The Annals of Statistics
1993, Vol. 21, No. 1, 503–519

ASYMPTOTIC EXPANSIONS FOR THE MOMENTS OF A RANDOMLY STOPPED AVERAGE

BY GIRISH ARAS AND MICHAEL WOODROOFE[1]

University of California, Santa Barbara and University of Michigan

Let $\mathbf{S}_1, \mathbf{S}_2, \ldots$ denote a driftless random walk with values in an inner product space $\mathscr{W}$; let $Z_1, Z_2, \ldots$ denote a perturbed random walk of the form $Z_n = n + \langle \mathbf{c}, \mathbf{S}_n \rangle + \xi_n$, $n = 1, 2, \ldots$, where $\xi_1, \xi_2, \ldots$ are slowly changing, $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\mathbf{c} \in \mathscr{W}$; and let $t = t_a = \inf\{n \geq 1\colon Z_n > a\}$ for $0 \leq a < \infty$. Conditions are developed under which the first four moments of $\overline{\mathbf{X}}_t := \mathbf{S}_t/t$ have asymptotic expansions, and the expansions are found. Stopping times of this form arise naturally in sequential estimation problems, and the main results may be used to find asymptotic expansions for risk functions in such problems. Examples of such applications are included.

1. Introduction. Let $\mathscr{W}$ denote a finite-dimensional inner product space, with inner product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$; and let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ denote i.i.d., $\mathscr{W}$-valued random vectors with common distribution F. Suppose that F has mean $\mu = 0$, covariance operator Σ and higher moments as needed. Let $\xi_1, \xi_2, \ldots$ be random variables for which ξ_n is independent of $\mathbf{X}_{n+1}, \mathbf{X}_{n+2}, \ldots$ for all $n = 1, 2, \ldots$; let $\mathbf{c} \in \mathscr{W}$; and let

$$Z_n = n + \langle \mathbf{c}, \mathbf{S}_n \rangle + \xi_n, \qquad n \geq 1, \tag{1}$$

and

$$t = t_a = \inf\{n \geq 1\colon Z_n > a\}, \qquad a \geq 1, \tag{2}$$

where $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$ for $n \geq 1$ and the infimum of the empty set is ∞. (That $t_a < \infty$ w.p. 1 for all $a \geq 1$ under mild conditions is shown below.) The main results provide asymptotic expansions as $a \to \infty$ for the first four moments of $\overline{\mathbf{X}}_t := \mathbf{S}_t/t$ and the first two moments of a smooth, suitably bounded function of $\overline{\mathbf{X}}_t$.

Stopping times of the form t_a arise naturally in sequential estimation, and risk functions often involve the second moment of $\overline{\mathbf{X}}_t$ in such problems. See, for example, Woodroofe (1977), Martinsek (1983), Aras (1989) and Sriram (1990), where special cases of the main results of this paper may be found. The purpose of this paper is to develop expansions under weak moment conditions, in a form which may be applicable to other problems. The results of Martinsek and Aras are compared to those of Theorems 2 and 4 in Examples 2 and 3.

Received May 1991; revised April 1992.
[1]Supported by NSF Grant DMS-89-02188.
AMS 1991 *subject classification*. 62L12.
Key words and phrases. Anscombe's theorem, martingales, maximal inequalities, nonlinear renewal theory, sequential estimation, stopping times, risk functions.

2. Conditions and preliminaries. It is convenient to regard the moments of F as multilinear functionals. If k is a positive integer and $\int_{\mathscr{W}} \|\mathbf{x}\|^k F(d\mathbf{x}) < \infty$, then the kth moment of F is defined by

$$\mu_k(\mathbf{b}_1, \dots, \mathbf{b}_k) = \int_{\mathscr{W}} \prod_{i=1}^{k} \langle \mathbf{b}_i, \mathbf{x} \rangle F(d\mathbf{x}) \tag{3}$$

for $\mathbf{b}_1, \dots, \mathbf{b}_k \in \mathscr{W}$. Of course, μ_1 and μ_2 may be identified with an element of $\mathscr{W}$ and a linear operator, respectively; and μ_1 is denoted by μ. It is convenient to use the notation (3) whenever the integral is finite and to write

$$v_\alpha(\mathbf{b}) = \left\{ \int_{\mathscr{W}} |\langle \mathbf{b}, \mathbf{x} \rangle|^\alpha F(d\mathbf{x}) \right\}^{1/\alpha}$$

for $\mathbf{b} \in \mathscr{W}$ and $0 < \alpha < \infty$, finite or infinite. If μ is any k-linear functional on $\mathscr{W}$, let $\tilde{\mu}(\mathbf{b}) = \mu(\mathbf{b}, \dots, \mathbf{b})$ for $\mathbf{b} \in \mathscr{W}$; and recall that if μ is symmetric then $\tilde{\mu}$ determines μ. In fact, if μ is symmetric, then

$$\mu(\mathbf{b}_1, \dots, \mathbf{b}_k) = \frac{1}{k!} \times \frac{\partial^k}{\partial s_1 \dots \partial s_k} \tilde{\mu}(\mathbf{b}_1 s_1 + \cdots + \mathbf{b}_k s_k) \tag{4}$$

for all $\mathbf{b}_1, \dots, \mathbf{b}_k \in \mathscr{W}$.

The following conditions are needed: for some $3 \le p < \infty$, and $0 < \varepsilon_0, \varepsilon_1 < 1$,

$$\mu = 0, \quad \int_{\mathscr{W}} \|\mathbf{x}\|^2 F(d\mathbf{x}) < \infty \quad \text{and} \quad v_p(\mathbf{c}) < \infty, \tag{C1}$$

$$\left[\left(Z_n - \frac{n}{\varepsilon_0} \right)^+ \right]^p, \ n \ge 1, \text{ are uniformly integrable}, \tag{C2}$$

$$\sum_{n=1}^{\infty} nP\{\xi_n < -(1 - \varepsilon_1)n\} < \infty, \tag{C3}$$

$$\lim_{\delta \searrow 0} \sup_{n \ge 1} P\left\{ \max_{k \le n\delta} |\xi_{n+k} - \xi_n| > \varepsilon \right\} = 0, \qquad \forall\, 0 < \varepsilon < \infty. \tag{C4}$$

In addition, it is assumed that there are events A_n, $n = 1, 2, \dots,$ and a $3/2 \le \alpha < \infty$ for which

$$\sum_{n=1}^{\infty} nP\left(\bigcup_{k=n}^{\infty} A'_k \right) < \infty, \tag{C5}$$

$$\max_{k \le n} \left| \xi_{n+k} I_{A_{n+k}} \right|^\alpha, \ n \ge 1, \text{ are uniformly integrable}.$$

The condition (C4) is called slow change by Lai and Siegmund (1977). It follows easily from (C5), Markov's inequality and the Borel–Cantelli lemmas that

$$P\left\{ \lim_{n \to \infty} \frac{\xi_n}{n^r} = 0 \right\} = 1, \qquad \forall\, 0 < r < \infty. \tag{5}$$

Conditions (C1)–(C5) and (C6), below, are assumed throughout Sections 2, 3, and 5–8. They are repeated in the statements of the main results, but not the supporting ones.

It is convenient to develop some simple preliminaries before stating the main results. Extensive use is made of the inequality

$$E\left\{\sup_{n\geq m}\left|\langle \mathbf{b},\overline{\mathbf{X}}_n\rangle\right|^q\right\} \leq \left(\frac{q}{q-1}\right)^q E\left\{\left|\langle \mathbf{b},\overline{\mathbf{X}}_m\rangle\right|^q\right\} \tag{6}$$

for all $m = 1,2,\dots,1 < q < \infty$, and $\mathbf{b} \in \mathscr{W}$, which follows easily from the Doob's [(1953), pages 317–318] maximal inequality applied to the reverse martingale $\langle \mathbf{b},\overline{\mathbf{X}}_n\rangle$, $n \geq 1$. Extensive use is also made of the following result which is a simple consequence of Lemma 5 of Chow and Lai (1978) and Theorem 3 of Chow, Hsiung and Lai (1979).

