CONFIDENCE INTERVALS WITH FIXED PROPORTIONAL ACCURACY*

Michael WOODROOFE

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

Abstract: The problem of setting a confidence interval with fixed proportional accuracy for the mean of a normal distribution is considered. A sequential procedure is proposed; and asymptotic expansions for its average coverage probabilities and its expected sample size are obtained. The procedure is shown to have asymptotically minimal expected sample size, subject to a constraint on the error probabilities.

1. Introduction

The possibility of using sequential methods to set a fixed width confidence interval for the mean of a distribution with an unknown variance has attracted substantial interest in theoretical statistics. This interest was stimulated by Stein's (1945) two-stage procedure, the fully sequential procedures of Anscombe (1953) and Chow and Robbins (1965) for the normal and non-parametric cases, and Hall's (1981) three stage procedure. The recent paper by Finster (1985) describes multiparameter extensions and may be consulted for further references.

Much of this research exploits the independence of the sample mean and variance in the normal case. The independence provides a simple relation between the coverage probability of a sequential procedure and the first two moments of the stopping time; and this relation permits a second order asymptotic expansion for the coverage probability as the width of the interval shrinks to zero. See, for example, Anscombe (1953), Hall (1981), and Woodroofe (1982, Section 10.2). The second order terms in these expansions are especially interesting, since the effect of optional stopping appears there.

The objective of this paper is to provide a simple example of a related problem in which the estimator and the stopping time are much more highly dependent, but a weak asymptotic expansion is still possible. Suppose that potential observations $X_1, X_2, \ldots$ are independent and normally distributed with unknown mean θ, $-\infty < \theta < \infty$, and unit variance, and a confidence interval with fixed proportional accuracy is sought. Let

$$S_n = X_1 + \cdots + X_n \quad \text{and} \quad \bar{X}_n = S_n/n, \quad n \geq 1,$$

denote the partial sums and averages of $X_1, X_2, \ldots$; and consider estimators of θ of the form

$$\hat{\theta}_n = \bar{X}_n - \hat{b}_n/n, \quad n \geq 1, \tag{1}$$

where

$$\hat{b}_n = b_n(\bar{X}_n), \quad n \geq 1,$$

and b_n, $n \geq 1$, is a convergent sequence of absolutely continuous functions on $(-\infty, \infty)$. (In the sequential case, the sample mean may be a biased estimator, and an appropriate choice of b_n, $n \geq 1$, may reduce bias.)

Next, let Δ denote a continuous, almost everywhere positive function on $(-\infty, \infty)$, the proportionality function; and let $\hat{\Delta}_n = \Delta(\bar{X}_n)$ denote the maximum likelihood estimators of $\Delta(\theta)$ for $n > 0$. If $t = t(X_1, X_2, \ldots)$ is a stopping time, let

$$\gamma_h(t; \theta) = P_\theta\{|\hat{\theta}_t - \theta| \leq h\sqrt{\hat{\Delta}_t}\} \tag{2}$$

denote the probability that $\hat{\theta}_t$ differs from θ by at most $h\sqrt{\hat{\Delta}_t}$. Given $0 < \gamma < 1$ and a small $h > 0$, a stopping timne t is sought for which $\gamma_h(t; \theta) \approx \gamma$ for all $-\infty < \theta < \infty$. The closely related forulation in which $\sqrt{\hat{\Delta}_t}$ is replaced by $\sqrt{\Delta(\theta)}$ in (2) is considered too.

Motivation for the formulation(s) is provided by supposing that θ is the mean difference between two treatments which have been administered to a pair of patients in a clinical trial. If $|\theta|$ is large, then ethical considerations demand that the experiment be terminated as soon as the sign of θ is determined, since then one number of each pair is receiving a greatly inferior treatment. While, if $|\theta|$ is small, then the ethical demands are less, and more data is needed to determine the sign of θ. Proportionality functions of the form $\Delta(\theta) = 1 + a\theta^2$, with $a > 0$, are suggested in such a case.

The exact meaning of $\approx$ here is complicated. If ξ is a density on $(-\infty, \infty)$, and t is a stopping time, then the *average confidence level* of t unde ξ is defined to be

$$\gamma_h(t; \xi) = \int_{-\infty}^{\infty} \gamma_h(t; \theta)\xi(\theta)\, d\theta, \tag{3}$$

and the approximate equality $\approx$ is interpreted to mean that

$$\gamma_h(t; \xi) = \gamma + o(h^2) \tag{4}$$

as $h \to 0$ for a large class of densities ξ, twice continuously differentiable densities with compact support.

Average confidence levels have been considered recently by Stein (1981) and by Woodroofe (1986) in a sequential setting. Rationale for replacing confidence levels by their averages is provided in Section 1 and 6 of the latter.

The paper proceeds as follows. A sequential procedure is defined and studied in Section 2. Some properties of average confidence levels are described in Section 3. The main result, an asymptotic expansion for the average confidence levels of the sequential procedure, is presented in Section 4. The sequential procedure is shown to have asymptotically minimal expected sample size, subject to (4), in Section 5 and 6.

2. The procedure

To understand the procedure, it is useful to have a first approximation to the sample size required. If $\hat{\Delta}_t$ were replaced by $\Delta(\theta)$ and t were replaced by a fixed sample size n, in the formulation then

$$P_\theta\{|\bar{X}_n - \theta| \le h\sqrt{\Delta(\theta)}\} \ge \gamma \quad \text{iff} \quad n \ge N = c^2/h^2\Delta(\theta), \tag{5}$$

where $2\Phi(c) - 1 = \gamma$ and Φ denotes the standard normal distribution function. Here N represents the first approximation.

