The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 704–709

A GENERAL METHOD OF DETERMINING FIXED-WIDTH CONFIDENCE INTERVALS

BY RASUL A. KHAN

Columbia University

0. Summary. A general method for determining stopping rules to obtain a fixed-width confidence interval of prescribed coverage probability for an unknown parameter of a distribution is obtained. Asymptotic theory in the sense of Chow and Robbins [4] is discussed. The sequential procedure obtained is asymptotically consistent and efficient in the sense of Chow and Robbins [4].

1. Introduction. Fixed-width confidence interval estimation for the mean of a normal distribution has been considered by Ray [6] and Starr [7] etc. The analogous problem for the variance of a normal population has been considered by Graybill and Connell [5] by using two stage sampling. Chow and Robbins [4] have considered the problem of determining a confidence interval of prescribed width and prescribed coverage probability for the unknown mean of a population with unknown finite variance. They constructed a stopping rule and thereby developed an asymptotic theory in a certain sense. When there are some nuisance parameters present, presumbaly unknown, fixed sample size procedure will usually not work to obtain a fixed-width interval with a given coverage probability. But there are examples where there are no nuisance parameters and still the fixed sample size procedure does not work, e.g., for the variance of a normal population with zero mean. In all such cases a stopping rule can be adopted which will provide a bounded length confidence interval of given coverage probability. However, bounded length confidence intervals with prescribed coverage probability have been treated in few special cases. The object of this note is to give a general method of constructing sequential procedure for obtaining fixed-width confidence intervals of prescribed coverage probability for an unknown parameter of a distribution involving possibly some unknown nuisance parameters. The distribution involved will be assumed to be known except for the parameters. For the sake of simplicity, the discussion is restricted to the case of a single nuisance parameter since the case of several nuisance parameters is immediate.

Let $p(x, \theta_1, \theta_2)$ be the probability density function of a random variable X (for convenience with respect to Lebesque measure) with real valued parameters θ_1 and θ_2 where θ_2 is regarded as nuisance parameter. We want to determine a confidence interval of fixed-width $2d$ $(d > 0)$ for θ_1 when both θ_1 and θ_2 are unknown, with preassigned coverage probability $1 - \alpha$ $(0 < \alpha < 1)$.

ASSUMPTION. We assume that all the regularity assumptions of maximum likelihood estimation are satisfied.

Received 27 July 1967; revised 13 November 1968.

NOTATION AND PRELIMINARIES. N will denote a bonafide stopping variable (i.e. N is a positive integer-valued random variable such that the stopping set $\{N = n\} \varepsilon \mathfrak{F}_n$ where $\mathfrak{F}_n \uparrow$ sub-σ-algebras are the σ-algebras of subsets generated by $X^n = (X_1, \cdots, X_n)$ and $P(N < \infty) = 1$). n will denote the fixed size of a random sample. Fisher's information matrix is

$$I(n) = n(l_{ij}), \qquad i, j = 1, 2,$$

where

$$l_{ij} = -E(\partial^2 \log p(x, \theta_1, \theta_2)/\partial\theta_i\partial\theta_j).$$

We assume (l_{ij}) to be positive definite. And $(l_{ij})^{-1} = (\lambda_{ij}) = \Lambda$, i.e. $I^{-1}(n) = \Lambda/n$.

$\hat{\theta}_1(n)$ and $\hat{\theta}_2(n)$ will denote maximum likelihood estimators (possibly unbiased which can be done in most of the cases by trivial modification of the MLE, though unbiasedness is not essential for our discussion) of θ_1 and θ_2 respectively based on a random sample of size n. It should be noted that $\hat{\theta}_1(n)$ is asymptotically normal $N(\theta_1, \lambda_{11}/n)$ where $\lambda_{11} = \lambda_{11}(\theta_1, \theta_2)$ since in general the l_{ij}'s are functions of θ_1 and θ_2.