PROPOSITION 1. *Let $Y_1, Y_2, \dots$ denote i.i.d. random variables with a distribution function G with mean 0 and finite pth moment for some $2 \leq p < \infty$. Then*

$$\max_{k\leq n}\left|\frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_k)\right|^p, \ n \geq 1, \textit{ are uniformly integrable}.$$

Moreover, if $0 < \varepsilon < \infty$, then there is a nonincreasing, Lebesgue integrable function G_ε on $(0,\infty)$ for which

$$P\left\{\max_{k\leq n}|Y_1 + \cdots + Y_k| > y\right\} \leq \frac{n}{y^{p-1}}G_\varepsilon(y)$$

for all $y \geq n\varepsilon$ and all $n = 1,2,\dots$.

COROLLARY. *There are constants $\kappa_1, \kappa_2, \dots$ for which $\sum_{n=1}^\infty n\kappa_n < \infty$ and $P\{t_a > n\} \leq \kappa_n$, for all $n \geq 2a/\varepsilon_1$ and $a \geq 1$.*

PROOF. For $n > 2a/\varepsilon_1$, $a \leq \varepsilon_1 n/2$ and, therefore,

$$\begin{aligned} P\{t_a > n\} \leq P\{Z_n \leq a\} &\leq P\left\{\langle \mathbf{c},\mathbf{S}_n\rangle + \xi_n < -\left(1 - \tfrac{1}{2}\varepsilon_1\right)n\right\} \\ &\leq P\left\{\langle \mathbf{c},\mathbf{S}_n\rangle < -\tfrac{1}{2}\varepsilon_1 n\right\} + P\left\{\xi_n < -(1-\varepsilon_1)n\right\}. \end{aligned}$$

So, the corollary follows directly from Proposition 1, (C1) and (C3). □

PROPOSITION 2.

(i) $$\lim_{a\to\infty}\frac{t_a}{a} = 1 \quad w.p.1,$$

(ii) $$\lim_{a\to\infty} E\left[\left(\frac{t_a}{a}\right)^2\right] = 1,$$

(iii) $$\lim_{a\to\infty}\int_{t_a > 2a/\varepsilon_1} t_a^2\, dP = 0.$$

PROOF. (i) follows easily from (5) and the strong law of large numbers, by a standard argument; (ii) is then an easy consequence of (iii); and (iii) follows from the last corollary and an integration by parts, as in Woodroofe [(1982), page 46]. □

It is implicit in the statement and proof of Proposition 2 that $E(t_a^2) < \infty$ for all $a \geq 1$.

COROLLARY 1. *$E\{\langle \mathbf{b}, \mathbf{S}_t \rangle\} = 0$ and $E\{\langle \mathbf{b}, \mathbf{S}_t \rangle^2\} = \langle \mathbf{b}, \Sigma \mathbf{b} \rangle E(t)$ for all $\mathbf{b} \in \mathscr{W}$. If $\mathbf{b} \in \mathscr{W}$ and $v_3(\mathbf{b}) < \infty$, then*

$$E\{\langle \mathbf{b}, \mathbf{S}_t \rangle^3\} = 3\langle \mathbf{b}, \Sigma \mathbf{b} \rangle E(t\langle \mathbf{b}, \mathbf{S}_t \rangle) + \tilde{\mu}_3(\mathbf{b}) E(t).$$

COROLLARY 2. *If $\mathbf{b} \in \mathscr{W}$ and $v_4(\mathbf{b}) < \infty$, then $a^{-2}\langle \mathbf{b}, \mathbf{S}_t \rangle^4$, $a \geq 1$, are uniformly integrable.*

PROOFS. The first corollary follows directly from Wald's lemmas [for example, Chow, Robbins, and Teicher (1965)]. The second then follows from Lemma 5 of Chow and Yu (1981). □

Observe that $\mathbf{S}_n^* := \mathbf{S}_n / \sqrt{n} \Rightarrow \mathbf{W}$ as $n \to \infty$, where $\mathbf{W}$ has the normal distribution with mean 0 and covariance operator Σ, the covariance operator of $\mathbf{X}_1$, and $\Rightarrow$ denotes convergence in distribution. Suppose that $(\mathbf{S}_n^*, \xi_n)$ have a limiting joint distribution, say

$$\text{(C6)} \qquad (\mathbf{S}_n^*, \xi_n) \Rightarrow (\mathbf{W}, \xi) \quad \text{as } n \to \infty;$$

and let

$$R_a = Z_t - a, \qquad a \geq 1.$$

PROPOSITION 3. *As $a \to \infty$, $(\mathbf{S}_t^*, \xi_t) \Rightarrow (\mathbf{W}, \xi)$; and if $\langle c, \mathbf{X}_1 \rangle$ has a nonarithmetic distribution, then $(\mathbf{S}_t^*, \xi_t, R_a) \to (\mathbf{W}, \xi, R)$, where R is independent of $(\mathbf{W}, \xi)$ and*

$$P\{r \leq R \leq r + dr\} = \frac{1}{E(\tau)} P\{\tau + \langle \mathbf{c}, \mathbf{S}_\tau \rangle > r\}\, dr, \qquad 0 < r < \infty,$$

with

$$\tau = \inf\{n \geq 1 : n + \langle \mathbf{c}, \mathbf{S}_n \rangle > 0\}.$$

PROOF. The first assertion follows directly from Anscombe's theorem. The second may be established along the lines of the proof of Theorem 2 of Lai and Siegmund (1977), and is actually a consequence of that theorem if $\xi_t - g(\mathbf{S}_t^*) \to 0$ in probability for some continuous function $g: \mathscr{W} \to \mathbb{R}$. □

3. Statement of results. In Theorems 1 and 2, let ρ and ν denote the means of R and ξ, say

$$\rho = E(R) \quad \text{and} \quad \nu = E(\xi).$$

THEOREM 1. *If conditions* (C1)–(C6) *hold, then* $E(t-a) = O(1)$ *as* $a \to \infty$; *and if* $\langle c, \mathbf{X}_1 \rangle$ *has a nonarithmetic distribution, then* $\lim_{a\to\infty} E(t-a) = \rho - \nu$.

PROOF. These assertions are proved by Hagwood and Woodroofe (1982), though only the second is explicitly stated. Alternatively, Theorem 1 may be deduced from Theorems 1 and 2 of Zhang (1988). □

In the statement of Theorems 2 and 3, let

$$\nu_2(\mathbf{b}_1, \mathbf{b}_2) = E\{\xi\langle \mathbf{b}_1, \mathbf{W}\rangle\langle \mathbf{b}_2, \mathbf{W}\rangle\} - \nu\langle \mathbf{b}_1, \Sigma\mathbf{b}_2\rangle,$$

$$\Delta_2^a(\mathbf{b}_1, \mathbf{b}_2) = a^2 E\{\langle \mathbf{b}_1, \overline{\mathbf{X}}_t\rangle\langle \mathbf{b}_2, \overline{\mathbf{X}}_t\rangle\} - \langle \mathbf{b}_1, \Sigma\mathbf{b}_2\rangle a$$

and

$$\Delta_k^a(\mathbf{b}_1, \dots, \mathbf{b}_k) = a^2 E\left\{\prod_{i=1}^k \langle \mathbf{b}_i, \overline{\mathbf{X}}_t\rangle\right\},$$

for $\mathbf{b}_1, \dots, \mathbf{b}_k \in \mathscr{W}$ for which the expectations are finite for $k = 3, 4$ and $1 \le a < \infty$, where $\Sigma = \mu_2$ is the covariance operator of $\mathbf{X}_1$.