The sequential procedure effectively estimates N and stops as soon as the sample size exceeds the current estimate. Let g_n, $n \ge 1$, denote convergent sequence of continuous functions on $(-\infty, \infty)$ for which $g_n > -n$ for all n; and let

$$c_n = c\sqrt{\{1 + \hat{g}_n/n\}},$$

where $\hat{g}_n = g_n(\bar{X}_n)$, for all $n = 1, 2, \ldots$. Then the sequential procedure $t = t_h$ is defined by

$$t = \inf\{n \ge m : n > c_n^2/h^2\hat{\Delta}_n\}, \tag{6}$$

where $m \ge 1$ denotes an initial sample size and the infimum of the empty set is understood to be ∞. In fact, by the Strong Law of Large Numbers and the Law of the Iterated Logarithm, $t < \infty$ w.p. 1 (P_θ) unless Δ and its first two derivatives all vanish at θ.

Properties of t may be deduced from the non-linear renewal theorem of Lai and Siegmund (1977, 1979), under modest conditions. To see how, write

$$t = \inf\{n \ge m; Z_n > c^2/h^2\},$$

where

$$Z_n = n\hat{\Delta}_n c^2/c_n^2 \quad \text{for } n \ge 1,$$

and observe that if Δ is differentiable at a fixed θ then Z_n, $n \geq 1$, may be written in the form

$$Z_n = T_n + U_n, \tag{7}$$

where

$$T_n = T_{n,\theta} = n\Delta(\theta) + \Delta'(\theta)(S_n - n\theta)$$

and

$$U_n = U_{n,\theta} = n[\Delta(\bar{X}_n) - \Delta(\theta) - \Delta'(\theta)(\bar{X}_n - \theta)]$$

$$+ n\Delta(\bar{X}_n)[c^2/c_n^2 - 1] = U_n' + U_n'', \quad \text{say}, \tag{8}$$

for $n = 1, 2, \ldots$. Since each T_n, $n \geq 1$, is a random walk, this is of the form considered in the non-linear renewal theorem.

To exploit this observation, some conditions are needed. In their statement, I denotes a compact interval,

$$E_I(W) = \sup_{\theta \in I} E_\theta(W) \tag{9}$$

for non-negative random variables W, possibly depending on θ, and $P_I(B) = E_I(I_B)$ for events B, possibly depending on θ.

Condition G on I: g_n, $n \geq 1$, are continuous functions on $(-\infty, \infty)$ for which $g_n \geq -\frac{1}{2}n$ on $(-\infty, \infty)$, $g_n \to g$ uniformly on I, and

$$E_I\left[\sup_n \hat{g}_n^4\right] < \infty. \tag{10}$$

Observe that (10) holds, if g_n, $n \geq 1$, are dominated by a polynomial.

Condition D on I: Δ is a continuous, almost everywhere positive function on $(-\infty, \infty)$; Δ is positive and twice continuously differentiable on some neighborhood of I; $|\Delta'| > 0$ a.e. on I; and (10) holds with $\hat{g}_n$ replaced by $\hat{\Delta}_n$ for $n \geq 1$.

Theorem 1. *Let I denote a compact interval; and suppose that conditions* D *and* G *are satisfied on I. Let*

$$t^* = t_h^* = (t - N)/\sqrt{N} \quad and \quad R_h = Z_t - c^2/h^2, \quad h > 0.$$

Then

$$t^* \Rightarrow N[0, \Delta'(\theta)^2/\Delta(\theta)^2]$$

as $h \to 0$ for all $\theta \in I$, where $\Rightarrow$ denotes convergence in distribution. If $\theta \in I$ and $\Delta'(\theta) \neq 0$, then

$$R_h \Rightarrow R \sim H_\theta,$$

where

$$H_\theta\{dr\} = [1/E_\theta(T_\sigma)]P_\theta\{T_\sigma > r\} \, dr, \quad r > 0,$$

with

$$\sigma = \inf\{n \geq 1: T_n > 0\},$$

and R_h is asymptotically independent of t^* as $h \rightarrow 0$, under P_θ. Finally, if $\theta \in I$ and $\Delta'(\theta) \neq 0$, then

$$\lim E_\theta(t - N) = \alpha(\theta)/\Delta(\theta)$$

where

$$\alpha = \varrho + \Delta g - \tfrac{1}{2} \Delta'', \quad and \quad \varrho(\theta) = \int_0^\infty r H_\theta\{dr\}.$$

Proof. The theorem follows directly from Theorems 4.2 and 4.5 of Woodroofe (1982), with A_k equal to the sample space and $l_k = 0$ for all k in the latter. The conditions imposed in these theorems are verified in Lemmas 1 and 2 below.

The calculation of quantities of the form ϱ is discussed by Woodroofe (1982, p. 33).

