The Annals of Statistics
1983, Vol. 11, No. 3, 827–836

SECOND ORDER APPROXIMATION TO THE RISK OF A SEQUENTIAL PROCEDURE

BY ADAM T. MARTINSEK

University of Illinois

Given $X_1, X_2, \cdots$, i.i.d. with mean μ and variance σ^2, suppose that at stage n one wishes to estimate μ by the sample mean $\bar{X}_n$, subject to the loss function $L_n = A(\bar{X}_n - \mu)^2 + n, A > 0$. If σ is known, the optimal fixed sample size $n_0 \approx A^{1/2}\sigma$ can be used, with corresponding risk R_{n_0}, but if σ is unknown there is no fixed sample size procedure that will achieve the risk R_{n_0}. For the sequential estimation procedure with stopping rule $T = \inf\{n \geq n_A: n^{-1} \sum_1^n (X_i - \bar{X}_n)^2 \leq A^{-1}n^2\}$, the second order approximation of Woodroofe (1977) to the risk R_T for normal X_i is extended to the distribution-free case. Specifically, if the X_i have finite moments of order greater than eight and are non-lattice, under certain conditions on the delay n_A it is shown that the regret $R_T - R_{n_0} = c + o(1)$ as $A \to \infty$, where c depends on the first four moments of the distribution of the X_i. For the lattice case, bounds of the form $c_1 + o(1) \leq R_T - R_{n_0} \leq c_2 + o(1)$ are obtained, where the c_j are $c \pm 3$. It follows from these approximations that the regret can take arbitrarily large negative values as the distribution of the X_i varies, in contrast to previous results for normal and gamma cases.

1. Introduction. Let $X_1, X_2, \cdots$ be independent observations from a population with mean μ and variance $\sigma^2 \in (0, \infty)$. Given a sample of size n, one wishes to estimate μ by the sample mean $\bar{X}_n$, subject to the loss function

$$L_n = A(\bar{X}_n - \mu)^2 + n, \quad A > 0. \tag{1.1}$$

For a fixed sample size n, the risk is

$$R_n = AE(\bar{X}_n - \mu)^2 + n = A\sigma^2 n^{-1} + n, \tag{1.2}$$

which is minimized (when σ is known) by using the optimal fixed sample size

$$n_0 \approx A^{1/2}\sigma \tag{1.3}$$

(that is, the minimizing n_0 is one of the two integers closest to $A^{1/2}\sigma$). The corresponding minimum fixed sample size risk is

$$R_{n_0} \approx 2A^{1/2}\sigma. \tag{1.4}$$

When σ is unknown, the optimal fixed sample size n_0 cannot be used, and there is no fixed sample size rule that will achieve the risk R_{n_0}. For this case the stopping rule

$$\begin{aligned} T = T_A &= \inf\{n \geq n_A : n \geq A^{1/2}[n^{-1} \textstyle\sum_1^n (X_i - \bar{X}_n)^2]^{1/2}\} \\ &= \inf\{n \geq n_A : n^{-1} \textstyle\sum_1^n (X_i - \bar{X}_n)^2 \leq A^{-1}n^2\}, \end{aligned} \tag{1.5}$$

where n_A is a positive integer which may depend on A, can be used, and the population mean μ is then estimated by $\bar{X}_T$. This type of sequential estimation procedure was first considered by Robbins (1959) in the normal case.

When the distribution of the X_i is normal, asymptotic risk efficiency (i.e., $R_T/R_{n_0} \to 1$ as $A \to \infty$) has been established by Starr (1966), and the much stronger result $R_T - R_{n_0} = O(1)$ as $A \to \infty$ (bounded regret of the sequential estimation procedure with stopping rule T) has been proved by Starr and Woodroofe (1969). Furthermore, Woodroofe (1977)

Received October 1982; revised December 1982.
AMS 1980 *subject classifications.* Primary 62L12; secondary 60G40.
Key words and phrases. Sequential estimation, regret, nonlinear renewal theory, uniform integrability.

has given second order approximations to the expected sample size and risk of this procedure, showing in particular that

$$R_T - R_{n_0} = \tfrac{1}{2} + o(1) \tag{1.6}$$

as $A \to \infty$. In all three papers the delay n_A does not depend on A.

For the gamma and Poisson cases, bounded regret has been obtained by Starr and Woodroofe (1972) and Vardi (1979), respectively, using stopping rules which differ from (1.5) and which only make sense in these special cases.

For the general (distribution-free) case Ghosh and Mukhopadhyay (1979) and Chow and Yu (1981) have proved the asymptotic risk efficiency of the sequential procedure with stopping rule T, under the assumptions that (Ghosh-Mukhopadhyay)

$$E\,|X_1|^8 < \infty \quad \text{and} \quad \delta A^{1/(2\gamma+2)} \le n_A = o(A^{1/2}) \quad \text{as} \quad A \to \infty,$$
$$\text{for some} \quad \delta > 0 \quad \text{and} \quad 0 < \gamma < \tfrac{1}{4}$$

and (Chow-Yu)

$$E\,|X_1|^{2p} < \infty, p > 1, \quad \text{and} \quad K\log(A) \le n_A = o(A^{1/2}) \quad \text{as} \quad A \to \infty,$$
$$\text{for some} \quad K > K_p.$$

Similar results of asymptotic risk efficiency have been obtained by Sen and Ghosh (1981) for sequential estimation of symmetric parametric functions using U-statistics.