THEOREM 2. *Suppose that* (C1)–(C6) *hold. If* $\mathbf{b} \in \mathscr{W}$ *and* $v_q(\mathbf{b}) < \infty$ *for some* $q \ge \max\{4, 2\alpha/(\alpha-1), 2p/(p-2)\}$, *then*

$$\lim_{a\to\infty} aE\{\langle \mathbf{b}, \overline{\mathbf{X}}_t\rangle\} = \langle \mathbf{b}, \Sigma\mathbf{c}\rangle \tag{7}$$

and

$$\tilde{\Delta}_2^a(\mathbf{b}) = O(1) \tag{8}$$

as $a \to \infty$; *and if* $\langle \mathbf{c}, \mathbf{X}_1\rangle$ *has a nonarithmetic distribution, then*

$$\begin{aligned}\lim_{a\to\infty} \tilde{\Delta}_2^a(\mathbf{b}) &= 2\nu_2(\mathbf{b}, \mathbf{b}) + (\nu - \rho)\langle \mathbf{b}, \Sigma\mathbf{b}\rangle \\ &\quad + \langle \mathbf{b}, \Sigma\mathbf{b}\rangle\langle \mathbf{c}, \Sigma\mathbf{c}\rangle + 2\langle \mathbf{b}, \Sigma\mathbf{c}\rangle^2 + 2\mu_3(\mathbf{b}, \mathbf{b}, \mathbf{c}).\end{aligned} \tag{9}$$

PROOF (Outline). The structure of the proof of (7) is easily described. By Wald's lemma,

$$aE\{\langle \mathbf{b}, \overline{\mathbf{X}}_t\rangle\} = E\left\{\frac{1}{t}(a-t)\langle \mathbf{b}, \mathbf{S}_t\rangle\right\}.$$

Then, using the relation $a - t = \langle \mathbf{c}, \mathbf{S}_t\rangle + (\xi_t - R_a)$, which follows from (2), and Propositions 2 and 3,

$$\frac{1}{t}(a-t)\langle \mathbf{b}, \mathbf{S}_t\rangle = \frac{1}{t}\{\langle \mathbf{c}, \mathbf{S}_t\rangle\langle \mathbf{b}, \mathbf{S}_t\rangle + (\xi_t - R_a)\langle \mathbf{b}, \mathbf{S}_t\rangle\} \Rightarrow \langle \mathbf{c}, \mathbf{W}\rangle\langle \mathbf{b}, \mathbf{W}\rangle$$

as $a \to \infty$. So, if uniform integrability could be established, then $aE\{\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle\} = E\{\langle \mathbf{c}, \mathbf{W} \rangle \langle \mathbf{b}, \mathbf{W} \rangle\} + o(1) = \langle \mathbf{b}, \Sigma \mathbf{c} \rangle + o(1)$ as $a \to \infty$. The proofs of (8) and (9) use similar techniques.

Uniform integrability is considered in Section 5, and the proof of Theorem 2 is presented in Section 6.

The following corollary may be used to compute the regret of several sequential estimation procedures.

COROLLARY 1. *If* $\langle \mathbf{b}, \Sigma \mathbf{b} \rangle = 1$, *then* $a^2 E\{\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^2\} + E(t) = 2a + O(1)$ *as* $a \to \infty$; *and if* $\langle \mathbf{c}, \mathbf{X}_1 \rangle$ *has a nonarithmetic distribution, then*

$$\lim_{a \to \infty} \left\{ a^2 E\left[\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^2 \right] + E(t) - 2a \right\}$$
$$= 2\nu_2(\mathbf{b}, \mathbf{b}) + \langle \mathbf{c}, \Sigma \mathbf{c} \rangle + 2\langle \mathbf{b}, \Sigma \mathbf{c} \rangle^2 + 2\mu_3(\mathbf{b}, \mathbf{b}, \mathbf{c}).$$

PROOF. This follows from Theorems 1 and 2 and some simple algebra. □

COROLLARY 2. *For all* $\mathbf{b}_1, \mathbf{b}_2 \in \mathscr{W}$,

$$\lim_{a \to \infty} \Delta_2^a(\mathbf{b}_1, \mathbf{b}_2) = 2\nu_2(\mathbf{b}_1, \mathbf{b}_2) + (\nu - \rho)\langle \mathbf{b}_1, \Sigma \mathbf{b}_2 \rangle + \langle \mathbf{b}_1, \Sigma \mathbf{b}_2 \rangle \langle \mathbf{c}, \Sigma \mathbf{c} \rangle$$
$$+ 2\langle \mathbf{b}_1, \Sigma \mathbf{c} \rangle \langle \mathbf{b}_2, \Sigma \mathbf{c} \rangle + 2\mu_3(\mathbf{b}_1, \mathbf{b}_2, \mathbf{c}).$$

PROOF. This is clear, since $4\Delta_2^a(\mathbf{b}_1, \mathbf{b}_2) = \tilde{\Delta}_2^a(\mathbf{b}_1 + \mathbf{b}_2) - \tilde{\Delta}_2^a(\mathbf{b}_1 - \mathbf{b}_2)$ for all $\mathbf{b}_1, \mathbf{b}_2 \in \mathscr{W}$ and $a \geq 1$. □

THEOREM 3. *Suppose that* (C1)–(C6) *are satisfied. If* $\mathbf{b} \in \mathscr{W}$ *and* $v_q(\mathbf{b}) < \infty$ *for some* $q \geq \max\{4, 6\alpha/(2\alpha - 1), 3p/(p - 2)\}$, *then*

$$\lim_{a \to \infty} \tilde{\Delta}_3^a(\mathbf{b}) = 6\langle \mathbf{b}, \Sigma \mathbf{b} \rangle \langle \mathbf{b}, \Sigma \mathbf{c} \rangle + \tilde{\mu}_3(\mathbf{b}) = \overline{\Delta}_3(\mathbf{b}), \quad \textit{say}; \tag{10}$$

and if $v_q(\mathbf{b}) < \infty$ *for some* $q \geq \max\{4, 4p/(p - 2)\}$, *then*

$$\lim_{a \to \infty} \tilde{\Delta}_4^a(\mathbf{b}) = 3\langle \mathbf{b}, \Sigma \mathbf{b} \rangle^2 = \overline{\Delta}_4(\mathbf{b}), \quad \textit{say}. \tag{11}$$

The proof of Theorem 3 uses techniques similar to those in the proof of (7). The details are presented in Section 7.

The space $\mathscr{W}_k$ of all k-linear functionals is itself an inner product space with inner product

$$\langle \mu, \nu \rangle_k = \sum_{e_1 \in \mathscr{E}} \cdots \sum_{e_k \in \mathscr{E}} \mu(e_1, \ldots, e_k) \nu(e_1, \ldots, e_k)$$

for $\mu, \nu \in \mathscr{W}_k$, where $\mathscr{E}$ denotes an orthonormal basis for $\mathscr{W}$. Let $\|\mu\|_k = \sqrt{\langle \mu, \mu \rangle_k}$ for $\mu \in \mathscr{W}_k$.

COROLLARY. *If there is a* $q \geq \max\{4, 6\alpha/(2\alpha - 1), 4p/(p-2)\}$ *for which*

$$\int_{\mathscr{W}} \|x\|^q F(dx) < \infty, \tag{12}$$

then there are symmetric multilinear functionals Δ_3 *and* Δ_4 *for which* $\tilde{\Delta}_k = \overline{\Delta}_k$ *for* $k = 3, 4$ *and*

$$\lim_{a \to \infty} \Delta_k^a = \Delta_k \in \mathscr{W}_k, \qquad k = 3, 4.$$

PROOF. This follows directly from Theorem 3 and (4). □

Recall that the derivatives of a function $h\colon \mathscr{W} \to \mathbb{R}$ at a given $\mathbf{w}_0 \in \mathscr{W}$ may be regarded as multilinear functionals on $\mathscr{W}$. See, for example, Edwards [(1973), page 414]. Let $\mathscr{H}$ be the class of all $h\colon \mathscr{W} \to \mathbb{R}$ for which $h(\mathbf{0}) = 0$ and h is twice continuously differentiable on some neighborhood N_h of $\mathbf{0} \in \mathscr{W}$; and let $\mathscr{K}$ denote the class of all $h \in \mathscr{H}$ for which $Dh_{\mathbf{0}} = \mathbf{0}$ and h is four times continuously differentiable on N_h. Observe that if $h \in \mathscr{H}$ has four continuous derivatives near $\mathbf{0}$, then $h^2 \in \mathscr{K}$.

