THE PERFORMANCE OF A SEQUENTIAL PROCEDURE FOR THE FIXED-WIDTH INTERVAL ESTIMATION OF THE MEAN[1]

By Norman Starr

Columbia University and the University of Minnesota

1. The problem. Let

$$X_1, X_2, \cdots \tag{1}$$

be independent $N(\theta, \sigma^2)$ with $\sigma^2 < \infty$. Suppose we require a confidence interval for θ of width at most $2d$ $(d > 0)$ and with probability of coverage at least α $(0 < \alpha < 1)$, irrespective of the values of θ and σ^2. Define for $n \geqq 2$

$$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i, \qquad S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

and for $x > 0$ let

$$\varphi(x) = (2\pi)^{-\frac{1}{2}} \int_{-x}^{x} e^{-t^2/2}\, dt;$$

a = constant for which $\varphi(a) = \alpha$, and

$$\varphi_n(x) = \Gamma((n+1)/2)/(n\pi)^{\frac{1}{2}}\Gamma(n/2) \int_{-x}^{x} (1 + t^2/2)^{-(n+1)/2}\, dt; \tag{2}$$

a_n = constant for which $\varphi_n(a_n) = \alpha$ $(n = 1, 2, \cdots)$; then

$$\lim_{n\to\infty} a_n = a. \tag{3}$$

Observe that if σ is *known* a confidence interval $I_n = [\bar{X}_n - d, \bar{X}_n + d]$ for θ of width $2d$ and with coverage probability $\geqq \alpha$ is assured provided n is chosen so that

$$n \geqq a^2\sigma^2/d^2, \tag{4}$$

since then $P(\theta \varepsilon I_n) = \varphi(n^{\frac{1}{2}} d/\sigma) \geqq \alpha$. However, it is clear that no procedure based on a fixed number n of observations of (1) satisfies the requirements when σ is *unknown*. In this circumstance one recourse is to two-stage sampling [10]. The Stein procedure leads to an n which approximately satisfies (4) with σ^2 estimated from the initial sample of size n_1 (and a increased to a_{n_1} to reflect the limited degrees of freedom). It seems intuitively inefficient not to utilize all of the sample; we shall investigate the performance of a sequential procedure Λ which does just this, leading to an n satisfying (4) with σ^2 estimated on $n - 1$ degrees of freedom. Accordingly, we prescribe the rule

Λ: Observe the sequence (1) term by term, stopping with X_N, where

$$N \text{ is the first integer } n \geqq n_0 \text{ such that } S_n^2 \leqq n\, d^2/a_{n-1}^2, \tag{5}$$

Received 9 November 1964; revised 20 August 1965.

[1] Research supported in part by National Science Foundation Grant NSF-GP-2074 at Columbia University. Revised at the University of Minnesota with the partial support of National Science Foundation Grant NSF-GP-3813.

with $n_0 \geqq 2$ a fixed integer; then form the interval $I_N = [\bar{X}_N - d, \bar{X}_N + d]$ which has the required width.

Heuristically, the interval I_N will be an approximate solution of the problem:

$$(6) \qquad P(\theta \varepsilon I_N) \cong \alpha.$$

Furthermore, the expected sample size using Λ will be

$$(7) \qquad EN \cong a^2\sigma^2/d^2.$$

To see why (6) should hold we remark that if we set

$$J_n = [\bar{X}_n - (a_{n-1}S_n/n^{\frac{1}{2}}), \bar{X}_n + (a_{n-1}S_n/n^{\frac{1}{2}})],$$

then for fixed values of θ, σ, and n

$$(8) \qquad P(\theta \varepsilon J_n) = P[n^{\frac{1}{2}} |\bar{X}_n - \theta|/S_n) \leqq a_{n-1}] = \varphi_{n-1}(a_{n-1}) = \alpha.$$

Moreover, the N defined by (5) is such that I_N contains J_N; hence "neglecting the excess", $P(\theta \varepsilon I_N) \cong P(\theta \varepsilon J_N)$. If we could infer from (8) that $P(\theta \varepsilon J_N) \cong \alpha$, (6) would follow. Since $\lim_{n\to\infty} S_n^2 = \sigma^2$ with probability 1, it follows from (5) by neglecting the excess that $\sigma^2 \cong d^2N/a^2$, at least for N large; thus EN should satisfy (7). Finally, we remark that a comparison of (4) and (7) suggests that not much should be lost due to ignorance of σ if we use Λ instead of a sample of fixed size.

It would therefore seem to be of considerable interest to establish rigorously the values of the functions

$$(9) \qquad C(\lambda) = P(\theta \varepsilon I_N), \qquad D(\lambda) = EN$$

for values $0 < \lambda < \infty$ of the parameter

$$(10) \qquad \lambda = \sigma/d$$

upon which C and D are easily seen to depend, and thereby to determine whether (6) and (7) are reasonably exact, for all σ. Thus, we define

$$(11) \qquad \tau_\lambda = C(\lambda)/\alpha,$$

the ratio of the coverage probability using Λ in ignorance of σ to the prescribed coverage probability α, and

$$(12) \qquad \eta_\lambda = D(\lambda)/a^2\lambda^2,$$

the ratio of the average sample size using Λ in ignorance of σ to (approximately) the smallest value of n satisfying (4) which would be used when σ is known. Following the terminology of [3], we shall say that τ_λ and η_λ are measures, at the parameter value λ, of the *consistency* and *efficiency* of Λ, respectively. Therefore, in order to implement practically a recommendation of Λ we should establish that the procedure is "more or less" consistent ($\tau_\lambda \cong 1$) and efficient ($\eta_\lambda \cong 1$), for all λ.

