The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 667–671

AN EXTENSION OF A THEOREM OF CHOW AND ROBBINS ON SEQUENTIAL CONFIDENCE INTERVALS FOR THE MEAN[1]

BY ARTHUR NÁDAS[2]

IBM Corporation

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent observations from a population with mean μ and finite nonzero variance σ^2. We wish to estimate the unknown μ by confidence intervals of prescribed "accuracy" and prescribed probability of coverage α. Let

$$(1) \qquad \bar{X}_n = n^{-1} \textstyle\sum_{k=1}^n X_k, \qquad (n = 1, 2, \cdots).$$

We speak of "absolute accuracy" when estimating μ by

$$(2) \qquad I_n = (\mu : |\bar{X}_n - \mu| \leqq d), \qquad (d > 0),$$

and, if $\mu \neq 0$, we speak of "proportional accuracy" when estimating μ by

$$(3) \qquad J_n = (\mu : |\bar{X}_n - \mu| \leqq p\,|\mu|), \qquad (0 < p < 1).$$

Denote by ρ the coefficient of variation $\sigma/|\mu|$ and define

$$(4) \qquad n(d) = \min_{n \geqq 1} (n : \sigma^2 \leqq n(d/a)^2),$$

$$(5) \qquad m(p) = \min_{n \geqq 1} (n : \rho^2 \leqq n(p/a)^2)$$

where a is the $\frac{1}{2}(1 - \alpha)$th fractile of the standard normal distribution. Then (4) and (5) increase without bound as the arguments tend to zero. Hence for small arguments we can achieve (at least approximately) the required probability of coverage α by taking the "sample size" n no smaller than $n(d)$ (for absolute accuracy) or $m(p)$ (for proportional accuracy).

If, however, σ^2 (the "appropriate parameter" for absolute accuracy) is unknown, or if ρ^2 (the "appropriate parameter" for proportional accuracy) is unknown then (4) or (5) are not available. On the other hand if we let

$$(6) \qquad V_n^2 = n^{-1}(1 + \textstyle\sum_{k=1}^n (X_k - \bar{X}_n)^2), \qquad (n = 1, 2, \cdots),$$

then the stopping rules

$$(7) \qquad N = \min_{n \geqq 1} (n : V_n^2 \leqq n(d/a)^2)$$

and

$$(8) \qquad M = \min_{n \geqq 1} (n : (V_n/\bar{X}_n)^2 \leqq n(p/a)^2)$$

Received 27 November 1967.

[1] This paper is part of the author's Ph.D. Thesis at Columbia University.

[2] Now at the Polytechnic Institute of Brooklyn.

are well defined. In the event that ρ^2 is known (but not σ^2) and one insists on absolute accuracy, or if σ^2 is known (but not ρ^2) in the proportional case, then one arrives at

$$N^* = \min_{n \geqq 1} (n : \bar{X}_n^2 + n^{-2} \leqq n(d/\rho a)^2) \tag{9}$$

and

$$M^* = \min_{n \geqq 1} (n : \bar{X}_n^2 \geqq n^{-1}(a\sigma/p)^2) \tag{10}$$

as the sequential analogues of (4) and (5). (The purpose of the $o(1)$ terms in (6) and (9) is merely to ensure that the sample sizes diverge as d or p tend to zero.)

Denote by K any one of (7) thru (10), let k be the corresponding fixed "sample size" (4) or (5) and let H_K be the corresponding interval estimate (2) or (3)).

THEOREM.

$$\lim K/k = 1 \quad \text{a.s.,} \tag{11}$$

$$\lim P(\mu \varepsilon H_K) = \alpha \qquad \textit{"asymptotic consistency,"} \tag{12}$$

$$\lim EK/k = 1 \qquad \textit{"asymptotic efficiency,"} \tag{13}$$

as d or p tend to zero.

Chow and Robbins [4] have proved this theorem in the case of $K = N$. In this case we improve on (13) by giving a uniform upper bound on $EN - n(d)$. We also show that $EM^* - m(p)$ is bounded above. Under the additional assumption that the distribution function F of X_1 is continuous with a finite fourth moment, we show that $EM - m(p)$ is also bounded above. We do not know of reasonably general conditions on F that would ensure that $EK - k$ is bounded below (see remark (b)).

REMARKS. (a) In practice one may wish to replace the constant "a" in (7) thru (10) by positive convergents $(a_n : n = 1, 2, \cdots)$ in the hope of improving the coverage probability for non-infinitesmal d or p. It is shown in [6] that that may be done without invalidating the asymptotic results. The same is true if the range of K is restricted to certain infinite subsets of the positive integers ([4]).
(b) Anscombe [2] made asymptotic expansions for the distribution of (essentially) $N - n(d)$ on the assumption that F is not purely discrete, that F has a finite eighth moment and that the tails of the distribution of $N - n(d)$ satisfy certain order of magnitude conditions. The latter are not given in terms of F.