In Chow and Martinsek (1982), it is proved that the sequential procedure T is of bounded regret, under the assumptions

$$E\,|X_1|^{6p} < \infty \quad \text{for some} \quad p > 1$$

and

$$\delta A^{1/4} \le n_A = o(A^{1/2}) \quad \text{as} \quad A \to \infty,$$

for some $\delta > 0$ (but without any other assumption about the type of distribution of the X_i). However, the only second order approximation to the risk R_T known so far is (1.6), given by Woodroofe (1977) for normal X_i. Since such second order approximations give a much better idea of the size of the regret than the mere fact of boundedness in A, it is desirable to have these approximations in a more general setting than just the normal case. The purpose of this paper is to give such approximations, assuming only that the X_i satisfy rather mild distributional requirements. The main results are given in the following two theorems, which separate the non-lattice and lattice cases.

Theorem 1. *If $E\,|X_1|^{8p} < \infty$ for some $p > 1$, the distribution of the X_i is non-lattice, and $\delta A^{1/4} \le n_A = o(A^{1/2})$ as $A \to \infty$, for some $\delta > 0$, then*

$$\begin{aligned} R_T - R_{n_0} &= 2 - (\tfrac{3}{4})E\{(Z_1^2 - 1)^2\} + 2E^2(Z_1^3) + o(1) \\ &= 2 - (\tfrac{3}{4})E(Z_1^4) + (\tfrac{3}{4}) + 2E^2(Z_1^3) + o(1) \\ &= 2 - (\tfrac{3}{4})\mathrm{Var}(Z_1^2) + 2E^2(Z_1^3) + o(1) \end{aligned} \tag{1.7}$$

as $A \to \infty$, where $Z_1 = (X_1 - \mu)/\sigma$.

Theorem 2. *Under the hypotheses $E\,|X_1|^{8p} < \infty$ for some $p > 1$ and $\delta A^{1/4} \le n_A = o(A^{1/2})$ as $A \to \infty$ (but without assuming the X_i are non-lattice),*

$$\begin{aligned} -1 - (\tfrac{3}{4})E\{(Z_1^2 - 1)^2\} + 2E^2(Z_1^3) + o(1) &\le R_T - R_{n_0} \\ &\le 5 - (\tfrac{3}{4})E\{(Z_1^2 - 1)^2\} + 2E^2(Z_1^3) + o(1) \end{aligned} \tag{1.8}$$

as $A \to \infty$, where $Z_1 = (X_1 - \mu)/\sigma$.

The proofs of Theorems 1 and 2 are given in Section Two. The approximations (1.7) and (1.8) have interesting consequences, some of which are discussed in Section Three. In particular, it is shown that the regret can take arbitrarily large negative values as the distribution of the X_i varies, providing an answer to a question suggested by Starr and Woodroofe (1972) and pursued further by Woodroofe (1977). The expression (1.7) is also compared with the second order approximation obtained by Woodroofe (1977) for the gamma case (using the stopping rule of Starr and Woodroofe, 1972). It is shown that the non-vanishing term in (1.7) is always larger than the non-vanishing term in Woodroofe's case (not surprisingly, since the procedure considered by Starr and Woodroofe takes advantage of the knowledge that the distribution is gamma, whereas the procedure T does not). However, the difference is often small: for example, when the X_i are exponential the two non-vanishing terms differ by only 1.

REMARK 1. Although the assumption that the X_i have finite moments of order higher than eighth may seem unusual, it should be noted that the resulting second order approximations depend on the first four moments of the distribution. It is therefore fairly clear that one needs at least finite fourth moment to obtain such approximations (even though the problem, and the sequential estimation procedure with stopping rule T, make sense when only the variance is assumed finite).

REMARK 2. In the context of sequential estimation of a non-zero mean with loss function

$$A(\bar{X}_n - \mu)^2/\mu^2 + n \tag{1.9}$$

at stage n, Martinsek (1981, page 75) has given a distribution-free second order approximation to the risk of a sequential procedure when the variance is known. The stopping rule in that case is

$$\inf\{n \geq 1 : |S_n| > A^{1/2}\sigma\},$$

where $S_n = X_1 + \cdots + X_n$, the optimal fixed sample size is $A^{1/2}\sigma/|\mu|$, and the regret (assuming $E|X_1|^6 < \infty$) is

$$2E\{(X_1 - \mu)^3\}/(\sigma^2\mu) + 3\sigma^2/\mu^2 + o(1)$$

as $A \to \infty$. The sequential procedure studied in the present paper is much more complicated, as it is defined in terms of sample variances rather than sample sums, and the methods used in Section Two below are consequently more involved than (and quite different from) those used in the case of sequential estimation with the loss function (1.9) and known variance.

