REMARKS ON SEQUENTIAL POINT ESTIMATION

BY NORMAN STARR AND MICHAEL B. WOODROOFE

UNIVERSITY OF MICHIGAN

Communicated by Paul A. Smith, March 12, 1969

Abstract.—We consider utilizing a sequential experiment for estimating the mean of a normal distribution when the variance of the distribution is unknown. For a suitable choice of the design constants of the experiment and for loss structures of practical interest, it is shown that the difference in the expected loss with optional stopping and the expected loss which would be incurred if σ were known is a bounded function of σ.

1. *Introduction.*—H. Robbins[1] considered the problem of estimating the mean of a normal distribution and suggested a sequential procedure that Starr[2] subsequently studied in some detail. Since this procedure is easy to use in practice and applies to a rather frequently encountered statistical problem, results relating to its performance should have considerable practical importance.

2. *The Problem.*—The results of this section are given in reference 2, and we refer the reader to that paper for proofs. Let $x_1,x_2, \ldots$ be independent $N(\mu,\sigma^2)$ random variables; and, for given $s > 0$, suppose the loss incurred when we estimate μ by $\bar{x}_n = 1/n \sum_{n=1}^{n} x_i$ on the basis of a sample of size $n > 0$ is defined by

$$L_n = A|\bar{x}_n - \mu|^s + n \qquad (A > 0).$$

For fixed n the resulting risk is

$$\nu_n(\sigma) = E(L_n) = \frac{2}{s} K\sigma^s n^{-s/2} + n, \tag{1}$$

where the constant $K = 2^{(s-1)/2} A\Gamma(s + 1/2)\pi^{-1/2}$ depends on the choice of A,s only. Define the constant $\beta = K^{2/s}$; then, treating $n > 0$ as a continuous variable, the risk (1) is minimized by taking n to be n_0, where by definition

$$n_0 = (\beta\sigma^2)^{s/(s+2)} \qquad (0 < \sigma < \infty). \tag{2}$$

Replacing this optimal choice of n in (1), the minimum risk is easily seen to be

$$\nu(\sigma) = \nu_{n_0}(\sigma) = \left(\frac{2}{s} + 1\right) n_0. \tag{3}$$

When σ is unknown, a poor choice of n in relation to σ in advance of experimentation will magnify (1). Consequently, we consider, as in references 1 and 2, determining a random sample size N by means of the following sequential procedure. Define

$$s_n^2 = 1/n - 1 \sum_{i=1}^{n} (x_i - \bar{x}_n)^2, \quad n \geq 2,$$

and let

$$N = \text{least integer } n \geq m \text{ for which } n \geq (\beta s_n^2)^{s/(s+2)},$$

where the *starting sample size* $m \geq 2$ is a given integer.

The risk $\bar{\nu}$ when the sample size N is sequentially determined can be shown to be

$$\bar{\nu}(\sigma) = E(L_N) = \frac{2}{s} n_0^{(s+2)/2} E(N^{-s/2}) + EN. \tag{4}$$

As possible measures of the usefulness of this procedure, we define for each σ, $0 < \sigma < \infty$, the *risk efficiency*

$$\eta(\sigma) = \frac{\bar{\nu}(\sigma)}{\nu(\sigma)},$$

and the *regret*

$$\omega(\sigma) = \bar{\nu}(\sigma) - \nu(\sigma).$$

It is known[2] that

$$\lim_{\sigma \to \infty} \eta(\sigma) = \begin{cases} 1 & \text{when } m > \dfrac{s^2}{s+2} + 1 \\ 1 + c & \phantom{\text{when }} m = \dfrac{s^2}{s+2} + 1, \\ \infty & \phantom{\text{when }} m < \dfrac{s^2}{s+2} + 1 \end{cases} \tag{5}$$

where the constant $c > 0$ depends on A,s. Thus the procedure is asymptotically risk-efficient, provided that the starting sample size m is chosen so that $m > s^2/(s + 2) + 1$.

Moreover, Robbins' computations[1] suggest that the procedure should be satisfactory for all σ. We shall here support that thesis by proving that the regret suffered in using the sequential procedure in ignorance of σ is *uniformly* bounded, provided that $m \geq s + 1$.

3. *Results.*—We begin by stating two useful lemmas.

LEMMA 1. $P(N = m) = O_e(\sigma^{-m+1})$, *where* O_e *denotes exact order.*

LEMMA 2. *For fixed* θ, $0 < \theta < 1$,

$$P(m < N < \theta n_0) = O(\sigma^{-m}).$$

COROLLARY 1. $P(N \leq \theta n_0) = O(\sigma^{-m+1})$. The lemmas were proved by Starr[2] and have been rederived by G. Simons[3] for a related problem.

THEOREM. *As* $\sigma \to \infty$,

$$\omega(\sigma) = O(1) \tag{6}$$

if and only if

$$m \geq s + 1. \tag{7}$$

Proof: For simplicity in what follows, we let c with or without affixes denote a positive constant that does not depend on σ and is not necessarily the same from one usage to the next.