2. Proof of the theorem. For every definition (7) thru (10) of K, (11) and (12) follow from (a trivial modification of) Lemma 1 of [4] and a result of Anscombe [1], just as in [4]. We have from (11) and Fatou's lemma that

$$\liminf EK/k \geqq 1. \tag{14}$$

To prove that

$$\limsup EK/k \leqq 1 \tag{15}$$

and to find upper bounds, we consider the separate cases.

2.1 *The case* $K = N$. Let $t = (a\sigma/d)^2$ and for $n = 1, 2, \cdots$ let

$$\sigma^2 S_n = 1 + \sum_{k=1}^{n} (X_k - \mu)^2. \tag{16}$$

Define

$$N(t) = \min_{n\geqq 1} (n: S_n \leqq n^2/t). \tag{17}$$

Then $N(t)$ is well defined, it is no smaller than N and it can be shown (e.g. by truncating $N(t)$ and proceeding as below) that $EN(t) < \infty$. By Wald's theorem for cumulative sums

$$\begin{aligned} \sigma^{-2} + EN(t) &= ES_{N(t)} \\ &\geqq \int_{(N(t)=1)} S_1 + \int_{(N(t)>1)} S_{N(t)-1} \\ &\geqq t^{-1} \int_{(N(t)>1)} (N(t) - 1)^2 = t^{-1}E(N(t) - 1)^2 \geqq t^{-1}E^2(N - 1)\cdot \end{aligned} \tag{18}$$

Thus

$$EN(t) \; \varepsilon \; \{x: x^2 - (t + 2)x + (1 - t\sigma^{-2}) \leqq 0\} \tag{19}$$

whence for all $t > 0$ we have $EN(t) - t \leqq 2 + \sigma^{-2}$. If F is continuous then the n^{-1} term in (6) can be dropped and we obtain the same way that for all $t > 0$ and all $\sigma > 0$

$$EN(t) - t \leqq 2. \tag{20}$$

2.2 *The cases* $K = N^*$ *and* $K = M^*$. In these cases (15) is a consequence of the following

LEMMA. *Let* $Z_1, Z_2, \cdots$ *be a sequence of independent and identically distributed random variables such that* $0 < EZ_1 < \infty$. *Denote by* S_n $(n = 1, 2, \cdots)$ *their* nth *cumulative sum. Let* c *be a positive number.*

(i) *Assume* $Z_1 \geqq 0$ a.s. *Let* $\gamma > 0$ *and* $t = (EZ_1/c)^{1/\gamma}$. *Define*

$$T = \min_{n\geqq 1} (n: S_n \leqq cn^{1+\gamma}). \tag{21}$$

Then T *is well defined,* $ET^{1+\gamma} < \infty$ *and the family* $(T/t: 0 < c < 1)$ *is uniformly integrable.*

(ii) *Let* $0 < \gamma < 1$ *and* $t = (c/EZ_1)^{1/(1-\gamma)}$. *Define*

$$T = \min_{n\geqq 1} (n: |S_n| \geqq cn^{\gamma}). \tag{22}$$

Then T *is well defined,* $ET < \infty$ *and* $\lim_{c\to\infty} ET/t = 1$.

The lemma can be proved by truncating both the stopping rule T (to T') and the Z_n in each case (i) and (ii) and then approximating the roots of a polynomial in the variable ET'. The lemma is related to a result of Chow [3] and it has been generalized by Siegmund [7].

In the case $K = N^*$ (13) is established by using (i) above to prove the uniform integrability of the family $(N^*/n(d): 0 < d < 1)$. In the case $K = M^*$ (13) follows by a change of notation in (ii) above. The boundedness (from above) of $EM^* - m(p)$ can be shown by the method of 2.3 below. The same is true for

$EN^* - n(d)$ provided we make the (highly unsatisfactory) assumption that the X_n are positive. We omit the details.

2.3 *The case* $K = M$. With no loss of generality we assume $\mu > 0$. Let $t = (a\rho/p)^2$ and put, for $n = 1, 2, \cdots$,

$$(23) \qquad S_n = \textstyle\sum_{k=1}^n X_k \quad \text{and} \quad Q_n = 1 + \sum_{k=1}^n (X_k - \mu)^2.$$

The random variable

$$(24) \qquad M(t) = \min_{n\geqq 1} (n : S_n \geqq \rho^{-1}(tQ_n)^{\frac{1}{2}})$$

is well defined and it is no smaller than M. For $r = 1, 2, \cdots$ define

$$(25) \qquad R = \min (M(t), r) \quad \text{and} \quad B = (1 < M(t) \leqq r).$$