Λ appears to have been first discussed by Stein [10], who stated the second order limiting performance of the procedure. Subsequently, the consistent and efficient asymptotic character of Λ was proved along somewhat different lines in [1], [2] and [5]. The computation of C and D for several moderate values of λ was carried out by Ray [7] using a slight modification of (5) for ease of computation; Ray incorrectly conjectured on the basis of his fragmentary computations that this modified version of Λ is asymptotically inconsistent.

However, while Λ has attracted theoretical attention, it appears that the procedure has been scrupulously avoided by practicing statisticians; owing perhaps, at least in part, to the fact that the several investigations of its performance have been only partial. The object of this article is to confirm that the procedure (together with several variants of interest) is indeed reasonably consistent and efficient, for *all* σ. Moreover, we will show that while Λ has a slightly reduced minimum probability of coverage compared with the Stein procedure [10], it is always more efficient than two-stage sampling; the difference in efficiencies being sizeable whenever, in ignorance of the variance, the first stage sample size is chosen poorly.

Preliminary theory is given in Section 2, following [7] and [8] in certain respects. Asymptotic results for small and large λ are given in Sections 3 and 4, respectively, while computational results for intermediate λ are presented in Section 5. Comparisons with Stein's procedure are presented in Section 6.

2. Preliminary theory. At the outset we remark that N is a *geniune* sampling variable; that is,

LEMMA 2.1. $P(N < \infty) = 1$.

PROOF. By the strong law of large numbers, $\lim_{n\to\infty} S_n^2 = \sigma^2$ with probability 1, which implies

$$P(N = \infty) = P\{(S_n^2/n) > (d^2/a_{n-1}^2) \quad \text{for all} \quad n \geqq n_0\} = 0.$$

Define the $N(0, 1)$ random variables $W_n = (\sum_{i=1}^n X_i - nX_{n+1})/\sigma[n(n+1)]^{\frac{1}{2}}$, $(n \geqq 1)$, and observe that $(n-1)S_n^2/\sigma^2 = \sum_{i=1}^{n-1} W_i^2$, $(n \geqq 2)$. Therefore (5) can be rewritten in the form

(13) N is the first integer $n \geqq n_0$ such that $\sum_{i=1}^{n-1} W_i^2 \leqq n(n-1)/\lambda^2 a_{n-1}^2$,

and the probability distribution of N is defined for $n \geqq n_0$ by

$$p_n(\lambda) = P(N = n) = P\{\textstyle\sum_{i=1}^k W_i^2 \leqq (k+1)k/\lambda^2 a_k^2 \text{ for } k = n-1, \text{ but not for any } n_0 - 1 \leqq k < n-1\}. \tag{14}$$

The joint probability density of $\{W_i^2; i = 1, \cdots, n-1\}$ is easily seen to be

$$dF(W_1^2, \cdots, W_{n-1}^2) = [2^{(n-1)/2}\Gamma^{n-1}(\tfrac{1}{2})]^{-1} \exp(-\tfrac{1}{2}\textstyle\sum_{i=1}^{n-1} w_i^2)[\prod_{i=1}^{n-1} w_i^2]^{-\frac{1}{2}}\prod_{i=1}^{n-1} d(w_i^2)$$

for $0 \leqq w_i$; $i = 1, \cdots, n-1$. Noting that

$$V_k = \textstyle\sum_{i=1}^k W_i^2 \qquad (k \geqq 1) \tag{15}$$

has the chi-squared distribution with k degrees of freedom, the joint probability density of $\{V_k ; k = 1, \cdots, n-1\}$ is

$$dF(V_1, \cdots V_{n-1})$$
$$= [2^{(n-1)/2}\Gamma^{n-1}(\tfrac{1}{2})]^{-1} \exp\left(-\tfrac{1}{2}v_{n-1}\right)[v_1 \textstyle\prod_{i=2}^{n-1} (v_i - v_{i-1})]^{-\frac{1}{2}} \textstyle\prod_{i=1}^{n-1} dv_i$$

for $0 \leqq v_1 \leqq v_2 \leqq \cdots \leqq v_{n-1}$. Therefore, subject to Λ, the probability that sampling is terminated at the nth observation is from (14) for $n \geqq n_0$

$$(16) \quad p_n(\lambda) = \int \cdots \int_{[0,\infty)} \int_{B_{n_0-1}} \int_{B_{n_0}} \cdots \int_{B_{n-2}} \int_{\bar{B}_{n-1}} dF(V_1, \cdots V_{n-1})$$

where for $n_0 - 1 \leqq k \leqq n - 1$, B_k is the interval $((k + 1)k/\lambda^2 a_k^2, \infty)$, $\bar{B}_k$ is the complement of B_k, and where the multiple integral over $[0, \infty)$ represents an $n_0 - 2$ fold integration while the remainder represents an $n - n_0 + 1$ fold integration. Then with $p_n(\lambda)$ defined by (16), from (9) we have

$$(17) \qquad D(\lambda) = \textstyle\sum_{n=n_0}^{\infty} np_n(\lambda),$$

and, since $\bar{X}_n$ is independent of the vector $(W_1, \cdots, W_{n-1})$,

$$\begin{aligned} C(\lambda) &= \textstyle\sum_{n=1}^{\infty} p_n(\lambda)P(\theta \varepsilon I_N \mid N = n) \\ (18) \qquad &= \textstyle\sum_{n=n_0}^{\infty} p_n(\lambda)P(n^{\frac{1}{2}}|\bar{X}_n - \theta|/\sigma \leqq n^{\frac{1}{2}} d/\sigma \mid N = n) \\ &= \textstyle\sum_{n=n_0}^{\infty} p_n(\lambda)\varphi(n^{\frac{1}{2}}/\lambda). \end{aligned}$$

With this background we are prepared to provide a conspectus of the performance of the rule Λ.