Recall that random variables V_n, $n \geq 1$, are said to be *slowly changing* iff $\max\{|V_k|: k < n\} = o_p(n)$ and for every $\varepsilon > 0$, there is a $\delta > 0$ with

$$P\left\{\max_{k \leq n\delta} |V_{n+k} - V_n| > \varepsilon\right\} < \varepsilon \quad \text{for all } n \geq 1.$$

Lemma 1. *Suppose that conditions* D *and* G *are satisfied on a compact interval* I. *Then, for each* $\theta \in I$,

$$U_n \Rightarrow \tfrac{1}{2}\Delta''(\theta)\chi_1^2 - \Delta(\theta)g(\theta) \tag{11}$$

as $n \rightarrow \infty$ *under* P_θ, *where* χ_1^2 *denotes a random variable which has the chi-squared distribution on* 1 *degree of freedom; and* U_n, $n \geq 1$, *are slowly changing under* P_θ. *Moreover,*

$$\sup_n E_I\left\{\max_{k < n} |U_{n+k}|^2\right\} < \infty \tag{12}$$

and

$$\sum_{n=1}^\infty P_I\{|U_n| > n\varepsilon\} < \infty \quad \text{for all } \varepsilon > 0. \tag{13}$$

Proof. Relation (11) is clear; and the slow change of U_n, $n \geq 1$, follows directly from Example 4.1 of Woodroofe (1982, pp. 41–42). To establish (12), write $U_n = U_n' + U_n''$, as in (8). Then

$$|U_n''| \leq 2\hat{\Delta}_n |\hat{g}_n|, \quad \text{for } n \geq 1;$$

so, U_n'', $n \geq 1$, satisfy (12) by Schwarz' Inequality, since $\hat{\Delta}_n$ and $\hat{g}_n$, $n \geq 1$, both satisfy (10). Next, let J denote a compact neighborhood of I on which Δ is twice continuously differentiable; let δ denote the distance from I to J', the complement of J; and let $K = \max_J |\Delta''|$. Then

$$\max_{k < n} |U_{n+k}'| \leq K \max_{k \leq 2n} (S_k - k\theta)^2/n$$

$$+ \sup_{k \geq n} |U_k'| \, I\{|\bar{X}_k - \theta| > \delta, \exists k \geq n\}$$

for all $n \geq 1$. Here the first term has a distribution which is independent of $\theta \in I$; and its square is uniformly integrable by standard maximal inequalities (cf. Woodroofe (1982, pp. 48–49)). The square of the second term is dominated, since Δ satisfies (10) and $k^4 P_0\{|\bar{X}_k| > \delta\}$ is summable for all $\delta > 0$. Finally, (13) follows from (12) and Markov's Inequality.

Lemma 2. *If conditions* D *and* G *are satisfied on a compact interval* I, *then there is an* $\varepsilon = \varepsilon(I)$ *for which*

$$P_I\{t \leq \varepsilon h^{-2}\} = o(h^2) \quad and \quad E_I\{tI\{t > 1/\varepsilon h^2\}\} = o(1)$$

as $h \to 0$.

Proof. Let J and δ be as in Lemma 1; let K and K' be upper bounds for Δ and Δ' on J; and let $a = c^2/h^2$ and $\varepsilon = 1/4K$. Then $t \leq \varepsilon a$ implies $\bar{X}_t \in J'$,

$$P_I\{t \leq \log^2 a\} \leq \sum_{k < \log^2 a} P_I\{2k\hat{\Delta}_k > a\}$$

$$\leq \left[\sum_{k < \log^2 a} (2k/a)^2 \right] E_I\left\{ \sup_n \hat{\Delta}_n^2 \right\}$$

$$= O[a^{-2}\log^6(a)] = o(h^2),$$

and

$$P_I\{\log^2 a < t < \varepsilon a\} \leq P_0\{|\bar{X}_k| > \delta, \exists k > \log^2 a\},$$

which is of smaller order of magnitude than h^2, by elementary calculations. This estabishes the first assertion of the lemma.

For the second, let $K_0 = \inf_J \Delta$ and $\varepsilon = K_0/2c^2$. If $n > 1/\varepsilon h^2$, $|\hat{g}_n/n| \leq 1$, and $\bar{X}_n \in J$, then $t \leq n$; so,

$$P_I\{t > n\} \leq P_I\{|\hat{g}_n/n| > 1\} + P_0\{|\bar{X}_n| > \delta\}$$

for all $n > 1/\varepsilon h^2$. Since the sequence on the right is summable, the second assertion follows easily by a simple summation by parts.

3. On average confidence levels

The key to computing average confidence levels is the following observation: if ζ is a density and t is stopping time, then

$$\gamma_h(t; \zeta) = P\{|\theta - \hat{\theta}_t| \leq h\sqrt{\hat{\Delta}_t}\}, \tag{14}$$

where $P = P^\zeta$ denotes probability in the Bayesian model in which θ has density ζ and $X_1, X_2, \ldots$ are conditionally independent normal random variables with mean θ and unit variance, given θ. The measure P may have to be defined on the product of the original space with $(-\infty, \infty)$; if so, the random variables $X_1, X_2, \ldots$ are to be

injected into the large space. Of course, optional stopping has a much smaller effect in the Bayesian model than it does on frequentist calculations.

Throughout this section, ξ denotes a fixed density with satisfies the following conditions: for some $-\infty < \theta_0 < \theta_1 < \infty$ and $q \geq 2$,

$$\xi(\theta) = (\theta - \theta_0)_+^q (\theta_1 - \theta)_+^q \zeta(\theta) \tag{15}$$

for $-\infty < \theta < \infty$, where ζ is a positive twice continuously differentiable function on $(-\infty, \infty)$ and $(x)_+ = \max\{x, 0\}$ for $-\infty < x < \infty$.

