The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 1946–1952

ON THE COST OF NOT KNOWING THE VARIANCE WHEN MAKING A FIXED-WIDTH CONFIDENCE INTERVAL FOR THE MEAN

By Gordon Simons[1]

Stanford University

1. Summary. It is shown that the mean of a normal distribution with unknown variance σ^2 may be estimated to lie within an interval of given fixed width at a prescribed confidence level using a procedure which overcomes ignorance about σ^2 with no more than a finite number of observations. That is, the expected sample size exceeds the (fixed) sample size one would use if σ^2 were known by a finite amount, the difference depending on the confidence level α but *not* depending on the values of the mean μ, the variance σ^2 and the interval width $2d$. A number of unpublished results on the moments of the sample size are presented. Some do not depend on an assumption of normality.

2. Introduction. Let $X, X_1, X_2, \cdots$ be iid random variables with unknown mean μ and unknown variance $\sigma^2 < \infty$. Let $\bar{X}_n \equiv n^{-1} \sum_1^n X_i$. We desire to find a confidence interval for μ of width $2d$ $(d > 0)$ for which the probability of coverage is at least as large as $\alpha(0 < \alpha < 1)$ for all values μ and σ^2. N. Starr [6] and Chow-Robbins [4] have proposed using the interval $(\bar{X}_N - d, \bar{X}_N + d)$ where sample size N is to be sequentially determined. Let a be defined by $2\Phi(a) - 1 = \alpha$ where $\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-u^2/2}\, du$. With X normally distributed, if σ^2 were known, one could use a fixed sample size $N \geqq C \equiv a^2\sigma^2/d^2$. They reason that when σ^2 is unknown one might estimate σ^2 by some good estimator s_n^2 and use a sequential procedure of the basic form

$$(1) \qquad N \equiv \text{smallest index } n \geqq n_0 \geqq 2 \quad \text{for which} \quad n \geqq a_n^2 s_n^2/d^2,$$

where n_0 is an integer constant and where the a_n are chosen to be either identical to a or such that $0 < a_n \to a$. With $s_n^2 \equiv (n-1)^{-1} \sum_1^n (X_i - \bar{X}_n)^2$, Chow and Robbins showed that no matter what continuous distribution X might have,

$$(2) \qquad \lim_{d\to 0} P\{|\bar{X}_N - \mu| < d\} = \alpha \qquad (\text{"asymptotic consistency"})$$

and

$$(3) \qquad \lim_{d\to 0} EN/C = 1 \qquad (\text{"asymptotic efficiency"}).$$

H. Chernoff and the author (unpublished) have shown a stronger efficiency result which holds when $a_n \equiv a$, namely,

$$(4) \quad EN \leqq C + n_0 + 1 \qquad (\text{independent of } d, a, \text{ and the distribution of } X).$$

One is tempted to claim that the "cost of ignorance" in not knowing σ^2 is at

Received 6 February 1968.
[1] Now at the University of North Carolina, Chapel Hill.

most $n_0 + 1$ observations. However, the objective is not to achieve asymptotic consistency but rather to achieve

$$(5) \qquad P\{|\bar{X}_N - \mu| < d\} \geqq \alpha \qquad \text{for all values of } \mu \text{ and } \sigma^2.$$

One must be able to satisfy (5) before one can properly assess the true cost of ignorance. Starr conducted a numerical study for normally distributed X using a particular sequence a_n (of the form $a + O(n^{-1})$). It appears he nearly achieves objective (5). One can show for $a_n = a + O(n^{-1})$ that

$$(6) \qquad EN \leqq C + O(1) \quad \text{as} \quad d \to 0.$$

Thus it seems likely and we shall verify for normally distributed X that there exist stopping rules N for which (5) holds and for which the cost of ignorance, $EN - C$, is uniformly bounded for all μ, σ^2 and $d > 0$. Specifically, for some integer k, we can achieve these objectives by taking k more observations after rule (1) says to stop. We shall need $n_0 \geqq 3$ (not 2) and we shall be satisfied with $a_n \equiv a$. Stopping rules of this type are suggested in Starr's paper but they were not analyzed mathematically.