3. Theory for small λ.

THEOREM 3.1. *With λ defined by* (10),

$$(19) \qquad \lim_{\lambda\to 0} P(N = n_0) = 1,$$

$$(20) \qquad \lim_{\lambda\to 0} C(\lambda) = 1,$$

$$(21) \qquad \lim_{\lambda\to 0} D(\lambda) = n_0.$$

Relation (19) follows immediately from (14). Since from (18) and (19)

$$\lim_{\lambda\to 0} C(\lambda) = \lim_{\lambda\to 0}\textstyle\sum_{n=n_0}^{\infty} p_n(\lambda)\varphi(n^{\frac{1}{2}}/\lambda) = \lim_{\lambda\to 0}\varphi(n_0^{\frac{1}{2}}/\lambda) = \varphi(\infty) = 1,$$

relation (20) is established.

To prove (21) we require

LEMMA 2.1. *Define* $S_\lambda = \sum_{n=n_0-1}^{\infty} P(V_n/n > (n + 1)/\lambda^2 a_n^2)$, *where* V_n *is defined in* (15). *Then*

$$(22) \qquad \lim_{\lambda\to 0} S_\lambda = 0.$$

PROOF OF LEMMA. Setting $A = 1 + \sup_{n\geqq 1} a_n$, we have for $\lambda < A^{-2}$

$$\begin{aligned} (23) \quad S_\lambda &\leqq \textstyle\sum_{n=n_0-1}^{\infty} P\{(V_n/n) - 1 > [(n + 1)A^2/\lambda a_n^2] - 1\} \\ &\leqq \textstyle\sum_{n=n_0-1}^{\infty} P((V_n/n) - 1 > n/\lambda). \end{aligned}$$

Noting that $E(V_n - n)^2 = 2n$, and applying the Chebyshev inequality, we obtain from (23)

$$S_\lambda \leqq \sum_{n=n_0-1}^{\infty} [\lambda^2 E((V_n/n) - 1)^2/n^2] = 2\lambda^2 \sum_{n=n_0-1}^{\infty} n^{-3},$$

from which it follows that (22) holds.

To establish (21) we remark that

$$(24)\ D(\lambda) = \sum_{n=1}^{\infty} P(N \geqq n) = \sum_{n=1}^{n_0} P(N \geqq n) + \sum_{n=n_0-1}^{\infty} P(N > n+1)$$
$$= n_0 + \sum_{n=n_0-1}^{\infty} P(N > n + 1).$$

Since it is easily seen from (14) that $P(N > n + 1) = P(\sum_{i=1}^{k} W_i^2 > (k + 1)k/\lambda^2 a_k^2$ for every $n_0 - 1 \leqq k \leqq n) \leqq P(V_n/n > (n + 1)/\lambda^2 a_n^2)$, we have

$$(25)\qquad \lim_{\lambda\to 0} \sum_{n=n_0-1}^{\infty} P(N > n + 1) \leqq \lim_{\lambda\to 0} S_\lambda,$$

and (21) follows by combining (24) and (25) and applying Lemma 2.1.

4. Theory for large λ. Because $p_n(\lambda)$ given by (16) is relatively intractable, the computations of C and D are extremely difficult; to simplify them, modify Λ to Λ^* by amending definition (5) to read

(5*) N^* is the first *odd* integer $n \geqq n_0^*$ such that $S_n^2 \leqq n\, d^2/a_{n-1}^2$,

with n_0^* a fixed odd integer $\geqq 3$. Then, for the modified rule Λ^*, in analogy with (9), (11) and (12), define

$$C^*(\lambda) = P(\theta \varepsilon I_{N^*}) \quad\text{and}\quad D^*(\lambda) = EN^*,$$
$$\tau_\lambda^* = C^*(\lambda)/\alpha, \quad\text{and}$$
$$\eta_\lambda^* = D^*(\lambda)/a^2\lambda^2.$$

The computations of C^* and D^* have been carried out for $\alpha = .95$ and $\lambda = 0.5, 1, 1.25, 2, 2.5$ (with $n_0^* = 3$) by W. D. Ray [7], and are summarized in Table 1.

We observe that as λ increases the values of $C^*(\lambda)$ appear to steadily decrease; indeed Ray conjectures ([7], p. 240) that as λ becomes infinite, $C^*(\lambda)$ "appears to tend to a value smaller than 95 per cent, in contradistinction to Anscombe's result."

The reference is to a theorem first stated by Stein [10] and proved in somewhat

TABLE 1
Λ^*, $n_0^* = 3$, $\alpha = .95$, $a = 1.96$

λ	0.5	1.0	1.25	2.0	2.5
$C^*(\lambda)$	.99975	.9775	.941	.931	.929
$D^*(\lambda)$	4.1	6.6	8.4	16.7	24.9
$a^2\lambda^2$	.96	3.84	6.0	15.4	24.0

different forms by Anscombe [1], [2] and by Gleser, Robbins, and Starr [5]. The version proved in [5] is as follows:

THEOREM 4.1. *Let N be the sampling variable defined by* (5), *where* $\{a_n\}$ *is any sequence of positive constants (not necessarily those defined in* (2)) *such that* (3) *holds. Then with* λ *defined by* (10), τ_λ *by* (11), *and* η_λ *by* (12),

$$(26) \qquad \lim_{\lambda\to\infty} \tau_\lambda = 1 \qquad (\textit{asymptotic consistency}),$$

$$(27) \qquad \lim_{\lambda\to\infty} \eta_\lambda = 1 \qquad (\textit{asymptotic efficiency}).$$

Moreover, (26) *and* (27) *hold when* N, τ_λ, *and* η_λ *are replaced by their starred equivalents.*

Now Ray reasons that while his conjecture is *apparently* inimical to (26) there are, after all, differences in the stipulations of Λ, for he is willing to admit (26), and of Λ^*, which dictates his computations; in particular the starting sample sizes n_0^*(equal to 3 in Ray's computations) and n_0 (equal to 2 in Anscombe's [1] proof of (26)) differ.