2. Proofs of Theorems 1 and 2.

Assume throughout this section that the n_A are as in Theorems 1 and 2. Without loss of generality, take $\mu = 0$, $\sigma = 1$, and define $V_n = \sum_1^n (X_i - \bar{X}_n)^2$. The relation $V_n \leq V_{n+1}$ for all n will be used occasionally in the two proofs. For simplicity we will always assume $T > n_A$, so that in particular $V_T > 0$. This assumption causes no harm, because as shown in Chow and Martinsek (1982), the probability of the event $\{T = n_A\}$ vanishes so quickly as $A \to \infty$ that the expectations of all relevant random variables on this event also vanish.

The following uniform integrability results are needed in the proofs below. They are consequences of Lemmas 2, 4 and 5 of Chow and Yu (1981).

$$E|X_1|^{2t} < \infty,\ t \geq 1 \Rightarrow \{(A^{-1/2}T)^t : A \geq 1\} \text{ is uniformly integrable;} \tag{2.1}$$

$$E(X_1^2) < \infty \Rightarrow \{(A^{-1/2}T)^{-q} : A \geq 1\} \text{ is uniformly integrable for all } q > 0; \tag{2.2}$$

$$E|X_1|^{2t} < \infty,\ t \geq 1 \Rightarrow \{|A^{-1/4}S_T|^{2t} : A \geq 1\} \text{ is uniformly integrable;} \tag{2.3}$$

$$E|X_1|^{2t} < \infty,\ t \geq 2 \Rightarrow \{|A^{-1/4}(\textstyle\sum_1^T X_i^2 - T)|^t : A \geq 1\} \text{ is uniformly integrable.} \tag{2.4}$$

We will also need the nonlinear renewal theory results of Lai and Siegmund (1977, 1979), as applied by Chang and Hsiung (1979). Specifically, if the X_i are non-lattice with finite fourth moment, and U_A is the overshoot in the sense of Lai and Siegmund (1977, 1979),

$$\begin{aligned} U_A &= T(TV_T^{-1})^{1/2} - A^{1/2} \\ &= T - A^{1/2} - (1/2)(\textstyle\sum_1^T X_i^2 - T) + (1/2)T^{-1}S_T^2 + (3/8)\lambda_T^{-5/2}T^{-1}(V_T - T)^2, \end{aligned} \tag{2.5}$$

λ_T a random variable lying between 1 and $T^{-1}V_T$, then Chang and Hsiung (1979) have shown that

$$U_A \to_{\mathscr{L}} U \text{ as } A \to \infty, \tag{2.6}$$

where the random variable U has distribution defined in terms of the first time that the random walk

$$n - (1/2)(\textstyle\sum_1^n X_i^2 - n) = (3/2)n - \sum_1^n X_i^2$$

is positive.

REMARK. The paper of Chang and Hsiung (1979) assumes an additional continuity condition on the distribution function of the X_i. This condition is used only to obtain uniform integrability of negative powers of $A^{-1/2}T$, and in view of (2.2) the condition is not needed here. Also, although the results of Lai and Siegmund (1977, 1979) and Chang and Hsiung (1979) are stated for stopping rules without a variable delay such as n_A, it is easily checked that they apply in this case as well. The same comment holds for the Lemma given below.

To simplify notation, let

$$\xi_n = (1/2)n^{-1}S_n^2 + (3/8)\lambda_n^{-5/2}n^{-1}(V_n - n)^2. \tag{2.7}$$

Because

$$\begin{aligned} T(TV_T^{-1})^{1/2} - A^{1/2} &\le T(TV_T^{-1})^{1/2} - (T-1)\{(T-1)V_{T-1}^{-1}\}^{1/2} \le V_T^{-1/2}\{(3/2)(T-1)^{1/2} + 3/8\} \\ &\le V_{T-1}^{-1/2}\{(3/2)(T-1)^{1/2} + 3/8\} \le A^{1/2}(T-1)^{-3/2}\{(3/2)(T-1)^{1/2} + 3/8\} \\ &= (3/2)A^{1/2}(T-1)^{-1} + (3/8)A^{1/2}(T-1)^{-3/2}, \end{aligned}$$

from (2.2) we have

$$E(X_1^2) < \infty \Rightarrow \{U_A^q : A \ge 1\} \text{ is uniformly integrable for all } q > 0. \tag{2.8}$$

Furthermore, since

$$(T^{-1}V_T)^{-5/2} \le 2^{5/2}\{(T-1)^{-1}V_{T-1}\}^{-5/2} \le 2^{5/2}\{A^{-1}(T-1)^2\}^{-5/2},$$

from (2.2), (2.3), (2.4), (2.7) and Hölder's inequality,

$$E|X_1|^{4t+\varepsilon} < \infty \text{ for } t \ge 1 \text{ and } \varepsilon > 0 \Rightarrow \{\xi_T^t : A \ge 1\} \text{ is uniformly integrable.} \tag{2.9}$$

The results (2.8) and (2.9), together with (2.5), will be used below to re-write parts of the regret in terms of expected products of powers of S_T and $\sum_1^T X_i^2 - T$. These expected products will then be evaluated as differences of moments of S_T, $\sum_1^T X_i^2 - T$, and $\sum_1^T X_i^2 - T \pm S_T$, using Anscombe's Theorem, as in (2.22) and (2.25).