THEOREM 4. *Suppose that* (12) *holds for some* $q \geq 4$ *and that conditions* (C1)–(C6) *hold with* $p = q$ *and* $\alpha \geq q/(q-2)$. *Let* $h \in \mathscr{H}$; *and let* h_k, $k = 1, 2, \dots$, *be functions for which* $h_k = h$ *on* N_h *for all* $k = 1, 2, \dots$ *and, for some* $r \geq (q-1)/(q-2)$,

$$E\left\{\sup_{k \geq 1} \left|h_k(\overline{\mathbf{X}}_k)\right|^r\right\} < \infty, \tag{13}$$

then

$$\lim_{a \to \infty} aE\{h_t(\overline{\mathbf{X}}_t)\} = \langle Dh_{\mathbf{0}}, \Sigma \mathbf{c}\rangle + \tfrac{1}{2}\langle D^2 h_{\mathbf{0}}, \Sigma\rangle_2. \tag{14}$$

If $h \in \mathscr{K}$, $q \geq 6$ *and* (13) *holds for some* $r \geq (q-1)/(q-3)$, *then*

$$a^2 E\{h_t(\overline{\mathbf{X}}_t)\} - \tfrac{1}{2}\langle D^2 h_{\mathbf{0}}, \Sigma\rangle_2 a = O(1); \tag{15}$$

and if, in addition, $\langle \mathbf{c}, \mathbf{X}_1\rangle$ *has a nonarithmetic distribution, then*

$$\begin{aligned} &\lim_{a \to \infty} a^2 E\{h(\overline{\mathbf{X}}_t)\} - \tfrac{1}{2}\langle D^2 h_{\mathbf{0}}, \Sigma\rangle_2 a \\ &\quad = \tfrac{1}{2}\langle D^2 h_{\mathbf{0}}, \Delta_2\rangle_2 + \tfrac{1}{6}\langle D^3 h_{\mathbf{0}}, \Delta_3\rangle_3 + \tfrac{1}{24}\langle D^4 h_{\mathbf{0}}, \Delta_4\rangle_4. \end{aligned} \tag{16}$$

The proof of Theorem 4 uses a Taylor series expansions together with Theorems 2 and 3. The details are presented in Section 8.

4. Examples. The examples considered are all of the following form. Let $g\colon \mathscr{W} \to \mathbb{R}$ be a function for which $g(\mathbf{0}) = 1$ and g is twice continuously differentiable on some neighborhood of $\mathbf{0} \in \mathscr{W}$; and let g_k, $k = 1, 2, \dots$, be functions for which $g_k = g$ for all $k = 1, 2, \dots$ on some neighborhood of $\mathbf{0}$. Then

$$Z_n = n g_n(\overline{\mathbf{X}}_n), \qquad n \geq 1, \tag{17}$$

are of the form (1) with $\mathbf{c} = Dg_{\mathbf{0}}$, the derivative of g at zero, and $\xi_n = Z_n - (n + \langle \mathbf{c}, \mathbf{S}_n \rangle)$, $n = 1, 2, \dots$.

PROPOSITION 4. *If* (17) *holds, with the conditions of the previous paragraph, and* (12) *holds for some* $q \geq 3$, *then* (C4), (C5) *and* (C6) *hold, with* $\alpha = q/2$ *and* $\xi = \frac{1}{2}\langle \mathbf{W}, D^2 g_{\mathbf{0}} \mathbf{W} \rangle$, *where* $D^2 g_{\mathbf{0}}$ *denotes the second derivative of* g *at* $\mathbf{0}$.

PROOF. For (C5), let $0 < \delta < \infty$ be so small that g is twice continuously differentiable on $N = \{\mathbf{w} \in \mathscr{W}\colon \|\mathbf{w}\| \leq \delta\}$ and $g_k = g$ on N for all $k = 1, 2, \dots$; and let $A_n = \{\|\mathbf{S}_n/n\| \leq \delta\}$ for all $n = 1, 2, \dots$. Then

$$\sum_{n=1}^{\infty} nP\left(\bigcup_{k=n}^{\infty} A'_k\right) \leq \sum_{n=1}^{\infty} nP\left\{\sup_{k \geq n} \left\| \frac{\mathbf{S}_k}{k} \right\| > \delta\right\} < \infty$$

by the Baum–Katz (1965) inequalities. Of course, if A_n occurs, then

$$\xi_n = \frac{1}{2n}\left\langle \mathbf{S}_n, D^2 g_{\mathbf{Y}_n} \mathbf{S}_n \right\rangle,$$

where $\mathbf{Y}_n$ denotes an intermediate point between $\mathbf{0}$ and $\overline{\mathbf{X}}_n$, by Taylor's theorem. So, if C denotes an upper bound for the operator norm of $Dg_{\mathbf{w}}$ for $\|\mathbf{w}\| \leq \delta$, then

$$\max_{k \leq n} \left| \xi_{n+k} I_{A_{n+k}} \right|^{\alpha} \leq \left\{ \frac{C}{2n} \max_{k \leq 2n} \|\mathbf{S}_k\|^2 \right\}^{\alpha}$$

for all $n \geq 1$; and the sequence on the right is uniformly integrable, by Proposition 1. It follows easily that

$$\max_{k \leq n} \left| \xi_{n+k} - \tfrac{1}{2}\left\langle \mathbf{S}^*_{n+k}, D^2 g_{\mathbf{0}} \mathbf{S}^*_{n+k} \right\rangle \right| \to 0$$

in probability as $n \to \infty$. So, (C6) holds with $\xi = \frac{1}{2}\langle \mathbf{W}, D^2 g_{\mathbf{0}} \mathbf{W} \rangle$; and (C4) may be established as in Woodroofe [(1982), pages 41–42]. □

PROPOSITION 5. *Suppose that* (17) *holds with* $g_k = g$ *on* $\mathscr{W}$ *for all* $k = 1, 2, \dots$, *where* g *is a convex function on* $\mathscr{W}$. *If* $p \geq 3$ *and* $E\{[g(\mathbf{X}_1)^+]^p\} < \infty$, *then* (C2) *and* (C3) *are satisfied.*

PROOF. In this case (C3) is clear, since $\xi_n \geq 0$ for all $n = 1, 2, \ldots$. For (C2), observe that $Z_n \leq g(\mathbf{X}_1) + \cdots + g(\mathbf{X}_n) \leq n\vartheta + Y_1 + \cdots + Y_n$, where $\vartheta = E[g(\mathbf{X}_1)]$ and $Y_k = g(\mathbf{X}_k) - \vartheta$ for all $k = 1, 2, \ldots$. If $\varepsilon_0 = 1/2\vartheta$, then

$$P\left\{Z_n - \frac{n}{\varepsilon_0} > z\right\} \leq P\{Y_1 + \cdots + Y_n > n\vartheta + z\} \leq \frac{G_\vartheta(n\vartheta + z)}{(n\vartheta + z)^{p-1}}$$

for all $z \geq 0$ and $n = 1, 2, \ldots$, where G is the distribution of $g(\mathbf{X}_1)$ (and G_ϑ is as in Proposition 1). So, (C2) holds. □

EXAMPLE 1 (The linear case). If $g(\mathbf{x}) = 1 + \langle \mathbf{c}, \mathbf{x} \rangle$ for all $\mathbf{x} \in \mathscr{W}$, then $\xi_k = 0$ for all $k = 1, 2, \ldots$, so that (C3), (C4), (C5) and (C6) are clearly satisfied for all $0 < \alpha < \infty$; and if (C1) holds, then (C2) is satisfied with $\varepsilon_0 = 1/2$, by Proposition 5. So, if (C1) holds and if $v_q(\mathbf{b}) < \infty$ for some $q \geq \max\{4, 2p/(p-2)\}$, then the conditions of Theorem 2 are satisfied. For $p = q$, this requires $p \geq 4$. For the case $\mathbf{b} = \mathbf{c}$, a finite third moment in the direction $\mathbf{b} = \mathbf{c}$ is clearly necessary for (9).

