SEQUENTIAL CONFIDENCE INTERVALS FOR THE MEAN OF A NORMAL DISTRIBUTION WITH KNOWN VARIANCE

By Charles Stein and Abraham Wald

Columbia University

1. Summary. We consider sequential procedures for obtaining confidence intervals of prescribed length and confidence coefficient for the mean of a normal distribution with known variance. A procedure achieving these aims is called optimum if it minimizes the least upper bound (with respect to the mean) of the expected number of observations. The result proved is that the usual non-sequential procedure is optimum.

2. Introduction. The problem of sequential confidence sets in general has been considered briefly by one of the authors [1]. Let $\{X_i\}$, $(i = 1, 2, \cdots)$, be a sequence of random variables whose distribution is specified except for the value of a parameter θ whose range is a space Ω. Sequential confidence sets are determined by a rule as to when to stop sampling, together with a function of the sample whose value is one of a specified class of subsets of Ω. The class of subsets is chosen in advance depending on the purpose of the estimation. For example, it may be the class of all intervals of prescribed length or the class of all sets whose diameter does not exceed a given value. It is required that the probability that this (random) set covers θ should be greater than or equal to a specified confidence coefficient α for all θ. A procedure for finding sequential confidence intervals is considered optimum if it minimizes some specified function of the expected numbers of observations. Here this function is taken to be the least upper bound. In contrast with the result of this paper, a case where sequential confidence intervals may have an advantage over non-sequential procedures has been given by one of the authors [2]. The X_i are independently normally distributed with unknown mean and unknown variance, and the problem is to find confidence intervals of fixed length for the unknown mean. As was first shown by Dantzig [3] this cannot be accomplished by a non-sequential procedure. Another case where this is true is the problem of finding confidence intervals of the form (p_0, kp_0) where k is a specified number greater than 1, for the probability in a binomial distribution.

Let $\{X_i\}$, $(i = 1, 2, \cdots)$, be independently normally distributed with unknown mean ξ and known variance σ_1^2. It is desired to specify a sequential procedure for obtaining confidence intervals of fixed length l for the mean ξ. This is provided by a rule according to which at each stage of the experiment, after obtaining the first m observations $X_1, \cdots, X_m$ for each integral value m, one makes one of the following decisions:

a) Take an $(m + 1)$st observation.

b) Terminate the procedure and state that the mean lies in the interval

$(Y - \frac{1}{2}l, Y + \frac{1}{2}l)$, where $Y = \mathfrak{C}_m(X_1, \cdots, X_m)$, $\mathfrak{C}_m$ being a measurable real-valued function. The serial number m of the observation on which the procedure terminates is, of course, a random variable and will be denoted by n.

For any relation R the symbol $P(R \mid \xi)$ will denote the probability that R holds when ξ is the true mean of X_i. The confidence coefficient of a sequential procedure S is defined by

$$\alpha(S) = \underset{\xi}{g.l.b.}\, P(Y - \tfrac{1}{2}l < \xi < Y + \tfrac{1}{2}l \mid \xi). \tag{1}$$

Denote by $n_0(S)$ the maximum expected number of observations, i.e.

$$n_0(S) = \underset{\xi}{\text{l.u.b.}}\, E(n \mid \xi, S) \tag{2}$$

where $E(n \mid \xi, S)$ denotes the expected value of n when ξ is the true mean and the procedure S is used.

A procedure S will be considered optimum if, for all S' such that $\alpha(S') = \alpha(S)$,

$$n_0(S) \leq n_0(S'). \tag{3}$$

It will be shown that an optimum procedure $S(\nu, c)$ can be obtained as follows:

a) For all $m < \nu$, a fixed positive integer, take another observation.

b) For $m = \nu$, terminate the procedure if

$$\sum_1^{\nu} X_i^2 - \frac{1}{\nu}\left(\sum_1^{\nu} X_i\right)^2 > c\sigma_1^2 \tag{4}$$

and let $Y = \frac{1}{\nu}\sum_1^{\nu} X_i$. (The inequality (4) is used merely as a device for fixing the probability of taking ν observations, this random event to be independent of whether $(Y - \frac{1}{2}l, Y + \frac{1}{2}l)$ covers ξ, given ν.)

c) Otherwise take a $(\nu + 1)$st observation, terminating the process, and let

$$Y = \frac{1}{\nu + 1}\sum_1^{\nu+1} X_i,$$

When $c = 0$, this is the usual non-sequential procedure.