A slight modification of one of the nonlinear renewal theory results of Lai and Siegmund (1977) is also needed. This modification is given in the following lemma, which reflects the fact that under the general conditions of Lai and Siegmund's work the limit distribution of the overshoot does not depend on the slowly varying term ξ_n.

LEMMA. *If the X_i are non-lattice with finite fourth moment, then $T^{-1}S_T^2$ and U_A are* *asymptotically independent as $A \to \infty$. That is, for all $x, y > 0$,*

$$P\{T^{-1}S_T^2 \le x,\ U_A \le y\} \to F(x)G(y) \text{ as } A \to \infty,$$

where G is the distribution function of U, and F is the distribution function of a Chi squared random variable with one degree of freedom.

PROOF. To simplify notation, let

$$W_n = n - (\tfrac{1}{2})(\textstyle\sum_1^n X_i^2 - n).$$

As proved by Chang and Hsiung (1979), for each $\eta > 0$, there exists $\rho > 0$ and an integer n' such that for all $n \ge n'$,

$$P\{\max_{n \le j \le n+\rho n} |\xi_j - \xi_n| \ge \eta\} < \eta, \tag{2.10}$$

and also

$$P\{\max_{n \le j \le n+\rho n} |j^{-1}S_j^2 - n^{-1}S_n^2| \ge \eta\} < \eta. \tag{2.11}$$

Furthermore, since

$$\{|A^{-1/4}(T - A^{1/2})| : A \ge 1\}$$

is uniformly integrable by Theorem IV-3 of Yu (1978),

$$A^{-1/2}(T - A^{1/2}) \to_P 0 \text{ as } A \to \infty. \tag{2.12}$$

Proceeding as in the proof of Theorem 1 of Lai and Siegmund (1977), let

$$n_1 = [A^{1/2} - \rho A^{1/2}/4], \quad n_2 = [A^{1/2} + \rho A^{1/2}/4]$$

(so that $n_1 + \rho n_1 > n_2$ and $n_A < n_1$ for large A, since ρ in (2.10) and (2.11) can clearly be chosen less than 2). Also let

$$B_A = \{\max_{n_A \le n \le n_1}(W_n + \xi_n) < A^{1/2} - A^{1/4}\}.$$

Then from (2.12), (2.10), and the Strong Law of Large Numbers, if first $A \to \infty$ and then $\eta \to 0$,

$$P(B_A^c) + P(T \ge n_2) \to 0, \tag{2.13}$$

where B_A^c denotes the complement of B_A. Putting $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$, and defining

$$t = t_A(\beta) = \inf\{n \ge n_A : W_n + \xi_{n_1} > A^{1/2} + \beta\}$$

for $-\infty < \beta < \infty$, by the renewal theorem (Feller, 1966, page 354), for A sufficiently large and all $y > 0$,

$$|P\{W_{t(\beta)} + \xi_{n_1} - (A^{1/2} + \beta) \le y \mid \mathscr{F}_{n_1}\} - G(y)| < \eta \tag{2.14}$$

on B_A. As shown by Lai and Siegmund (1977, page 948), if $y > 2\eta$, on the event

$$B_A \cap \{T < n_2, \max_{n_1 \le k \le n_2} |\xi_k - \xi_{n_1}| < \eta\},$$

$$\{U_A > y\} \subseteq \{t(\eta) = T,\ W_{t(\eta)} + \xi_{n_1} - (A^{1/2} + \eta) > y - 2\eta\} \tag{2.15}$$

and

$$\{W_{t(-\eta)} + \xi_{n_1} - (A^{1/2} - \eta) > y + 2\eta\} \subseteq \{t(-\eta) = T,\ U_A > y\}. \tag{2.16}$$

Hence, on the event

$$B_A^* = B_A \cap \{T < n_2, \max_{n_1 \le k \le n_2} |\xi_k - \xi_{n_1}| < \eta\} \cap \{\max_{n_1 \le j \le n_2} |j^{-1}S_j^2 - n_1^{-1}S_{n_1}^2| < \eta\},$$

if $x > \eta$ and $y > 2\eta$, from (2.16)

$$P\{U_A > y,\, T^{-1}S_T^2 \le x\}$$
$$\ge P[B_A^* \cap \{W_{t(-\eta)} + \xi_{n_1} - (A^{1/2} - \eta) > y + 2\eta\} \cap \{n_1^{-1}S_{n_1}^2 \le x - \eta\}]$$
$$\ge \int_{B_A \cap \{n_1^{-1}S_{n_1}^2 \le x-\eta\}} P\{W_{t(-\eta)} + \xi_{n_1} - (A^{1/2} - \eta) > y + 2\eta \mid \mathscr{F}_{n_1}\}\, dP$$
$$- P(T \ge n_2) - P\{\max_{n_1 \le k \le n_2} |\xi_k - \xi_{n_1}| \ge \eta\} - P\{\max_{n_1 \le j \le n_2} |j^{-1}S_j^2 - n_1^{-1}S_{n_1}^2| \ge \eta\}.$$

Letting $A \to \infty$ and then $\eta \to 0$, from (2.13), (2.14), (2.10) and (2.11), since $n_1^{-1}S_{n_1}^2 \to_{\mathscr{L}} \chi_1^2$,

$$\liminf_{A\to\infty} P\{T^{-1}S_T^2 \le x,\, U_A > y\} \ge F(x)\{1 - G(y)\}.$$

A similar argument, using (2.15), shows

$$\limsup_{A\to\infty} P\{T^{-1}S_T^2 \le x,\, U_A > y\} \le F(x)\{1 - G(y)\},$$

and hence as $A \to \infty$,

$$P\{T^{-1}S_T^2 \le x,\, U_A > y\} \to F(x)\{1 - G(y)\},$$

proving the lemma.

