The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 2179–2183

ON THE FIRST TIME $|S_n| > cn^{\frac{1}{2}}$ (1)

By Michael Woodroofe

University of Michigan

1. Introduction. Let $X_1, X_2, \cdots$ be an infinite sequence of independent, identically distributed (i.i.d.) random variables having a finite mean μ and a finite, positive variance σ^2 and consider the stopping time N defined by

$$N = \text{least } n \geqq 1 \text{ for which } |S_n| > cn^{\frac{1}{2}} \text{ or } +\infty \text{ if no such } n \text{ exists,} \tag{1.1}$$

where c is a positive constant and $S_n = X_1 + \cdots + X_n$, $n \geqq 1$. Obviously, $N < \infty$ w.p. one (by the Strong Law of Large Numbers and the Law of the Iterated Logarithm), but if $\mu = 0$, then $E(N) < \infty$ if and only if $c^2 < \sigma^2$ ([1], [3]). Here we will consider the case $c^2 > \sigma^2$ and will investigate the rate at which $E(N)$ diverges to infinity as $\mu \to 0$. Our results assert the existence of positive constants b_1, b_2, γ_1, and γ_2 for which $0 < \gamma_1 < \gamma_2 < 1$ and

$$b_1 |\mu|^{-(1+\gamma_1)} \leqq E(N) \leqq b_2 |\mu|^{-(1+\gamma_2)} \tag{1.2}$$

for all sufficiently small values of μ. The constants b_1 and γ_1 depend only on c^2 and σ^2 and exist when $c^2 > 2\sigma^2$; the constants b_2 and γ_2 depend also on the distribution of $(X_i - \mu)/\sigma$ and require higher moments. Explicit values are given for all constants, and it is shown that γ_1 may be made arbitrarily close to one by taking c sufficiently large.

The left side of (1.2) is established in Section 2 and the right side in Section 3. An application to testing the sign of a bias is given in Section 4.

2. The lower bound. Throughout this section and the next we will assume the X's to be i.i.d. with mean μ and finite, positive variance σ^2. We begin with a variant on Wald's Lemma.

Lemma 2.1. *Let* $0 < \alpha \leqq 1$ *and let* $\beta = 1 - \alpha$; *then*

$$E(N^{-\beta} S_N{}^2) \leqq 4c|\mu|(1+2\alpha)^{-1} E(N^{\frac{1}{2}+\alpha}) + \alpha^{-1}(\sigma^2 + \mu^2) E(N^\alpha).$$

Proof. Without loss of generality, we may assume that $E(N^\alpha) < \infty$, in which case

$$n^{-\beta} \int_{N>n} S_n{}^2 \, dP \leqq c^2 n^\alpha P(N > n) \to 0 \tag{2.1}$$

as $n \to \infty$. Now for any $k \geqq 2$ we may write

$$\int_{N \leqq k} N^{-\beta} S_N{}^2 \, dP = \int_{N=1} S_1{}^2 \, dP + \sum_{n=2}^k [n^{-\beta} \int_{N>n-1} S_n{}^2 \, dP - n^{-\beta} \int_{N>n} S_n{}^2 \, dP]. \tag{2.2}$$

Received November 3, 1969; revised June 26, 1970.
[1] Research supported by NSF Grant GP-11769.

Moreover, for $2 \leqq n \leqq k$, we have

$$n^{-\beta} \int_{N>n-1} S_n^2 \, dP \leqq (n-1)^{-\beta} \int_{N>n-1} S_{n-1}^2 \, dP + [2c|\mu|(n-1)^{\frac{1}{2}} n^{-\beta} + (\sigma^2+\mu^2) n^{-\beta}] P(N \geqq n) \tag{2.3}$$

since the event $N > n-1$ is independent of X_n and implies $|S_{n-1}| \leqq c(n-1)^{\frac{1}{2}}$. Substituting (2.3) into (2.2), letting $k \to \infty$, and using (2.1), we now obtain

$$2c|\mu| \sum_{n=2}^{\infty} (n-1)^{\frac{1}{2}} n^{-\beta} P(N \geqq n) + (\sigma^2+\mu^2) \sum_{n=1}^{\infty} n^{-\beta} P(N \geqq n) \tag{2.4}$$

as an upper bound for $E(N^{-\beta} S_N^2)$. The lemma now follows on writing $P(N \geqq n) = \sum_{j=n}^{\infty} P(N=j)$ and reversing the orders of summation in (2.4). (The cases $\alpha \leqq \frac{1}{2}$ and $\alpha > \frac{1}{2}$ should be considered separately in this step.)