For each $n = 2, 3, \ldots$, let A_n denote the event

$$A_n = \{\theta_0 + \log n / \sqrt{n} < \bar{X}_n < \theta_1 - \log n / \sqrt{n}\}.$$

Then an easy exercise shows the existence of a constant $K = K(\xi)$, depending only on ξ, for which

$$P\left(\bigcup_{k=n}^{\infty} A_k'\right) \leq K(\log n / \sqrt{n})^3 \tag{16}$$

for all $n = 2, 3, \ldots$. For the details, see Woodroofe (1986), where it is also shown that

$$E\left\{\sup_{n>1} [(\bar{X}_n - \theta_0)_+^{-2} + (\theta_1 - \bar{X}_n)_+^{-2}] I_{A_n}\right\} < \infty. \tag{17}$$

For each $n = 1, 2, \ldots$, let $\mathcal{D}_n$ denote the sigma algebra generated by $X_1, \ldots, X_n$. Then, the conditional (posterior) distribution of θ given $\mathcal{D}_n$ has density

$$\xi_n(s) \propto \exp\{-n(s - \bar{X}_n)^2 / 2\} \xi(s)$$

for $-\infty < s < \infty$. The conditional distribution function of $\sqrt{n}(\theta - \bar{X}_n)$ given $\mathcal{D}_n$ is needed. Define e_1, e_2, and $r_n = r_n(z, \bar{X}_n)$ by

$$e_1 = \xi'/\xi, \qquad e_2 = \xi''/\xi \tag{18}$$

and

$$P\{\sqrt{n}(\bar{X}_n - \theta) \leq z \mid \mathcal{D}_n\} = \Phi(z) + (1/\sqrt{n})\hat{e}_{1n}\phi(z)$$

$$- (1/2n)\hat{e}_{2n} z\phi(z) + (1/n)r_n \tag{19}$$

for $-\infty < z < \infty$ and $n = 1, 2, \ldots$, where $0/0$ is to be interpreted as 0 in (18) and $\hat{e}_{in} = e_i(\bar{X}_n)$ for $i = 1, 2$ in (19). Then a Taylor series expansion and some algebra show that

$$r_n \to 0 \quad \text{w.p. } 1 \ (P),$$

$$|r_n| \leq K\{(\bar{X}_n - \theta_0)_+^{-2} + (\theta_1 - \bar{X}_n)_+^{-2}\} \quad \text{on } A_n \tag{20}$$

for all $-\infty < z < \infty$ and $n \geq 2$ for some K depending only on ξ. In addition, it is easily seen that $\hat{e}_{1n}^2$ and $|\hat{e}_{2n}|$ are bounded by the right side of (20) on A_n for some K.

With $\hat{\theta}_n$ as in (1), it follows from (19) and (20) that

$$P\{|\theta - \hat{\theta}_n| \leq h\sqrt{\hat{\Delta}_n} \mid \mathcal{D}_n\}$$

$$= 2\Phi(C_n) - 1 - (C_n/n)\phi(C_n)[\hat{b}_n^2 + 2\hat{b}_n\hat{e}_{1n} + \hat{e}_{2n}] + (1/n)r_{n,0}, \tag{21}$$

where

$$C_n = h\sqrt{(n\hat{\Delta}_n)} \tag{22}$$

and

$$|r_{n,0}| \leq K\{(\bar{X}_n - \theta_0)_+^{-2} + (\theta_1 - \bar{X}_n)_+^{-2}\}\{1 + \hat{b}_n^4\}/\sqrt{n} + 2\sup_z |r_n(z)|$$

on A_n for all $n = 2, 3, \ldots$ for some constant K.

In order to integrate these expansions, additional conditions are needed. Again, let I denote a compact interval.

Condition B on I: b_n, $n \geq 1$, are continuous functions on $(-\infty, \infty)$ which converge to an absolutely continuous limit, uniformly on I.

In the next lemma, $\mathscr{D}_t$ denotes the sigma algebra of events determined prior to a stopping time t.

Lemma 3. *Let ξ denote a density of the form* (15): *suppose that conditions* B, D, *and* G, *are satisfied on the support I of ξ; let $t = t_h$, $h > 0$, be any stopping times for which*

$$t/N \to 1 \quad \text{in probability}, \tag{23}$$

$$P\{t \leq \varepsilon h^{-2}\} = o(h^2) \tag{24}$$

as $h \to 0$ for some $\varepsilon > 0$; and let $F = F_h$ denote any $\mathscr{D}_t$-measurable events for which

$$P(F') = o(h^2) \quad \text{as } h \to 0. \tag{25}$$

Then

$$\gamma_h(t; \xi) = \int_B [2\Phi(C_t) - 1] \, dP - h^2 c^{-1} \phi(c) \int_{-\infty}^{\infty} \beta\xi \, d\theta + o(h^2),$$

where

$$B = F \cap A_t \cap \{t > \varepsilon h^{-2}\} \quad \text{and} \quad \beta = \Delta b^2 - 2(\Delta b)' + \Delta''.$$

