The Annals of Statistics
1977, Vol. 5, No. 5, 984–995

SECOND ORDER APPROXIMATIONS FOR SEQUENTIAL POINT AND INTERVAL ESTIMATION[1]

By Michael Woodroofe

The University of Michigan

Several stopping times which arise from problems of sequential estimation may be written in the form $t_c = \inf\{n \geqq m: S_n < cn^\alpha L(n)\}$ where S_n, $n \geqq 1$, are the partial sums of i.i.d. positive random variables, $\alpha > 1$, $L(n)$ is a convergent sequence, and c is a positive parameter which is often allowed to approach zero. In this paper we find the asymptotic distribution of the excess $R_c = ct_c^\alpha - S_{t_c}$ as $c \to 0$ and use it to obtain sharp estimates for $E\{t_c\}$. We then apply our results to obtain second order approximations to the expected sample size and risk of some sequential procedures for estimation.

1. Introduction. Several stopping times which arise from problems in sequential point and interval estimation may be written in the form

$$t_c = \inf\{n \geqq m: S_n < cn^\alpha L(n)\}, \tag{1.1}$$

where S_n, $n \geqq 1$, are the partial sums of i.i.d. positive random variables X_1, $X_2, \cdots, L(n)$ is a convergent sequence, $\alpha > 1$, $m \geqq 1$, and c is a positive parameter (which is often allowed to approach zero). In particular, the stopping times for the sequential procedures proposed by Robbins (1959), Chow and Robbins (1965), Starr (1966a, 1966b), Starr and Woodroofe (1972), and Ghosh et al. (1976) are all of this form.

The analyses of these and related sequential procedures often use the inequality

$$c(t_c - 1)^\alpha L(t_c - 1) \leqq S_{t_c} < ct_c^\alpha L(t_c),$$

which is valid on $t_c > m$, to estimate the risk function and/or expected sample size (see, in particular, [10], [14] and [15]). In this paper we will find the asymptotic distribution of the difference

$$R_c = ct_c^\alpha L(t_c) - S_{t_c} \tag{1.2}$$

as $c \to 0$, under some modest conditions on L and the distribution of X_1. This asymptotic distribution is then used to refine some of the approximations mentioned above.

In Section 2 we find and study the asymptotic distribution of R_c and use it to compute $E\{t_c\}$ up to terms which are $o(1)$ as $c \to 0$. In Sections 3 and 4 we apply the results of Section 2 to some specific problems in sequential point

Received July 1975; revised Jaunary 1977.

[1] Research supported by the National Science Foundation under NSF-GP-38487, Amendment 01, and the National Institute of Health GM-20507.

AMS 1970 *subject classifications.* Primary 62L12; Secondary 60F05.

Key words and phrases. Excess under the boundary, sequential estimation, fixed width confidence intervals.

and interval estimation. Here is one of the applications. Let $Y_1, Y_2, \cdots$ be independent normally distributed random variables with unknown mean θ and unknown variance σ^2. We suppose that we may observe as many of $Y_1, Y_2, \cdots$ as we please, that at some time we must stop observing the process and report an estimate $\hat{\theta}$ of θ, and that if we stop at time n and report the estimate $\hat{\theta}$ we incur the loss $L_n = A(\hat{\theta} - \theta)^2 + n$. If σ^2 were known, it would be optimal (best invariant) to take $n_0 = \sigma A^{\frac{1}{2}}$ observations and to estimate θ by $\bar{Y}_n$, in which case our expected loss is $E\{L_{n_0}\} = 2n_0$. For the case of unknown σ^2, Robbins (1959) proposed the following sequential procedure: let

$$N = \inf\{n \geqq m : n > \hat{\sigma}_n A^{\frac{1}{2}}\},$$

where $m \geqq 2$ and $\hat{\sigma}_n^2$ denotes the unbiased sample variance, take N observations and estimate θ by $\bar{Y}_N$. We show that if $m \geqq 4$, then

$$E\{N\} = n_0 + \tfrac{1}{2}\sigma^{-2}\nu - \tfrac{3}{4} + o(1) \qquad \text{and} \qquad E\{L_N\} = 2n_0 + \tfrac{1}{2} + o(1) \tag{1.3}$$

as $A \to \infty$, where ν is a constant whose computation is discussed in Section 2. The terms of order n_0 in (1.3) were known, as was the fact that $E\{N - n_0\}$ and $E\{L_N\} - 2n_0$ are bounded. However, the evaluation of the $O(1)$ terms is (to the best of our knowledge) new.

2. Preliminaries. In this section $X_1, X_2, \cdots$ will denote i.i.d. positive random variables. We suppose that the mean $\mu = E\{X_1\}$ and the variance $\tau^2 = E\{X_1^2\} - \mu^2$ are both finite and positive. We suppose also that X_1 has a density f which is continuous a.e. and that some power of the characteristic function of X_1 is integrable. Finally, we suppose that L is a positive continuous function on $[0, \infty)$ for which

$$L(x) = 1 + L_0 x^{-1} + o(x^{-1}) \tag{2.1}$$

as $x \to \infty$, where $-\infty < L_0 < \infty$. These are standing assumptions.

