $x_{\nu+1}, x_{\nu+2}, \cdots$ with probability density function $\exp(\theta x - \Psi(\theta))$. We write P_0 for $P_{0,\theta}^{(\infty)}$ and P_θ for $P_{0,\theta}^{(0)}$. We would like to find a stopping rule t which takes on large values with high probability under P_0 and yet yields small values of $E^{(\nu)}_{\omega,\theta}(t-\nu \mid t>\nu)$ uniformly in (ω,ν) for a wide range of values of θ. One criterion, which differs only slightly from those considered in [1], [5], and [7], is the following: given $a>0$, subject to the condition that

$$E^{(\infty)}_{\omega,\theta}t \geqq a \qquad \text{for all} \quad \omega \leqq 0 \tag{33}$$

choose t insofar as possible to minimize

$$\sup_{\omega,\nu} E^{(\nu)}_{\omega,\theta}(t-\nu \mid t>\nu) \tag{34}$$

over a wide range of values of θ. A special case of the class of stopping rules suggested in [5], which in spirit go back to [7], is given by

$$t = \inf\{n : \max_{0\leqq k<n} f(s_n - s_k, n-k) \geqq a\}, \tag{35}$$

where f is defined as in Section 2 relative to a measure F assigning unit measure to $J \cap (0,\infty)$. By (2) and Theorem 2 of [5], (33) holds for all $a>1$. Corresponding to our Theorem 1, we obtain

THEOREM 3. *Let $\theta>0$. Assume that F' exists in a neighborhood of θ and is positive and continuous at θ. Then as $a \to \infty$*

$$\begin{aligned} \sup_{\omega,\nu} & E^{(\nu)}_{\omega,\theta}(t-\nu \mid t>\nu) \\ & \cong \{2\log a + \log((\log a)/I(\theta)) - \log(2\pi(F'(\theta))^2/\Psi''(\theta)) \\ & \quad - 1 + 2E_\theta[\min_{0\leqq k<\infty}(\theta s_k - k\Psi(\theta))]\}/2I(\theta) + o(1). \end{aligned} \tag{36}$$

The symbol $\cong$ has the same meaning as in Theorem 1. It may be shown (cf. Lemma 11) that $E_\theta\{\min_{0\leqq k<\infty}[\theta s_k - k\Psi(\theta)]\} \in [-1, 0)$ and $\cong -1$.

The proof of Theorem 3 is similar to but somewhat more difficult than that of Theorem 1. Several technical lemmas are deferred until the end of the proof.

PROOF OF THEOREM 3. It is easy to see from the definition (35) that

$$\sup_{\omega,\nu} E^{(\nu)}_{\omega,\theta}(t-\nu \mid t>\nu) = E^{(0)}_{0,\theta}(t).$$

Hence we may assume that $x_1, x_2, \cdots$ are independent with probability density function $\exp(\theta x - \Psi(\theta))$, and write $P_\theta(E_\theta)$ for $P^{(0)}_{0,\theta}(E^{(0)}_{0,\theta})$.

The numbers $\delta_1 < \delta_2$, ε, and η will play the same roles as in the proof of Theorem 1. In addition let $0<\varepsilon<\varepsilon_0<1$, $0<\beta<1$,

$$\begin{aligned} n_0 &= (1-\varepsilon_0)\log a/I(\theta), \qquad n_1 = (1-\varepsilon)\log a/I(\theta), \\ n_2 &= (1+\varepsilon)\log a/I(\theta). \end{aligned} \tag{37}$$

and

$$A_1 = \left\{\max_{n\geqq n_0}\left|\frac{s_n}{n} - \mu\right| < \delta_1\right\}, \qquad A_2 = \{n_1 < t < n_2\},$$

$$A_3 = \{\max_{1\leqq k\leqq n_0} |s_k - k\mu| < \delta_1(n_1 - n_0)\},$$

$$A_4 = \{\max_{n_0<n\leqq n_2} \max_{n_0\leqq k<n} f(s_n - s_k, n-k) < a^\beta\},$$

and $A = A_1 \cap A_2 \cap A_3 \cap A_4$. By Lemmas 2 and 3 and Lemmas 8–10 below, for sufficiently small ε_0 there exists $\lambda > 0$ such that

$$P_\theta(A^c) = O(a^{-\lambda}) \qquad (a \to \infty) . \tag{38}$$

Corresponding to (4) we have

$$\begin{aligned} \log a \leqq \theta s_t - t\Psi(\theta) - 1/2 \log t \\ + \log \max\nolimits_{0 \leqq k < t} t^{\frac12} \textstyle\int_{J \cap (0,\infty)} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta)) \\ - ys_k + k\Psi(y)]\, dF(y) . \end{aligned} \tag{39}$$