Clearly,

$$\alpha[S(\nu, c)] = P\{\chi_{\nu-1}^2 > c\}H\left(\frac{\sqrt{\nu}\,l}{2\sigma_1}\right) + [1 - P\{\chi_{\nu-1}^2 > c\}]H\left(\frac{\sqrt{\nu+1}\,l}{2\sigma_1}\right), \tag{5}$$

where

$$H(u) = \frac{1}{\sqrt{2\pi}}\int_{-u}^{u} e^{-\frac{1}{2}x^2}\,dx = \sqrt{\frac{2}{\pi}}\int_0^u e^{-\frac{1}{2}x^2}\,dx. \tag{6}$$

Also

$$n_0[S(\nu, c)] = \nu + 1 - P\{\chi_{\nu-1}^2 > c\}, \tag{7}$$

By a proper choice of ν and c we can achieve any desired confidence coefficient

$\alpha \geq H\left(\frac{l}{\sqrt{2}\sigma_1}\right)$. There is no essential loss of generality in considering only the case $\sigma_1 = 1$, and this will be done in the remainder of this paper.

3. A lower bound for $n_0(S)$ and an upper bound for $\alpha(S)$. Consider any sequential procedure S for obtaining confidence intervals of length l. Put

$$\alpha(\xi, S) = P\{Y - \tfrac{1}{2}l < \xi < Y + \tfrac{1}{2}l \mid \xi\}. \tag{8}$$

That is, $\alpha(\xi, S)$ is the probability that the confidence interval will cover the true mean ξ when the procedure S is used. According to (1),

$$\alpha(S) = \underset{\xi}{\text{g.l.b.}}\ \alpha(\xi, S). \tag{9}$$

In order to obtain a lower bound for $n_0(S)$ and an upper bound for $\alpha(S)$, we suppose that the procedure S is applied when ξ is not a fixed number but a random variable normally distributed with mean 0 and variance σ^2. Then the probability that the confidence interval covers ξ is

$$\bar{\alpha}(\sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} \alpha(\xi, S)\, d\xi \geqq \alpha(S) \tag{10}$$

and the expected number of observations is

$$\bar{E}(n \mid \sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} E(n \mid \xi, S)\, d\xi \leqq n_0(S). \tag{11}$$

Let $p_m(\xi, S)$, ($m = 1, 2, \cdots$, ad. inf.), denote the probability that $n = m$ when ξ is the true mean and procedure S is used. Put

$$\bar{p}_m(\sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} p_m(\xi, S)\, d\xi. \tag{12}$$

Since

$$\bar{E}(n \mid \sigma, S) = \sum_{m=1}^{\infty} m\bar{p}_m(\sigma, S) \tag{13}$$

we obtain from (11)

$$\sum_{m=1}^{\infty} m\bar{p}_m(\sigma, S) \leqq n_0(S). \tag{14}$$

We shall now derive an upper bound for $\bar{\alpha}(\sigma, S)$. Since $X_i = \xi + \epsilon_i$ where the ϵ_i are independently normally distributed with mean 0 and variance 1, the joint distribution of ξ and X_i, ($i = 1, \cdots, m$), is a multivariate normal distribution with

$$E\xi = EX_i = 0 \tag{15}$$

and covariance matrix

$$E\begin{pmatrix}\xi\\ X_1\\ \vdots\\ X_m\end{pmatrix}(\xi, X_1, \cdots, X_m) = \begin{pmatrix}\sigma^2 & \sigma^2 & \cdots & \cdots & \sigma^2\\ \sigma^2 & \sigma^2+1 & \sigma^2 & \cdots & \sigma^2\\ \vdots & \sigma^2 & \sigma^2+1 & \cdots & \sigma^2\\ \vdots & \vdots & & & \vdots\\ \sigma^2 & \sigma^2 & \cdots & \cdots & \sigma^2+1\end{pmatrix}. \tag{16}$$