Proof. It follows directly from (16) and (24) that

$$P(B') = o(h^2) \quad \text{as } h \to 0. \tag{26}$$

Since posterior distributions are unaffected by optional stopping, it follows from (21) and (26) that

$$\gamma_h(t; \xi) = P\{B, \ |\theta - \hat{\theta}_t| \leq h\sqrt{\hat{\Delta}_t}\} + o(h^2)$$

$$= \int_B P\{|\theta - \hat{\theta}_t| \leq h\sqrt{\hat{\Delta}_t}| \ \mathscr{D}_t\} \, dP + o(h^2)$$

$$= \int_B [2\Phi(C_t) - 1] \, dP - h^2 \int_B (1/h^2 t) C_t \Phi(C_t)[\hat{b}_t^2 + 2\hat{b}_t \hat{e}_{1t} + \hat{e}_{2t}] \, dP$$

$$+ \int_B (1/t) r_{t,0} \, dP + o(h^2) \tag{27}$$

as $h \to 0$. Since $\hat{b}_t$ are bounded B, it follows easily from (17) and (20) that the last line in (27) is $o(h^2)$ as $h \to 0$. Next, the integrand on the middle line converges to $c^{-1} \phi(c) \Delta [b^2 + 2be_1 + e_2]$ in probability as $h \to 0$; and the integrand is dominated on B, as above. So, the integral converges to

$$c^{-1} \phi(c) \int_{-\infty}^{\infty} \Delta [b^2 \xi + 2b\xi' + \xi''] \, d\theta = c^{-1} \phi(c) \int_{-\infty}^{\infty} [\Delta b^2 - 2(\Delta b)' + \Delta''] \xi \, d\theta,$$

by the dominated convergence theorem and a simple integration by parts. This completes the proof of the lemma.

Remark 1. The proofs of (26) and (27) used only conditions (24) and (25), not (23).

4. Average confidence of the sequential procedure

The main result is presented next.

Theorem 2. *Let I denote a compact interval; suppose that conditions B, D, and G hold on I; and define t by (6). Then*

$$\gamma_h(t; \xi) = \gamma + h^2 c^{-1} \phi(c) \int [\varrho + \Delta g - \beta] \xi \, d\theta + o(h^2) \tag{28}$$

as $h \to 0$, for all ξ of the form (15) with support in I, where β is as in Lemma 3.

Proof. Let ξ denote a fixed density, as described in the theorem; let ε be as in Lemma 2; and let $F = F_h$ be the event $F = \{ t \le 1/\varepsilon h^2 \}$. Then conditions (23), (24), and (25) are satisfied, by Theorem 1 and Lemma 2. So, Lemma 3 is applicable, and it suffices to consider the integral of $2\Phi(C_t) - 1$. Now,

$$C_t^2 = h^2 Z_t c_t^2 / c^2$$

$$= c^2 + h^2 R_h + c^2 \hat{g}_t / t + h^2 R_h \hat{g}_t / t.$$

Let
$$\psi(x) = 2\Phi(\sqrt{x}) - 1 \quad \text{for } x > 0.$$

Then

$$\int_B [2\Phi(C_t) - 1] \, dP = \gamma + \int_B [\psi(C_t^2) - \psi(c^2)] \, dP + o(h^2)$$

$$= \gamma + \int_B \psi'(C_t^{*2})(C_t^2 - c^2) \, dP + o(h^2)$$

for some intermediate point C_t^* between C_t and c. Now, $h^2 t \to c^2 / \Delta(\theta)$ and $C_t \to c$ w.p. 1 (P_θ) as $h \to 0$ for all $\theta \in I$; and $R_h \Rightarrow R \sim H_\theta$ under P_θ as $h \to 0$ for a.e. $\theta \in I$. It follows easily that

$$h^{-2} \psi'(C_t^{*2})(C_t^2 - c^2) \Rightarrow \psi'(c^2)[R + \Delta(\theta)g(\theta)]$$

under P as $h \to 0$, where the conditional distribution of R given θ is H_θ. Moreover, $C_t \geq \frac{1}{2}c$ on B, $\hat{g}_t^2$ are dominated, and $E[R_h^2 I_B]$ remains bounded as $h \to 0$, by Lemma 4 below. So,

$$h^{-2} \int_B \psi'(C_t^{*2})(C_t^2 - c^2) \, dP \to c^{-1} \phi(c) \int_{-\infty}^{\infty} [\varrho + \Delta g] \xi \, d\theta$$

as $h \to 0$. This completes the proof of the theorem, except for Lemma 4; the latter is presented below.

The function $\hat{g}_n$, $n \geq 1$, which appear in the definition of t, are design parameters. The first corollary indicates how they may be chosen to make the average confidence levels be at least $\gamma + o(h^2)$ for all ξ of the form (15).

Corollary 1. *If $\varrho + \Delta g \geq$ (=) β, then*

$$\gamma_h(t; \xi) \geq \ (=) \ \gamma + o(h^2) \tag{29}$$

as $h \to 0$, for all ξ of the form (15) with support in I.

Corollary 2. *If Conditions B, D, and G are satisfies on every compact interval I, then (28) holds for all ξ of the form (15); and if $\varrho + \Delta g \geq$ (=) β, then (29) holds for all ξ of the form (15).*

Example 1. Suppose that $b_n = 0$ for all $n \geq 1$, so that $\hat{\theta}_n = \bar{X}_n$ for all $n \geq 1$. Then the function g of Corollary 1 is $g = [\Delta'' - \varrho]/\Delta$. Let $g_n = \min\{n, \max\{-\frac{1}{2}n, g\}\}$ for $n \geq 1$. If Δ satisfies condition D on an interval I, then it is easily seen that g_n, $n \geq 1$, satisfy condition G on I.