The first three terms on the right-hand side of (39) may be treated as in the proof of Theorem 1. To estimate the final term we split the integral into three parts according as $0 < y < \theta - \delta_2$, $|y - \theta| \leqq \delta_2$, and $y > \theta + \delta_2$. Now

$$\begin{aligned} \max\nolimits_{0 \leqq k < t} \textstyle\int_{|y-\theta|<\delta_2} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta)) - ys_k + k\Psi(y)]\, dF(y) \\ \leqq \textstyle\int_{|y-\theta|<\delta_2} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta))] \\ \times \exp\left[\max\nolimits_{0 \leqq k < t} y\left(k\frac{\Psi(y)}{y} - s_k\right)\right] dF(y) . \end{aligned} \tag{40}$$

By the convexity of Ψ, $\Psi(y)/y$ is increasing for positive y as is the nonnegative quantity $\max_{0 \leqq k < t} y[k(\Psi(y)/y) - s_k]$. Hence the right-hand side of (40) is majorized by

$$\begin{aligned} \exp[-\min\nolimits_{0 \leqq k < \infty} \{(\theta + \delta_2)s_k - k\Psi(\theta + \delta_2)\} \\ \times \textstyle\int_{|y-\theta|<\delta_2} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta))]\, dF(y) . \end{aligned} \tag{41}$$

The integral in (41) is estimated as in the proof of Theorem 1, and we conclude that on A

$$\begin{aligned} \max\nolimits_{0 \leqq k < t} t^{\frac12} \textstyle\int_{|y-\theta|<\delta_2} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta)) \\ - ys_k + k\Psi(y)]\, dF(y) \\ \leqq \exp\left[-\min\nolimits_{0 \leqq k < \infty} \{(\theta + \delta_2)s_k - k\Psi(\theta + \delta_2)\} + \frac{(s_t - \mu t)^2}{2\sigma^2 t(1 - \eta)}\right] \\ \times [2\pi(F'(\theta)(1 + \eta))^2/\sigma^2(1 - \eta)]^{\frac12} . \end{aligned} \tag{42}$$

The same monotonicity argument applied to the integral over $0 < y < \theta - \delta_2$ together with the inequality (24) shows that on $A_1 \cap A_2$

$$\begin{aligned} \max\nolimits_{0 \leqq k < t} t^{\frac12} \textstyle\int_{0<y<\theta-\delta_2} \exp[(y - \theta)s_t - t(\Psi(y) - \Psi(\theta)) \\ - ys_k + k\Psi(y)]\, dF(y) \\ \leqq \exp[-\min\nolimits_{0 \leqq k < \infty} (\theta s_k - k\Psi(\theta))] n_1^{\frac12} \exp(-\delta_1 \delta_2 n_1) . \end{aligned} \tag{43}$$

We now turn to the integral over the range $y > \theta + \delta_2$. On $A_1 \cap A_2 \cap A_4$

$$\begin{aligned} \max\nolimits_{n_0 \leqq k < t} t^{\frac12} \textstyle\int_{y>\theta+\delta_2} \exp(y - \theta)s_t - t(\Psi(y) - \Psi(\theta)) \\ - ys_k + k\Psi(y)]\, dF(y) \\ \leqq n_2^{\frac12} \exp[-n_1\{\theta(\mu - \delta_1) - \Psi(\theta)\}] \\ \times \max\nolimits_{n_1 < n \leqq n_2} \max\nolimits_{n_0 \leqq k < n} f(s_n - s_k, n - k) \\ \leqq n_2^{\frac12} \exp[-n_1(I(\theta) - \delta_1\theta)] a^\beta \to 0 \qquad \text{as } a \to \infty \end{aligned} \tag{44}$$

provided δ_1 and ε are so small that $(1-\varepsilon) - \delta_1\theta/I(\theta) - \beta > 0$. On $A_1 \cap A_3$, for all $n_1 \leqq n \leqq n_2$ and $0 \leqq k \leqq n_0$

$$\left|\frac{s_n - s_k - (n-k)\mu}{n-k}\right| \leqq \left|\frac{s_n}{n} - \mu\right| \frac{n}{n-k} + \left|\frac{s_k - k\mu}{n_1 - n_0}\right| \leqq \delta_1\left(\frac{1-\varepsilon}{\varepsilon_0 - \varepsilon} + 1\right) = \delta_3\,, \quad \text{say.} \tag{45}$$

