The Annals of Probability
1983, Vol. 11, No. 1, 123–142

CONVERGENCE RATES RELATED TO THE STRONG LAW OF LARGE NUMBERS[1]

By James Allen Fill

Stanford University

Let $X_1, X_2, \cdots$ be independent random variables with common distribution function F, zero mean, unit variance, and finite moment generating function, and with partial sums S_n. According to the strong law of large numbers,

$$p_m \equiv P\left\{\frac{S_n}{n} > c_n \quad \text{for some} \quad n \geq m\right\}$$

decreases to 0 as m increases to ∞ when $c_n \equiv c > 0$. For general c_n's the Hewitt-Savage zero-one law implies that either $p_m = 1$ for every m or else $p_m \downarrow 0$ as $m \uparrow \infty$. Assuming the latter case, we consider here the problem of determining p_m up to asymptotic equivalence.

For constant c_n's the problem was solved by Siegmund (1975); in his case the rate of decrease depends heavily on F. In contrast, Strassen's (1967) solution for smoothly varying $c_n = o(n^{-2/5})$ is independent of F.

We complete the solution to the convergence rate problem by considering c_n's intermediate to those of Siegmund and Strassen. The rate (Theorem 1.1) in this case depends on an ever increasing number of terms in the Cramér series for F the more slowly c_n converges to zero.

1. Objective. Throughout this paper we suppose that $X_1, X_2, \cdots$ is a sequence of independent random variables with common distribution function F. Denote the random walk of partial sums by $S_n = \sum_{k=1}^n X_k$, with $S_0 \equiv 0$. The distribution F is assumed to be standardized in the sense that $EX = 0$, $\text{Var } X = 1$, where, to facilitate notation, we have introduced another random variable X distributed according to F. Assume throughout that the moment generating function (mgf) $E \exp(\xi X)$ for F is finite for ξ in some neighborhood of 0 and write

$$K(\xi) = \log(Ee^{\xi X}) \tag{1.1}$$

for the cumulant generating function (cgf). This assumption, which restricts attention to the so-called mgf case, is stronger than required for the more elementary results (for example, the laws of large numbers) discussed in this paper. However, the main result of this work deals only with the mgf case.

Our main goal will be to estimate the probability of the event

$$\left\{\frac{S_n}{n} > c_n \quad \text{for some} \quad n \geq m\right\}$$

when m is large for a specified sequence $c = (c_n)$ of positive numbers. It is natural to think of the sequence c as a "boundary" on the growth of the sequence (S_n/n) of sample means as the "time" n increases. Often it will be more convenient to deal either with the

Received September 1981; revised June 1982.

[1] Supported in part by ONR Grant N00014-77-C-0306 and NSF Grants SOC 72-05228 A04 and MCS 72-04364 A04 from the University of Chicago, where the author's research was carried out.

AMS 1970 *subject classifications.* Primary 60F15; secondary 60F10, 60G50, 60J15.

Key words and phrases. Random walk, laws of large numbers, convergence rates, boundary crossing probabilities, invariance principles, large deviations, law of iterated logarithm, Brownian motion.

standardized process $(S_n/n^{1/2})$ or with the random walk S. The corresponding boundaries will be denoted as follows:

TABLE 1

Process	Value at time n	Boundary	Value at time n
sample means	$\frac{S_n}{n}$	c	c_n
standardized	$\frac{S_n}{\sqrt{n}}$	Ψ	$\Psi(n) = \sqrt{n}c_n$
random walk	S_n	g	$g(n) = \sqrt{n}\psi(n)$

We write Z for a standard normal random variable. Our result can be stated roughly as follows.

THEOREM 1.1. *If* $g:(0, \infty) \to (0, \infty)$ *has a smooth derivative and satisfies the monotonic growth conditions*

$$\frac{g(t)}{t^{1/2+\delta}} \uparrow \tag{1.2}$$

for some $\delta > 0$ *and*

$$\frac{g(t)}{t} \downarrow 0, \tag{1.3}$$

then with $\psi(t) = g(t)/\sqrt{t}$

$$p_m \sim I_m \equiv \frac{g'(m)}{\sqrt{m}\Psi'(m)} \cdot P\{S_m > g(m)\}. \square \tag{1.4}$$

The rate of convergence of $P\{S_m > g(m)\}$ is given by Cramér's Theorem (2.7) below.

2. Weak law of large numbers. Although the present work concerns itself with convergence rates related to the strong law of large numbers, we begin with an examination of convergence rates related to the weak law. There are two reasons for this review: the weak-law results (1) provide motivation for, and (2) are used in the proofs of, the corresponding strong-law theorems.

The *convergence rate problem for the weak law of large numbers* (WLLN) is to determine

$$P\left\{\left|\frac{S_m}{m}\right| > c\right\}, \tag{2.1}$$

or equivalently, the upper tail probability

$$P\left\{\frac{S_m}{m} > c\right\}, \tag{2.2}$$

up to a factor $(1 + o(1))$.