Thus the conditional distribution of ξ given $X_1, \cdots, X_m$ is normal with mean

$$E(\xi \mid X_1, \cdots, X_m) = (\sigma^2, \cdots, \sigma^2)\begin{pmatrix}\sigma^2+1 & \sigma^2 & \cdots & \sigma^2\\ \sigma^2 & \sigma^2+1 & \cdots & \sigma^2\\ \vdots & \vdots & & \vdots\\ \sigma^2 & \sigma^2 & \cdots & \sigma^2+1\end{pmatrix}^{-1}\begin{pmatrix}X_1\\ \vdots\\ X_m\end{pmatrix}$$

$$= \sigma^2(1, 1, \cdots, 1)\begin{pmatrix}\frac{(m-1)\sigma^2+1}{m\sigma^2+1} & -\frac{\sigma^2}{m\sigma^2+1} & \cdots & -\frac{\sigma^2}{m\sigma^2+1}\\ -\frac{\sigma^2}{m\sigma^2+1} & \frac{(m-1)\sigma^2+1}{m\sigma^2+1} & \cdots & -\frac{\sigma^2}{m\sigma^2+1}\\ \vdots & \vdots & & \vdots\\ -\frac{\sigma^2}{m\sigma^2+1} & -\frac{\sigma^2}{m\sigma^2+1} & \cdots & \frac{(m-1)\sigma^2+1}{m\sigma^2+1}\end{pmatrix} \tag{17}$$

$$\times\begin{pmatrix}X_1\\ \vdots\\ X_m\end{pmatrix} = \frac{\sigma^2}{m\sigma^2+1}\sum_1^m X_i$$

and variance

$$\sigma^2 - \frac{\sigma^4}{(m\sigma^2+1)^2}E\left(\sum_{i=1}^m X_i\right)^2 = \frac{\sigma^2}{m\sigma^2+1}. \tag{18}$$

If $X_1, \cdots, X_m$ is a sequence for which the process is terminated on the mth trial, the conditional probability that the interval of length l will cover ξ is clearly maximized by taking

$$Y = E(\xi \mid X_1, \cdots, X_m) = \frac{\sigma^2}{m\sigma^2+1}\sum_1^m X_i \tag{19}$$

and, by (18) this probability has the value $H(c_m)$ where H is defined by (6) and

$$c_m = \sqrt{m + \frac{1}{\sigma^2}}\,\frac{l}{2}. \tag{20}$$

Hence,

$$\bar{\alpha}(\sigma, S) \leq \sum_{m=1}^{\infty} \bar{p}_m(\sigma, S)H(c_m). \tag{21}$$

From this and (10) we obtain

$$\alpha(S) \leq \sum_{1}^{\infty} \bar{p}_m(\sigma, S)H(c_m). \tag{22}$$

This upper limit of $\alpha(S)$ and the lower limit of $n_0(S)$ given in (14) will be used later to prove that $S(\nu, c)$ is an optimum procedure.

4. Maximum value of $\sum_{1}^{\infty} \bar{p}_m(\sigma, S)H(c_m)$ subject to the condition that $\sum_{1}^{\infty} m\bar{p}_m(\sigma, S)$ does not exceed a given bound. We shall show that the maximum of $\sum_{1}^{\infty} \bar{p}_m(\sigma, S)H(c_m)$ subject to

$$E(n \mid \sigma, S) = \sum_{1}^{\infty} m\bar{p}_m(\sigma, S) \leqq \nu + a,$$

where ν is a positive integer and $0 \leq a < 1$, is obtained by choosing $\bar{p}_m(\sigma, S) = p_m^*$ defined by

$$\begin{aligned} p_m^* &= 0 \text{ for } m < \nu \text{ or } m > \nu + 1 \\ p_\nu^* &= 1 - a \\ p_{\nu+1}^* &= a. \end{aligned} \tag{23}$$

For, suppose to the contrary that there exists a sequence $\{p_m\}$ such that the following conditions hold:

$$\begin{gathered} p_m \geq 0, \qquad \sum_{1}^{\infty} p_m = 1 \\ \sum_{1}^{\infty} mp_m \leq \nu + a = \sum_{1}^{\infty} mp_m^* \\ \sum_{1}^{\infty} p_m H(c_m) > \sum_{1}^{\infty} p_m^* H(c_m). \end{gathered} \tag{24}$$

We have

$$H(u) = \sqrt{\frac{2}{\pi}} \int_0^u e^{-\frac{1}{2}x^2}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^{u^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}\, dy. \tag{25}$$