Let $\{a_n, n \geqq 1\}$ be a sequence of positive constants converging to a constant a such that

$$(2\pi)^{-1/2} \int_{-a}^{a} e^{-x^2/2}\, dx = 1 - \alpha.$$

Let

$$I_n = [\hat{\theta}_1(n) - d, \hat{\theta}_1(n) + d] \quad \text{and}$$

$$n_d = \text{smallest integer} \geqq a^2\lambda_{11}(\theta_1, \theta_2)/d^2 = n_0. \tag{1.1}$$

From (1.1) it follows that $\lim_{d\to 0} n_d = \infty$ and

$$\lim_{d\to 0} [d^2 n_d/a^2\lambda_{11}(\theta_1, \theta_2)] \geqq 1.$$

Therefore,

$$\begin{aligned}\lim_{d\to 0} \Pr\{\theta_1 \varepsilon I_{n_d}\} &= \lim_{d\to 0} \Pr\{n_d^{1/2}|\hat{\theta}_1(n_d) - \theta_1|/\lambda_{11}^{1/2} \leqq d(n_d/\lambda_{11})^{1/2}\} \\ &= \Pr\{|N(0, 1)| \leqq a'\}, a' \geqq a \\ &\geqq 1 - \alpha.\end{aligned}$$

We will treat n_0 as the optimum sample size if θ_1 and θ_2 were known. This is not justified in the strict sense but will serve as a standard for comparison with the stopped random variable in the sequential procedure to be adopted. And in some cases n_0 might turn out to be optimum if only θ_2 were known and $\lambda_{11}(\theta_1, \theta_2) = \lambda_{11}(\theta_2)$, for example, in the case of a normal distribution $N(\mu, \sigma^2)$ where $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.

2. Stopping rule and the asymptotic theory. Now when θ_1 and θ_2 are unknown, a fixed n such as that determined by (1.1) will not be available to guarantee

fixed-width $2d$ and coverage probability $1 - \alpha$. So in analogy to (1.1) we adopt the following sequential rule. Let m be a given fixed positive integer.
R: Starting with $n \geqq m$, stop whenever for the first time

$$n \geqq a_n^2 \lambda_{11}(\hat{\theta}_1(n), \hat{\theta}_2(n))/d^2, \tag{2.1}$$

i.e. $N = \inf \{n \geqq m : n \geqq a_n^2 \hat{\lambda}_{11}(n)/d^2\}$ where $\hat{\lambda}_{11}(n) = \lambda_{11}(\hat{\theta}_1(n), \hat{\theta}_2(n))$.

LEMMA. *Under the sole assumption $\lambda_{11}(\theta_1, \theta_2) < \infty$, the sequential process terminates with probability* 1.

PROOF. Under the regularity assumptions, $\hat{\lambda}_{11}(n) \to \lambda_{11}(\theta_1, \theta_2)$ with probability 1. Hence the lemma follows from the fact that the right hand member of (2.1) $\to n_0$ with probability 1 which in turn implies $\Pr\{N = \infty\} = 0$.

Before discussing the asymptotic theory in the sense of Chow and Robbins [4] we state the following due to Anscombe [1].

Let $\{Y_n, n \geqq 1\}$ be an infinite sequence of random variables and suppose that there exist a real number θ, a sequence of positive nubmers $\{\omega_n\}$ and a distribution function $\mathfrak{F}(x)$, such that the following conditions are satisfied:

(A1) Convergence in law of Y_n : For any continuity point x of $\mathfrak{F}(x)$,

$$\Pr\{Y_n - \theta \leqq x\,\omega_n\} \to_{\mathfrak{L}} \mathfrak{F}(x) \qquad \text{as} \quad n \to \infty.$$

(A2) Uniform continuity in probability of Y_n : Given any $\epsilon > 0$ and $\eta > 0$, there is large ν and small positive c such that, for any $n > \nu$,

$$\Pr\{|Y_{n'} - Y_n| < \epsilon\omega_n \quad \text{simultaneously for all integers } n' \text{ such that } |n' - n| < cn\} > 1 - \eta$$