A more general problem is to determine the asymptotic behavior of

$$P\left\{\frac{S_m}{m} > c_m\right\} \tag{2.3}$$

for an arbitrary sequence of positive numbers c_m. In view of the central limit theorem (CLT) for S, it is convenient to express (2.3) in the standardized form

$$P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\}. \tag{2.3a}$$

Indeed, the CLT states that when $\Psi(m) \equiv \Psi_0$ is constant,

$$P\left\{\frac{S_m}{\sqrt{m}} > \Psi_0\right\} \to P\{Z > \Psi_0\}. \tag{2.4}$$

The CLT is thus an *invariance principle* in the sense that the right side here is independent of F.

The Berry-Esséen Theorem (see Feller, 1971, Theorem XVI, 5.1) bounds the error in approximating the left side of (2.4) by the right, uniformly in Ψ_0. As a particular consequence,

$$P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} \to 0 \tag{2.5a}$$

if and only if

$$\Psi(m) \to \infty. \tag{2.5b}$$

The general *problem of convergence rates related to the* WLLN is to determine (2.3) up to a factor $(1 + o(1))$ when (2.5) is in force.

Return to the case $\Psi(m) \equiv cm^{1/2}$ of (2.1). Any unified theory for handling this case must require c to be small in some sense. For example, if S has symmetric Bernoulli components, i.e., if X assumes the values ± 1 with probability ½ each, then for any $c \geq 1$

$$P\left\{\frac{S_m}{m} > c\right\} = 0 \quad \text{for every} \quad m. \tag{2.6}$$

Making precise the condition that c be small, Bahadur and Ranga Rao (1960) solved the WLLN convergence rate problem. The criterion of smallness for c is that there exist a unique nonzero value ξ_1, necessarily positive, for which $K(\xi_1) = c\xi_1$. In marked contrast to the invariance principle (2.4), the rate of convergence in (2.5a) in this case depends heavily on F.

The case $\Psi(m) = o(m^{1/2})$ with $\Psi(m) \to \infty$, intermediate to the CLT case of constant Ψ and the WLLN case $\Psi(m) = cm^{1/2}$, was resolved by Cramér (1938):

$$P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} = (1 + o(1))P\{Z > \Psi(m)\} \cdot \exp\left[\Psi^2(m)\frac{\Psi(m)}{\sqrt{m}}\lambda\left(\frac{\Psi(m)}{\sqrt{m}}\right)\right]. \tag{2.7}$$

Here

$$\lambda(\xi) = \textstyle\sum_{k\geq 0} \lambda_k \xi^k$$

is a certain power series, the so-called Cramér series for F, which converges for ξ in a neighborhood of 0. For each k the coefficient λ_k depends on the moments of F of orders up to and including $k + 3$; for example,

$$\lambda_0 = \frac{1}{6}EX^3, \quad \lambda_1 = \frac{1}{24}EX^4 - \frac{1}{8} - \frac{1}{8}(EX^3)^2.$$

For a precise definition of λ, see (6.1).

For the normal tail probability on the right in (2.7) we have the standard estimate

$$P\{Z > \Psi(m)\} \sim [\sqrt{2\pi}\Psi(m)]^{-1}\exp\left[-\frac{1}{2}\Psi^2(m)\right] \tag{2.8}$$

(as usual, $a_m \sim b_m$ means the same as $a_m = (1 + o(1))b_m$). On a logarithmic scale the correction in (2.7) to the normal approximation becomes negligible:

$$\log P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} \sim -\frac{1}{2}\Psi^2(m).$$

Even on the probability scale of (2.7), the correction is unnecessary if Ψ does not grow too rapidly:

$$\text{(2.9a)} \qquad \Psi(m) = o(m^{1/6}) \quad \text{implies} \quad P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} \sim P\{Z > \Psi(m)\}.$$

If Ψ is allowed to increase somewhat more quickly, the correction requires only the constant term λ_0 from the Cramér series:

$$\text{(2.9b)} \quad \Psi(m) = o(m^{1/4}) \quad \text{implies} \quad P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} \sim P\{Z > \Psi(m)\} \cdot \exp\left[\lambda_0 \frac{\Psi^3(m)}{\sqrt{m}}\right].$$

In general, if $\Psi(m) \to \infty$ and $\Psi(m) = o(m^{1/2-\eta})$ with $0 < \eta \le ⅓$, then only the moments of F of orders up to and including $\lceil 1/\eta \rceil - 1$ need be known to identify the convergence rate (2.7). ($\lceil x \rceil$, the "ceiling" of x, denotes the smallest integer at least as large as x.)