Example 2. For the sequential procedure t, $\bar{X}_t$ may be a biased estimator of θ. In fact, if conditions D and G are satisfied on an interval I, then it may be shown that

$$E_\theta\{\bar{X}_t\} = \theta + h^2 c^{-2} \Delta'(\theta) + o(h^2)$$

for all $\theta \in I$, by using techniques described by Siegmund (1978). This suggests letting $b = \Delta'/\Delta$ and $\varrho + \Delta g = -\Delta(\log \Delta)''$. As above, if Δ satisfies condition D on an interval I, then appropriately truncated versions of b and g satisfy conditions B and G on I.

If $b = \Delta'/\Delta$, and if conditions B, D, and G are satisfied on an interval I, then the relation

$$P\{\theta - \hat{\theta}_t < -h\sqrt{\Delta_t}\} = P\{\theta - \hat{\theta}_t > h\sqrt{\Delta_t}\} + o(h^2),$$

for all ξ of the form (15) with support in I, may be proved along the lines of Theorem 2.

Remark 2. The function ϱ may be computed as in Example 2.4 of Woodroofe (1982) with $\mu = \Delta(\theta)$ and $\sigma = |\Delta'(\theta)|$.

Remark 3. Alternatively, if ϱ is replaced by 0 in (28), then the $=$ is replaced by $\geq$; and if $g = \beta/\Delta$, then the average coverage probabilities are $\geq \gamma + o(h^2)$ as $h \to 0$. The proofs of these relations are simpler than that of Theorem 1: one simply notes that $C_t^2 > c^2\{1 + \hat{g}_t/t\}$ and proceeds as in the proof of the theorem without the complicatiang references to R_h.

Remark 4. The conditions on the prior ϱ may be relaxable, although some smoothness is required. At present, the most general such conditions are not known (by the author).

Remark 5. Theorem 2 does not assert that $\gamma_h(t; \theta) - \gamma = h^2 c^{-1} \phi(c)[\varrho + \Delta g - \beta](\theta) + o(h^2)$ as $h \to 0$ for fixed θ. There are small oscillations, of order h^2, which are damped by the integration. See Woodroofe and Keener (1986).

For the alternative formulation, in which $\hat{\Delta}_n$ is replaced by $\Delta(\theta)$, let

$$\gamma^h(t; \theta) = P_\theta\{|\hat{\theta}_t - \theta| \leq h\sqrt{\Delta(\theta)}\} \tag{30}$$

for stopping times t, $h > 0$, and $-\infty < \theta < \infty$; and define average confidence levels $\gamma^h(t; \xi)$ by (2) with γ_h replaced by γ^h for densities ξ.

Theorem 3. *Let I be a compact interval; suppose that conditions* B, D, *and* G *are satisfied; and define t by* (6). *Then* (28) *with* γ_h *replaced by* γ^h *and* β *by*

$$\beta^* = \beta + \tfrac{1}{4}(c^4 - c^2)\Delta'^2/\Delta + \tfrac{1}{2}c^2\Delta'' - c^2 b\Delta'$$

holds as $h \to 0$, *for all* ξ *of the form* (15) *with support in* I.

The proof of Theorem 3 is similar to that of Theorem 2 and is, therefore, omitted. In addition, Corollaries 1 and 2 have obvious analogues.

In the proof of Lemma 4, it is convenient to let $N' = \varepsilon h^{-2}$ and $N'' = 1/\varepsilon h^2$, where ε is as in Lemma 2.

Lemma 4. *Under the assumptions of Theorem 2,*

$$E\{R_h^2 I_B\} = O(1) \quad as \ h \to 0.$$

Proof. Let $a = c^2/h^2$, let T_n, $n \geq 1$, be as in (7), and let

$$V = \max\{|U_n| : N' \leq n \leq N''\}.$$

Then $E(V^2)$ is bounded, by Lemma 1. Now $R_h = (T_t - a) + U_t$ and $|U_t| \leq V$ on B, so that

$$P\{B, R_h > 2r\} \leq P\{B, R_h > 2r, V \leq r\} + P\{V > r\}$$

for all $r>0$ and $h>0$. For fixed $\theta \in I$, let σ denote the first passage time for the unperturbed random walk,

$$\sigma = \inf\{n \geq N': T_n > a + r\}.$$

Then $B, R_h > 2r$, and $V \leq r$ imply $t = \sigma$. So, for all $r>0$, $h>0$, and $\theta \in I$,

$$P_\theta\{B, R_h > 2r, V \leq r\} \leq P_\theta\{T_\sigma - (a+r) > r\}$$

$$\leq \sup_{s>0} P_\theta\{T_1 > r+s \mid T_1 > s\}.$$

That

$$E\{R_h^2 I_B\} = 2 \int_0^\infty r P\{B, R_h > r\} \, dr$$

remains bounded as $h \to 0$ follows easily. This completes the proof of Lemma 4 and, therefore, that of Theorem 2 too.

5. Two lemmas

Two lemmas are needed to establish optimality.