THEOREM (Anscombe). *Let $\{n_t\}$ be an increasing sequence of positive integers $\to \infty$ and let $\{N(t)\}$ be a sequence of positive integer-valued proper random variables such that $N(t)/n_t \to 1$ in probability as $t \to \infty$. Then if the sequence of random variables Y_n satisfies conditions* (A1) *and* (A2),

$$\Pr\{Y_{N(t)} - \theta \leqq x\omega_{n_t}\} \to_{\mathfrak{L}} \mathfrak{F}(x) \qquad \text{as} \quad t \to \infty.$$

In our case, $Y_n = \hat{\theta}_1(n)$, $\mathfrak{F}(x) = \Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-u^2/2}\,du$ and note that the condition (A1) is evidently satisfied by $\hat{\theta}_1(n)$ upon taking $\omega_n = (\lambda_{11}/n)^{\frac{1}{2}}$ and $\theta = \theta_1$. As demonstrated by Anscombe [1], $\hat{\theta}_1(n)$ also satisfies (A2) (Theorem 4 of [1]).

THEOREM 1. *Under the assumption $E(\sup_n \hat{\lambda}_{11}(n)) < \infty$,*

(i) $\lim_{d\to 0} N/n_0 = 1$ a.s.,
(ii) $\lim_{d\to 0} \Pr.\{\theta_1 \,\varepsilon\, I_N\} = 1 - \alpha$, *asymptotic consistency,*
(iii) $\lim_{d\to 0} E(N)/n_0 = 1$, *asymptotic efficiency.*

PROOF. To prove (i) let $y_n = \hat{\lambda}_{11}(n)/\lambda_{11}$, $f(n) = na^2/a_n^2$ and $t = a^2\lambda_{11}(\theta_1, \theta_2)/d^2 = n_0 \to \infty$ as $d \to 0$. Then the conditions of Lemma 1 of Chow and Robbins [4] are satisfied and hence

$$\lim_{t\to\infty} N/t = \lim_{d\to 0} N/n_0 = 1 \quad \text{a.s.}$$

To prove (ii) we observe that $N(t)/t \to 1$ a.s. as $t \to \infty$ and hence $N(t)/n_t \to 1$

a.s. as $t \to \infty$ where $n_t = [t]$ greatest integer $\leqq t$. We have already noted that $\hat{\theta}_1(n)$ satisfies conditions (A1) and (A2) and hence by the theorem of Anscombe it follows that

$$(n_t/\lambda_{11}(\theta_1, \theta_2))^{\frac{1}{2}}(\hat{\theta}_1(N(t)) - \theta_1) \to_{\mathcal{L}} N(0, 1) \quad \text{as} \quad t \to \infty.$$

Further,

$$(N(t)/\lambda_{11}(\theta_1, \theta_2))^{\frac{1}{2}}(\hat{\theta}_1(N(t)) - \theta_1) = (N(t)/n_t)^{\frac{1}{2}}(n_t/\lambda_{11})^{\frac{1}{2}}(\hat{\theta}_1(N(t)) - \theta_1).$$

Now since $N(t)/n_t \to 1$ a.s. as $t \to \infty$, therefore a well known theorem of Cramér [3] implies

$$(N(t)/\lambda_{11}(\theta_1, \theta_2))^{\frac{1}{2}}(\hat{\theta}_1(N(t)) - \theta_1) \to_{\mathcal{L}} N(0, 1) \quad \text{as} \quad t \to \infty.$$

Also from (i) it follows that $d(N/\lambda_{11})^{\frac{1}{2}} \to_{\text{a.s.}} a$ as $d \to 0$. Therefore

$$\lim_{d\to 0} \Pr\{\theta_1 \varepsilon I_N\} = \lim_{t\to\infty} \Pr\{(N(t)/\lambda_{11})^{\frac{1}{2}}|\hat{\theta}_1(N(t)) - \theta_1| \leqq d(N(t)/\lambda_{11})^{\frac{1}{2}}\}$$
$$= \Pr\{|N(0, 1)| \leqq a\} = 1 - \alpha.$$

And finally, (iii) follows from Lemma 2 of Chow and Robbins [4].