To allay these doubts regarding the asymptotic consistency of the general procedure it is necessary to stress two points.

(i) Theorem 4.1 holds irrespective of the choice of the starting sample size $n_0 \geqq 2$.

(ii) It is immaterial to the asymptotic theory which of the rules Λ and Λ^* is used, the relations (26) and (27) holding in either case.

Ray has apparently been misled because his computations are fragmentary and inadequate. In fact, the author has performed more extensive computations using Λ^* which do indeed indicate that $C^*(\lambda)$ begins to increase steadily as λ becomes large, apparently tending to α as in fact Theorem 4.1 requires. These computations are discussed in the following section.

5. Theory and computations for moderate λ. We have from (18) that $C(\lambda) \geqq \varphi(n_0^{\frac{1}{2}}/\lambda)$, $0 < \lambda < \infty$, which, with (20) and (26), shows that the constant β,

$$(28) \qquad \beta = \inf_{0<\lambda<\infty} C(\lambda) \leqq \varphi(a),$$

is positive. The value of β depends on the whole sequence $\{a_n\}$ which defines N through (5) and not just on the limit (3). It is obvious that we can choose the sequence $\{a_n\}$ subject to (3) so that the value τ,

$$(29) \qquad \tau = \inf_{0<\lambda<\infty} \tau_\lambda = \beta/\alpha$$

is arbitrarily close to 1, but only at the expense of making the quantity $\eta = \sup_{0<\lambda<\infty} \eta_\lambda$ unreasonably large.

In this section we restrict our attention to the sequence $\{a_n\}$ defined in (2) and Λ^* given by (5*), precisely the rule evaluated by Ray.

Taking $n = 2m + 1$, we have in analogy with (21)

(30) N^* is the first $n = 2m + 1$ with $m \geqq m_0^*$ such that

$$\sum_{i=1}^{2m} W_i^2 \leqq 2m(2m+1)/\lambda^2 a_m^2 = 2b_{m+1},$$

where

(31) $m_0^* = (n_0^* - 1)/2 \geqq 1$ and $b_m = (m-1)(2m-1)/\lambda^2 a_{2m-2}^2$.

We observe that $V_2 = W_1^2 + W_2^2$ has probability density $dF(V_2) = \frac{1}{2}\exp(-\frac{1}{2}v_2)\, dv_2$, $(v_2 > 0)$, and define $Z = \frac{1}{2}V_2$, which has probability density $dG(Z) = e^{-z}\, dz$, $(z > 0)$.

Then for the rule Λ^* we have in analogy with (14) for $m \geqq m_0^*$

$$\begin{aligned} p_m^*(\lambda) &= P(N^* = 2m+1) \\ &= P\{\textstyle\sum_{i=1}^{k} Z_i > b_{2k+1} \text{ for every } m_0^* \leqq k \leqq m-1, \\ &\qquad \text{but not for } k = m\}, \end{aligned} \tag{32}$$

and in analogy with (18) and (17)

$$\begin{aligned} C^*(\lambda) &= \textstyle\sum_{m=m_0^*}^{\infty} p_m^*(\lambda)\varphi[(2m+1)^{\frac{1}{2}}/\lambda], \\ D^*(\lambda) &= \textstyle\sum_{m=m_0^*}^{\infty} (2m+1)p_m^*(\lambda) = 1 + 2\sum_{m=m_0^*}^{\infty} mp_m^*(\lambda). \end{aligned} \tag{33}$$

The probability (32) can be evaluated by a method following [8] which is summarized in Appendix 2.

One further minor modification of the procedure is of interest; namely, that the experimenter using Λ^* be required to make a *fixed* number j of additional observations after having decided (nominally) to terminate sampling. We call the amended rule

$\Lambda^*(j)$: Observe the sequence (1) term by term, stopping with X_{N^*+j} $(j \geqq 0)$, where N^* satisfies (5^*).

We remark that

$$\begin{aligned} R^*(0) &= R^*, \\ C_j^*(\lambda) &= \textstyle\sum_{m=m_0^*}^{\infty} p_{m^*}(\lambda)\varphi[(2m+j+1)^{\frac{1}{2}}/\lambda] && (j \geqq 0), \\ D_j^*(\lambda) &= j + D^*(\lambda) && (j \geqq 0), \end{aligned}$$

where C_j^* and D_j^* are respectively the coverage probability and expected sample size using $\Lambda^*(j)$. We further remark that taking any *finite* number of additional observations after stopping does not affect the asymptotic theory, nor indeed the theory for small λ, save in the respect that (21) becomes

$$\lim_{\lambda\to 0} D_j^*(\lambda) = n_0^* + j \qquad (j \geqq 0). \tag{34}$$

C_j^* and D_j^* have been computed for several hundred values of the parameter $0 < \lambda < \infty$ with $\alpha = .95$ ($n_0^* = 3, j = 0, 2, 4$, and $n_0^* = 5, j = 0, 2$) and with $\alpha = .99$ ($n_0^* = 3, j = 0, 2$) subject to the requirements of $\Lambda^*(j)$. The computations were carried out on an IBM 7094 at the Columbia University Computer Center. A summary of these computations for representative values of λ is given in Table I ($\alpha = .95$) and Table II ($\alpha = .99$) of Appendix 1. Figure 1 ($\alpha = .95$) and Figure 2 ($\alpha = .99$) provide a diagramatic representation of the behavior of $C_j^*(\lambda)$ and, in analogy with (12), of $\eta_\lambda^*(j) = D_j^*(\lambda)/a^2\lambda^2$, for the several