A useful random variable related to N (defined by (1)) is the variable

$$(7) \qquad \begin{aligned} M &\equiv \text{last index } m \geqq n_0 \text{ for which } m < a_m^2 s_m^2/d^2 \text{ if such an } m \text{ exists,} \\ &\equiv n_0 - 1 \text{ if } m \geqq a_m^2 s_m^2/d^2 \text{ for all } m \geqq n_0, \\ &\equiv \infty \text{ if } m < a_m^2 s_m^2/d^2 \text{ infinitely often.} \end{aligned}$$

Such a random variable is not a stopping variable but rather a *reverse stopping variable*, one that depends on the future and not on the past. If M and N are usually close we can hope to learn something about N by studying M.

In Section 3, we define and relate reverse stopping variables to (reverse) martingales. In Section 4, we derive some preliminary results involving moments of M and N, and in Section 5, we prove the true cost of ignorance concerning σ^2 is a finite number of observation.

It may be recalled that C. Stein [7] showed that (5) could be accomplished for normal X using a two-stage procedure but for his procedure $EN - C$ is not bounded.

3. Reverse stopping variables and some martingale lemmas. Let $(\Omega, \mathfrak{F}, P)$ be a probability space and $\{\mathfrak{F}_j : \mathfrak{F}_j \subset \mathfrak{F}, j \in J\}$ be a non-increasing sequence of σ-fields where J is a continuous sequence of integers including possibly $\pm\infty$. We recall that a family $Z = \{Z_j, \mathfrak{F}_j, j \in J\}$ is called a *reverse martingale* if for all $j \in J$ (i) Z_j is an $\mathfrak{F}_j$-measurable random variable, (ii) $E|Z_j| < \infty$, and (iii) $\int_A Z_j = \int_A Z_k$ for all $k \in J$, $k \geqq j$, $A \in \mathfrak{F}_k$. The appropriateness of the term "reverse" comes from the observation that if Z is a reverse martingale, then by "reversing" the usual ordering of the indexing set J, we obtain a martingale. Extending this terminology, we say that a random variable M with values a.s. in J is a *reverse stopping variable* if $\{M = j\} \in \mathfrak{F}_j$ for all $j \in J$. We shall use the following trivial generalization of a result of Doob ([5], p. 300):

LEMMA 1. *Let Z be a reverse martingale and M a reverse stopping variable. If J has a first element j_0 (possibly $-\infty$), then*

$$E|Z_M| \leqq E|Z_{j_0}| < \infty \quad \text{and} \quad EZ_M = EZ_{j_0}. \tag{8}$$

A well known reverse martingale is $n^{-1}S_n$ where $S_n \equiv \sum_1^n X_i$ is a sum of n iid random variables and $E|X_1| < \infty$. More generally, sequences of U-statistics form reverse martingales.

LEMMA 2. (Berk [2]). *Let $X_1, X_2, \cdots$ be* iid *and $U_{j_0}, U_{j_0+1}, \cdots$ a sequence of U-statistics for some $j_0 \geqq 1$. If $E|U_{j_0}| < \infty$ and $\mathfrak{F}_j$ is the Borel field $\mathcal{B}(U_j, U_{j+1}, \cdots)$ for $j \geqq j_0$, then $\{U_j, \mathfrak{F}_j, j \geqq j_0\}$ is a reverse martingale.*

PROOF. For $j \geqq j_0$, let Y_j be the order statistic for the first j X's and let $\mathcal{A}_j$ be the Borel field $\mathcal{B}(Y_j, X_{j+1}, X_{j+2}, \cdots)$. Since $U_j = E^{\mathcal{A}_j}U_{j_0}$ and $\{\mathcal{A}_j\}$ is a non-increasing sequence of σ-fields, it follows from [5], pg. 293, that $\{U_j, \mathfrak{F}_j, j \geqq j_0\}$ is a reverse martingale.