Let $\beta = 1/(\alpha - 1)$ and $\lambda = \mu^\beta c^{-\beta}$; it is known that $\lambda^{-1}t_c \to 1$ w.p. 1 and that

$$t_c^* = \lambda^{-\frac{1}{2}}\{t_c - \lambda\}$$

is asymptotically normal with mean 0 and variance $\beta^2\tau^2\mu^{-2}$ as $c \to 0$ (see, for example, [2]). Our second order approximations require this result and the asymptotic distribution of R_c. In fact, we give the asymptotic joint distribution of R_c and t_c^*.

THEOREM 2.1. *R_c and t_c^* are asymptotically independent as $c \to 0$. The asymptotic distribution of t_c^* is normal with mean 0 and variance $\beta^2\tau^2\mu^{-2}$; and the asymptotic distribution H of R_c has density $h = H'$, where*

$$h(y) = \beta\mu^{-1}P\{S_j \leqq j\alpha\mu - y\,, \ \textit{for all } j \geqq 1\}\,, \qquad y > 0\,. \tag{2.2}$$

The proof of Theorem 2.1 is similar to that of Theorem 4.3 of [18]. Alternatively, the asymptotic distribution of R_c may be deduced from Theorem 1 of Lai and Siegmund (1975). We omit the details.

It is possible to derive a useful expression for the characteristic function $\hat{H}$ of H. Let $X_i' = X_i - \alpha\mu$ and let F_α and ϕ_α denote the distribution function and characteristic function of X_1', respectively. Also let $S_n' = X_1' + \cdots + X_n'$, $n \geqq 1$, and

$$M = \max\{S_k' : k \geqq 1\}.$$

We use a result of Spitzer (1960) which asserts that the characteristic function of $M^+ = \max\{M, 0\}$ is

$$w(t) = \exp\{\textstyle\sum_{n=1}^\infty n^{-1} \int_0^\infty (e^{itx} - 1)\, dF_\alpha{}^{*n}(x)\}, \qquad t \in R,$$

where * denotes convolution. (See also Feller (1966), page 576.)

THEOREM 2.2. *The characteristic function and mean of H are given by*

$$\hat{H}(t) = \left(\frac{\beta}{i\mu t}\right)[\phi_\alpha(-t) - 1]w(-t) \qquad t \neq 0 \tag{2.3}$$

$$\nu = \frac{\beta}{2\mu}[(\alpha - 1)^2\mu^2 + \tau^2] - \textstyle\sum_{n=1}^\infty n^{-1}E\{(S_n - n\alpha\mu)^+\}. \tag{2.4}$$

PROOF. Let G denote the distribution function of $M^- = \max\{-M, 0\}$. Then $h(y) = \beta\mu^{-1}[1 - G(y)]$ for $y > 0$, so that

$$\hat{H}(t) = \left(\frac{\beta}{i\mu t}\right) E\{e^{itM^-} - 1\}, \qquad t \neq 0,$$

by an integration by parts. Since $M = X_1' + M_1^+$, where M_1 is independent of X_1 and has the same distribution as M, we find

$$E\{e^{itM^-} - 1\} = E\{e^{-itM} - e^{-itM^+}\} = [\phi_\alpha(-t) - 1]w(-t).$$

This establishes (2.3), and (2.4) then follows by differentiation.

EXAMPLE. If X_1 has the gamma distribution, say

$$f(x) = \Gamma(a)^{-1}b^a x^{a-1}e^{-bx}, \qquad x > 0,$$

where $a > 0$ and $b > 0$, then it is possible to relate ν to the incomplete gamma function

$$G(a; x) = \textstyle\int_x^\infty y^{a-1}e^{-y}\, dy, \qquad x > 0.$$

Let us write $\nu_\alpha(a, b)$ for ν in (2.4). Then $\nu_\alpha(a, b) = b^{-1}\nu_\alpha(a, 1)$. Moreover, when $b = 1$, $\mu = a$, $\tau^2 = a$, and

$$E\{(S_n - n\alpha\mu)^+\} = \Gamma(na)^{-1}\{(na\alpha)^{na}e^{-na\alpha} - na(\alpha - 1)G(na; na\alpha)\}.$$

TABLE 2.1
Values of $\nu_\alpha(a, 1)$

a/α	2.0	2.5	3.0	4.0
.5	.410	.520	.634	.864
1.0	.747	.963	1.187	1.647
2.9	1.343	1.777	2.229	3.163

Thus, $\nu_\alpha(a, 1)$ may be computed to any desired degree of accuracy from tables of the incomplete gamma function. We include some typical values.

We will now develop an asymptotic expression for $E\{t_c\}$. We need some auxiliary results on uniform integrability.

LEMMA 2.1. *If* $E\{X_1^r\} < \infty$, *where* $r \geqq 2$, *then* $(\lambda^{-1}t_c)^{r(\alpha-1)}$ *and* R_c^r *are dominated.*

PROOF. On $\{t_c > m\}$, we have $ct_c^{\alpha-1} \leqq Bc(t_c-1)^{\alpha-1}L(t_c-1) \leqq B \sup\{n^{-1}S_n : n \geqq 1\}$ for all $c > 0$ for some constant B. The first assertion now follows from the fact that $n^{-1}S_n$ is a backward martingale and the maximal inequality (Doob (1953), pages 317–318). The second assertion then follows from $R_c \leqq ct_c^\alpha L(t_c) - c(t_c - 1)^\alpha L(t_c - 1) \leqq Bct_c^{\alpha-1}$ on $\{t_c > m\}$ for some (possibly new) constant B and all $c > 0$.