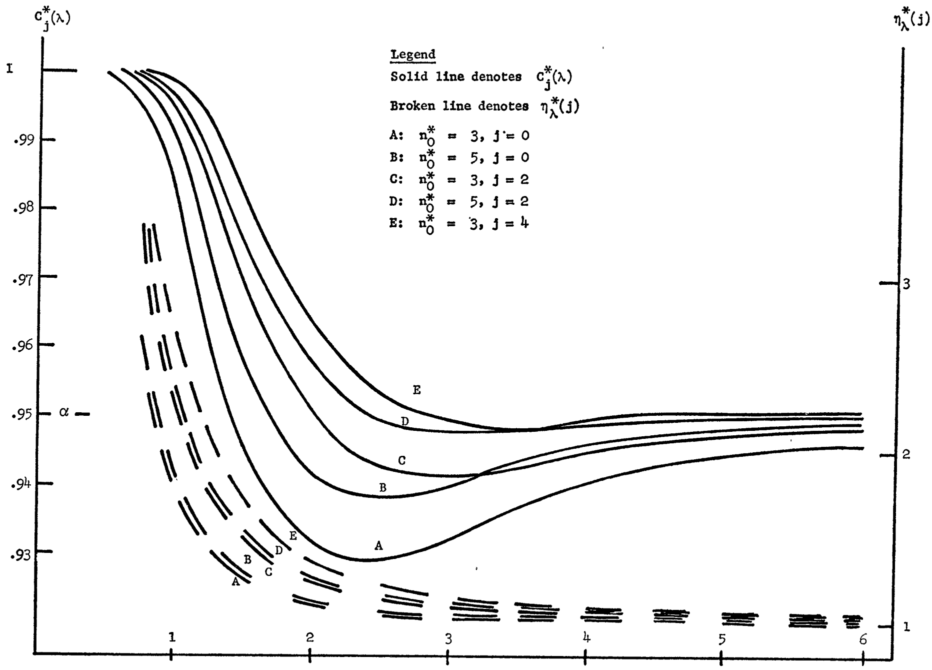

FIG. 1. Efficiency and coverage probability using $\Lambda^*(j)$ with α = .95

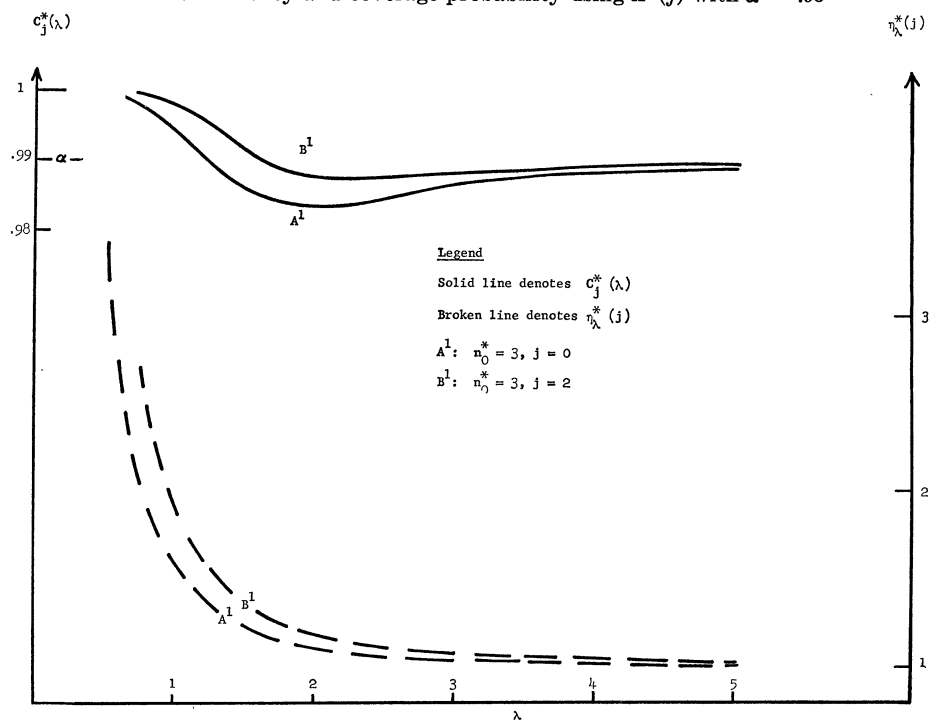

FIG. 2. Efficiency and coverage probability using $\Lambda^*(j)$ with α = .99

TABLE 2

β^*

$\alpha = .95$

	n_0^*	j	β^*
A	3	0	.92848
B	5	0	.93792
C	3	2	.94146
D	5	2	.94783
E	3	4	.94873
	$\alpha = .99$		
A^1	3	0	.98347
B^1	3	2	.98752

versions of Λ^*. We remark in passing that the computed values differ sometimes from those of Ray.

Of special interest of course are the values (in analogy with (28)) of $\beta^* = \inf_{0<\lambda<\infty} C^*(\lambda)$, which are displayed in Table 2.

We observe that τ^* defined in analogy with (29) is not appreciably less than 1 in any of the versions, so that the cost in coverage probability due to ignorance of σ is "next to nothing." Moreover, we remark that for any of the several versions of Λ^*, $\eta_\lambda{}^*$ appears to decrease monotonically (and rapidly) to 1, being large only for λ near zero when the expected number of observations $D^*(\lambda)$ is small anyway (and depends largely on the starting sample size together with the additional number, if any, of observations take subsequent to stopping). Therefore the cost in observations due to ignorance of σ is similarly not appreciable.

It is obvious that the basic rule Λ will have a minimum coverage probability β not very much less than .928 (when $\alpha = .95$) and efficiency η exceedingly close to 1 for λ bounded somewhat away from zero.