Many sequential estimation procedures call for taking a sample size n for which $n \geq \hat{\sigma}_n / \sqrt{c_0}$, where $\hat{\sigma}_n^2$ is an estimate of a variance and c_0 is a cost parameter; and it is often desirable to truncate $\hat{\sigma}_n$ below to avoid problems with early stopping. If $\hat{\sigma}_n$ is truncated below at $1/n$, say, then the resulting sample size may be written in the form (2) with $a = \sigma / \sqrt{c_0}$ and

$$Z_n = \frac{n\sigma}{\max(\hat{\sigma}_n, 1/n)}, \qquad n \geq 1. \tag{18}$$

EXAMPLE 2 [Martinsek's (1983) problem]. Let $Y_1, Y_2, \ldots$ denote i.i.d. random variables with an unknown distribution function G, having mean θ and variance $0 < \sigma^2 < \infty$; and define Z_n by (18) with

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{Y}_n\right)^2, \qquad n \geq 1.$$

Then Z_n is of the form (17) with $\mathbf{X}_k = [(Y_k - \theta), (Y_k - \theta)^2 - \sigma^2]$, $k = 1, 2, \ldots$, and $g(x_1, x_2) = \sigma / \sqrt{(\sigma^2 + x_2 - x_1^2)}$ for all $\mathbf{x} = (x_1, x_2) \in \mathscr{W}$ for which $x_2 - x_1^2 > -\sigma^2$, in which case $\mathbf{c} = (0, -1/2\sigma^2)$.

It is shown that conditions (C1)–(C6) are satisfied with $p = 3$ provided that $E\{|Y_1|^6\} < \infty$. In the verification, there is no loss of generality in supposing that $\theta = 0$ and $\sigma = 1$. Then (C1) and (C3) are clear, since $\langle \mathbf{c}, \mathbf{X}_1 \rangle = (1 - Y_1^2)/2$ and $\xi_n \geq 0$ for all $n \geq 2$ (essentially, since g is convex); and (C4), (C5) and (C6) follow from Proposition 4 with $q = 3$. For (C2), observe that $Z_n \leq n^2$ and that $Z_n \leq 2n$ on $\{\hat{\sigma}_n \geq 1/2\}$ w.p. 1 for all $n = 2, 3, \ldots$. So, $E\{[(Z_n - 2n)^+]^r\} \leq n^{2r} P\{\hat{\sigma}_n \leq \frac{1}{2}\}$ which approaches zero as $n \to \infty$ for all $r > 0$, as in the proof of

Lemma 4 of Chow and Yu (1981). With $\mathbf{b} = (1, 0)$ and $q = 6$, Corollary 1 now supplies an expansion for the regret of Martinsek's procedure. This expansion was obtained by Martinsek (1983), under the assumption that $E|Y_1|^{8+\varepsilon} < \infty$ for some $\varepsilon > 0$.

EXAMPLE 3 (Aras' problem). Let $L_1, L_2, \dots$ denote i.i.d. exponentially distributed random variables (lifetimes) with an unknown mean $0 < \theta < \infty$; and let $C_1, C_2, \dots$ denote i.i.d. positive random variables (censoring times). In the censored data problem, one observes the random variables, $\delta_i = I\{L_i \le C_i\}$ and $Y_i = \min(C_i, L_i)$, $i = 1, 2, \dots$. Let $K_n = \delta_1 + \cdots + \delta_n$ and $T_n = Y_1 + \cdots + Y_n$, $n = 1, 2, \dots$. Then the maximum likelihood estimator of θ is T_n/K_n, provided that $K_n > 0$. Let $\hat{\theta}_n = T_n/\max(1, K_n)$ for $n = 1, 2, \dots$. Then it may be shown that $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically normal with mean 0 and variance $\sigma^2 = \theta^2/p$, where p is the probability that $L_1 \le C_1$. Let $\hat{\sigma}_n = \hat{\theta}_n/\sqrt{\hat{p}_n}$, where $\hat{p}_n = \max(1, K_n)/n$ for $n \ge 1$. Then (a slight variation on) Aras' (1989) stopping time is of the form (2) with Z_n defined by (18). Let $\mathbf{X}_k = (\delta_k - p, Y_k - p\theta)$, $k = 1, 2, \dots$. Then Z_n is of the form (17) with $g(x_1, x_2) = \sigma\sqrt{(p + x_1)^3}/(p\theta + x_2)$ for $x_2 > -p\theta$. Since $\mathbf{X}_1$ has moments of all orders, it is easily seen that (C1), (C4), (C5) and (C6) are satisfied; and it may be shown that (C2) and (C3) are satisfied, as in the last example.

In this case $\hat{\theta}_n - \theta = h_n(\overline{\mathbf{X}}_n)$, $n \ge 1$, where $h(x_1, x_2) = (p\theta + x_2)/(p + x_1) - \theta$ for all $x_1 > -p$; and an expansion for $E[(\hat{\theta}_t - \theta)^2]$ may be deduced from Theorem 4.

5. Uniform integrability. Estimates are needed for the probability that t_a is small.

LEMMA 1. *There is a function Δ on $[1, \infty)$ for which $\lim_{a \to \infty} \Delta(a) = 0$ and $P\{t_a \le n\} \le na^{-p}\Delta(a)$, for all $n \le \varepsilon_0 a/2$ and $a \ge 1$.*

PROOF. Let $\Delta(a) = \sup_{n \ge 1} a^p P\{Z_n - n/\varepsilon_0 > a/2\}$ for $a \ge 1$. Then $\Delta(a) \to 0$ as $a \to \infty$, by (C2); and, for all $n \le a\varepsilon_0/2$ and $a \ge 1$,

$$\begin{aligned} P\{t_a \le n\} &= P\left\{\max_{1 \le k \le n} Z_k - \frac{k}{\varepsilon_0} > \frac{a}{2}\right\} \\ &\le \sum_{k=1}^{n} P\left\{Z_k - \frac{k}{\varepsilon_0} > \frac{a}{2}\right\} \le \frac{n}{a^p}\Delta(a). \end{aligned} \qquad \square$$

PROPOSITION 6. *Let $\mathbf{b} \in \mathscr{W}$, $0 < \eta < \varepsilon_0/4$ and $2 \le r < \infty$. If $v_q(\mathbf{b}) < \infty$ for some $r < q < \infty$, then*

$$\int_{t \le \eta a} \left|\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle\right|^r dP \le C v_q(\mathbf{b})^r \times \left(\frac{1}{a}\right)^{p(1-r/q)} \Delta(a)^{1-r/q}, \tag{19}$$

where C is a constant depending only on r and q.

PROOF. Let K be the least integer for which $2^K > \eta a$. Then, using (6), the Marcinkiewicz–Ziegmund inequality, and Lemma 1,

$$\begin{aligned}
\int_{t \le \eta a} \left|\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle\right|^r dP &\le \sum_{k=1}^{K} \int_{2^{k-1} \le t \le 2^k} \left|\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle\right|^r dP \\
&\le \sum_{k=1}^{K} E\left\{ \sup_{n \ge 2^{k-1}} \left|\langle \mathbf{b}, \overline{\mathbf{X}}_n \rangle\right|^q \right\}^{r/q} P\{t \le 2^k\}^{1-r/q} \\
&\le \sum_{k=1}^{K} \left[C_q \frac{v_q(\mathbf{b})}{\sqrt{2^{k-1}}} \right]^r \left[2^k \left(\frac{1}{a}\right)^p \Delta(a) \right]^{1-r/q} \\
&\le 2^r C_q^r v_q(\mathbf{b})^r \left(\frac{1}{a}\right)^{p(1-r/q)} \Delta(a)^{1-r/q} \times \sum_{k=1}^{K} \left(\frac{1}{2}\right)^{rk/q}
\end{aligned}$$

for all $a \ge 1$, where C_q depends only on q. (19) follows. □

COROLLARY. *If $q \ge rp/(p-2)$, then $\int_{t \le \eta a} |\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle|^r \, dP = o(a^{-2})$ as $a \to \infty$.*

For the next result, let $\beta = \min(2, \alpha)$,

$$\tau = \tau_a = \inf\{n \ge 1 \colon n + \langle \mathbf{c}, \mathbf{S}_n \rangle > a\},$$

$$E_a = \tau_a + \langle \mathbf{c}, \mathbf{S}_{\tau_a} \rangle - a.$$

Then, comparing (C1), (C3), (C5) and Lemma 1 with the conditions of Theorem 2(ii) of Zhang (1988),

$$|t_a - \tau_a|^\beta,\ a \ge 1, \text{ are uniformly integrable.} \tag{20}$$

Let $\beta' = \min(2\beta, p) = \min(4, 2\alpha, p)$. Then it follows that $|\langle c, S_t - S_\tau \rangle|^{\beta'}$, $a \ge 1$, are uniformly integrable. See Lemma 5 of Chow and Yu (1981).