LEMMA 2.2. *Suppose that* $E\{X_1^r\} < \infty$, *where* $r \geqq 2$. *Then for* $y \geqq 2$,

$$P\{t_c > \lambda y\} \leqq y^{-\frac{1}{2}(2\alpha-1)r} \cdot o(\lambda^{-\frac{1}{2}r}),$$

where $o(\lambda^{-\frac{1}{2}r})$ *is uniform in* y.

PROOF. Let k be the greatest integer in λy. Then $t_c > \lambda y$ implies that $S_k - k\mu \geqq ck^\alpha L(k) - k\mu$. It is easily seen that $ck^\alpha L(k) - k\mu \geqq \frac{1}{2}k\mu[y^{\alpha-1} - 1]$ for $y \geqq 2$ and c sufficiently small. Thus, letting $S_k^* = (S_k - k\mu)/\tau k^{\frac{1}{2}}$, we have

$$P\{t_c > \lambda y\} \leqq P\{S_k^* \geqq \tfrac{1}{2}\mu\tau^{-1}k^{\frac{1}{2}}[y^{\alpha-1} - 1]\} \leqq B(k^{\frac{1}{2}} \cdot y^{\alpha-1})^{-r} \int_{A_k} |S_k^*|^r \, dP,$$

where A_k denotes the event that $|S_k^*| > \frac{1}{2}\mu\tau^{-1}k^{\frac{1}{2}}[2^{\alpha-1} - 1]$. The lemma now follows from the uniform integrability of $|S_k^*|^r$, which may be deduced from Theorem 2 of [17].

Let F denote the distribution function of X_1. In our next results we will impose the condition

$$F(x) \leqq Bx^a, \qquad \text{for all} \quad x > 0, \tag{2.5}$$

for some $B > 0$ and $a > 0$. Of course, if (2.5) holds for all sufficiently small x, then it holds for all x with a possibly new B but the same a.

LEMMA 2.3. *Suppose that* $E\{X_1^r\} < \infty$, *where* $r \geqq 2$, *and that* (2.5) *holds. Then for* $0 < \delta, \gamma < 1$, *we have* $P\{t_c \leqq \delta\lambda\} = O(c^{ma}) + O(\lambda^{-r\gamma/2})$ *as* $c \to 0$.

THEOREM 2.3. *Suppose that* $E\{X_1^r\} < \infty$, *where* $r \geqq 2$, *and that* (2.5) *holds. If*

$$0 < s < \min\{r, \tfrac{1}{2}(2\alpha - 1)r\} \qquad \text{and} \qquad ma > \tfrac{1}{2}\beta s, \tag{2.6}$$

then $|t_c^*|^s$ *are uniformly integrable.*

The proofs of Lemma 2.3 and Theorem 2.3 are technical. They will be given in Section 5.

THEOREM 2.4. *Suppose that* (2.5) *holds and that* $E\{X_1^r\} < \infty$ *for some* $r > 2$.

If $r(2\alpha - 1) > 4$ *and* $ma > \beta$, *then*

$$E\{t_c\} = \lambda + \beta\mu^{-1}\nu - \beta L_0 - \tfrac{1}{2}\alpha\beta^2\tau^2\mu^{-2} + o(1)$$

as $c \to 0$, *where* ν *is as in* (2.4).

PROOF. By Wald's lemma and some simple algebra, we have

$$\mu E\{t_c\} = cE\{t_c^{\alpha}L(t_c) - R_c\} = cE\{t_c^{\alpha}\} + \mu L_0 - \nu + o(1)\,. \tag{2.7}$$

Here we also use Theorem 2.2 and Lemma 2.1 to estimate $E\{R_c\}$ and Lemma 2.1 to estimate $E\{ct_c^{\alpha-1}\}$. Subtracting $\mu\lambda$ from both sides of (2.7) and expanding t_c^{α} in a Taylor series about λ^{α}, we now find that

$$\begin{aligned}\mu E\{t_c - \lambda\} &= cE\{t_c^{\alpha} - \lambda^{\alpha}\} + \mu L_0 - \nu + o(1)\\ &= \alpha\mu E\{t_c - \lambda\} + \tfrac{1}{2}\alpha(\alpha - 1)cE\{\lambda_1^{\alpha-2}(t_c - \lambda)^2\} + \mu L_0 - \nu + o(1)\,,\end{aligned}$$

where $|\lambda_1 - \lambda| \leqq |t_c - \lambda|$. Let $W = c\mu^{-1}\lambda_1^{\alpha-2}(t_c - \lambda)^2 = (\lambda_1\lambda^{-1})^{\alpha-2}t_c^{*2}$. Then W converges in distribution to $(\beta^2\tau^2\mu^{-2})\chi_1^2$, and

$$E\{t_c - \lambda\} = \beta\mu^{-1}\nu - \beta L_0 - \tfrac{1}{2}\alpha E\{W\} + o(1)\,,$$

so it will suffice to show that $\lim E\{W\} = \beta^2\tau_2\mu^{-2}$.

Suppose first that $\alpha \geqq 2$ and let A be the event that $t_c \leqq 2\lambda$. Then $\lim E\{WI_A\} = \beta^2\tau^2\mu^{-2}$ by Theorem 2.3, since $\lambda_1\lambda^{-1}$ is bounded on A and t_c^{*2} is uniformly integrable. Moreover, on A' we have $W \leqq \lambda^{1-\alpha}t_c^{\alpha}$, so that

$$\int_{A'} W\,dP \leqq \lambda \int_{\{t_c > 2\lambda\}} (\lambda^{-1}t_c)^{\alpha}\,dP\,,$$

which tends to zero as $c \to 0$ by Lemma 2.2.