For given $\varepsilon_0 > \varepsilon$, we can take δ_1 yet smaller if need be so that (20) holds with δ_3 in place of δ_1. Then on $A_1 \cap A_2 \cap A_3$, by (45) and (20) with δ_3 in place of δ_1 we obtain as in the proof of (23)

$$\begin{aligned} \max\nolimits_{0\leqq k<n_0} t^{\frac12} \textstyle\int_{y>\theta+\delta_2} &\exp[(y-\theta)s_t - t(\Psi(y) - \Psi(\theta)) \\ &- ys_k + k\Psi(y)]\, dF(y) \\ &= \max\nolimits_{0\leqq k<n_0} \{\exp[-\theta s_k + k\Psi(\theta)] t^{\frac12} \textstyle\int_{y>\theta+\delta_2} \exp[(y-\theta)(s_t - s_k) \\ &\quad - (t-k)(\Psi(y) - \Psi(\theta))]\, dF(y)\} \\ &\leqq \exp[-\min\nolimits_{0\leqq k<\infty}(\theta s_k - k\Psi(\theta))] n_2^{\frac12} \exp[-\delta_2\delta_3(n_1 - n_0)]\,. \end{aligned} \tag{46}$$

From (42), (43), (44), (46), and Lemma 11 it may be seen that if we separate off the quantity

$$(s_t - \mu t)^2/2\sigma^2 t(1-\eta) - \min\nolimits_{0\leqq k<\infty}[(\theta+\delta_2)s_k - k\Psi(\theta+\delta_2)]$$

from the last term in (39), then the remaining quantity is asymptotically (as $a \to \infty$) no larger than

$$1/2 \log[2\pi(F'(\theta)(1+\eta))^2/\sigma^2(1-\eta)]$$

and is majorized (on A) for all a by a random variable having finite expectation. Hence we may integrate (39) over A, treat the first three terms as well as that part of the fourth term which involves $(s_t - \mu t)^2/2\sigma^2 t(1-\eta)$ as in the proof of Theorem 1, and interchange the limit as $a \to \infty$ and the integral over A of the remaining part of the fourth term to obtain

$$\begin{aligned} \log a \leqq I(\theta)E_\theta t - n_1/2 &+ \tfrac12 \log[2\pi(F'(\theta)(1+\eta))^2/\sigma^2(1-\eta)] \\ &+ [2(1-\eta)(1-\varepsilon)]^{-1} \\ &- E_\theta[\min\nolimits_{0\leqq k<\infty}((\theta+\delta_2)s_k - k\Psi(\theta+\delta_2))] + o(1)\,. \end{aligned} \tag{47}$$

Letting $\delta_2 \to 0$, then ε and $\eta \to 0$, from Lemma 11 we obtain (36) with the inequality $\geqq$.

To prove the reverse inequality (ignoring the excess $\max_{0\leqq k<t} \log f(s_t - s_k, t-k) - \log a$), observe that in place of (40), for each $0 \leqq k < t$ by the mean value theorem there exists a number $y_k(|y_k - \theta| < \delta_2)$ such that

$$\begin{aligned} \textstyle\int_{|y-\theta|<\delta_2} &\exp[(y-\theta)s_t - t(\Psi(y) - \Psi(\theta)) - ys_k + k\Psi(y)]\, dF(y) \\ &= \exp[-y_k s_k + k\Psi(y_k)] \textstyle\int_{|y-\theta|<\delta_2} \exp[(y-\theta)s_t \\ &\quad - t(\Psi(y) - \Psi(\theta))]\, dF(y)\,. \end{aligned} \tag{48}$$

Taking the max over $0 \leqq k < t$ in (48), we may use monotonicity as before to obtain an appropriate lower bound which allows the argument to proceed as previously, completing the proof.

LEMMA 8. *For some* $\lambda > 0$ $P_\theta\{t \leqq n_1\} + P_\theta\{t > n_2\} = O(a^{-\lambda})$ *as* $a \to \infty$. (n_1 *and* n_2 *are defined in* (37).)