PROPOSITION 7. *$|\xi_t - R_a|^\beta$, $a \ge 1$, are uniformly integrable.*

PROOF. This is clear from the preceding remarks, since $R_a - \xi_t = E_a + t_a - \tau_a + \langle c, S_t - S_\tau \rangle$, and E_a^{p-1} is uniformly integrable. (Note that $p - 1 \ge 2 \ge \beta$ and $\beta' \ge 3 \ge \beta$.) □

Let $0 < \eta < \varepsilon_1/4$; and let $t_a^* = (t_a - a)/\sqrt{a}$ for $1 \le a < \infty$.

PROPOSITION 8. *$|t_a^*|^{\beta'} I_{\{t \le a/\eta\}}$, $a \ge 1$, are uniformly integrable.*

PROOF. It is easily seen that $\{t \le a/\eta,\ |t_a^*| > 2x\} = \varnothing$ for $x > \sqrt{a}/\eta$ and

that $\{t \le a/\eta,\ |t_a^*| > 2x\} \subseteq \{|R_a - \xi_t| > x\sqrt{a}\} \cup \{\max_{n \le a/\eta} |\langle c, \mathbf{S}_n\rangle| > x\sqrt{a}\}$ for all $x \ge 1$ and $a \ge 1$. So,

$$P\left\{t \le \frac{a}{\eta},\ |t_a^*| > 2x\right\} \le P\{|(R_a - \xi_t)| > \eta x^2\} + P\left\{\max_{n \le a/\eta} |\langle c, \mathbf{S}_n\rangle| > x\sqrt{a}\right\}$$

for all $x \ge 1$ and $a \ge 1$. The proposition follows. □

COROLLARY. *t_a^{*2}, $a \ge 1$, are uniformly integrable.*

PROOF. This follows easily from the Propositions 2(iii) and 8. □

6. The first two moments. Now let $0 < \eta < \min(\varepsilon_0, \varepsilon_1)/4$. Then, by Wald's lemma and Proposition 2(iii),

$$\int_{t > a/\eta} \langle \mathbf{c}, \mathbf{S}_t\rangle^2\, dP \le \langle \mathbf{c}, \Sigma \mathbf{c}\rangle \int_{t > a/\eta} t\, dP \to 0. \tag{21}$$

Similarly, if $\mathbf{b} \in \mathscr{W}$, $v_4(\mathbf{b}) < \infty$ and $0 < r \le 4$, then

$$\begin{aligned} \int_{t > a/\eta} |\langle \mathbf{b}, \mathbf{S}_t\rangle|^r\, dP &\le P\left\{t > \frac{a}{\eta}\right\}^{1-r/4} \times \left\{\int_{t > a/\eta} |\langle \mathbf{b}, \mathbf{S}_t\rangle|^4\, dP\right\}^{r/4} \\ &= O[a^{-2(1-r/4)}]o[a^{r/2}] = o(a^{r-2}) \end{aligned} \tag{22}$$

by Hölder's inequality, Proposition 2 and its corollaries.

LEMMA 2.

$$\lim_{a \to \infty} \frac{1}{a} E(t\langle \mathbf{b}, \mathbf{S}_t\rangle) = -\langle \mathbf{b}, \Sigma \mathbf{c}\rangle, \qquad \forall\, \mathbf{b} \in \mathscr{W}.$$

PROOF. By Wald's lemma, $E(t\langle \mathbf{b}, \mathbf{S}_t\rangle) = E[(t - a)\langle \mathbf{b}, \mathbf{S}_t\rangle]$ for all $a \ge 1$. Now, $(t - a)\langle \mathbf{b}, \mathbf{S}_t\rangle/a = \{(R_a - \xi_t) - \langle \mathbf{c}, \mathbf{S}_t\rangle\}\langle \mathbf{b}, \mathbf{S}_t\rangle/a \Rightarrow -\langle \mathbf{c}, \mathbf{W}\rangle\langle \mathbf{b}, \mathbf{W}\rangle$ as $a \to \infty$; and $(t - a)\langle \mathbf{b}, \mathbf{S}_t\rangle/a$, $a \ge 1$, are uniformly integrable, by Propositions 2 and 8 and Hölder's inequality, since t_a^{*2}, $a \ge 1$, and $\langle \mathbf{b}, \mathbf{S}_t\rangle^2/a$ are. So, $\lim_{a \to \infty} E\{(t - a)\langle \mathbf{b}, \mathbf{S}_t\rangle\}/a = -E\{\langle \mathbf{c}, \mathbf{W}\rangle\langle \mathbf{b}, \mathbf{W}\rangle\} = -\langle \mathbf{b}, \Sigma \mathbf{c}\rangle$. □

PROOF OF (7). The proof of (7), described after the statement of Theorem 2, may be justified in a similar manner. □

PROOF OF (8) AND (9). To begin, write $a^2\langle \mathbf{b}, \overline{\mathbf{X}}_t\rangle^2 = \langle \mathbf{b}, \mathbf{S}_t\rangle^2 + (a^2/t^2 - 1)\langle \mathbf{b}, \mathbf{S}_t\rangle^2$ and $(a^2/t^2 - 1) = -2(t/a - 1) + (3/s^4)(t/a - 1)^2$, where s is an intermediate point, $|s - 1| \le |t_a/a - 1|$. Combining these relations with $a - t = \langle \mathbf{c}, \mathbf{S}_t\rangle + \xi_t - R_a$ yields

$$a^2\langle \mathbf{b}, \overline{\mathbf{X}}_t\rangle^2 = \langle \mathbf{b}, \mathbf{S}_t\rangle^2 + \frac{2}{a}\langle \mathbf{c}, \mathbf{S}_t\rangle\langle \mathbf{b}, \mathbf{S}_t\rangle^2 + Y_1 + Y_2,$$

where

$$Y_1 = \frac{2}{a}(\xi_t - R_a)\langle \mathbf{b}, \mathbf{S}_t\rangle^2$$

and

$$Y_2 = \frac{3}{s^4}\left(\frac{t-a}{a}\right)^2 \langle \mathbf{b}, \mathbf{S}_t\rangle^2.$$

Here $E\{\langle \mathbf{b}, \mathbf{S}_t\rangle^2\} = \langle \mathbf{b}, \Sigma\mathbf{b}\rangle E(t)$ by Wald's lemma. So, by Theorem 1, $E\{\langle \mathbf{b}, \mathbf{S}_t\rangle^2\} = \langle \mathbf{b}, \Sigma\mathbf{b}\rangle a + O(1)$, and

$$E\left\{\langle \mathbf{b}, \mathbf{S}_t\rangle^2\right\} = \langle \mathbf{b}, \Sigma\mathbf{b}\rangle(a + \rho - \nu) + o(1)$$

as $a \to \infty$, if $\langle \mathbf{c}, \mathbf{X}_1\rangle$ has a nonarithmetic distribution. Similarly,

$$\begin{aligned} E\left\{\frac{2}{a}\langle \mathbf{c}, \mathbf{S}_t\rangle\langle \mathbf{b}, \mathbf{S}_t\rangle^2\right\} &= \frac{2}{a}\left\{\langle \mathbf{b}, \Sigma\mathbf{b}\rangle E\left[t\langle \mathbf{c}, \mathbf{S}_t\rangle\right] + 2\langle \mathbf{b}, \Sigma\mathbf{c}\rangle E\left[t\langle \mathbf{b}, \mathbf{S}_t\rangle\right]\right. \\ &\qquad \left. + \mu_3(\mathbf{b}, \mathbf{b}, \mathbf{c})E(t)\right\} \\ &\to 2\left\{\mu_3(\mathbf{b}, \mathbf{b}, \mathbf{c}) - \langle \mathbf{b}, \Sigma\mathbf{b}\rangle\langle \mathbf{c}, \Sigma\mathbf{c}\rangle - 2\langle \mathbf{b}, \Sigma\mathbf{c}\rangle^2\right\} \end{aligned}$$