PROOF OF THEOREM 1. From the main theorem of Chang and Hsiung (1979), since the X_i have finite moments of order greater than six, as $A \to \infty$,

$$E(T - A^{1/2}) = \nu - \tfrac{1}{2} - (\tfrac{3}{8})E\{(X_1^2 - 1)^2\} + o(1),$$

where $\nu = E(U)$. By Theorem 2 of Chow, Robbins and Teicher (1965),

$$\begin{aligned} R_T - R_{n_0} &= E(AT^{-2}S_T^2) + ET - 2A^{1/2} \\ &= E\{S_T^2(AT^{-2} - 1)\} + 2E(T - A^{1/2}) \\ &= E\{S_T^2(AT^{-2} - 1)\} + 2\nu - 1 - (\tfrac{3}{4})E\{(X_1^2 - 1)^2\} + o(1). \end{aligned} \tag{2.17}$$

Concentrating on the first term on the right-hand side of (2.17) we have

$$\begin{aligned} E\{S_T^2(AT^{-2} - 1)\} &= E\{S_T^2(AT^{-2} - TV_T^{-1})\} + E\{S_T^2(TV_T^{-1} - 1)\} \\ &= I + II, \quad \text{say}. \end{aligned} \tag{2.18}$$

Since

$$S_T^2(AT^{-2} - TV_T^{-1}) = -T^{-1}S_T^2(U_A)\{A^{1/2}T^{-1} + (TV_T^{-1})^{1/2}\},$$

and $A^{1/2}T^{-1} + (TV_T^{-1})^{1/2} \to 2$ a.s. as $A \to \infty$, by the Lemma together with the uniform integrability results (2.2), (2.3) and (2.8), because $TV_T^{-1} \le TV_{T-1}^{-1} \le T\{A^{-1}(T-1)^3\}^{-1} \le 8AT^{-2}$,

$$I = -2E(\chi_1^2)\nu + o(1) = -2\nu + o(1). \tag{2.19}$$

By Anscombe's Theorem (Anscombe, 1952), (2.2), (2.3) and Wald's Lemma,

$$\begin{aligned} II = E\{S_T^2(TV_T^{-1} - 1)\} &= E\{(S_T^2 - V_T)V_T^{-1}(T - V_T)\} + E(T^{-1}S_T^2) \\ &= A^{-1/2}E\{(S_T^2 - V_T)(T - V_T)\} \\ &\quad + E\{(S_T^2 - V_T)(T - V_T)(A^{1/2} - V_T)A^{-1/2}V_T^{-1}\} + 1 + o(1) \\ &= IIa + IIb + 1 + o(1). \end{aligned} \tag{2.20}$$

From (2.17) of Chow and Martinsek (1982), together with (2.1), (2.2), (2.3), (2.4) and Anscombe's Theorem, since $A^{-1/2}\sum_1^T X_i^2 \to 1$ and $A^{-1/2}T \to 1$ as $A \to \infty$,

$$
\begin{aligned}
IIa &= A^{-1/2}E\{(S_T^2 - V_T)(T - V_T)\} \\
&= A^{-1/2}E\{(S_T^2 - \textstyle\sum_1^T X_i^2)(T - \sum_1^T X_i^2)\} + A^{-1/2}E(T^{-1}S_T^4) \\
&\quad - A^{-1/2}E\{(\textstyle\sum_1^T X_i^2)T^{-1}S_T^2\} + o(1) \\
&= A^{-1/2}E\{(S_T^2 - \textstyle\sum_1^T X_i^2)(T - \sum_1^T X_i^2)\} + 3 - 1 + o(1) \\
&= A^{-1/2}E\{(S_T^2 - T)(T - \textstyle\sum_1^T X_i^2)\} \\
&\quad + A^{-1/2}E\{(T - \textstyle\sum_1^T X_i^2)^2\} + 2 + o(1) \qquad (2.21)\\
&= -2A^{-1/2}E(X_1^3)E\{(T - A^{1/2})S_T\} \\
&\quad - A^{-1/2}E\{(X_1^2 - 1)^2\}E(T) + E\{(X_1^2 - 1)^2\} + 2 + o(1) \\
&= -2A^{-1/2}E(X_1^3)E\{(T - A^{1/2})S_T\} - E\{(X_1^2 - 1)^2\} \\
&\quad + E\{(X_1^2 - 1)^2\} + 2 + o(1) \\
&= -2A^{-1/2}E(X_1^3)E\{(T - A^{1/2})S_T\} + 2 + o(1).
\end{aligned}
$$