PROOF. The asserted bound for $P_\theta\{t \leqq n_1\}$ follows at once from Lemma 2 and the observation that by (35) $P_\theta\{t \leqq n_1\} \leqq n_1 P_\theta\{T \leqq n_1\}$, where T is defined by (1).

Given $\varepsilon > 0$, let $\delta > 0$ be so small that $(1 + \varepsilon)[(\theta - \delta)(\mu - \delta) - \Psi(\theta + \delta)] > I(\theta)$. Then for $n \geqq n_2 = (1 + \varepsilon) \log a/I(\theta)$ on the event $\{\max_{n \geqq n_1} |n^{-1}s_n - \mu| < \delta\}$

$$\begin{aligned} f(s_n, n) &\geqq \int_{|y-\theta|<\delta} \exp[ys_n - n\Psi(y)]\, dF(y) \\ &\geqq \exp[n((\theta - \delta)(\mu - \delta) - \Psi(\theta + \delta))]F(\theta - \delta, \theta + \delta) \\ &\geqq \exp[\log a(1 + \varepsilon)[(\theta - \delta)(\mu - \delta) - \Psi(\theta + \delta)]/I(\theta)]F(\theta - \delta, \theta + \delta) \\ &> a \end{aligned}$$

for all sufficiently large a. Hence for large a

$$\{t > n_2\} \subset \left\{\max_{n \geqq n_1} \left|\frac{s_n}{n} - \mu\right| \geqq \delta\right\},$$

and Lemma 3 now completes the proof.

LEMMA 9. *For arbitrary* $\delta > 0$ *there exists a* $\lambda > 0$ *such that*

$$P_\theta\{\max_{1 \leqq k \leqq n} |s_k - \mu k| \geqq \delta n\} = O(e^{-\lambda n}). \tag{49}$$

PROOF. For ease of exposition assume that $\theta = 0$ (and hence $\mu = 0$). By Kolmogorov's inequality for submartingales (cf. [2], page 24), for all $\xi > 0$, $\xi \in J$

$$\begin{aligned} P_0\{\max_{1 \leqq k \leqq n} s_k \geqq \delta n\} &= P_0\{\max_{1 \leqq k \leqq n} \exp(\xi s_k) \geqq \exp(\xi \delta n)\} \\ &\leqq \exp(-\xi\delta n)E_0(\exp(\xi s_n)) = \exp(-n(\xi\delta - \Psi(\xi)))\,. \end{aligned}$$

Now $\Psi(\xi) \sim \xi^2\Psi''(0)/2$ as $\xi \to 0$ and hence for small positive ξ $\xi\delta - \Psi(\xi) = \lambda > 0$. This completes the proof.

LEMMA 10. *For arbitrary* $\beta > 0$, *for* ε_0 *sufficiently small (and any* $\varepsilon < \varepsilon_0$) *there exists a* $\lambda > 0$ *such that*

$$P_\theta\{\max_{n_0 < n \leqq n_2} \max_{n_0 \leqq k < n} f(s_n - s_k, n - k) \geqq a^\beta\} = O(a^{-\lambda}) \qquad \text{as } a \to \infty\,. \tag{50}$$

PROOF. If we write $T(a)$ to emphasize the dependence on a of the stopping rule defined by (1), then the probability on the left-hand side of (50) is majorized by

$$\sum_{n_0 < n \leqq n_2} P_\theta\{\max_{n_0 \leqq k < n} f(s_n - s_k, n - k) \geqq a^\beta\} = \sum_{n_0 < n \leqq n_2} P_\theta\{T(a^\beta) \leqq n - n_0\},$$

from which (50) follows by Lemma 3.

LEMMA 11. *Let* $\theta > 0$ *and* $y_\theta = \sup\{y : y \in J, \Psi(y) < y\Psi'(\theta)\}$. *Then* $y_\theta > \theta$ *and* $E_\theta[\min_{0 \leqq k < \infty} (ys_k - k\Psi(y))]$ *is a decreasing continuous function on* $(0, y_\theta)$.