LEMMA 3. *Let $Y_1, Y_2, \cdots$ be a sequence of independent random variables with a common (two parameter) gamma distribution (having a density of the form $c(\theta, \beta)x^{\beta-1}e^{-x/\theta}$; $\beta, \theta > 0$). For given $\lambda > 0$, let $S_n \equiv \sum_1^n Y_i$, $Z_n \equiv (S_n)^\lambda/E(S_n)^\lambda$, and $\mathfrak{F}_n$ be the Borel field $\mathcal{B}(Z_n, Z_{n+1}, \cdots)$ for $n = 1, 2, \cdots$. Then $\{Z_n, \mathfrak{F}_n; n \geqq 1\}$ is a reverse martingale.*

The proof is routine if one first derives the conditional distribution of S_n given S_{n+1}.

4. Some preliminary results involving M and N. Let M and N be defined by (7) and (1), respectively, with $a_n \equiv a$. We proceed with the notation of Section 2.

THEOREM 1. *The following results do not depend on X being normally distributed.*

$$EM \leqq C + (n_0 - 1)P\{M = n_0 - 1\} \leqq C + n_0 - 1; \tag{9}$$

$$EN \leqq C + 1 + (n_0 - 1)P\{M = n_0 - 1\} \leqq C + n_0; \tag{10}$$

$$EM \geqq C - 2 - 2n_0^{-1}. \tag{11}$$

PROOF. The fact $N \leqq M + 1$ and (9) imply (10). Now, using indicator functions,

$$M \leqq a^2 s_M^2/d^2 + (n_0 - 1)I_{[M=n_0-1]}. \tag{12}$$

(Defining $s_1^2 \equiv s_2^2$, s_M^2 is well defined for $M = 1$. This can occur when $n_0 = 2$. The event $[M = \infty]$ is null.)

Using Lemma 2 and then Lemma 1, we conclude first that $s_1^2, s_2^2, \cdots$ is a reverse martingale and then $Es_M^2 = \sigma^2$.

Since $C \equiv a^2\sigma^2/d^2$, (9) follows. A reverse martingale argument has been used by Starr and Woodroofe to prove the extremes of (10) in a similar way. Now define $M' = \max(M, n_0 + 1)$, a reverse stopping variable. If $M = n_0 - 1$ or n_0, then $M' = n_0 + 1$ and $M + 2 \geqq a^2s^2_{M'}/d^2$. If $M \geqq n_0 + 1$, then $M' = M$ and

$$M + 1 \geqq a^2s^2_{M+1}/d^2 = a^2M^{-1}\textstyle\sum_1^{M+1}(X_i - \bar{X}_{M+1})^2/d^2 \geqq a^2M^{-1}(M-1)s^2_{M'}/d^2.$$

In general,

$$(13)\qquad M \geqq a^2 s^2{}_{M'}/d^2 - 2 - 2n_0^{-1}$$

from which (11) follows.

We remark that $P\{M = n_0 - 1\} = o(1)$ as $d \searrow 0$. When X is normally distributed $P\{M = n_0 - 1\} = o(d^k)$ for any k as $d \searrow 0$. These lead to strong asymptotic upper bounds for EM and EN.

THEOREM 2. *The following results apply to normally distributed* X.

$$(14)\quad EM^\lambda \leqq C^\lambda + O(C^{\lambda-1}) \quad as \quad C \to \infty \quad for \quad \lambda = 1, 2, \cdots;$$

$$(15)\quad EN^\lambda \leqq C^\lambda + O(C^{\lambda-1}) \quad as \quad C \to \infty \quad for \quad \lambda = 1, 2, \cdots;$$

$$(16)\quad E(M - N) = O(1) \quad for\ any \quad n_0 \geqq 3;$$

$$(17)\quad EM^\lambda \geqq C^\lambda + O(C^{\lambda-1}) \quad as \quad C \to \infty \quad for \quad \lambda = 1, 2, \cdots;$$

$$(18)\quad EN^\lambda \geqq C^\lambda + O(C^{\lambda-1}) \quad as \quad C \to \infty \quad for\ any \quad n_0 \geqq 3, \quad for \quad \lambda = 1, 2, \cdots.$$