It should be noted that (i) and (ii) are universally valid and the assumption $E(\sup_n \hat{\lambda}_{11}(n)) < \infty$ is required only for the validity of (iii). However, in some cases it might be possible to establish (iii) without the hypothesis

$$E(\sup_n \hat{\lambda}_{11}(n)) < \infty$$

by using Lemma 3 of Chow and Robbins [4].

3. Examples. (a) $N(\mu, \sigma^2)$ $(0 < \sigma^2 < \infty)$: Taking $\theta_1 = \mu$, $\theta_2 = \sigma^2$, it is easily found that the information matrix is

$$(l_{ij}) = \begin{bmatrix} \theta_2^{-1} & 0 \\ 0 & \frac{1}{2}\theta_2^{-2} \end{bmatrix} \quad \text{and} \quad \lambda_{11}(\theta_1, \theta_2) = \theta_2 = \sigma^2.$$

The maximum likelihood estimators of θ_1 and θ_2 are, respectively,

$$\hat{\theta}_1(n) = \bar{X}_n = (n)^{-1}\textstyle\sum_{i=1}^n x_i \quad \text{and} \quad \hat{\theta}_2(n) = n^{-1}\textstyle\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Instead of using $\hat{\theta}_2(n)$, we can use $\hat{\theta}_2^*(n) = (n\hat{\theta}_2(n))/(n - 1) = S_n^2$ which is an unbiased and consistent estimator of θ_2. Hence we obtain the following stopping rule.

$$R{:}N = \inf\{n \geqq 2{:}n \geqq a_n^2 S_n^2/d^2\}$$
$$\text{and} \quad n_0 = (a^2\sigma^2)/d^2.$$

(b) $N(\mu, \sigma^2)(0 < \sigma^2 < \infty){:}\theta_1 = \sigma^2$, $\theta_2 = \mu$. Then the information matrix is

$$(l_{ij}) = \begin{bmatrix} \theta_1^{-2}/2 & 0 \\ 0 & \theta_1^{-1} \end{bmatrix} \quad \text{and} \quad \lambda_{11}(\theta_1, \theta_2) = 2\theta_1^2 = 2\sigma^4.$$

$$R\!:\!N = \inf\{n \geqq 2 : n \geqq (2a_n^2 S_n^4)/d^2\}$$
$$\text{and} \qquad n_0 = (2a^2\sigma^4)/d^2.$$

(c)' $p(x' \mid \theta) = \theta \exp(-\theta x)$, $x \geqq 0$, $0 < \theta < \infty$. Then

$$i(\theta) = -E(\partial^2 \log p(x \mid \theta)/\partial\theta^2) = \theta^{-2} \qquad \text{and}$$
$$R\!:\!N = \inf\{n \geqq 1 : n \geqq a_n^2/(\bar{X}_n^2 d^2)\} \quad \text{and} \quad n_0 = a^2\theta^2/d^2.$$

That the hypothesis $E(\sup_{n \geqq 2} \hat{\lambda}_{11}(n)) < \infty$ is true in (a) and (b) follows from the following lemma which is proved from Wiener's theorem. However, the hypothesis is not true in (c) and hence (iii) cannot be concluded from Lemma 2 of [4]. We first state Wiener's theorem without proof which can be found in [8].

THEOREM (Wiener's special case). *Let $\{X_n, n \geqq 1\}$ be a sequence of* iid *random variables with $E|X_n|^r < \infty$ or $E|X_n|^r \log^+|X_n| < \infty$ according as $r > 1$ or $r = 1$. Then*

$$E(\sup_{n \geqq 1} n^{-r} |\textstyle\sum_{i=1}^n X_i|^r) < \infty$$

and conversely.

LEMMA. *Under the assumption $0 < \sigma^2 < \infty$,*

$$E(\sup_{n \geqq 2} S_n^q) < \infty \quad \textit{for} \quad q \geqq 2$$

where S_n^2 is defined in (a).