PROOF. That $y_\theta > \theta$ follows at once from the (strict) convexity of Ψ. For $0 < y < y_\theta$, by the strong law of large numbers

$$\min_{0 \leq k < \infty} (ys_k - k\Psi(y)) \tag{51}$$

is a.e. continuous and decreasing in y. (Its being decreasing in y follows from the non-positivity of (51) so that for the k minimizing at y, $s \leqq k\Psi(y)/y \leqq k\Psi'(\tilde{y})$ for $\tilde{y} \geqq y$ (i.e., $d(\tilde{y}s_k - k\Psi(\tilde{y}))/d\tilde{y} \leqq 0$) whence $\tilde{y}s_k - k\Psi(\tilde{y})$ is decreasing in $\tilde{y}$ on $[y, y_\theta)$ and the possibility of a different minimizing k for $\tilde{y} > y$ only helps.) Hence by the monotone convergence theorem it suffices to show that the random variable (51) has finite expectation for arbitrary $0 < y < y_\theta$. Let $0 < \xi < \theta$. Then for $x > 0$

$$\begin{aligned} &P_\theta\{\min_{0 \leq k < \infty} (ys_k - k\Psi(y)) < -x\} \\ &\quad = P_\theta\{s_k < -x/y + k\Psi(y)/y \text{ for some } k\} \\ &\quad = P_\theta\{\exp[(\xi - \theta)s_k - k(\Psi(\xi) - \Psi(\theta))] \geqq \exp[k[(\xi - \theta)\Psi(y)/y \\ &\qquad - (\Psi(\xi) - \Psi(\theta))] + (\theta - \xi)x/y] \text{ for some } k\}\,. \end{aligned} \tag{52}$$

Now since $y < y_\theta$ and $(\Psi(\xi) - \Psi(\theta))/(\xi - \theta) \to \Psi'(\theta)$ as $\xi \to \theta$, we may choose $\xi < \theta$ so that

$$(\xi - \theta)\Psi(y)/y - (\Psi(\xi) - \Psi(\theta)) \geqq 0\,.$$

Then by Lemma 6 the right-hand side of (52) is $\leqq \exp(-(\theta - \xi)x/y)$, and the proof is completed by integration over x.

REMARK. If we set $y = \theta$ and $\xi = 0$ in (52), we obtain

$$P_\theta\{\min_{0 \leq k < \infty} [\theta s_k - k\Psi(\theta)] \leqq -x\} \leqq \exp(-x) \qquad (x > 0)\,,$$

which after integration over x substantiates the remark following Theorem 3.

6. Detecting a change of distribution under a bounded error probability restriction. Under the same assumptions as in the preceding section, there may arise situations in which the restriction (33) that we not stop "too soon" is not strong enough. One may, for example, wish to consider only stopping rules σ for which for specified $0 < \alpha < 1$

$$P_{\omega,\theta}^{(\nu)}\{\sigma \leqq \nu\} \leqq \alpha \tag{53}$$

for all $\omega \leqq 0$, $\nu < \infty$. Under the restriction (53) it is impossible that

$$E_{\omega,\theta}^{(\nu)}(\sigma - \nu \mid \sigma > \nu) \tag{54}$$

remain bounded as $\nu \to \infty$, but one might look for a stopping rule σ for which (54) is asymptotically a minimum as $\nu \to \infty$.

Given α $(0 < \alpha < 1)$, let $\alpha_0, \alpha_1, \alpha_2, \cdots$ be nonnegative numbers such that

$$\sum \alpha_k = \alpha \tag{55}$$

and define

$$\sigma = \inf\{n : \textstyle\sum_{k=0}^{n-1} \alpha_k f(s_n - s_k, n - k) + \sum_{k=n}^{\infty} \alpha_k \geqq 1\}\,, \tag{56}$$

where f is defined as in Section 1. It is easy to see that the process within braces in (56) is a supermartingale (martingale) under $P^{(\infty)}_{\omega,\theta}(P^{(\infty)}_{0,\theta})$, and since $P^{(\nu)}_{\omega,\theta}\{\sigma \leqq \nu\} = P^{(\infty)}_{\omega,\theta}\{\sigma \leqq \nu\} \leqq P^{(\infty)}_{\omega,\theta}\{\sigma < \infty\}$ for all $\omega \leqq 0$ and $\nu < \infty$, (53) follows from Lemma 6.