By (2.3), (2.8), (2.9) and Hölder's inequality, using $E|X_1|^{5+\varepsilon} < \infty$ for some $\varepsilon > 0$,

$$E\{|(U_A - \xi_T)S_T|\} \le E^{4/5}(|U_A - \xi_T|^{5/4})E^{1/5}(|S_T|^5) = O(A^{1/4}) \quad \text{as} \quad A \to \infty,$$

so from (2.3), (2.4) and Anscombe's Theorem, since $E(X_1^4) < \infty$,

$$
\begin{aligned}
&2A^{-1/2}E\{(T - A^{1/2})S_T\} \\
&\quad = A^{-1/2}E\{(\textstyle\sum_1^T X_i^2 - T + 2U_A - 2\xi_T)S_T\} = A^{-1/2}E\{(\sum_1^T X_i^2 - T)S_T\} + o(1) \\
&\quad = (½)A^{-1/2}[E\{(\textstyle\sum_1^T X_i^2 - T + S_T)^2\} - E(S_T^2) - E\{(\sum_1^T X_i^2 - T)^2\}] + o(1) \qquad (2.22)\\
&\quad = ½[E\{(X_1^2 - 1 + X_1)^2\} - 1 - E\{(X_1^2 - 1)^2\}] + o(1) \\
&\quad = E(X_1^3) + o(1).
\end{aligned}
$$

It follows from (2.21) and (2.22) that as $A \to \infty$,

$$IIa = -E^2(X_1^3) + 2 + o(1). \tag{2.23}$$

From (2.2), (2.3), (2.4), (2.8), (2.9), Hölder's inequality and $V_T^{-1} \le V_{T-1}^{-1} \le A(T-1)^{-3}$, since $E|X_1|^{8p} < \infty$,

$$
\begin{aligned}
&E\{|(S_T^2 - V_T)(T - V_T)V_T^{-1}(U_A - \xi_T + T^{-1}S_T^2)|\} \\
&\quad \le E^{1/4}\{(S_T^2 - V_T)^4\}E^{1/4}[\{A(T-1)^{-3}(T - V_T)\}^4] \times E^{1/2}\{(U_A - \xi_T + T^{-1}S_T^2)^2\} \\
&\quad = o(A^{1/2}) \quad \text{as} \quad A \to \infty,
\end{aligned}
$$

hence by (2.3), (2.4) and Anscombe's Theorem,

$$
\begin{aligned}
IIb &= A^{-1/2}E\{(S_T^2 - V_T)(T - V_T)(A^{1/2} - V_T)V_T^{-1}\} \\
&= A^{-1/2}E[(S_T^2 - V_T)(T - V_T)\{T + (½)(T - \textstyle\sum_1^T X_i^2) - V_T - U_A + \xi_T\}V_T^{-1}] \\
&= \tfrac{3}{2}A^{-1/2}E\{(S_T^2 - V_T)(T - V_T)(T - \textstyle\sum_1^T X_i^2)V_T^{-1}\} + o(1) \qquad (2.24)\\
&= \tfrac{3}{2}A^{-1/2}E\{S_T^2(T - \textstyle\sum_1^T X_i^2)^2V_T^{-1}\} - (\tfrac{3}{2})A^{-1/2}E\{(T - \sum_1^T X_i^2)^2\} + o(1) \\
&= \tfrac{3}{2}A^{-1/2}E\{S_T^2(T - \textstyle\sum_1^T X_i^2)^2V_T^{-1}\} - (\tfrac{3}{2})E\{(X_1^2 - 1)^2\} + o(1).
\end{aligned}
$$

By a Hölder's inequality argument similar to those above, since $E|X_1|^{8p} < \infty$,

$$E\{|S_T^2(T - \textstyle\sum_1^T X_i^2)^2(A^{1/2} - V_T)V_T^{-1}A^{-1/2}|\} = O(A^{1/4}),$$

so from (2.3), (2.4) and Anscombe's Theorem, using $E(X_1^8) < \infty$,

$$
\begin{aligned}
&A^{-1/2}E\{S_T^2(T-\textstyle\sum_1^T X_i^2)^2V_T^{-1}\}\\
&\quad = A^{-1}E\{S_T^2(T-\textstyle\sum_1^T X_i^2)^2\}+o(1)\\
&\quad = \tfrac{1}{12}A^{-1}[E\{(\textstyle\sum_1^T X_i^2-T+S_T)^4\}+E\{(\textstyle\sum_1^T X_i^2-T-S_T)^4\}\\
&\qquad\qquad -2E(S_T^4)-2E\{(\textstyle\sum_1^T X_i^2-T)^4\}]+o(1)\\
&\quad = \tfrac{1}{12}[3E^2\{(X_1^2-1+X_1)^2\}+3E^2\{(X_1^2-1-X_1)^2\}\\
&\qquad\qquad -6-6E^2\{(X_1^2-1)^2\}]+o(1)\\
&\quad = \tfrac{1}{12}[6E^2\{(X_1^2-1)^2\}+6+12E\{(X_1^2-1)^2\}\\
&\qquad\qquad +24E^2(X_1^3)-6-6E^2\{(X_1^2-1)^2\}]+o(1)\\
&\quad = 2E^2(X_1^3)+E\{(X_1^2-1)^2\}+o(1).
\end{aligned}
\tag{2.25}
$$

Hence by (2.24) and (2.25),

$$IIb = 3E^2(X_1^3)+\tfrac{3}{2}E\{(X_1^2-1)^2\}-(\tfrac{3}{2})E\{(X_1^2-1)^2\}+o(1)=3E^2(X_1^3)+o(1) \tag{2.26}$$

as $A\to\infty$.