Put

$$C = H(c_{\nu+1}) - H(c_\nu) = \frac{1}{\sqrt{2\pi}} \int_{c_\nu^2}^{c_\nu^2+1} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}\, dy. \tag{26}$$

With the aid of $p_\nu = 1 - \sum_{m \neq \nu} p_m$, we obtain from the last two inequalities in (24)

$$0 < \sum_1^\infty (p_m - p_m^*)H(c_m) - C \sum_1^\infty (p_m - p_m^*)m = \sum_{m \neq \nu} (p_m - p_m^*)K_m \tag{27}$$

where

$$K_m = H(c_m) - H(c_\nu) - (m - \nu)[H(c_{\nu+1}) - H(c_\nu)]. \tag{28}$$

Clearly $K_{\nu+1} = 0$. Also, for $m < \nu$, since the integrand is a strictly decreasing function of y,

$$\begin{aligned} K_m &= (\nu - m) \int_{c_\nu^2}^{c_{\nu+1}^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}\, dy - \int_{c_m^2}^{c_\nu^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}\, dy \\ &< (\nu - m) \frac{l^2}{4} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \Big|_{y=c_\nu^2} - (\nu - m) \frac{l^2}{4} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \Big|_{y=c_\nu^2} = 0. \end{aligned} \tag{29}$$

Similarly for $m > \nu + 1$, $K_m < 0$. But $p_m^* = 0$ for $m \neq \nu, \nu + 1$ so that

$$\sum_{m \neq \nu, \nu+1} (p_m - p_m^*)K_m \leq 0 \tag{30}$$

which contradicts (27) since $K_{\nu+1} = 0$.

Thus, we have shown that the inequality

$$\bar{E}(n \mid \sigma, S) \leq \nu + a \tag{31}$$

implies the inequality

$$\sum_1^\infty \bar{p}_m(\sigma, S)H(c_m) \leq (1 - a)H(c_\nu) + aH(c_{\nu+1}). \tag{32}$$

5. Proof that $S(\nu, c)$ is an optimum procedure. Since, according to (14) and (22)

$$n_0(S) \geq \bar{E}(n \mid \sigma, S) \quad \text{and} \quad \alpha(S) \leq \sum_1^\infty \bar{p}_m(\sigma, S)H(c_m), \tag{33}$$

it follows from the result expressed in (31) and (32) that, for any procedure S satisfying the inequality

$$n_0(S) \leq \nu + a, \tag{34}$$

we must have

$$\alpha(S) \leq (1 - a)H(c_\nu) + aH(c_{\nu+1}) \tag{35}$$

identically in σ. Since $H(u)$ is continuous, it follows that

$$\alpha(S) \leq (1 - a)H\left(\sqrt{\nu}\, \frac{l}{2}\right) + aH\left(\sqrt{\nu + 1}\, \frac{l}{2}\right) \tag{36}$$

for any procedure S satisfying (34).

The right hand side of (36) is $\alpha[S(\nu, c)]$ where c is chosen so that

$$1 - a = P\{\chi^2_{\nu-1} > c\}. \tag{37}$$

We use an indirect proof to show that $S(\nu, c)$ is an optimum procedure. Suppose to the contrary that there is a procedure S' such that

$$\alpha(S') = \alpha[S(\nu, c)] \tag{38}$$

but

$$n_0(S') < n_0[S(\nu, c)]. \tag{39}$$

By (5) and (7), $\alpha[S(\nu, c)]$ is a continuous strictly increasing function of

$$\nu + 1 - P\{\chi^2_{\nu-1} > c\}$$

and this latter is $n_0[S(\nu, c)]$. If we choose ν', c' so that

$$\begin{aligned} n_0(S') &< \nu' + 1 - P\{\chi^2_{\nu-1} > c'\} \\ &< \nu + 1 - P\{\chi^2_{\nu-1} > c\}, \end{aligned} \tag{40}$$

it follows that

$$\alpha[S(\nu', c')] < \alpha[S(\nu, c)] = \alpha(S'). \tag{41}$$

But (41) and the first part of (40) contradict the result expressed in (34) and (36).

REFERENCES

[1] A. Wald, *Sequential Analysis*, John Wiley and Sons, 1947, section 11.2.
[2] Charles Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 243–258.
[3] G. B. Dantzig, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ", *Annals of Math. Stat.*, Vol. 11 (1940), p. 186.