as $a \to \infty$, by Wald's lemmas, Proposition 1 and Lemma 2. It is clear that $Y_2 \Rightarrow 3\langle \mathbf{c}, \mathbf{W}\rangle^2\langle \mathbf{b}, \mathbf{W}\rangle^2$ and $Y_1 \Rightarrow 2(\xi - R)\langle \mathbf{b}, \mathbf{W}\rangle^2$, if $\langle \mathbf{c}, \mathbf{X}_1\rangle$ has a nonarithmetic distribution; and it is shown below that Y_1 and Y_2 are uniformly integrable. So, $E(Y_1 + Y_2) = O(1)$, and

$$E(Y_1) = 2\nu_2(\mathbf{b}, \mathbf{b}) + 2(\nu - \rho)\langle \mathbf{b}, \Sigma\mathbf{b}\rangle + o(1)$$

and

$$\lim_{a\to\infty} E(Y_2) = 3E\left\{\langle \mathbf{c}, \mathbf{W}\rangle^2\langle \mathbf{b}, \mathbf{W}\rangle^2\right\} = 3\langle \mathbf{b}, \Sigma\mathbf{b}\rangle\langle \mathbf{c}, \Sigma\mathbf{c}\rangle + 6\langle \mathbf{b}, \Sigma\mathbf{c}\rangle^2$$

if $\langle \mathbf{c}, \mathbf{X}_1\rangle$ has a nonarithmetic distribution. Relations (8) and (9) then follow by substitution. □

UNIFORM INTEGRABILITY OF Y_1 AND Y_2. On $\{t \le a/\eta\}$,

$$|Y_1| \le |\xi_t - R_a| \times \left\{\frac{2}{a}\max_{n \le a/\eta}\langle \mathbf{b}, \mathbf{S}_n\rangle^2\right\},$$

which are uniformly integrable by Propositions 1 and 8 and Hölder's inequality, since $|\xi_t - R_a|^\beta$, $a \ge 1$, are uniformly integrable and $q \ge 2\beta/(\beta - 1) = \max\{4, 2\alpha/(\alpha - 1)\}$. On $(t > a/\eta\}$, $|\xi_t - R_a| \le t + |\langle \mathbf{c}, \mathbf{S}_t\rangle|$. So,

$$\int_{t>a/\eta} |Y_1|\, dP \le \frac{1}{a}\sqrt{\left\{\int_{t>a/\eta}\left[t + |\langle \mathbf{c}, \mathbf{S}_t\rangle|\right]^2 dP\right\}}\sqrt{\left\{\int_{t>a/\eta}\langle \mathbf{b}, \mathbf{S}_t\rangle^4\, dP\right\}} = o(1)$$

by Proposition 2(iii), (21) and (22). So, Y_1 is uniformly integrable.

On $\{a\eta \le t \le a/\eta\}$,

$$|Y_2| \le \frac{3}{\eta^4} \times (t_a^*)^2 \times \left\{\frac{1}{a} \max_{k \le a/\eta} \langle \mathbf{b}, \mathbf{S}_k \rangle^2\right\},$$

which are uniformly integrable by Propositions 1 and 8 and Hölder's inequality, since $q \ge 2\beta'/(\beta' - 2)$. On $\{t > a/\eta\}$, $3/s^4 \le 4a/t$, so that

$$\int_{t>a/\eta} Y_2 \, dP \le \frac{4}{a} \int_{t>a/\eta} \frac{(t-a)^2}{t} \langle \mathbf{b}, \mathbf{S}_t \rangle^2 \, dP \le \frac{4}{a} \int_{t>a/\eta} t \langle \mathbf{b}, \mathbf{S}_t \rangle^2 \, dP \to 0$$

as $a \to \infty$, by Proposition 2(iii), (22) and Schwarz's inequality. Finally, on $t \le a\eta$, $3/s^4 \le 4a^2/t^2$, so that

$$\int_{t \le a\eta} Y_2 \, dP \le 4a^2 \int_{t \le a\eta} \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^2 \, dP \to 0$$

by Proposition 6. So, Y_2 is uniformly integrable. □

7. Higher moments.

PROOF OF (11). Let $0 < \eta < \min(\varepsilon_0, \varepsilon_1)/4$; and let M be the least integer which exceeds a/η. Then, by Proposition 6, (6) and the Marcinkiewicz–Ziegmund inequalities,

$$\lim_{a \to \infty} a^2 \int_{t \le \eta a} \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^4 \, dP = 0$$

and

$$a^2 \int_{t>a/\eta} \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^4 \, dP \le 4a^2 E\{\langle \mathbf{b}, \overline{\mathbf{X}}_M \rangle^4\} \le 8C\eta^2 v_4(\mathbf{b})^4$$

for all $1 \le a < \infty$, where C is an absolute constant. Moreover, since $\sqrt{a}\,\overline{\mathbf{X}}_t \Rightarrow \mathbf{W}$, by Proposition 3, and $a^2 \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^4$, $a \ge 1$, are uniformly integrable on $\{\eta a < t \le a/\eta\}$, by Proposition 1,

$$\lim_{a \to \infty} a^2 \int_{\eta a < t \le a/\eta} \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^4 \, dP = 3 \langle \mathbf{b}, \Sigma \mathbf{b} \rangle^2.$$

Relation (11) follows by letting $a \to \infty$ and $\eta \to 0$, in that order. □

PROOF OF (10). For the third moment,

$$a^2 E\{\langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle^3\} = \frac{1}{a} E\{\langle \mathbf{b}, \mathbf{S}_t \rangle^3\} + E(Y),$$

where

$$Y = \frac{1}{a}\left(\frac{a^3}{t^3} - 1\right) \langle \mathbf{b}, \mathbf{S}_t \rangle^3 = \frac{3}{a^2 s^4}(a - t) \langle \mathbf{b}, \mathbf{S}_t \rangle^3$$

and $|s-1| \le |t/a - 1|$. By Wald's lemma, Lemma 2 and Proposition 2,

$$\frac{1}{a}E\{\langle \mathbf{b}, \mathbf{S}_t\rangle^3\} = \frac{1}{a}\{3\langle \mathbf{b}, \Sigma\mathbf{b}\rangle E[t\langle \mathbf{b}, \mathbf{S}_t\rangle] + \tilde{\mu}_3(\mathbf{b})E(t)\}$$
$$\to -3\langle \mathbf{b}, \Sigma\mathbf{b}\rangle\langle \mathbf{b}, \Sigma\mathbf{c}\rangle + \tilde{\mu}_3(\mathbf{b})$$

as $a \to \infty$. It is clear that $Y \Rightarrow 3\langle \mathbf{c}, \mathbf{W}\rangle\langle \mathbf{b}, \mathbf{W}\rangle^3$ as $a \to \infty$ and may be shown that Y is uniformly integrable, using (22) and Propositions 1, 6 and 8, as above. So,

$$\lim_{a\to\infty} E(Y) = 3E\{\langle \mathbf{c}, \mathbf{W}\rangle\langle \mathbf{b}, \mathbf{W}\rangle^3\} = 9\langle \mathbf{b}, \Sigma\mathbf{b}\rangle\langle \mathbf{b}, \Sigma\mathbf{c}\rangle. \qquad \square$$

LEMMA 3. *If* (12) *holds for some* $q \ge \max\{4, 4p/(p-2)\}$, *then* $a^2\|\overline{\mathbf{X}}_t\|^4$, $a \ge 1$, *are uniformly integrable.*

PROOF. The proof of Theorem 3 shows that $a^2\langle \mathbf{b}, \overline{\mathbf{X}}_t\rangle^4$, $a \ge 1$, are uniformly integrable. $\square$

8. Smooth functions. The proofs of (15) and (16) are presented in this section. The similar, simpler proof of (14) is omitted. Thus it is assumed throughout this section that $h \in \mathscr{K}$, so that $Dh_{\mathbf{0}} = \mathbf{0}$, that (12) holds for some $q \ge 6$ and that (C1)–(C6) hold with $p = q$ and $\alpha \ge q/(q-2)$.