Lemma 5. *Suppose that ξ is of the form* (15) *with support I. Suppose also that h is twice continuously differentiable on I, and let*

$$H_n = \hat{h}_n + (1/n)\hat{h}_n' E(e_1 \mid \mathcal{D}_n) + (1/2n)\hat{h}_n'',$$

when $\bar{X}_n \in I$, where $\hat{h}_n = h(\bar{X}_n)$, $\hat{h}_n' = h'(\bar{X}_n)$ and $\hat{h}_n'' = h''(\bar{X}_n)$ for $n \geq 1$. Then

$$\lim_{n \to \infty} E\left\{ \sup_{k>n} k \, |E(h(\theta) \mid \mathcal{D}_k) - H_k| \, I_{A_k} \right\} = 0.$$

Proof. By Taylor's theorem,

$$h(\theta) - \hat{h}_n = \hat{h}_n'(\theta - \bar{X}_n) + \tfrac{1}{2}\hat{h}_n''(\theta - \bar{X}_n)^2 + \tfrac{1}{2}[h''(\theta_n^*) - h''(\bar{X}_n)](\theta - \bar{X}_n)^2$$

for some intermediate point θ_n^* between θ and $\bar{X}_n$, whenever $\bar{X}_n \in I$ and $n \geq 1$. Moreover, if $\bar{X}_n \in I$, then a simple integration by parts shows that

$$E[(\theta - \bar{X}_n) \mid \mathcal{D}_n] = (1/n)E(e_1 \mid \mathcal{D}_n),$$

$$E[(\theta - \bar{X}_n)^2 \mid \mathcal{D}_n] = 1/n + (1/n^2)E(e_2 \mid \mathcal{D}_n).$$

See Lemma 1 of Woodroofe (1985). The lemma now follows easily from the fact that the conditional expectations on the right are uniformly integrable martingales and some simple algebra.

Lemma 6. *Suppose that ξ is of the form* (15) *and that condition D is satisfied on*

the support I of ξ; let $t = t_h$ be any stopping times for which (23) and (24) are satisfied; and define B as in Lemma 3 with any F for which (25) holds. Then

$$\int_B t[\Delta(\theta) - \hat{\Delta}_t] \, dP \to -\tfrac{1}{2} \int_{-\infty}^{\infty} \Delta''\xi \, d\theta \tag{31}$$

as $h \to 0$; and if t is the sequential procedure (6), then

$$\int_B \Delta(t - N) \, dP \to \int_{-\infty}^{\infty} \alpha\xi \, d\theta, \tag{32}$$

where $\alpha = \varrho + \Delta g - \tfrac{1}{2}\Delta''$, as in Theorem 1.

Proof. By Lemma 5, the integral in (31) is

$$\int_B tE(\Delta - \hat{\Delta}_t \mid \mathcal{D}_t] \, dP \to \int_{-\infty}^{\infty} (\Delta'\xi' + \tfrac{1}{2}\Delta''\xi) \, d\theta = -\tfrac{1}{2} \int_{-\infty}^{\infty} \Delta''\xi \, d\theta$$

as $h \to 0$, establishing the first assertion of the lemma. If t is the sequential procedure (6), then

$$\int_B \Delta(\theta)(t - N) \, dP = \int_B t[\Delta(\theta) - \hat{\Delta}_t] \, dP + \int_B (t\hat{\Delta}_t - c^2/h^2) \, dP.$$

The first integral is given by (31); and the second converges to $\int(\varrho + \Delta g)\xi \, d\theta$, as in the proof of Theorem 2.

Remark 6. Alternatively, (32) may be established by showing that the convergence in Theorem 1 is bounded.

6. Asymptotic optimality

In this section it is shown that the sequential procedure (6) is asymptotically optimal up to terms which are small compared to the cost of a single observation, among all stopping times for which (24) is satisfied. The proof studies a related decision problem in which the statistician must select a sample size n and incur a loss

$$L_h(n, \theta) = Ac^2h^{-2}I\{|\theta - \theta_n| > h\sqrt{\hat{\Delta}_t}\} + n\Delta(\theta) - c^2h^{-2}[A(1 - \gamma) + 1],$$

for $n \geq 1$, $-\infty < \theta < \infty$, and $h > 0$, where $A = 1/c\phi(c)$, ϕ denotes the standard normal density, and c denotes the $\tfrac{1}{2}(1 + \gamma)$ quantile of the standard normal distribution. If t is a stopping time and ξ is a density, then the risk function of t and the integrated risk of t with respect to ξ are denoted

$$r_h(t; \theta) = E_\theta[L_h(t; \theta)] \quad \text{and} \quad r_h(t; \xi) = \int_{-\infty}^{\infty} r_h(t; \theta)\xi(\theta) \, d\theta.$$

Observe that $r(t; \xi)$ may be written in the following two ways:

$$r_h(t; \xi) = E[L_h(t; \theta)],$$

$$r_h(t; \xi) = Ac^2 h^{-2}\{[1 - \gamma_h(t; \xi)] - [1 - \gamma]\} + \int_{-\infty}^{\infty} E_\theta(t - N)(\Delta\xi)(\theta)\, d\theta. \tag{33}$$

The three terms in L represent the cost of not covering θ, the cost of sampling, and a normalization. With this interpretation, the cost of a single observation is $\Delta(\theta)$.

Theorem 4. *Let I denote a compact interval; suppose that conditions* B *and* D *hold on* I; *and let* ξ *be any density of the form* (15) *with support in* I. *If* $t = t_h$, $h > 0$, *are any stopping times for which* (24) *holds, then*

$$\liminf_{h \to 0} r_h(t; \xi) \geq \int_{-\infty}^{\infty} (\beta - \tfrac{1}{2}\Delta'')\xi\, d\theta, \tag{34}$$

where $\beta = b^2 - 2(b\Delta)' + \Delta''$, *as in Theorem 1. Moreover, the limit exists and there is equality in* (34) *when* t *is the sequential procedure* (6) *with* $g = (\beta - \varrho)/\Delta$.