PROOF. The fact $N \leqq M + 1$ and (14) imply (15). Jensen's inequality and (11) imply (17). Jensen's inequality, (11) and (16) imply (18). If X is normally distributed, we can write $\sum_1^n (X_i - \bar{X}_n)^2 = \sigma^2 \sum_2^n u_i$ where $u_2, u_3, \cdots$ are iid chi-square random variables with one degree of freedom. According to Lemma 3,

$$Z_n^{(\lambda)} \equiv \left(\textstyle\sum_2^n u_i\right)^\lambda / E\left(\textstyle\sum_2^n u_i\right)^\lambda = \Gamma((n-1)/2)\left(\textstyle\sum_2^n u_i\right)^\lambda / (2^\lambda \Gamma((n-1)/2 + \lambda))$$

is a reverse martingale with $EZ_n^{(\lambda)} = 1$ for $n = 2, 3, \cdots$, and fixed $\lambda > 0$. It easily follows (by definition), for (positive) integer valued λ, that

$$(19)\qquad (s_n^2)^\lambda = \sigma^{2\lambda} Z_n^{(\lambda)}(1 + O(1/n)).$$

(12) and (19) combine to give $M^\lambda \leqq C^\lambda Z_M^{(\lambda)} + O(M^{\lambda-1})$. Using Lemma 1 and trivial induction we derive (14). Finally (16) follows directly from

LEMMA 4. *For* $0 < \theta < 1$,

$$(20)\qquad P\{N \leqq \theta C\} = O_e(C^{-(n_0-1)/2}) \quad as \quad C \to \infty,$$

where O_e *denotes exact order*;

$$(21)\qquad E(M - N \mid N = n) \leqq C + 1 \quad for\ all \quad n \geqq n_0;$$

and for $n > \theta C$, $\theta > \frac{1}{2}$,

$$(22)\qquad E(M - N \mid N = n) \leqq K(\theta), \text{ a constant (independent of } C).$$

PROOF. Let $u_1, u_2, \cdots$ be a sequence of iid random variables distributed as χ_1^2 (chi-square with one degree of freedom) which we will use in various contexts below.

PROOF of (20). Let $0 < \theta < 1$.

$$P\{N \leqq \theta C\} \leqq P_1 + P_2 + P_3,$$

where

$$P_1 \equiv P\{n_0 \leqq N \leqq 2n_0\},$$
$$P_2 \equiv P\{2n_0 < N \leqq C^{\frac{1}{2}}\},$$
$$P_3 \equiv P\{C^{\frac{1}{2}} < N \leqq \theta C\}.$$

Now for $n \geqq n_0$,

$$P\{N = n\} \leqq P\{n \geqq a^2 s_n^2/d^2\} = P\{n(n-1) \geqq C\chi^2_{n-1}\} = O_e(C^{-(n-1)/2}) \quad \text{as} \quad C \to \infty.$$

(The last equality may be easily verified.) Thus for large C,

$$P\{N \leqq \theta C\} > P\{N = n_0\} = O_e(C^{-(n_0-1)/2}),$$

and

$$P_1 = O_e(C^{-(n_0-1)/2}).$$

Using a fairly well-known result concerning the probability of a random walk crossing a linear boundary (e.g., Section 2.1 Bartlett [1]), we find for $\gamma < 0 < \beta < 1$, that

$$(23) \qquad P\{\textstyle\sum_1^m u_i \leqq \beta\, m + \gamma \text{ for some (positive) } m\} \leqq e^{-\gamma h},$$

where $e^{-2\beta h} = 1 - 2h$ defines negative valued h.