For each ν for which $\alpha_\nu > 0$, on $\{\sigma > \nu\}$ we have

$$\sigma \leqq \inf\{n : n > \nu, f(s_n - s_\nu, n - \nu) \geqq \alpha_\nu^{-1}\}\,. \tag{57}$$

Hence if we neglect the excess over the stopping boundary as in Theorems 1–3, under the same conditions on F we obtain from (57) and Theorem 1 an asymptotic upper bound for $E^{(\nu)}_{\omega,\theta}(\sigma - \nu \mid \sigma > \nu)$ as $\nu \to \infty$ along any sequence of values for which $\alpha_\nu > 0$. If the sequence $\{\alpha_\nu\}$ is sufficiently regular it is possible to show that the dominant term of this bound is asymptotically correct, i.e.,

$$E^{(\nu)}_{\omega,\theta}(\sigma - \nu \mid \sigma > \nu) \sim \log \alpha_\nu^{-1}/I(\theta) \qquad (\nu \to \infty)\,; \tag{58}$$

but we have not been able to obtain a more detailed asymptotic expansion comparable to those of Theorems 1–3.

It is natural to conjecture that for any stopping rule σ satisfying (53)

$$\limsup_{\nu\to\infty} E^{(\nu)}_{\omega,\theta}(\sigma - \nu \mid \sigma > \nu)/\log \nu \geqq 1/I(\theta)\,, \tag{59}$$

but we have not been able to prove this result. (By (58) equality is attained in (59) for σ of the form (56) with α_ν proportional to $1/\nu(\log \nu)^{1+\varepsilon}$ for large ν.)

7. Remarks. (a) As noted in Section 1, the right-hand side of (3) is strictly speaking only an asymptotic lower bound for $E_\theta T$. For particular distributions one may use the method of proof of Theorem 1 together with arguments of Wald [13] (which provide a bound for $E_\theta(x_T - \mu)$) to obtain asymptotic upper bounds for $E_\theta(T)$. For example, in the case of normal variables ($\Psi(\theta) = \theta^2/2$) if we add $(1 + 2\varphi(\theta)/\theta\Phi(\theta))$ to the right-hand side of (3), we obtain an asymptotic upper bound. For variables assuming the values $+1$ and -1 with probabilities p and $q = 1 - p$ respectively, for which $\theta = 1/2 \log(p/q)$, $\Psi(\theta) = -\log 2(pq)^{\frac{1}{2}}$, and $I(\theta) = p \log p + q \log q + \log 2$, it suffices to add $(\theta - \Psi(\theta))/I(\theta)$ to the right-hand side of (3).

(b) For applications in statistics it is important that $E_\theta T$ should be as small as possible for a wide range of values of θ. For a given $\theta \neq 0$, by using the stopping rule (1) with F a measure assigning unit mass to the single point θ, one may obtain $E_\theta T = \log a/I(\theta) + O(1)$, which is smaller than (3) by a term which is $O(\log\log a)$ as $a \to \infty$. However, this stopping rule requires prior knowledge of θ and hence is impossible to implement in general. Thus it is an interesting problem to describe precisely in what sense (3) is minimal as $a \to \infty$ for stopping rules T satisfying (2). This question has been studied in [8], and the results will be published soon.

A similar question of optimality arises in the analysis of the problem of Section 5. G. Lorden [5] has obtained some results in this direction, although his criterion is slightly different from ours.

(c) The inequality (33) is rather crude for stopping rules defined by (35) although it presumably does indicate the correct order of magnitude as $a \to \infty$. It would be interesting to obtain an asymptotic expression for $E_{0,\theta}^{(\infty)}(t)$ as $a \to \infty$.

REFERENCES

[1] AROIAN, L. A. and LEVENE, H. (1950). The effectiveness of quality control charts. *J. Amer. Statist. Assoc.* **45** 520–529.

[2] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston.

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

[4] LAI, T. L. and SIEGMUND, D. (1975). On the distribution of excess over a boundary with statistical applications. To appear.

[5] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

[6] LORDEN, G. (1973). Open-ended tests for Koopman–Darmois families. *Ann. Statist.* **1** 633–643.

[7] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–114.

[8] POLLAK, M. (1973). Asymptotic problems related to boundary crossings of one-dimensional shifted Brownian motion. Unpublished thesis.

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

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

[11] ROBBINS, H. and SIEGMUND, D. (1973). Statistical tests of power one and the integral representation of solutions of certain partial differential equations. *Bull. Inst. Math. Acad. Sinica* **1** 93–120.

[12] ROBBINS, H. and SIEGMUND, D. (1974). The expected sample size of some tests of power one. *Ann. Statist.* **2** 415–436.

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

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM, ISRAEL

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027