Hence the basic procedure and its several variations are remarkably consistent and efficient, *uniformly* in λ. τ can always be adjusted upward by using a different sequence from that defined in (2), increasing the starting sample size n_0, or taking j additional observations subsequent to nominally stopping, but only at the expense of increasing η. Alternatively, η can be arbitrarily decreased but only at the cost of concomitantly decreasing τ. There is no unique version of Λ which jointly reduces the loss in observations and minimum coverage probability, due to ignorance of σ. Presumably the statistical worker should let his particular requirements dictate the choice of the procedure to be used in applications.

6. Comparison of Λ with Stein's two-stage sampling procedure. B. M. Seelbinder [9] and others (e.g., J. Moshman [6]) have considered the problem of obtaining an appropriate determination of n_1, the size of the initial sample for C. Stein's [11] two-stage sampling procedure:

Let us define N_s = total sampling size using two-stage sampling, and $E_\lambda(N_s) =$

expectation of N_s at the parameter point $0 < \lambda < \infty$. Seelbinder proves a somewhat stronger limiting theorem than did Stein; namely

$$\lim_{\lambda \to \infty} [E_\lambda(N_s) - a_{n_1-1}^2 \lambda^2] = 0. \tag{35}$$

Noting that the sequence $\{a_n\}$ defined in (2) has the property: $a_1 > a_2 > a_3 > \cdots$, it follows from (3) that the Stein procedure is asymptotically *less* efficient than Λ; in fact, (35) implies

$$\lim_{\lambda \to \infty} E_\lambda(N_s)/a^2\lambda^2 = a_{n_1-1}^2/a^2 > 1 \qquad (n_1 \text{ fixed}),$$

compared to the asymptotic result (27) proved for Λ. Moreover, it is not difficult to prove that

$$\lim_{\lambda \to 0} E_\lambda(N_s) = n_1 . \tag{36}$$

Seelbinder suggests (p. 647) in using Stein's two-stage sampling procedure that if one is uncertain about the value of σ but believes it is such that $\lambda < 10$ "then the first part of the sample (n_1) should be taken to be 250 or thereabouts" so as to decrease the possibility of having $E_\lambda(N_s)$ inordinately large for λ bounded somewhat away from zero.

Therefore, comparing (36) with (21) or (34) (for which in practice none of the values n_0, n_0^*, or j will be large) we see that for λ near zero Λ and its several variants are distinctly more efficient than the Stein procedure.

Seelbinder has computed $E_\lambda(N_s)$ for $\alpha = .99, .98, .95, .90$ and $\lambda^{-1} = .01(.01).1(.1)1.0$, for a number of values of n_1. A portion of these data together with the corresponding values of $D^*(\lambda)$ (with $n_0^* = 3, j = 0$) is summarized in Table 3.

Now, for $0 < \lambda < \infty$, $E_\lambda(N_s)$ depends on the choice of the first-stage sample size n_1. Indeed, if one had *precise* knowledge of σ, the value $\xi_\lambda(n_1)$, $\xi_\lambda(n_1) = E_\lambda(N_s \mid n_1)$, could be minimized at λ by an inspection of Seelbinder's table (and interpolation). We remark from an inspection of Table 3 that

$$D^*(\lambda) < \min_{2 \leq n_1 \leq \infty} \xi_\lambda(n_1) \tag{37}$$

for all λ for which comparisons are possible. The author conjectures that in fact (37) holds for *all* λ. However this may be, it is clear that if we had precise knowledge of σ (or equivalently of λ) we would not rely on *any* sequential procedure, but simply preassign n the smallest integer value satisfying (4). Therefore, the only case of interest to sequential analysis is when σ is unknown. In this event an inappropriate (unlucky) choice of n_1 can have costly results.

By way of example, suppose we require a fixed-width ($2d > 0$) interval for θ, prescribing $\alpha = .95$, and that we have some prior reason to believe, or are prepared to assume, that σ is such that $\lambda = 5$, when *in fact* $\lambda = 1$. Then from Table 3 we would be inclined to take $n_1 = 61$ for which $\xi_5(n_1)$ is a minimum. Accordingly, the expected sample size using two-stage sampling is $\xi_1(61) = 61$, comparing very unfavorably with $D^*(1) = 6.6$, the expected sample size using Λ^* at the

TABLE 3

Expected sample size using Stein's two-stage sampling, condensed and revised from [9]
D^* and $a^2\lambda^2$ are included for comparison with Λ^* and the fixed sample size case, respectively.

				$\alpha = .95$				
λ		1	1.25	1.667	2.0	2.5	3.333	5.0
n_1	241	241	241	241	241	241	241	241
	121	121	121	121	121	121	121	121
	81	81	81	81	81	81	81	101
	61	61	61	61	61	61	61.1	100
	51	51	51	51	51	51	52.6	101
	41	41	41	41	41	41.1	47.8	102
	31	31	31	31	31	32	46.6	104
	21	21	21	21	21	28.2	48.4	109
	11	11	11	15.1	20.2	31.1	55.1	124
	6	7.9	10.9	18.5	46.4	41.3	73.4	165
$D^*(\lambda)$ (with $n_0^* = 3$, $j = 0$)		6.6	8.4	12.5	16.7	25.0	43.5	97.1
$a^2\lambda^2$		3.8	6.0	10.7	15.4	24.0	42.7	96.0
				$\alpha = .99$				
n_1	241	241	241	241	241	241	241	241
	121	121	121	121	121	121	121	171
	81	81	81	81	81	81	84.5	174
	61	61	61	61	61	61.1	79.1	177
	51	51	51	51	51.1	53.1	79.9	179
	41	41	41	41	41.2	48.1	82.3	183
	31	31	31	31.1	33.7	47.6	84.0	189
	21	21	21.2	24.7	32.8	50.6	89.9	202
	11	12.4	16.5	27.9	40.2	62.8	112	251
$D^*(\lambda)$ (with $n_0^* = 3$, $j = 0$)		10.4	13.9	21.5	29.5	44.3	76.5	168.9
$a^2\lambda^2$		6.6	10.4	18.4	26.5	41.5	73.7	166

parameter value $\lambda = 1$. Using the Stein procedure, a poor guess is a considerable extravagance.