Now for $0 < \beta < 1$,

$$(24) \qquad \text{(a)} \quad h < \beta - 1; \qquad \text{(b)} \quad h < (2\beta)^{-1} \log \beta.$$

The first inequality is immediate upon an expansion of $\log(1 - 2h)$ about $h = 0$. By substituting $(2\beta)^{-1}\log\beta$ for h in $e^{-2\beta h} = 1 - 2h$ one can easily verify (24b). Hence (for large C)

$$P_2 = P\{2n_0 < N \leqq C^{\frac{1}{2}}\} \leqq P\{\textstyle\sum_1^m u_i \leqq C^{-1}m(m+1) \text{ for some } m, 2n_0 \leqq m < C^{\frac{1}{2}}\} \leqq P\{\textstyle\sum_1^m u_i \leqq C^{-1}(2n_0 + C^{\frac{1}{2}} + 1)m - 2C^{-\frac{1}{2}}n_0 \text{ for some } m \geqq 1\}.$$

By (23) and (24b),

$$P_2 \leqq O(C^{-n_0/2}) \leqq O_e(C^{-(n_0-1)/2}).$$

P_3 is shown to be of a sufficiently small order in the same manner if we use (24a) instead of (24b).

PROOF OF (21). If $N = n$, the point $(n, \sum_2^n u_i)$ is below the parabola $C^{-1}x(x-1)$ and either $M - n = -1$ or $M - n =$ last time $k \geqq 1$ such that $(k, \sum_{n+1}^{n+k} u_i)$ is above the parabola $C^{-1}(x+n)(x+n-1) - \sum_2^n u_i$. In either case, $M - n \leqq M^*$ where $M^* \equiv$ last $k \geqq 1$ such that $(k, \sum_{n+1}^{n+k} u_i)$ is above the parabola $C^{-1}x(x-1)$. But $EM^* \leqq C + 1$ (cf. (9)) and (21) follows.

PROOF OF (22). Let $n > \theta C$, $\theta > \frac{1}{2}$.

$$\begin{aligned} P\{M = n + k \mid N = n\} &\leqq P\{n + k < a^2 s_{n+k}^2/d^2 \mid N = n\} \\ &\leqq P\{(n + k)(n + k - 1) < C \textstyle\sum_2^{n+k} u_i \mid n(n - 1) = C \sum_2^n u_i\} \\ &\leqq P\{C \textstyle\sum_{n+1}^{n+k} u_i > (n + k)(n + k - 1) - n(n - 1)\} \\ &\leqq P\{C \textstyle\sum_1^k u_i > 2nk\} \\ &\leqq P\{\textstyle\sum_1^k u_i > 2\theta k\}. \end{aligned}$$

Since $2\theta > EU_1 = 1$, we know from (for instance) H. Chernoff [3] that there exists a constant $b = b(\theta) > 0$ for which the latter probability is bounded above by e^{-bk}. Then

$$(25)\quad E(M - N \mid N = n) \leqq \textstyle\sum_{k=1}^{\infty} kP\{M = n + k \mid N = n\} \leqq \sum_{k=1}^{\infty} ke^{-bk} < b^{-2}$$

and hence (22) holds.

5. The cost of ignorance is a finite number of observations. Here we assume that $X, X_1, X_2, \cdots$ are normal iid random variables with mean μ and variance σ^2 and as before $d > 0$. Let $r \equiv d/\sigma$, so $C = a^2/r^2$.

MAIN THEOREM. *If the value of any stopping variable N is determined by $s_2^2, s_3^2, \cdots$, then*

$$(26)\qquad P\{|\bar{X}_N - \mu| < d\} = 2E\Phi(rN^{\frac{1}{2}}) - 1 \quad \textit{for all} \quad \mu, \sigma^2.$$

For N defined by (1) *with $a_n \equiv a$ and $n_0 \geqq 3$, we have for some finite integer $k \geqq 0$,*

$$(27)\quad E\Phi(r(N + k)^{\frac{1}{2}}) \geqq \Phi(a) = (1 + \alpha)/2 \quad \textit{for all} \quad \mu, \sigma^2, \quad \textit{and} \quad d.$$

Then

$$(28)\qquad P\{|\bar{X}_{N+k} - \mu| < d\} \geqq \alpha \quad \textit{for all} \quad \mu, \sigma^2 \quad \textit{and} \quad d,$$

and

$$(29)\qquad E(N + k) \leqq C + n_0 + k \quad \textit{for all} \quad \mu, \sigma^2 \quad \textit{and} \quad d.$$