We should perhaps remark that in the special case $\mu = 0$, Lemma 2.1 may be used to prove that $\alpha c^2 \geqq \sigma^2$ implies $E(N^\alpha) = \infty$, thus extending the results of [1]. Indeed, the finiteness of $E(N^\alpha)$ would imply $E(N^{-\beta} S_N^2) > c^2 E(N^\alpha)$, which contradicts Lemma 2.1 if $\mu = 0$ and $\alpha c^2 \geqq \sigma^2$.

COROLLARY 2.1. *If* $0 < \alpha \leqq \frac{1}{2}$, $\alpha c^2 > \sigma^2$, *and* $2\mu^2 \leqq \alpha c^2 - \sigma^2$, *then*

$$|\mu| E(N^{\alpha+\frac{1}{2}}) \geqq (1+2\alpha)(\alpha c^2 - \sigma^2)/8\alpha c, \qquad \mu \neq 0. \tag{2.5}$$

PROOF. Again we may assume that $E(N^\alpha) < \infty$. Then taking expectations in the inequality $c^2 N^\alpha < N^{-\beta} S_N^2$, which is true by (1.1), we obtain

$$c^2 E(N^\alpha) \leqq 4c|\mu|(1+2\alpha)^{-1} E(N^{\alpha+\frac{1}{2}}) + \alpha^{-1}(\sigma^2+\mu^2) E(N^\alpha). \tag{2.6}$$

The corollary now follows on subtracting $\alpha^{-1}(\sigma^2+\mu^2)E(N^\alpha)$ from both sides of (2.6) and observing that $E(N^\alpha) \geqq 1$.

COROLLARY 2.2. *Let* $c^2 > 2\sigma^2$ *and let* α *be any real number for which* $0 < \alpha < \frac{1}{2}$ *and* $\alpha c^2 > \sigma^2$, *e.g.* $4\alpha = 1 + 2\sigma^2 c^{-2}$. *Define* b_1 *and* γ_1 *by*

$$1+\gamma_1 = 2(1+2\alpha)^{-1} \quad \text{and} \quad b_1 = b^{2/(1+2\alpha)} \tag{2.7}$$

where b *is the right side of* (2.5). *Then* $|\mu|^{1+\gamma_1} E(N) \geqq b_1$ *for* $0 < \mu^2 \leqq (\alpha c^2 - \sigma^2)/2$.

PROOF. Apply Jensen's Inequality to the left side of (2.5).

COROLLARY 2.3. *Given* $0 < \varepsilon < 1$, *there exist constants* $c(\varepsilon)$ *and* $b(\varepsilon)$ *for which* $|\mu|^{2-\varepsilon} E(N) \geqq b(\varepsilon) > 0$ *for* $0 < |\mu| \leqq 1$ *if* $c \geqq c(\varepsilon)$.

PROOF. Let α_0 be the solution of $2/(1+2\alpha) = 2 - \varepsilon$ and let $c(\varepsilon)$ be the solution of $\alpha_0 c^2 - \sigma^2 = 2$. Taking $\alpha = \alpha_0$ in Corollary 2.2 then yields Corollary 2.3.

3. The upper bound.

LEMMA 3.1. *Let* $c' = 2(2^{\frac{1}{2}})c$ *and let* $\beta_0 = 2\Phi(c'/\sigma) - 1$ *where* Φ *denotes the standard normal distribution function. If* $\beta > \beta_0$, *then there is a constant* $M = M_\beta$ *for which*

$$P(N > 2^n) \leqq M\beta^{n+1} \tag{3.1}$$

for $n = 0, 1, 2, \cdots$.

PROOF. Let $T_0 = S_1$ and for $k \geqq 1$ let $T_k = S_{2^k} - S_{2^{k-1}}$; then

$$P(N > 2^n) \leqq \prod_{k=0}^{n} P(|T_k| \leqq c'(2^{\frac{1}{2}})^{k-1})$$

for $n \geqq 0$. Now given $\varepsilon > 0$, there is an integer $m = m(\varepsilon)$ for which $|P(T_k - \mu 2^{k-1} \leqq x\sigma(2^{\frac{1}{2}})^{k-1}) - \Phi(x)| \leqq \varepsilon/2$ for all x if $k \geqq m$. Taking $\varepsilon = \beta - \beta_0$ and using the fact that $\Phi(x+y) - \Phi(-x+y) \leqq 2\Phi(x) - 1$ for all $x \geqq 0$ and all y, we therefore obtain $P(|T_k| \leqq c'(2^{\frac{1}{2}})^{k-1}) \leqq \beta$ for $k \geqq m$. The lemma follows immediately with $M = \beta^{-m}$.