For the case $\alpha < 2$, we show that $(\lambda_1 \cdot \lambda^{-1})^{\alpha-2}$ is bounded. Let A be the event that $t_c > \frac{1}{2}\lambda$. Then $(\lambda_1 \cdot \lambda^{-1})^{\alpha-2} \leqq 4^{2-\alpha}$ on A; and on A',

$$\begin{aligned}0 \geqq t_c^{\alpha} - \lambda^{\alpha} &= \alpha\lambda^{\alpha-1}(t_c - \lambda) + \tfrac{1}{2}\alpha(\alpha - 1)\lambda_1^{\alpha-2}(t_c - \lambda)^2\\ &> -\alpha\lambda^{\alpha} + \tfrac{1}{8}\alpha(\alpha - 1)\lambda_1^{\alpha-2} \cdot \lambda^2\,,\end{aligned}$$

or $(\lambda_1 \cdot \lambda^{-1})^{\alpha-2} \leqq 8\beta$. That $\lim E\{W\} = \beta^2\tau^2\mu^{-2}$ now follows from the uniform integrability of t_c^{*2}.

3. Sequential point estimation.

3.1. *The normal case.* Let $Y_1, Y_2, \cdots$ be i.i.d. normally distributed random p-vectors with unknown mean $\theta \in R^p$ and unknown, nonsingular covariance matrix Σ. We suppose that we may observe as many of $Y_1, Y_2, \cdots$ as we please and that if we stop with Y_n then we incur the loss

$$L_n = A||\bar{Y}_n - \theta||^2 + n\,,$$

where $||\cdot||$ denotes the Euclidean norm and $A > 0$. It is easily seen that $E\{L_n\} = An^{-1}\operatorname{tr}(\Sigma) + n$ is minimized by letting $n = n_0 = (A\operatorname{tr}(\Sigma))^{\frac{1}{2}}$ in which case $E\{L_{n_0}\} = 2n_0$. For the case of unknown Σ Robbins (1959) suggested the following sequential procedure. Let

$$\hat{\Sigma}_n = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} (Y_i - \bar{Y}_n)(Y_i - \bar{Y}_n)' , \qquad n \geqq 2 , \tag{3.1}$$

and

$$N = \inf\{n \geqq m : n > k_n(A \operatorname{tr}(\hat{\Sigma}_n))^{\frac{1}{2}}\} ,$$

where $m \geqq 2$ is the initial sample size and $k_n \to 1$ as $n \to \infty$.

The random variable N may be written in the form $t_c + 1$, where t_c is as in (1.1). To see this, let

$$W_k = \frac{1}{(k(k+1))^{\frac{1}{2}}} \{\sum_{i=1}^{k} Y_i - kY_{k+1}\} , \qquad k \geqq 1 , \tag{3.2}$$

so that $W_1, W_2, \cdots$ are independent random variables which are normally distributed with mean 0 and covariance matrix Σ, and observe that $(n-1)\hat{\Sigma}_n = W_1 W_1' + \cdots + W_{n-1} W_{n-1}'$ for $n \geqq 2$. Let $X_k = ||W_k||^2$ for $k \geqq 1$. Then

$$(n-1) \operatorname{tr}(\hat{\Sigma}_n) = X_1 + \cdots + X_{n-1} = S_{n-1} , \quad \text{say.}$$

Thus, letting $c = 1/A$ and $L(n) = (n+1)^2/n^2 k_{n+1}^2$, we find that $N = t_c + 1$, where $t_c = \inf\{n \geqq m - 1 : S_n < cn^3 L(n)\}$. We observe that the distribution of X_1 satisfies the standing assumptions of Section 2. It also satisfies (2.5) with $a = \frac{1}{2}p$. This is clear if $p = 1$. For the general case, let C be an orthogonal matrix for which $C\Sigma C'$ is diagonal and let $Z = CW_1$. Then $X_1 = ||Z||^2 = Z_1^2 + \cdots + Z_p^2$, so that $P\{X_1 \leqq x\} \leqq \prod_{i=1}^{p} P\{Z_i^2 \leqq x\}$. Relation (2.5) follows easily.

In the following, denote the mean and variance of X_1 by

$$\mu = \operatorname{tr}(\Sigma) \qquad \text{and} \qquad \tau^2 = 2 \operatorname{tr}(\Sigma^2) .$$

THEOREM 3.1. *Suppose that $k_n = 1 + \Delta n^{-1} + o(n^{-1})$ as $n \to \infty$. If $m > 1 + p^{-1}$, then*

$$E\{N\} = n_0 + \Delta + \tfrac{1}{2}\mu^{-1}\nu - \tfrac{3}{8}\mu^{-2}\tau^2 + o(1) \tag{3.3}$$

as $A \to \infty$, where ν is as in (2.4). *Moreover, if $m > 1 + 2p^{-1}$, then, as $A \to \infty$,*

$$E\{L_N\} = 2n_0 + \tfrac{1}{4}\mu^{-2}\tau^2 + o(1) . \tag{3.4}$$

PROOF. (3.3) is an immediate consequence of Theorem 2.4. Indeed, since X_1 has moments of all orders, the relevant condition is $\frac{1}{2}p(m-1) > \beta = \frac{1}{2}$, and this is equivalent to $m > 1 + p^{-1}$.