Finally, combining (2.17), (2.18), (2.19), (2.20), (2.23) and (2.26),

$$
\begin{aligned}
R_T-R_{n_0} &= 2\nu-1-\tfrac{3}{4}E\{(X_1^2-1)^2\}-2\nu+1+2-E^2(X_1^3)+3E^2(X_1^3)+o(1)\\
&= 2-\tfrac{3}{4}E\{(X_1^2-1)^2\}+2E^2(X_1^3)+o(1),
\end{aligned}
$$

proving the Theorem.

PROOF OF THEOREM 2. The proof of Theorem 2 is similar to that of Theorem 1, except that one can no longer use the Lemma and the limiting behavior of $T(TV_T^{-1})^{1/2}-A^{1/2}$ as in the non-lattice case. However, from the argument leading to (2.8),

$$U_A\le(\tfrac{3}{2})A^{1/2}(T-1)^{-1}+(\tfrac{3}{8})A^{1/2}(T-1)^{-3/2},$$

and since $A^{1/2}(T-1)^{-1}\to 1$ a.s. as $A\to\infty$, we have from the various uniform integrability results and Anscombe's Theorem that

$$0\le E(U_A)\le \tfrac{3}{2}+o(1)$$

and

$$0\le E[T^{-1}S_T^2\{(TV_T^{-1})^{1/2}+A^{1/2}T^{-1}\}U_A]\le 3+o(1).$$

Using these bounds in the identity

$$E(T-A^{1/2})=E(U_A)-\tfrac{1}{2}-\tfrac{3}{8}E\{(X_1^2-1)^2\}+o(1)$$

(as in Chang and Hsiung, 1979), and in (2.18) above yields the result of Theorem 2.

REMARK. Results analogous to Theorems 1 and 2 can also be obtained for the more general loss function

$$L_n'=A\sigma^{2\beta-2}(\bar X_n-\mu)^2+n,\quad \beta>0,$$

considered by Chow and Yu (1981) and Chow and Martinsek (1982). In this general case, under the assumptions of Theorem 1 the regret can be written as

$$2\beta+\{(3\beta^2-6\beta)/4\}E\{(Z_1^2-1)^2\}+2\beta E^2(Z_1^3)+o(1),$$

which agrees with (1.7) when $\beta=1$. One also has the analogue of Theorem 2, with bounds

$$2\beta+\{(3\beta^2-6\beta)/4\}E\{(Z_1^2-1)^2\}+2\beta E^2(Z_1^3)\pm(2+\beta)+o(1).$$

The proofs are essentially the same as for the case $\beta = 1$, using Taylor series expansions as in Chow and Martinsek (1982).

3. Consequences of the second order approximation. It follows immediately from Theorem 1 that for symmetric distributions of the X_i (assuming the conditions of the theorem hold), the regret due to using the sequential procedure T in ignorance of σ is bounded above by $2 + o(1)$ as $A \to \infty$. That is, in the limit, one loses at most the cost of two observations when using the stopping rule T instead of n_0 (and the limit is taken as the cost of observations becomes insignificant compared to the cost of error). For the symmetric lattice case, under the moment assumption of Theorem 2 one has the same sort of result, with the upper bound $5 + o(1)$ instead of $2 + o(1)$.

By way of contrast, it also follows from Theorems 1 and 2 that the regret can take arbitrarily large negative values as the distribution of the X_i varies, even among symmetric distributions. To see this, let $X_1, X_2, \cdots$ be i.i.d. with probability density function

$$f(x) = 2\,|\,x\,|^{-5} I_{\{|x| \geq 1\}},$$

where I denotes indicator function, and for $M > 1$ define

$$X_{iM} = X_i I_{\{|X_i| \leq M\}}.$$

Then for each M, $X_{1M}, X_{2M}, \cdots$ are i.i.d. and their common distribution is symmetric about 0. Applying Theorem 1 (for fixed M, the X_{iM} are bounded and non-lattice), the non-vanishing term on the right-hand side of (1.7) is

$$\begin{aligned} 2 - \tfrac{3}{4}E(X_{1M}^4)/(\sigma_M^4) + \tfrac{3}{4} &= 2 - \tfrac{3}{4}\{4\log(M)\}/\{2(1 - M^{-2})\}^2 + \tfrac{3}{4} \\ &= 2 - \tfrac{3}{4}\log(M)/(1 - M^{-2})^2 + \tfrac{3}{4}, \end{aligned} \tag{3.1}$$

where σ_M^2 is the variance of the X_{iM}. Clearly, as $M \to \infty$ the expression (3.1) approaches $-\infty$, and hence the "regrets" $R_T - R_{n_0}$ corresponding to the sequences $X_{1M}, X_{2M}, \cdots$ take (for sufficiently large A) arbitrarily large negative values. This example provides an answer to the question raised by Starr and Woodroofe (1972) and discussed further by Woodroofe (1977), as to whether the regret can ever take negative values. Although the regret in the normal and gamma cases considered by Woodroofe is positive for sufficiently large A (since the constant terms in Woodroofe's second order approximations are positive), in general the regret need not be positive, and in fact for distributions with large fourth moments (as in the example above) arbitrarily large negative values can be achieved.