COROLLARY 3.1. *If* $\beta > \beta_0$, $\alpha > 0$, *and* $\beta 2^{\alpha} < 1$, *then* $E(N^{\alpha}) \leqq M(1 - \beta 2^{\alpha})^{-1} = M_1$, *say.*

PROOF.

$$\begin{aligned} E(N^{\alpha}) &= 1 + \sum_{n=0}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} k^{\alpha} P(N = k) \\ &\leqq 1 + \sum_{n=0}^{\infty} 2^{\alpha(n+1)} P(N > 2^n) \leqq M_1. \end{aligned}$$

The constant M appearing in (3.1) depends not only on β but also on the rate at which the distribution of $(S_n - n\mu)/\sigma n^{\frac{1}{2}}$ approaches normality, and therefore on the distribution function F of $(X_i - \mu)/\sigma$. It follows easily from the Berry–Esseen Theorem, however, that if F satisfies (3.2) below, then M may be chosen to depend only on d.

LEMMA 3.2. *Let* $Y = \sup_{n \geqq 1} (n^{-1})\,(X_n - \mu)^4/\sigma^4$; *if for some* $\theta > 1$

$$E[|X_i - \mu|^{4\theta} \sigma^{-4\theta}] \leqq d < \infty, \tag{3.2}$$

then $E(Y) \leqq 1 + d\theta(1-\theta)^{-2} = d_1$, *say.*

PROOF.

$$E(Y) = \int_0^{\infty} P(Y > y)\, dy \leqq 1 + d\theta(1-\theta)^{-1} \int_1^{\infty} y^{-\theta}\, dy = d_1.$$

COROLLARY 3.2. *Let* $N_k = \min(k, N)$; *then for each* $k \geqq 1$, *we have* $E[(X_{N_k} - \mu)^2] \leqq \sigma^2 d_1^{\frac{1}{2}} E(N_k)^{\frac{1}{2}}$.

PROOF. Write $(X_{N_k} - \mu)^2 \leqq \sigma^2 N_k^{\frac{1}{2}} Y^{\frac{1}{2}}$ and apply the Scharz Inequality.

THEOREM 3.1. *Let* $\beta > \beta_0$, $\alpha > 0$, *and* $\beta 4^{\alpha} < 1$. *If* (3.2) *is satisfied, then there is a constant* M_2 *depending only on* c, d, α, β, θ, *and* σ^2 *for which* $E(N^2) \leqq M_2 |\mu|^{-4/1+\alpha}$ *for* $|\mu| \leqq 1$.

PROOF. Let $N_k = \min(k, N)$; then by (1.1) we have $|S_{N_k - 1}| \leqq cN_k^{\frac{1}{2}}$ which implies

$$|\mu|\, N_k \leqq cN_k^{\frac{1}{2}} + |S_{N_k} - \mu N_k| + |X_{N_k} - \mu| + |\mu|.$$

Applying the Minkowski Inequality, Theorem 1 of [1], and Corollary 3.2, we therefore have

$$|\mu|\, E(N_k^2)^{\frac{1}{2}} \leqq d_2\, E(N_k)^{\frac{1}{2}} \tag{3.3}$$

where $d_2 = c + \sigma(1 + d_1) + 1$. Now by Corollary 3.1 and the remark which follows it, we have $E(N_k^{2\alpha}) \leqq E(N^{2\alpha}) \leqq M_1$ where M_1 depends only on α, β, and d. Therefore,

$$E(N_k) \leqq E(N_k^{2\alpha})^{\frac{1}{2}} E(N_k^{2(1-\alpha)})^{\frac{1}{2}} \leqq M_1^{\frac{1}{2}} E(N_k^2)^{\frac{1}{2}(1-\alpha)}. \tag{3.4}$$

Combining (3.3) and (3.4) now yields $\mu^2 E(N_k^2)^{\frac{1}{2}(1+\alpha)} \leqq d_2^2 M_1^{\frac{1}{2}}$ from which the theorem follows, with $M_2 = (d_2^4 M_1)^{1/1+\alpha}$, on letting $k \to \infty$.