From (3) and (4)

$$\omega(\sigma) = \frac{2}{s} n_0^{(s+2)/2} E(N^{-s/2} - n_0^{-s/2}) + E(N - n_0). \tag{8}$$

For fixed $t > 0$

$$N^{-t} - n_0^{-t} = -\frac{t}{n_0^{t+1}}(N - n_0) + \frac{t(t+1)}{2}\frac{(N - n_0)^2}{n_1^{t+2}}, \tag{9}$$

where n_1 is an intermediate value of n_0,N. Setting $t = s/2$, we obtain from (8)

$$\begin{aligned}\omega(\sigma) &= \left(\frac{s+2}{4}\right) n_0^{(s+2)/2} E[n_1^{-(s+4)/2}(N - n_0)^2] \\ &= O_e(\sigma^s)E[n_1^{-(s+4)/2}(N - n_0)^2].\end{aligned}$$

Since it can be verified from (9) that on the event $\{N = m\}$

$$n_1 \leq c n_0^{4/(s+4)} \qquad (n_0 \geq m),$$

the necessity of (7) for (6) follows from Lemma 1 by

$$\omega(\sigma) \geq O_e(\sigma^s)\frac{(m - n_0)^2}{n_0^2} P(N = m) = O_e(\sigma^{s+1-m}).$$

Moreover, for fixed θ, $0 < \theta < 1$, we can obtain from (9) that on the event $\{N \leq \theta n_0\}$

$$n_1 \geq c n_0^{4/(4+s)} \qquad (n_0 \geq m).$$

Thus,

$$\begin{aligned}\omega(\sigma) &= O_e(\sigma^s)\left[\int_{N \leq \theta n_0} n_1^{-(s+4)/2}(N - n_0)^2 dP + \int_{N > \theta n_0} n_1^{-(s+4)/2}(N - n_0)^2 dP\right] \\ &\leq O_e(\sigma^s)\left\{cP(N \leq \theta n_0) + O(\sigma^{-s})E\left[\frac{(N - n_0)^2}{n_0}\right]\right\} \\ &= O(\sigma^{s+1-m}) + c' E\left[\frac{(N - n_0)^2}{n_0}\right].\end{aligned}$$

Hence, the sufficiency of (7) follows from

LEMMA 3. *As* $\sigma \to \infty$

$$E\left[\frac{(N - n_0)^2}{n_0}\right] = O(1). \tag{10}$$

Proof of the lemma: It suffices to consider $n_0 \geq 10$. Integrating by parts, we have

$$E\left[\frac{(N - n_0)^2}{n_0}\right] \leq 1 + 2\int_1^{\sqrt{n_0}} \lambda P(N - n_0 < -\lambda\sqrt{n_0})d\lambda$$
$$+ 2\int_1^{\infty} \lambda P(N - n_0 > \lambda\sqrt{n_0})d\lambda. \tag{11}$$

The first integral in (9) does not exceed the sum of $n_0P(N \leq 1/2n_0)$ and

$$2\int_1^{1/2\sqrt{n_0}} \lambda P\left\{s_k^2 \leq \frac{(n_0 - \lambda\sqrt{n_0})^{(s+2)/s}}{\beta};\quad k \geq 1/2n_0\right\}d\lambda$$
$$\leq 2\int_1^{\infty} \lambda P\left\{s_k^2 - \sigma^2 \leq -\frac{s+2}{s\beta}(1/2n_0)^{2/s}(\lambda\sqrt{n_0});\quad k \geq 1/2n_0\right\}d\lambda$$
$$\leq 2\int_1^{\infty} \lambda\left(\frac{\alpha\beta}{\lambda}\right)^4 n_0^{-(8/s-2)} E|s_{n_2}^2 - \sigma^2|^4 d\lambda,$$

where we have let n_2 be the greatest integer less than or equal to $1/2n_0$, $\alpha = s/(s + 2)$, and have used a martingale inequality[4, 5] in the last step. Now for $m \geq s + 1$, $n_0P(N \leq 1/2n_0)$ is bounded by Corollary 1, while

$$n_0^{-(8/s+2)} E|s_{n_2}^2 - \sigma^2|^4 \leq cn_0^{-(8/s+2)} n_2^{-2}\sigma^8 \leq c'n_0^{-(8/s+4)}\sigma^8 = c''\sigma^{-8}\sigma^8 = c''$$

is bounded in any case. Thus, we have shown that the first integral in (11) is bounded if $m \geq s + 1$, and a similar (somewhat simpler) argument will establish the boundedness of the second, proving (10).

Remark: Robbins' computations[1] suggest that $O(1)$ in (6) is small, and useful estimates of this constant depending on m,σ could perhaps be found by refining the argument given above.

[1] Robbins, H., "Sequential estimation of the mean of a normal population," in *Probability and Statistics*, ed. Harold Cramér (Uppsala, Sweden: Almqvist and Wiksell, 1959), pp. 235–245.

[2] Starr, N., "On the asymptotic efficiency of a sequential procedure for estimating the mean," *Ann. Math. Statist.*, **37**, 1173–1185 (1966).

[3] Simons, G., "On the cost of not knowing the variance when making a fixed-width confidence interval for the mean," *Ann. Math. Statist.*, **39**, 1946–1952 (1968).

[4] Doob, J. L., *Stochastic Processes* (New York: John Wiley & Sons, 1953), p. 34.

[5] Starr, N., and M. B. Woodroofe, "Remarks on a stopping time," these PROCEEDINGS, **61**, 1215 (1968).