Reversing the situation, suppose σ is such that we believe $\lambda = 1$, when *in fact* $\lambda = 5$. Then we should be disposed to take $n_1 = 6$ or thereabouts for which $\xi_1(n_1)$ is a minimum. In this event $\xi_5(6) = 165$ and $D^*(5) = 97.1$. Again we pay dearly, using two-stage sampling, for our mistaken judgment regarding the value of σ.

We conclude that Λ^* is more efficient than two-stage sampling for λ very small and very large; that it *appears* that the former procedure is always more efficient than the latter for $0 < \lambda < \infty \cdot$ and finally that an inappropriate choice of the initial sample size n_1 in two-stage sampling can lead to an excessive cost in needless observations whereas Λ is always reasonably efficient irrespective of the value of λ and what we "think" this value may be.

On the other hand, two-stage sampling is somewhat more consistent than Λ, assuring a coverage probability $\geqq \alpha$, whereas Λ and its variants assure only $\beta < \alpha$.

However, in normal experimental situations it would seem prudent to settle for a minimum coverage probability (irrespective of the magnitude of σ) which is not conspicuously less than α, say .94146 (using $\Lambda^*(4)$ with $n_0^* = 3$) when $\alpha = .95$ or .98752 (using $\Lambda^*(2)$ with $n_0^* = 3$) when $\alpha = .99$, in the certainty that (at worst) our expected sample size does not sensibly exceed that of two-stage sampling and that we do not run the risk, in our ignorance of σ, of requiring extra (in extreme cases possibly thousands of) needless observations.

7. Further theoretical problems of interest. Y. S. Chow and H. Robbins [3] have proved the asymptotic properties (26) and (27) (with a slight modification of Λ in the discrete case) assuming of the sequence defined by (1) only that its members are independent, identically distributed, and have finite positive variance. This result has been extended to the case of the simultaneous estimation of several parameters by L. J. Gleser [4].

It would be of considerable practical interest to have an evaluation of the performance, over the entire range of values of λ, of this sequential procedure for observations which do not have the normal distribution. The author is investigating this problem when the underlying distribution of sequence (1) is binomial or Poisson, and in linear regression models for the simultaneous estimation of several parameters.

A test of a hypothesis about the mean of a normal population with unknown variance, which has prescribed error probabilities, can be constructed by combining Λ^* with a suitable decision procedure. H. Robbins and the author are evaluating this test and a study of its performance will soon be available.

8. Acknowledgment. I wish to express my gratitude to Professor Herbert Robbins, under whose guidance this research was conducted, for his advice and continuing encouragement, and to Professor Leon J. Gleser for his helpful comments and suggestions. I would like also to thank the referee for his useful suggestions.

APPENDIX 1

TABLE I

$\alpha = .95,\ a = 1.96$

λ	$n_0^* = 3$							$n_0^* = 5$				
	$A: \Lambda^*(0) = \Lambda^*$			$C: \Lambda^*(2)$		$E: \Lambda^*(4)$		$B: \Lambda^*(0) = \Lambda^*$			$D: \Lambda^*(2)$	
	$C^*(\lambda)$	$D^*(\lambda)$	η_λ^*	$C_2^*(\lambda)$	$\eta_\lambda^*(2)$	$C_4^*(\lambda)$	$\eta_\lambda^*(4)$	$C^*(\lambda)$	$D^*(\lambda)$	η_λ^*	$C_2^*(\lambda)$	$\eta_\lambda^*(2)$
0.50	.99974	4.1	4.277	.999996	6.359	.99999	8.442	.99992	5.1	5.278	.99999	7.361
0.75	.99336	5.2	2.403	.99908	3.329	.99986	4.254	.99795	5.8	2.664	.99970	3.590
1.00	.9777	6.6	1.720	.9930	2.241	.9977	2.761	.9872	7.0	1.831	.9958	2.352
1.25	.9600	8.4	1.402	.9815	1.735	.9911	2.069	.9715	8.8	1.465	.9865	1.799
1.50	.9457	10.7	1.235	.9688	1.467	.9816	1.698	.9574	11.0	1.278	.9750	1.509
1.75	.9361	13.4	1.142	.9579	1.312	.9718	1.482	.9473	13.8	1.174	.9648	1.344
1.90	.9324	15.3	1.106	.9528	1.251	.9665	1.395	.9433	15.7	1.134	.9598	1.278
2.00	.9307	16.7	1.088	.9500	1.219	.9634	1.349	.9413	17.1	1.114	.9571	1.244
2.05	.9301	17.4	1.081	.9488	1.205	.9620	1.329	.9406	17.8	1.105	.9559	1.229
2.10	.9296	18.2	1.074	.9477	1.192	.9601	1.310	.9399	18.6	1.098	.9548	1.216
2.15	.9292	19.0	1.068	.9467	1.180	.9595	1.293	.9393	19.4	1.091	.9538	1.203
2.20	.9289	19.8	1.062	.9459	1.170	.9583	1.277	.9389	20.2	1.084	.9529	1.192
2.25	.9287	20.6	1.057	.9451	1.160	.9573	1.263	.9385	21.0	1.079	.9521	1.181
2.30	.92853	21.4	1.053	.9444	1.151	.9563	1.250	.9383	21.8	1.073	.9514	1.172
2.35	.92848	22.3	1.049	.9438	1.143	.9554	1.238	.9381	22.7	1.069	.9508	1.163
2.40	.92850	23.2	1.045	.9433	1.136	.9546	1.226	.93797	23.6	1.065	.9502	1.155
2.45	.92859	24.0	1.042	.9429	1.129	.9538	1.216	.937916	24.5	1.061	.9498	1.148
2.50	.9287	25.0	1.039	.9425	1.122	.9532	1.206	.937919	25.4	1.058	.9493	1.141
2.55	.9289	25.9	1.037	.9422	1.117	.9526	1.197	.93797	26.3	1.054	.9490	1.135
2.60	.9292	26.9	1.035	.9420	1.112	.9520	1.189	.9381	27.3	1.052	.9487	1.129
2.80	.9305	31.0	1.028	.94146	1.094	.9504	1.161	.9388	31.4	1.043	.9480	1.109
3.00	.9321	35.4	1.023	.94156	1.081	.9494	1.139	.9399	35.9	1.037	.94783	1.095
3.50	.9363	47.9	1.018	.9430	1.060	.94873	1.103	.9430	48.4	1.029	.9485	1.071
4.00	.9397	62.4	1.015	.9446	1.048	.9489	1.080	.9454	62.9	1.024	.9495	1.056
4.50	.9422	78.8	1.013	.9459	1.039	.9492	1.065	.9470	79.4	1.020	.9501	1.046
5.00	.9438	97.1	1.011	.9467	1.032	.9494	1.053	.9480	97.7	1.017	.9505	1.038
5.85								.9488	133.2	1.013	.9506	1.028
6.00	.9457	139.5	1.008	.9478	1.023	.9496	1.037					
6.75	.9467	176.2	1.007	.9482	1.018	.94964	1.030					