To establish (3.4) we use the fact that N and $N^{\frac{1}{2}}(\bar{Y}_N - \theta)$ are independent (see [13], page 1175). It follows that

$$E\{L_N\} = E\{AN^{-1} \operatorname{tr}(\Sigma) + N\} = 2n_0 + E\{N^{-1}(N - n_0)^2\} .$$

A simple application of Lemma 2.2 and Theorem 2.3 (with $s = 2$) then yields $\lim E\{N^{-1}(N - n_0)^2\} = \tau^2/4\mu^2$ as $A \to \infty$, to complete the proof.

Equation (3.4) asserts that the asymptotic regret of the sequential procedure with respect to the optimal procedure for the case of known Σ is $\tau^2/4\mu^2$. It is interesting that this regret is always less than or equal to $\frac{1}{2}$, and less than $\frac{1}{2}$ for $p \geqq 2$.

3.2. *The gamma case.* Now let $Y_1, Y_2, \cdots$ be i.i.d random variables with common density

$$f_\theta(y) = \Gamma(a)^{-1}\left(\frac{a}{\theta}\right)^a y^{a-1}\exp\left\{-\frac{ay}{\theta}\right\}, \qquad y > 0,$$

where $\theta > 0$ is unknown and $a > 0$ is known. Again, we suppose that we may observe as many of $Y_1, Y_2, \cdots$ as we please, and that if we stop with Y_n then we incur the loss $L_n = A|\bar{Y}_n - \theta|^2 + n$, where $A > 0$. As above, it is easily seen that $E\{L_n\} = A\theta^2/an + n$ is minimized by letting $n = n_0 = \theta(Aa^{-1})^{\frac{1}{2}}$ in which case the resulting risk is $E\{L_{n_0}\} = 2n_0$. Of course, n_0 is unknown. Starr and Woodroofe (1972) proposed the sequential procedure

$$N = \inf\{n \geqq m : n > \bar{Y}_n(Aa^{-1})^{\frac{1}{2}}\}.$$

It is easily seen that N is of the form t_c with $X_i = Y_i/\theta$, $i \geqq 1$, $c = 1/n_0$, $\alpha = 2$, and $L = 1$. The mean and variance of X_1 are $\mu = 1$ and $\tau^2 = 1/a$, respectively.

Theorem 3.2, below, provides a partial answer to a question raised in [16], pages 1152–1153. There a Monte Carlo study of $E\{N\}$ and $E\{L_N\}$ was conducted and a few negative values of $R = E\{L_N\} - 2n_0$ were found. This raised the possibility that $E\{L_N\} < 2n_0$ for some values of n_0. Theorem 3.2 shows that, in fact, $E\{L_N\} > 2n_0$ for all sufficiently large values of n_0.

THEOREM 3.2. *If $ma > 1$, then*

$$E\{N\} = n_0 + \nu - a^{-1} + o(1) \tag{3.5}$$

as $n_0 \to \infty$, where ν is as in (2.4). *Moreover, if $ma > 2$, then as $n_0 \to \infty$,*

$$E\{L_N\} = 2n_0 + 3a^{-1} + o(1). \tag{3.6}$$

PROOF. Again, (3.5) is an immediate consequence of Theorem 2.4. The proof of (3.6), however, is more difficult than that of (3.4), since N and $(N(\bar{Y}_N - \theta))^{\frac{1}{2}}$ are no longer independent.

In the proof of (3.6) we will write t for N, and we recall that t is of the form t_c. Letting $S_n = X_1 + \cdots + X_n$, we find

$$A(\bar{Y}_t - \theta)^2 = a(S_t - t)^2 + a[c^{-2}t^{-2} - 1](S_t - t)^2.$$

By Theorem 1 of [4], $E\{(S_t - t)^2\} = a^{-1}E\{t\}$. Thus, $E\{L_t\} = 2E\{t\} + E\{v_t\}$, where $v_t = a[c^{-2}t^{-2} - 1](S_t - t)^2$. Observe that

$$v_t = a(1 - c^2t^2)(S_t - t)^2 + ac^{-2}t^{-2}(1 - c^2t^2)^2(S_t - t)^2 = I + II, \quad \text{say.}$$

We estimate I and II separately. By definition of t, we have $ct^2 - t = S_t - t + R_c$, so that

$$\begin{aligned}(1 - c^2t^2) &= (1 + ct)t^{-1}(t - ct^2)\\ &= (1 + ct)t^{-1}[(t - S_t) - R_c] = -2c(S_t - t) + O_p(c)\end{aligned}$$

as $c \to 0$. Thus, the asymptotic distribution of II is that of $4aZ^4$, where Z has the normal distribution with mean 0 and variance a^{-1}. Moreover, it can be shown that II is uniformly integrable (see below), and it follows that $\lim E\{II\} = 12a^{-1}$.