We apply Wald's theorem for cumulative sums to each of S_R, $Q_R - 1$ and $X_1^2 + \cdots + X_R^2$ to obtain

$$(26) \qquad \begin{aligned} \mu ER = ES_R &= \textstyle\int_{(R=1)} X_1 + \int_B S_{R-1} + \int_{(M(t)>r)} S_r + \int_B X_R \\ &\leqq E|X_1| + \rho^{-1} \textstyle\int_B (tQ_{R-1})^{\frac{1}{2}} + \rho^{-1} \int_{(M(t)>r)} (tQ_r)^{\frac{1}{2}} + \int_B |X_R| \\ &\leqq E^{\frac{1}{2}}X_1^2 + \rho^{-1}E(tQ_R)^{\frac{1}{2}} + E^{\frac{1}{2}} \textstyle\sum_{k=1}^R X_k^2 \\ &\leqq (\sigma^2 + \mu^2)^{\frac{1}{2}} + \rho^{-1}(t + t\sigma^2 ER)^{\frac{1}{2}} + (\sigma^2 + \mu^2)^{\frac{1}{2}} E^{\frac{1}{2}} R \\ &\leqq (\sigma^2 + \mu^2)^{\frac{1}{2}} + \rho^{-1} t^{\frac{1}{2}} + (\mu t^{\frac{1}{2}} + (\sigma^2 + \mu^2)^{\frac{1}{2}}) E^{\frac{1}{2}} R. \end{aligned}$$

Thus

$$(27) \qquad E^{\frac{1}{2}}R \ \varepsilon \ \{x : x^2 - (t^{\frac{1}{2}} + (\rho^2 + 1)^{\frac{1}{2}})x - (\sigma^{-1}t^{\frac{1}{2}} + (\rho^2 + 1)^{\frac{1}{2}}) \leqq 0\}$$

whence

$$(28) \qquad ER \leqq (t^{\frac{1}{2}} + o(t^{\frac{1}{2}}))^2.$$

It follows from the a.s. monotone convergence $R \uparrow M(t)$ as $r \to \infty$ that $EM(t) < \infty$ and that (28) holds with $M(t)$ in place of R.
Hence

$$(29) \qquad \limsup_{t\to\infty} EM(t)/t \leqq 1$$

so (15) holds in this case also.

To show that $EM - m(p)$ is bounded above, assume $EX_1^4 < \infty$. Assume also that F is continuous so that the constant term 1 in the definitions of nV_n^2 and Q_n can be omitted. Consider the line

$$(30) \qquad L(x; t) = \tfrac{1}{2}\mu(x + t), \qquad x > 0, \quad t > 0,$$

which is the tangent to the curve $\mu(tx)^{\frac{1}{2}}$, in an $x - y$ plane, at the point of intersection of the curve with the line μx.
Then

$$(31) \qquad \begin{aligned} M(t) &= \min_{n\geqq 1} (n : S_n \geqq \mu(tQ_n\sigma^{-2})^{\frac{1}{2}}) \\ &\leqq \min_{n=1} (n : S_n \geqq L(Q_n\sigma^{-2}; t)) \\ &= \min_{n\geqq 1} (n : \textstyle\sum_{k=1}^n Z_k \geqq t) \end{aligned}$$

where we have set

$$\mu Z_k = 2X_k - \mu(X_k - \mu)^2/\sigma^2, \qquad (k = 1, 2, \cdots). \tag{32}$$

Let φ_A denote the indicator of the set A. If we denote by

$$U(t) = E \sum_{n=0}^{\infty} \varphi_{(\Sigma_{k=1}^{n} Z_k < t)} \tag{33}$$

the "renewal measure" of the transient ($EZ_1 > 0$) random walk $\sum_{k=1}^{n} Z_k$ then it follows (see e.g. [5] p. 372) that

$$\limsup_{t\to\infty} EM - t \leqq \lim_{t\to\infty} U(t) - t = EZ_1^2/2E^2Z_1 < \infty. \tag{34}$$

REMARKS. (c) The idea of approximating a convex stopping boundary for a random walk by the appropriate tangent appears in [2], [7].

(d) We offer a minor insight into the nature of (13) by noting that in the case $K = N$, (13) is equivalent to

$$\lim_{d\to 0} EV_N^2 = \sigma^2. \tag{35}$$

(e) The stopping rules $U = \max(N, M)$ and $L = \min(N, M)$ are appropriate for some "mixed" accuracy criteria that arise in practice, i.e. for estimating the mean by the intervals

$$I_U \cap J_U = (\mu: |\bar{X}_U - \mu| \leqq \min(d, p\,|\mu|)), \tag{36}$$

$$I_L \cup J_L = (\mu: |\bar{X}_L - \mu| \leqq \max(d, p\,|\mu|)). \tag{37}$$

Let $u(d, p)$ and $l(d, p)$ be the corresponding fixed "sample sizes," and let the ratio d/p be fixed. It is fairly easy to see that the theorem holds for L, and that it holds for U is a straightforward consequence of the uniform integrability of $(N/n(d))$ and $(M/m(p))$.

3. Acknowledgments. I sincerely thank Professor David O. Siegmund for his interest and guidance in the preparation of my thesis, and Professor Herbert E. Robbins for a suggestion that led to (20).

REFERENCES

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

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

[3] CHOW, Y. S. (1966). On the moments of some one-sided stopping rules. *Ann. Math. Statist.* **37** 382–387.

[4] CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

[5] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* II. Wiley, New York.

[6] NÁDAS, A. (1967). On the asymptotic theory of estimating the mean by sequential confidence intervals of prescribed accuracy. Ph.D. Thesis, Columbia Univ.

[7] SIEGMUND, D. O. (1967). Some one-sided stopping rules. *Ann. Math. Statist.* **38** 1641–1646.