PROOF. The random variables $\bar{X}_n$ and $Y_n = (s_2^2, \cdots, s_n^2)$ are independent for $n = 2, 3, \cdots$ when X is normal. (See for instance N. Starr [6].) Thus, if the events $[N = n] \varepsilon \mathcal{B}(s_2^2, \cdots, s_n^2)$, then

$$\begin{aligned} P\{|\bar{X}_N - \mu| < d\} &= \textstyle\sum_{n=n_0}^{\infty} P\{|\bar{X}_n - \mu| < d \text{ and } N = n\} \\ &= \textstyle\sum_{n=n_0}^{\infty} P\{|\bar{X}_n - \mu| < d\}P\{N = n\} = 2E\Phi(rN^{\frac{1}{2}}) - 1. \end{aligned}$$

For $g(x) \equiv \Phi(rx^{\frac{1}{2}})$, $g'(x) = r\varphi(rx^{\frac{1}{2}})/(2x^{\frac{1}{2}})$ and $g''(x) = -r(r^2x + 1)\varphi(rx^{\frac{1}{2}})/(4x^{\frac{3}{2}})$, where $\varphi(y) \equiv (2\pi)^{-\frac{1}{2}}e^{-y^2/2}$. Expand $g(x)$ in a Taylor series about $x = C$ with a second degree remainder term. We find for arbitrary θ, $0 < \theta < 1$, that

$$\begin{aligned} &E\Phi(r(N + k)^{\frac{1}{2}}) \geqq E\Phi(r(N + k)^{\frac{1}{2}})I_{[N+k \geqq \theta^2 C]} \\ &\quad \geqq \Phi(a)P\{N + k \geqq \theta^2 C\} + a\varphi(a)(2C)^{-1}E(N + k - C)I_{[N+k>\theta^2 C]} \\ &\qquad -a(a^2\theta^2 + 1)\varphi(a\theta)(8\theta^3 C^2)^{-1}E(N + k - C)^2 I_{[N+k>\theta^2 C]} \end{aligned}$$

$$\geqq \Phi(a) + a\varphi(a)(2C)^{-1}E(N + k - C)$$
$$- a(a^2\theta^2 + 1)\varphi(a\theta)(8\theta^3C^2)^{-1}E(N + k - C)^2$$
$$+ \{-\Phi(a) + a\varphi(a)(1 - \theta^2)/2 + a(a^2\theta^2 + 1)\varphi(a\theta)(1 - \theta^2)^2/(8\theta^3)\}$$
$$\cdot P\{N + k \leqq \theta^2 C\}.$$

For small $\theta > 0$ the coefficient of $P\{N + k \leqq \theta^2C\}$ is positive and for such θ,

$$E\Phi(r(N + k)^{\frac{1}{2}}) \geqq \Phi(a) + a\varphi(a)(2C)^{-1}\{k + E(N - C)\}$$
$$- a(a^2\theta^2 + 1)\varphi(a\theta)(8\theta^3C^2)^{-1}\{k^2 + 2kE(N - C) + E(N - C)^2\}.$$

By (15) and (18),

$$E\Phi(r(N + k)^{\frac{1}{2}}) - \Phi(a) \geqq O(C^{-2})k^2 + \{a\varphi(a)(2C)^{-1} + O(C^{-2})\}k + O(C^{-1})$$
$$= \{a\varphi(a)(2C^{-1}) + O(C^{-2})\}\{O(C^{-1})k^2 + k + O(1)\}.$$

Thus for some large k and for all large C (say $C \geqq C_0$) (27) holds and clearly for some large k (27) holds for all $C < C_0$. Thus (27) holds for some integer $k \geqq 0$. (28) follows from (26) and (27), and (29) from (10).

6. Acknowledgment. I am indebted and wish to thank Professor Herman Chernoff for discussions which led to the proof of (20) and also to thank Professor W. J. Hall for directing my attention to the paper of R. Berk.

REFERENCES

[1] BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge Univ. Press.
[2] BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrest. *Ann. Math. Statist.* **37** 51–58.
[3] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.
[4] 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.
[5] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[6] STARR, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.
[7] STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.