To estimate I, we write $(1 - c^2t^2) = 2c(t - ct^2) + (1 - ct)^2$ and $t - ct^2 = (t - S_t) - R_c$. Thus,

$$I = -2ca(S_t - t)^3 - 2ca(S_t - t)^2 R_c + a(S_t - t)^2(1 - ct)^2 = -I_1 - I_2 + I_3 ,$$

say. The terms I_2 and I_3 may be estimated as above. The results are that $\lim E\{I_2\} = 2\nu$ and $\lim E\{I_3\} = 3a^{-1}$ as $c \to 0$. To estimate I_1, we use Theorem 8 of [4], which asserts that

$$\begin{aligned} 2caE\{(S_t - t)^3\} &= 4a^{-1}cE\{t\} + 6cE\{t(S_t - t)\} \\ &= 6cE\{t(S_t - t)\} + 4a^{-1} + o(1) . \end{aligned}$$

Finally,

$$\begin{aligned} ct(S_t - t) &= ct^2(ct - 1) - ctR_c \\ &= (t - c^{-1}) + c(t^2 - c^{-2})(ct - 1) - ctR_c = I_{11} + I_{12} - I_{13} , \end{aligned}$$

say. $E\{I_{11}\}$ is given by (3.5) and $E\{I_{13}\} \to \nu$ by Theorems 2.1 and 2.2 and Lemma 2.1. The estimation of I_{12} is similar to that of II. The result is that $\lim E\{I_{12}\} = 2a^{-1}$. Collecting terms, we find that

$$E\{v_t\} = 5a^{-1} - 2\nu + o(1) ,$$

from which (3.6) follows.

It remains to demonstrate the uniform integrability of the terms II, I_2, I_3, and I_{13}. We give the details only for II, since the treatment of the others is similar. Let A be the event that $2t < n_0$. On A, $S_t - t < ct^2 - t$ is negative, so that $(S_t - t)^2 \leqq t^2$. It follows that

$$\int_A II \, dP \leqq ac^{-2}P\{t \leqq \tfrac{1}{2}n_0\} ,$$

which tends to zero as $c \to 0$ by Lemma 2.3. On A', we have

$$\begin{aligned} II &\leqq 8a(ct^2 - t)^2(1 - c^2t^2)^2 + 8aR_c^2(1 - c^2t^2)^2 \\ &\leqq 8ac^2t^2(1 + ct)^2t^{*4} + 8acR_c^2(1 + ct)^2t_c^{*2} = II_1 + II_2 , \quad \text{say.} \end{aligned}$$

By Lemma 2.1, all powers of ct and R_c are uniformly integrable; and by Theorem 2.3, $|t^*|^s$ is uniformly integrable for some $s > 4$. Thus, by Hölder's inequality, $E\{II_2^2\} = O(c^2)$ and $E\{II_1^q\} = O(1)$ for some $q < 1$. The uniform integrability of II follows (see Loève (1963), page 184).

4. Fixed width confidence intervals. Let $Y_1, Y_2, \cdots$ be independent random variables which are normally distributed with unknown mean θ and variance $\sigma^2 > 0$. We desire a confidence interval for θ of fixed width $2d$, where $d > 0$. Let γ be the nominal confidence coefficient and $z = \Phi^{-1}((1 + \gamma)/2)$. If σ^2 were known, then we could simply take $n \geqq n_0 = z^2\sigma^2/d^2$ observations and use $I_n = (\bar{Y}_n - d, \bar{Y}_n + d)$. For the case of unknown σ^2, we consider the following sequential procedure: let $z_n \to z$ as $n \to \infty$, let

$$N = \inf\{n \geqq m : n > z_n^2\hat{\sigma}_n^2/d^2\}, \tag{4.1}$$

where $m \geqq 2$ and $\hat{\sigma}_n^2$ is as in (3.1), and use $I_N = (\bar{Y}_N - d, \bar{Y}_N + d)$. As in Section 3.1, it is easy to see that $N = t_c + 1$, where $t_c = \inf\{n \geqq m - 1 : S_n < cn^2L(n)\}$. Here $c = 1/n_0$, $L(n) = (n+1)z^2/nz_{n+1}^2$, and $X_k = W_k^2/\sigma^2$ have the chi-square distribution on one degree of freedom. Also as in Section 3.1, it can be shown that

$$P\{\theta \in I_N\} = E\{\psi(z^2n_0^{-1}N)\}, \tag{4.2}$$

where $\psi(x) = 2\Phi(x^{\frac{1}{2}}) - 1$. Finally, it is easy to see that the distribution of N depends only on n_0, so that (4.2) depends only on n_0.

It is known that $P\{\theta \in I_N\} \to \gamma$ as $n_0 \to \infty$. We now supplement this information by proving

THEOREM 4.1. *Suppose that $m \geqq 4$ and that $z_n = z\{1 + \Delta n^{-1} + o(n^{-1})\}$ as $n \to \infty$. Then*

$$E\{N\} = n_0 + \nu + 2\Delta - 2 + o(1) \tag{4.3}$$

as $n_0 \to \infty$. Moreover, if $m \geqq 7$, then as $n_0 \to \infty$

$$P\{\theta \in I_N\} = \gamma + n_0^{-1}\{z^2\psi'(z^2)[\nu + 2\Delta - 2] + z^4\psi''(z^2)\} + o(n_0^{-1}). \tag{4.4}$$

PROOF. Equation (4.3) is an immediate consequence of Theorem 2.4. To prove (4.4), we expand ψ in a Taylor series about z^2, and find

$$P\{\theta \in I_N\} - \gamma = n_0^{-1}z^2\psi'(z^2)E\{N - n_0\} + \tfrac{1}{2}n_0^{-1}z^4E\{\psi''(W)n_0^{-1}(N - n_0)^2\},$$

where $|z^2 - W| \leqq z^2|n_0^{-1}N - 1|$. Let $Z = \psi''(W)n_0^{-1}(N - n_0)^2$. Then the asymptotic distribution of Z is $2\psi''(z^2)\chi_1^2$. Let A be the event that $2N > n_0$; then $\lim E\{ZI_A\} = 2\psi''(z^2)$ by Theorem 2.3. Moreover, there is a constant B for which $\psi''(w) \leqq Bw^{-\frac{3}{2}}$ for all $w > 0$, so that

$$\int_{A'} Z\,dP \leqq z^{-3} \int_{A'} (N^{-1}n_0)^{\frac{3}{2}}\,dP,$$

which tends to zero as $c \to 0$ by Lemma 2.3. The theorem follows.