PROOF. For $q = 2$, $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = (n-1)^{-1}\sum_{i=1}^n X_i^2 - n(n-1)^{-1}\bar{X}_n^2$. Since

$$S_n^2 \leqq (n-1)^{-1}\textstyle\sum_{i=1}^n X_i^2 - n(n-1)^{-1}\bar{X}_n^2 + (n-1)^{-1}\bar{X}_n^2 + 2\bar{X}_n^2$$
$$\leqq X_1^2 + (n-1)^{-1}\textstyle\sum_{i=2}^n X_i^2 + \bar{X}_n^2.$$

Therefore,

$$\sup_{n \geqq 2} S_n^2 \leqq X_1^2 + \sup_{n \geqq 2}(n-1)^{-1}\textstyle\sum_{i=2}^n X_i^2 + \sup_{n \geqq 1} n^{-2}(\sum_{i=1}^n X_i)^2.$$

Therefore, $E(\sup_{n \geqq 2} S_n^2) < \infty$ if $EX^2 \log^+|X|^2 < \infty$ and $EX^2 < \infty$. But $EX^2 \log^+|X|^2 \leqq EX^4 < \infty$ and $EX^2 < \infty$ are true for normal distribution with finite variance. Now assume $q > 2$. Then,

$$S_n^q = (S_n^2)^{q/2} \leqq [(n-1)^{-1}\textstyle\sum_{i=1}^n X_i^2 + \bar{X}_n^2]^{q/2}$$

or

$$\leqq 2^{q/2}[(n-1)^{-q/2}(\textstyle\sum_{i=1}^n X_i^2)^{q/2} + |\bar{X}_n|^q]$$
$$\leqq 2^{q/2}[2^{q/2}\{|X_1^q| + (n-1)^{-q/2}(\textstyle\sum_{i=2}^n X_i^2)^{q/2}\} + |\bar{X}_n|^q].$$

Therefore,

$$\sup_{n \geqq 2} S_n^q \leqq 2^{q/2}[2^{q/2}\{|X_1^q| + \sup_{n \geqq 2}(n-1)^{-q/2}(\textstyle\sum_{i=2}^n X_i^2)^{q/2}\}$$
$$+ \sup_{n \geqq 1} n^{-q}|\textstyle\sum_{i=1}^n X_i|^q].$$

Therefore, $E(\sup_{n \geqq 2} S_n^{\,q}) < \infty$ if $E|X_1|^q < \infty$ which is true in a normal distribution with finite variance. This completes the proof of the lemma.

REMARK. It should be noted that in case of a single parameter family of distributions, the stopping rule R as determined by (2.1) takes the form:

$$R^*: N = \inf \{n \geqq m : n \geqq a_n^{\,2}/d^2 i(\hat{\theta}_n)\}$$

where $i(\theta) = -E(\partial^2 \log p(X \mid \theta)/\partial\theta^2)$ and $\hat{\theta}_n$ is the MLE of θ. However, if $i(\theta)$ is independent of θ, no sequential procedure is required since the bounded length confidence intervals of given coverage probability can be based on normal theory. More generally, no sequential procedure is required when $\lambda_{11}(\theta_1, \theta_2) = \lambda_{11}(\theta_2)$ and θ_2 is known. As an example where this is the case, is a normal distribution with the unknown mean and known variance. The discussion of the note has been restricted to the case of a single nuisance parameter. However, this was done only for the sake of notational simplicity and the generalization to the case of k nuisance parameters is immediate.

Acknowledgment. I am sincerely thankful to Professor Howard Levene, Chairman of the Department of Mathematical Statistics, Columbia University for providing financial assistance. I wish to express my sincere thanks to Professor Robert A. Wijsman for some useful discussions. I am also thankful to the referee for his helpful comments to improve the earlier version of the paper.

REFERENCES

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

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

[3] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[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] GRAYBILL, FRANKLIN A. and CONNELL, TERRENCE L. (1964). Sample size required for estimating the variance width in d units of the true value. *Ann. Math. Statist.* **35** 438–440.

[6] 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.

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

[8] WIENER, N. (1939). The ergodic theorem. *Duke Math. J.* **5** 1–18.