The proof of Theorem 4 is presented below. The following corollary shows that in order to have the average confidence levels be at least $\gamma + o(h^2)$, the average expected sample size must be at least that of the sequential procedure (6), up to (1).

Corollary 3. *If* $t = t_h$, $h > 0$, *are any stopping times for which* (24) *holds and*

$$\gamma(t; \xi) \geq \gamma + o(h^2) \quad as \ h \to 0, \tag{35}$$

for all ξ *of the form* (15) *with support in* I, *then*

$$\int_{-\infty}^{\infty} [E_\theta(t - N) - (\beta - \tfrac{1}{2}\Delta'')/\Delta]\xi\, d\theta \geq o(1) \tag{36}$$

as $h \to 0$ *for all* ξ *of the form* (15) *with support in* I. *Moreover, there is equality in* (35) *and* (36) *when* t *is the sequential procedure* (6) *described above.*

Proof. If $t = t_h$, $h > 0$, satisfy (24) and (35), then

$$\int_{-\infty}^{\infty} E_\theta(t - N)\Delta\xi\, d\theta \geq r_h(t; \xi) + o(1)$$

$$\geq \int_{-\infty}^{\infty} (\beta - \tfrac{1}{2}\Delta'')\xi\, d\theta + o(1)$$

for all ξ of the form (15) with support in I, by Theorem 4. The first assertion now follows by replacing ξ with $\xi_0 \propto \xi/\Delta$. The second follows easily from Theorem 2 and Lemma 6.

Proof of Theorem 4. Let $t = t_h$, $h > 0$, and ξ be as in the statement of the theorem.

In the proof, h approaches zero along a subsequence for which the lim inf in (32) is attained; and the lim inf is assumed to be less than ∞, without loss of generality.

From (26) and Lemma 6, the integrated risk may be written

$$r_h(t; \xi) = h^{-2}I_1 + I_2 + o(1),$$

where

$$I_1 = \int_B [K(C_t^2) - K(c^2)]\, dP,$$

$$K(x) = 2Ac^2[1 - \Phi(\sqrt{x})] + x, \quad x > 0,$$

$$I_2 = Ac^2 \int_B (1/h^2 t)C_t \phi(C_t)[\hat{b}_t^2 + 2\hat{b}_t \hat{e}_{1t} + \hat{e}_{2t}]\, dP + \int_B [\hat{A}_t \hat{e}_{1t} + \tfrac{1}{2}\hat{A}_t''] \, dP,$$

$C_t = h\sqrt{(t\hat{A}_t)}$, and B is as in Lemma 3, with F equal to the sample space.

The first two derivatives of K are

$$K'(x) = 1 - Ac^2 \phi(\sqrt{x})/\sqrt{x}, \qquad K''(x) = Ac^2(1 + x)\phi(x)/2\sqrt{x^3}$$

for $x > 0$. Thus, K is a convex function, since $K'' > 0$; and K attains its minimum at $x = c^2$, by definition of A. So, I_1 is non-negative.

Next, it is shown that (23) holds. Since $\hat{A}_t \to A(\theta)$ in probability as $h \to 0$. Given $\varepsilon > 0$, let δ denote the minimum of $K(x^2) - K(c^2)$ when $|x - c| > \varepsilon$. Then $\delta > 0$, and

$$r_h(t; \xi) \geq \delta h^{-2} P\{|C_t - c| > \varepsilon\} + I_2 + o(1)$$

for all $h > 0$. Since $r_h(t; \xi)$ is bounded above and I_2 is bounded as $h \to 0$ (along the subsequence), it follows easily that $C_t \to c$ in probability.

Finally, as in the proof of Lemma 3 and 5,

$$I_2 \to \int_{-\infty}^{\infty} (\beta - \tfrac{1}{2}A'')\xi \, d\theta$$

as $h \to 0$ (along the subsequence); and the first assertion of the theorem now follows easily from the non-negativity of I_1.

The second assertion of the theorem follows directly from Theorem 2 and Lemma 6.

Remark 7. The condition (24) imposed in Theorem 4 is inelegant and may limit the theorem's interest to some extent. The necessity of this condition is still unresolved.

References

Anscombe, F. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* 15, 1–21.

Chow, Y.S. and H. Robbins (1965). On the asymptotic theory of fixed width confidence intervals. *Ann. Math. Statist.* 36, 457–462.

Finster, M. (1985). Estimation in the general linear model when accuracy is specified before data is collected. *Ann. Statist.* 13, 663–675.

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* 9, 1229-1238.

Lai, T.L. and D. Siegmund (1977). A non-linear renewal theory with applications to sequential analysis I. *Ann. Statist.* 5, 946-954.

Lai, T.L. and D. Siegmund (1979). A non-linear renewal theory with applications to sequential analysis II. *Ann. Statist.* 7, 60-76.

Siegmund, D. (1978). Estimation following sequential tests. *Biometrika* 65, 341-349.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* 16, 243-258.

Stein, C. (1981). On the coverage probability of a confidence set based on a prior distribution. Technical Report, Statistics Department, Stanford University.

Woodroofe, M. (1982). *Non-linear Renewal Theory in Sequential Analysis.* S.I.A.M., Philadelphia, PA.

Woodroofe, M. (1985). Asymptotic local minimaxity in sequential estimation. *Ann. Statist.* 13, 676-688.

Woodroofe, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.,* to appear.

Woodroofe, M. and R. Keener (1986). Asymptotic expansions for the distribution of a randomly stopped sum. *Ann. Probab.,* to appear.