As a corollary, we see that $P\{\theta \in I_N\} > \gamma$ for all sufficiently large values of n_0 if

$$\nu + 2\Delta > 2 - z^2\psi''(z^2)/\psi'(z^2) = 2 + \tfrac{1}{2}(1 + z^2). \tag{4.5}$$

It is possible to phrase (4.4) in a different manner. Let N be defined by (4.1) and consider the procedure which takes $N + k$ observations where k is a fixed positive integer. Simons (1968) showed that there exists an integer k for which $P\{\theta \in I_{N+k}\} > \gamma$ for all n_0. Our techniques will show that (if the hypotheses of Theorem 4.1 are satisfied, then) $P\{\theta \in I_{N+k}\}$ is given by the right side of (4.4), but with the term 2Δ of (4.4) replaced by $2\Delta + k$. In particular, $P\{\theta \in I_{N+k}\} > \gamma$ for all sufficiently large values of n_0 if $\nu + 2\Delta + k$ exceeds the right side of (4.5).

If z_n is the $\frac{1}{2}(1+\gamma)$th fractile of the t-distribution on n degrees of freedom, then $z_n = z\{1 + \Delta_0 n^{-1} + o(n^{-1})\}$ with $\Delta_0 = (1+z^2)/4$. Moreover, from Table 2.1 we find $\nu \doteq .82$, so that $\nu + 2\Delta_0 < 2 + \frac{1}{2}(1+z^2)$. Thus, for this choice of z_n, we have $P\{\theta \in I_N\} < \gamma$ for all sufficiently large values of n_0. However, we have $P\{\theta \in I_{N+k}\} > \gamma$ for all large n_0 if

$$\nu + k > 2\,. \tag{4.6}$$

It is interesting that the condition (4.6) does not involve γ.

Starr (1966a) evaluated some exact coverage probabilities for a related procedure which allows stopping only after an odd number of observations. His computations indicate that (4.6) is not sufficient to make $P\{\theta \in I_N\} > \gamma$ for all n_0; but his computations were done with $m \leqq 5$.

5. Uniform integrability. In this section we prove Lemma 2.3 and Theorem 2.3. We use the standing assumptions of Section 2 without comment.

LEMMA 5.1. *Suppose that* (2.5) *holds. Then* $P\{t_c \leqq \lambda^\gamma\} = O(c^{ma})$ *as* $c \to 0$ *for any* $\gamma < 1$. *Here a is as in* (2.5).

PROOF. Let F_k denote the distribution function of S_k. Then (2.5) and a simple induction yield

$$F_k(x) \leqq \Gamma(1+ka)^{-1}B_1^k \cdot x^{ka}\,, \qquad x > 0,\, k \geqq 1\,,$$

where $B_1 = B\Gamma(1+a)$ with a and B as in (2.5). It follows easily that $P\{t_c = k\} \leqq P\{S_k \leqq ck^\alpha L(k)\} = O(c^{ka})$ for every $k \geqq 1$. In particular, $P\{t_c \leqq n\} = O(c^{ma})$ for every integer n. Now

$$P\{n < t_c \leqq \lambda^\gamma\} \leqq \sum_{n<k\leqq\lambda^\gamma} \Gamma(1+ka)^{-1}B_1^k[ck^\alpha L(k)]^{ka} \tag{5.1}$$

and by Stirling's formula

$$\Gamma(1+ka)^{-1}B_1^k k^{ka\alpha} \leqq ((2\pi k)^{-1})^{\frac{1}{2}} B_1^k a^{-ak} e^{ak} \cdot k^{(\alpha-1)ka}\,. \tag{5.2}$$

Since the right side of (5.2) does not exceed $B_2^k \cdot c^{-k\gamma a}$ for $n \leqq k \leqq \lambda^\gamma$ and n sufficiently large with $B_2 = 2B_1 a^{-a} e^a \mu^{\gamma a}$, it follows that the right side of (5.1) is of order $c^{na(1-\gamma)}$ as $c \to 0$. The lemma follows.

We will now prove Lemma 2.3, which asserts that if (2.5) holds and if $E\{X_1^r\} < \infty$ with $r \geqq 2$, then $P\{t_c \leqq \delta\lambda\} = O(c^{ma}) + O(\lambda^{-r\gamma/2})$ as $c \to 0$ for all $0 < \delta,\, \gamma < 1$. By Lemma 5.1, it will suffice to show that $P\{\lambda^\gamma < t_c \leqq \delta\lambda\} = O(\lambda^{-r\gamma/2})$ for all $0 < \delta,\, \gamma < 1$. For $\lambda^\gamma < k \leqq \delta\lambda$, we have $ck^\alpha L(k) - k\mu \leqq k\mu\{\delta^{\alpha-1}L(k) - 1\}$, which is $\leqq -k\mu\varepsilon$ for some $\varepsilon > 0$ for c sufficiently small. It follows that

$$P\{\lambda^\gamma \leqq t_c \leqq \delta\lambda\} \leqq P\{|\bar{X}_k - \mu| \geqq \varepsilon\mu\,, \text{ for some } k \geqq \lambda^\gamma\}\,,$$

which is $O(\lambda^{-r\gamma/2})$ by the martingale inequality.