It is also interesting to compare the second order approximation obtained by Woodroofe (1977) for a sequential estimation procedure designed for the gamma case, first considered by Starr and Woodroofe (1972), with the result of Theorem 1 for the sequential procedure with stopping rule T. Specifically, if $X_1, X_2, \cdots$ are i.i.d. with probability density function

$$f_\theta(x) = \Gamma(a)^{-1}(a/\theta)^a x^{a-1}\exp(-ax/\theta) I_{\{x \geq 0\}},$$

where $\theta > 0$ is unknown and $a > 0$ is known, and

$$N = \inf\{n > 2a^{-1} : n > \bar{X}_n(Aa^{-1})^{1/2}\},$$

then Woodroofe (1977) has shown that

$$R_N - R_{n_0} = 3a^{-1} + o(1) \tag{3.2}$$

as $A \to \infty$, where n_0 is as above. On the other hand, from Theorem 1, since $E(X_1) = \theta$ and $\mathrm{Var}(X_1) = \theta^2/a$,

$$\begin{aligned} R_T - R_{n_0} &= 2 - \tfrac{3}{4}E\{(X_1 - \theta)^4\}(\theta^4 a^{-2})^{-1} + \tfrac{3}{4} + 2E^2\{(X_1 - \theta)^3\}(\theta^6 a^{-3})^{-1} + o(1) \\ &= 2 - \tfrac{3}{4}(3a^{-2} + 6a^{-3})\theta^4(\theta^4 a^{-2})^{-1} + \tfrac{3}{4} + 2(2\theta^3 a^{-2})^2(\theta^6 a^{-3})^{-1} + o(1) \\ &= 2 - \tfrac{9}{4} - \tfrac{18}{4}a^{-1} + \tfrac{3}{4} + 8a^{-1} + o(1) = \tfrac{7}{2}a^{-1} + \tfrac{1}{2} + o(1). \end{aligned} \tag{3.3}$$

It follows from (3.2) and (3.3) that as $A \to \infty$, the difference in risks

$$R_T - R_N = (R_T - R_{n_0}) - (R_N - R_{n_0}) = \tfrac{1}{2}a^{-1} + \tfrac{1}{2} + o(1), \tag{3.4}$$

the non-vanishing term of which is positive and strictly decreasing as $a \to \infty$; this non-vanishing term approaches $+\infty$ as $a \to 0$ and approaches ½ as $a \to \infty$. In particular, for $a = 1$ (the exponential case) the difference $R_T - R_N$ is (asymptotically) 1.

REFERENCES

ANSCOMBE, F. (1952). Large sample theory of sequential estimation. *Proc. Cambr. Philos. Soc.* **48** 600–607.

CHANG, I. S. and HSIUNG, C. A. (1979). Approximations to the expected sample size of certain sequential procedures. *Proceedings of the Conference on Recent Developments in Statistical Methods and Applications*, Taipei, Taiwan, ROC, December 1979, Inst. of Math., Acad. Sinica, 71–82.

CHOW, Y. S. and MARTINSEK, A. T. (1982). Bounded regret of a sequential procedure for estimation of the mean. *Ann. Statist.* **10** 909–914.

CHOW, Y. S., ROBBINS, H. and TEICHER, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

CHOW, Y. S. and YU, K. F. (1981). On the performance of a sequential procedure for the estimation of the mean. *Ann. Statist.* **9** 184–189.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2.** Wiley, New York.

GHOSH, M. and MUKHOPADHYAY, N. (1979). Sequential point estimation of the mean when the distribution is unspecified. *Commun. Statist.-Theor. Meth.* **A8** 637–652.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7** 60–76.

MARTINSEK, A. T. (1981). Moments of two-sided stopping rules and the performance of some sequential estimation procedures. Ph.D. thesis, Columbia University. Unpublished.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics* (Harald Cramer Volume). Almquist and Wiksell, Uppsala, Sweden. 235–245.

SEN, P. K. and GHOSH, M. (1981). Sequential point estimation of estimable parameters based on U-statistics. *Sankhyā, Ser. A* **43** 331–344.

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

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

STARR, N. and WOODROOFE, M. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. U.S.A.* **63** 285–288.

STARR, N. and WOODROOFE, M. (1972). Further remarks on sequential point estimation: the exponential case. *Ann. Math. Statist.* **43** 1147–1154.

VARDI, Y. (1979). Asymptotic optimal sequential estimation: the Poisson case. *Ann. Statist.* **7** 1040–1051.

WOODROOFE, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probability* **4** 67–80.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* Society for Industrial and Applied Mathematics, Philadelphia.

YU, K. F. (1978). Limit theorems for a class of sequential procedures. Ph.D. thesis, Columbia University. Unpublished.

DEPARTMENT OF MATHEMATICS
1409 WEST GREEN STREET
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801-2975