Let $0 < \delta < 1$ be so small that $h_k = h$ for all $k = 1, 2, \ldots$ and h is four times continuously differentiable on $\{x \in \mathscr{W}: \|x\| \le \delta\}$; let $0 < \eta < \varepsilon_0/4$ and let $C = \{t > \eta a\} \cap \{\|\overline{\mathbf{X}}_t\| \le \delta\}$.

LEMMA 4.

$$\lim_{a\to\infty} a^2 \int_{C'} \sup_{k\ge 1} \left|h_k(\overline{\mathbf{X}}_k)\right| dP = 0$$

and

$$\lim_{a\to\infty} a^2 \int_{C'} \left(1 + \|\overline{\mathbf{X}}_t\|^4\right) dP = 0.$$

PROOF. Clearly, $P(C') \le P\{t \le \eta a\} + P(\sup_{k>\eta a}\|\overline{\mathbf{X}}_k\| > \delta\} = o(a^{-q+1})$ as $a \to \infty$, by Lemma 1, (6) and Proposition 1. So, letting $r = (q-1)/(q-3)$,

$$a^2 \int_{C'} \sup_{k\ge 1} \left|h_k(\overline{\mathbf{X}}_k)\right| dP \le a^2 P(C')^{1-1/r} E\left\{\sup_{k\ge 1}\left|h_k(\overline{\mathbf{X}}_k)\right|^r\right\}^{1/r} = o(1)$$

as $a \to \infty$. This establishes the first assertion; and the second then follows easily from Lemma 3. $\square$

PROOF OF (15) AND (16). By Lemma 4,

$$a^2 E\{h_t(\overline{\mathbf{X}}_t)\} = a^2 \int_C h(\overline{\mathbf{X}}_t)\, dP + o(1)$$

as $a \to \infty$. On C, $h(\overline{\mathbf{X}}_t)$ may be expanded in a Taylor series about $\mathbf{0}$, and

$$a^2 \int_C h(\overline{\mathbf{X}}_t)\, dP = a^2 \left\{ \frac{1}{2} \int_C \tilde{D}^2 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\, dP + \frac{1}{6} \int_C \tilde{D}^3 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\, dP + \frac{1}{24} \int_C \tilde{D}^4 h_{\mathbf{Y}}(\overline{\mathbf{X}}_t)\, dP \right\},$$

where $\mathbf{Y}$ denotes an intermediate point between $\mathbf{0}$ and $\overline{\mathbf{X}}_t$ [since $h(\mathbf{0}) = 0$ and $Dh_{\mathbf{0}} = \mathbf{0}$]. By Lemmas 3 and 4 and Theorem 3,

$$\lim_{a \to \infty} \frac{a^2}{24} \int_C \tilde{D}^4 h_{\mathbf{Y}}(\overline{\mathbf{X}}_t)\, dP = \frac{1}{24} E\{\tilde{D}^4 h_{\mathbf{0}}(\mathbf{W})\} = \frac{1}{24} \langle D^4 h_{\mathbf{0}}, \Delta_4 \rangle_4,$$

$$\begin{aligned} \frac{1}{6} a^2 \int_C \tilde{D}^3 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\, dP &= \frac{1}{6} a^2 E\{\tilde{D}^3 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\} + o(1) \\ &= \frac{1}{6} \langle D^3 h_{\mathbf{0}}, \Delta_3^a \rangle_3 + o(1) \to \frac{1}{6} \langle D^3 h_{\mathbf{0}}, \Delta_3 \rangle_3 \end{aligned}$$

and

$$\begin{aligned} \frac{1}{2} a^2 \int_C \tilde{D}^2 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\, dP &= \frac{1}{2} a^2 E\{\tilde{D}^2 h_{\mathbf{0}}(\overline{\mathbf{X}}_t)\} + o(1) \\ &= \frac{1}{2} \langle D^2 h_{\mathbf{0}}, \Sigma \rangle_2 a + \frac{1}{2} \langle D^2 h_{\mathbf{0}}, \Delta_2^a \rangle_2 + o(1). \end{aligned}$$

By Theorem 2, $\langle D^2 h_{\mathbf{0}}, \Delta_2^a \rangle_2$ is bounded and if $\langle \mathbf{c}, \mathbf{X}_1 \rangle$ has a nonarithmetic distribution, it converges to $\langle D^2 h_{\mathbf{0}}, \Delta_2 \rangle_2$ as $a \to \infty$. (15) and (16) now follow from simple algebra. □

9. Remarks. In some cases, $Z_1, Z_2, \ldots$ may be bounded random variables, in which case some of the results may be sharpened slightly. Suppose that $Z_1, Z_2, \ldots$ are bounded and let

$$l = l_a = \inf\{n \geq 1 \colon P\{Z_n > a\} > 0\},$$

so that $t \geq l_a$ w.p.1 for all $1 \leq a < \infty$. Suppose next that there are $0 < \kappa$, $\gamma \leq 1$ for which $l \geq \kappa a^\gamma$ for all $1 \leq a < \infty$. For example, if $Z_n \leq n^2$, as in Example 2, then $l \geq \sqrt{a}$ for $1 \leq a < \infty$. Then the conclusion of Proposition 6 may be sharpened to

$$\int_{t \leq \eta a} \left| \langle \mathbf{b}, \overline{\mathbf{X}}_t \rangle \right|^r dP \leq \frac{C}{\kappa^\delta} v_q(\mathbf{b})^r \times \left(\frac{1}{a} \right)^{p(1-r/q)+\gamma\delta} \Delta(a)^{1-r/q},$$

where $\delta = (r/2 + r/q - 1)$; and (11) holds provided $q \geq 4(p-\gamma)/(p-2+\gamma)$.

Acknowledgment. Thanks to an anonymous referee for suggestions which simplified the proofs of the main results.

REFERENCES

ARAS, G. (1989). Second order estimation of the mean exponential survival time under random censorship. *J. Statist. Plann. Inference* **21** 3–17.

BAUM, L. and KATZ, M. (1965). Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.* **120** 108–123.

CHOW, Y. S., HSIUNG, C. and LAI, T. L. (1979). Extended renewal theory and moment convergence in Anscombe's theorem. *Ann. Probab.* **7** 304–318.

CHOW, Y. S. and LAI, T. L. (1978). Paley-type inequalities and convergence rates related to the law of the iterated logarithm and extended renewal theory. *Z. Wahrsch. Verw. Gebiete* **45** 1–19.

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

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory*. Springer, New York.

CHOW, Y. S. and YU, K. F. (1981). The performance of a sequential procedure for estimating the mean. *Ann. Statist.* **9** 184–188.

DOOB, J. L. (1953). *Stochastic Processes*.Wiley, New York.

EDWARDS, C. (1973). *Advanced Calculus*. Academic, New York.

HAGWOOD, C. and WOODROOFE, M. (1982). On the expansion for expected sample size in nonlinear renewal theory. *Ann. Probab.* **10** 844–848.

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

MARTINSEK, A. (1983). Second order approximation to the risk of a sequential procedure. *Ann. Statist.* **11** 827–836.

SRIRAM, T. N. (1990). Sequential estimation of the ratio of normal parameters. *J. Statist. Plann. Inference* **26** 305–324.

WOODROOFE, M. (1977). Second order approximation for sequential point and interval estimation. *Ann. Statist.* **5** 985–995.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

ZHANG, C. (1988). A nonlinear renewal theory. *Ann. Probab.* **16** 793–824.

DEPARTMENT OF STATISTICS AND
APPLIED PROBABILITY
UNIVERSITY OF CALIFORNIA
AT SANTA BARBARA
SANTA BARBARA, CALIFORNIA

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1027