We will now prove Theorem 2.3, which asserts that if $E\{X_1^r\} < \infty$ with $r \geqq 2$, if $s < \min\{r, \frac{1}{2}r(2\alpha - 1)\}$, and if (2.5) holds with $ma > \beta s/2$, then $|t_c^*|^s$ is uniformly integrable. By Lemmas 2.2 and 2.3, it will suffice to show that there is a function J for which $y^{s-1}J(y)$ is integrable over $(0, \infty)$ and $P\{|t_c^*| > y, \delta\lambda \leqq t_c \leqq 2\lambda\} \leqq J(y)$ for all sufficiently large y and sufficiently small c. Let

$$H(y) = \sup_{k\geqq 1} P\left\{\left|\frac{S_k - k\mu}{k^{\frac{1}{2}}}\right| > y\right\}, \qquad y > 0.$$

Then it follows from Markov's inequality that $y^{s-1}H(y)$ is integrable over $(0, \infty)$. We will also need the easily verified inequality

$$P\{\min_{k\leqq n} S_k - k\mu < -y\} \leqq BP\{S_n - n\mu < -y\}, \qquad y > 0,$$

where $B^{-1} = \inf_{k\geqq 1} P\{S_k - k\mu < 0\}$.

Let $\alpha^{-\beta} < \delta < 1$, let $h(x) = cx^\alpha - \mu x$ for $x > 0$, and let n_0 be the greatest integer in $\lambda - y(\lambda)^{\frac{1}{2}}$. Observe that h is convex and that $h'(\delta\lambda) = \mu(\alpha\delta^{\alpha-1} - 1) = \varepsilon$, say, is independent of c. Now $t_c \geqq \delta\lambda$ and $t_c^* < -y$ imply that $S_k - k\mu < h(k) + ck^\alpha[L(k) - 1]$ for some k, $\delta\lambda \leqq k \leqq n_0$. For c sufficiently small this implies that $S_k - k\mu < h(n_0) + 2\mu|L_0|$ for some $k \leqq n_0$. Now

$$h(n_0) = h(n_0) - h(\lambda) \leqq h'(n_0)(n_0 - \lambda) \leqq -yn_0^{\frac{1}{2}}h'(\delta\lambda).$$

Thus,

$$\begin{aligned} P\{t_c \geqq \delta\lambda, t_c^* < -y\} &\leqq BP\{S_{n_0} - n_0\mu < -\varepsilon y n_0^{\frac{1}{2}} + 2\mu|L_0|\} \\ &\leqq BH(\varepsilon y - 2\mu|L_0|) \end{aligned}$$

for large y and small c. A similar argument will show that

$$P\{t_c \leqq 2\lambda, t_c^* > y\} \leqq H[\tfrac{1}{6}(\alpha - 1)\mu y - 2\mu|L_0|]$$

for $y \geqq 1$ and c sufficiently small to complete the proof.

Acknowledgment. I want to thank Norman Starr and Charles Keller for several interesting conversations; and I want to thank Tze Leung Lai and David Siegmund for allowing me to see their paper prior to its publication. I also want to thank the referee for a careful reading of my manuscript.

REFERENCES

[1] Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

[2] Bhattacharya, P. K. and Mallik, A. (1973). Asymptotic normality of the stopping times of some sequential procedures. *Ann. Statist.* **1** 1203–1211.

[3] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

[4] Chow, Y. S., Robbins, H. and Teicher, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

[5] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[6] Feller, William (1966). *An Introduction to Probability and Its Applications,* **2**. Wiley, New York.

[7] Ghosh, M., Sinha, B. and Mulckopadhyay, N. (1976). Multivariate sequential point estimation. *J. Multirariate Anal.* **6** 281–294.

[8] Lai, T. L. and Siegmund, David (1977). A nonlinear renewal theory. *Ann. Statist.* **5** 946–954.

[9] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics* (Harold Cramér Volume), Almquist and Wiksell, Upsala.

[10] SIMONS, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.* **39** 1946–1952.

[11] SPITZER, F. (1960). A Tauberian theorem and its probability interpretation. *Trans. Amer. Math. Soc.* **94** 150–169.

[12] STARR, N. (1966a). The performance of a sequential procedure for the fixed width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

[13] STARR, N. (1966b). On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* **37** 1173–1185.

[14] STARR, N. and WOODROOFE, M. (1968). Remarks on a stopping time. *Proc. Nat. Acad. Sci. USA* **61** 1215–1218.

[15] STARR, N. and WOODROOFE, M. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. USA* **63** 285–288.

[16] STARR, N. and WOODROOFE, M. (1972). Further remarks on sequential point estimation: the exponential case. *Ann. Math. Statist.* **43** 1147–1154.

[17] VON BAHR, B. (1965). On convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36** 808–818.

[18] WOODROOFE, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probability* **4** 67–80.

DEPARTMENT OF MATHEMATICS
M.I.T.
CAMBRIDGE, MASSACHUSETTS 02139