COROLLARY 3.3. *Let $\beta = (1+\beta_0)/2$ and let α be the solution of $\beta 4^\alpha = (1+\beta)/2$. Define b_2 and γ_2 by $b_2 = M_2^{\frac{1}{2}}$ and $1+\gamma_2 = 2/(1+\alpha)$; then*

$$E(N) \leqq b_2 |\mu|^{-(1+\gamma_2)} \quad \textit{for} \quad |\mu| \leqq 1.$$

PROOF. This follows trivially from Theorem 3.1 and the Scharz Inequality.

4. Testing the sign of a bias. Let $X_1, X_2, \cdots$ be i.i.d. $N(\theta, 1)$ where θ is an unknown parameter whose sign we wish to determine. We adopt what might be called a quasi-Bayesian approach to the problem. Specifically, we assume a prior density λ for θ, with posteriors denoted by $\lambda_n = \lambda(\cdot 1 X_1, \cdots, X_n)$ and continue sampling until

$$P(\theta < 0 \mid X_1, \cdots, X_n) = \textstyle\int_{-\infty}^{0} \lambda_n(\theta)\, d\theta \tag{4.1}$$

is either $\leqq \alpha$ or $\geqq 1-\alpha$ where α is the admissible probability of error, e.g. $\alpha = 0.01$ or 0.05; we then stop and decide that θ is negative if (4.1) is $\geqq 1-\alpha$. A Bayesian might object to our procedure on the grounds that we have not found the Bayes' solution with respect to some loss structure, while a non-Bayesian would probably object to the use of a prior distribution. Nevertheless, the procedure seemed simple and intuitive, at least to us, and we decided to investigate its properties.

Taking λ to be normal with mean μ and variance $1/h$, $h > 0$, we find that λ_n is again normal with mean $\mu_n = (\mu h + S_n)/(h+n)$ and variance $1/(h+n)$. It follows easily that our procedure calls for taking

$$N^* = \text{least } n \geqq 1 \text{ for which } |\mu h + S_n| \geqq c(h+n)^{\frac{1}{2}} \tag{4.2}$$

observations where c is the solution of $\Phi(-c) = \alpha$, and it is tacitly assumed that $|\mu h| < ch^{\frac{1}{2}}$. (4.2) of course, resembles (1.1) quite closely and, in fact, reduces to (1.1) if we formally set $h = 0$. Moreover, the development of Section 2 applies to N^* as well as N (with only trivial modifications) to yield

$$E_\theta(N) \geqq (c^2-2)/(4c|\theta|), \qquad 0 < |\theta| < (c^2-2)/4, \tag{4.3}$$

where E_θ denotes conditional expectation given θ. (See Corollary 2.1 with $\alpha = \frac{1}{2}$.) Thus, if $c^2 > 2$, and, in particular, for $\alpha \leqq 0.05$, we have

$$E(N^*) = \textstyle\int E_\theta(N^*)\lambda(\theta)\, d\theta = +\infty,$$

so that, from a Bayesian point of view, our procedure takes too many observations. This is, of course, a reflection of the fact that we have not assumed a loss structure which incorporates a cost of sampling. On the other hand, our procedure is computationally quite simple, whereas the Bayes (dynamic programming) solution is not, and moreover, the expected sample size of our procedure compares favorably with that of other procedures. Indeed, writing $N = N(c)$ and $N^* = N^*(c)$ in (1.1) and (4.2) respectively, we have $N^*(c) \leqq N(d)$ where $d = ch^{\frac{1}{2}} + h|\mu|$. It follows from Theorem 3.1 that

$$\lim_{\theta\to 0} \theta^2 E_\theta(N^*) = 0.$$

This compares favorably with the expected sample size for the procedure proposed by Darling and Robbins ([2]) which is $0(\theta^{-2}\log\log|\theta|^{-1})$ as $\theta \to 0$. Of course, the Darling–Robbins procedure has some properties which ours does not, in particular, P_θ (correct decision) $\geqq 1-\alpha$ for all $\theta \neq 0$.

Acknowledgment. I wish to thank my colleague Professor Norman Starr for suggesting the procedure of Section 4 to me and for his continuing interest in my research. I would also like to thank the referee for bringing reference [3] to my attention.

REFERENCES

[1] Chow, Y. S., Robbins, H. and Teicher, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

[2] Darling, D. and Robbins, H. (1967). Iterated logarithm inequalities. *Proc. Nat. Acad. Sci. USA* **57** 1188–1192.

[3] Gundy, R and Siegmund, B. (1968). On a stopping rule and the central limit theorem. *Ann. Math. Statist.* **38** 1915–1917.