TABLE II

$\alpha = .99,\ a = 2.5758,\ n_0^* = 3$

	A^1: $\Lambda^*(0) = \Lambda^*$			B^1: $\Lambda^*(2)$	
λ	$C^*(\lambda)$	$D^*(\lambda)$	η_λ^*	$C_2^*(\lambda)$	$\eta_\lambda^*(2)$
0.5	.999924	5.7	3.445	.999999	4.651
0.75	.99795	7.7	2.074	.99908	2.610
1.0	.9938	10.4	1.571	.9980	1.872
1.25	.9889	13.9	1.338	.9948	1.531
1.50	.9854	18.2	1.217	.9915	1.351
1.90	.98347	26.9	1.123	.9883	1.207
2.0	.98349	29.5	1.110	.9979	1.185
2.05	.98355	30.8	1.104	.9878	1.176
2.10	.9836	32.1	1.099	.98769	1.167
2.15	.9839	33.5	1.094	.98761	1.159
2.2	.9839	35.0	1.089	.98756	1.152
2.25	.9840	36.4	1.085	.98754	1.145
2.30	.9842	38.0	1.081	.98753	1.138
2.35	.9843	39.5	1.079	.98753	1.132
2.4	.9845	41.1	1:075	.98755	1.127
2.45	.9847	42.7	1.072	.98758	1.122
2.5	.9848	44.3	1.069	.9876	1.117
2.60	.9852	47.7	1.064	.9877	1.108
2.80	.9859	54.9	1.055	.9880	1.094
3.0	.9864	62.6	1.049	.9882	1.082
3.5	.9875	84.3	1.037	.9888	1.061
4.0	.9882	109.2	1.028	.9891	1.047
4.5	.9886	137.4	1.023	.9893	1.038
5.0	.9889	168.9	1.018	.9894	1.031

APPENDIX 2

The probability distribution of N^* defined by (30) was computed by the following method (see [8] for proof).

Define for the case $n_0^* = 3$ $h_1 \equiv 1$, $c_1 \equiv 1$, and compute recursively

$$h_m(b_n) = \textstyle\sum_{j=1}^{m-1} [(b_n - b_m)/j!]h_{m-j}(b_m) \quad (m = 2, 3, \cdots; n = m + 1, m + 2),$$

where b_m is given in (31), and then compute

$$c_m = \exp(-b_m) \cdot \textstyle\sum_{j=1}^{m-1} h_{m-j}(b_m) \qquad (m = 2, 3, \cdots).$$

Then for $m = 1, 2, \cdots$

$$p_m^*(\lambda) = P(N^* = 2m + 1) = c_m - c_{m+1}.$$

(We remark that for the case $n_0^* = 5$ we take $b_2 = 0$, otherwise computing as above.)

Then C^* is given in (33) and D^* becomes

$$D^*(\lambda) = 1 + 2 \textstyle\sum_{m=1}^{\infty} c_m.$$

REFERENCES

[1] Anscombe, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

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

[3] Chow, Y. S. and Robbins, H. (1964). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

[4] Gleser, L. J. (1964). On the asymptotic theory of fixed-width sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467.

[5] Gleser, L. J., Robbins, H. and Starr, N. (1964). Some asymptotic properties of fixed-width sequential confidence intervals. *Columbia Univ. Report.*

[6] Moshman, J. (1958). A method for selecting the size of the initial sample in Stein's two sample procedure. *Ann. Math. Statist.* **29** 1271–1275.

[7] Ray, W. D. (1957). Sequential confidence intervals for the mean of a normal distribution with unknown variance. *J. Roy. Statist. Soc. Ser. B* **19** 133–143.

[8] Robbins, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics—The Harald Cramér Volume*, 235–245, Almqvist and Wiksell, Uppsala, Sweden.

[9] Seelbinder, B. M. (1953). On Stein's two-stage sampling scheme. *Ann. Math. Statist.* **24** 640–649.

[10] Stein, C. (1949). Some problems in sequential estimation (abstract). *Econometrica* **17** 77–78.

[11] Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.