The transitions in form from the CLT to Cramér's result and from Cramér's result to the WLLN solution are smooth. When Ψ is nearly constant, as in (2.9a), Cramér's result is an invariance principle of the same form as the CLT. When, at the other extreme, $\Psi(m)$ grows nearly as quickly as $m^{1/2}$, the convergence rate (2.7) depends heavily on F. The Bahadur—Ranga Rao result for $\Psi(m) = cm^{1/2}$ can be stated in the form

$$\text{(2.10)} \qquad P\left\{\frac{S_m}{\sqrt{m}} > \Psi(m)\right\} \sim (1+\beta)P\{Z > \Psi(m)\}\exp\left[\Psi^2(m)\frac{\Psi(m)}{\sqrt{m}}\lambda\left(\frac{\Psi(m)}{\sqrt{m}}\right)\right]$$

as $m \to \infty$, where β depends on c and heavily on F but vanishes in the limit as $c \to 0$. Thus (2.7) may be regarded as the limiting form of (2.10) when $c \to 0$.

3. Strong law of large numbers. According to the strong law of large numbers (SLLN), $S_n/n \to 0$ with probability 1; equivalently, for any constant $c > 0$

$$\text{(3.1)} \qquad P\left\{\left|\frac{S_n}{n}\right| > c \quad \text{for some} \quad n \ge m\right\} \downarrow 0 \quad \text{as} \quad m \uparrow \infty.$$

We have the decomposition

$$\text{(3.2)} \qquad \begin{aligned} &P\left\{\frac{S_n}{n} > c \quad \text{for some} \quad n \ge m\right\} + P\left\{-\frac{S_n}{n} > c \quad \text{for some} \quad n \ge m\right\} \\ &- P\left\{\frac{S_p}{p} > c \quad \text{for some} \quad p \ge m \quad \text{and} \quad -\frac{S_q}{q} > c \quad \text{for some} \quad q \ge m\right\} \end{aligned}$$

for the probability in (3.1). In Fill (1980, Section 4.3) it is shown that the last term in (3.2) is asymptotically negligible when compared to the sum of the first two terms. So we consider the one-sided version

$$\text{(3.3)} \qquad P\left\{\frac{S_n}{n} > c \quad \text{for some} \quad n \ge m\right\} \downarrow 0 \quad \text{as} \quad m \uparrow \infty$$

of (3.1).

A more general problem is to determine the asymptotic behavior of

$$\text{(3.4)} \qquad p_m \equiv P\left\{\frac{S_n}{\sqrt{n}} > \Psi(n) \quad \text{for some} \quad n \ge m\right\}$$

for an arbitrary sequence of positive numbers $\Psi(n)$. No matter what the sequence Ψ,

$$\text{(3.5)} \qquad p_m \downarrow p \equiv P\left\{\frac{S_n}{\sqrt{n}} > \Psi(n) \text{ i.o. as } n \to \infty\right\}.$$

It follows from the Hewitt-Savage zero-one law (Feller, 1971, Theorem IV, 6.3) that

$$p = 0 \quad \text{or} \quad p = 1. \tag{3.6}$$

The case $p = 1$ is trivial from the convergence rate viewpoint, for then $p_m = 1$ for every m. The classification of boundaries Ψ according to the dichotomy (3.6) is effected by the Kolmogorov-Petrovski-Erdös-Feller integral test (cf. Jain, Jogdeo, and Stout, 1975);

KPEF integral test. If $0 < \Psi \uparrow$, then

$$p = \begin{matrix} 0 \\ 1 \end{matrix} \quad \text{according as} \quad \int^{\infty} \frac{\Psi(t)}{t} e^{-\psi^2(t)/2} \, dt \begin{matrix} < \\ = \end{matrix} \infty. \square \tag{3.7}$$

Note that the criterion (3.7) is, like its weak-law analogue (2.5), an invariance principle.

There are obvious counterparts to (3.4–3.7) for Brownian motion (BM). In fact, (3.7) for S is most easily proved from (3.7) for BM by showing that S can be closely approximated by a BM (cf. Komlós, Major and Tusńady, 1976, Theorem 1).

In the interesting case that $p = 0$ in (3.5) we say that g is an *upper class boundary* for the random walk S and write $g \in U(S)$. (Otherwise g is a *lower class boundary* and $g \in L(S)$.) If $\Psi \uparrow$, the KPEF test allows us to write $g \in U$ indifferently for $g \in U(S)$ (for any S). A similar comment applies to the notation $g \in L$.

The test (3.7) gives rise to the celebrated law of the iterated logarithm (LIL), which quite precisely describes the "interface" between U and L. Here L_k denotes k iterations of the natural log function $L \equiv \log$.

Law of the iterated logarithm. If

$$g(t) = \left[2t\left(L_2 t + \frac{3}{2} L_3 t + L_4 t + \cdots + L_{\rho-1} t + (1 + \beta) L_\rho t \right) \right]^{1/2} \tag{3.8}$$

with $\rho > 3$, then

$$g \in \begin{matrix} U \\ L \end{matrix} \quad \text{according as} \quad \beta \begin{matrix} > \\ \leq \end{matrix} 0. \square \tag{3.9}$$

The general *problem of convergence rates related to the* SLLN is to determine p_m up to a factor $(1 + o(1))$ when $g \in U$. Let $T_m = \inf\{n : n \geq m, S_n > g(n)\}$, the inf of the empty set being $+\infty$. Then $T_m \geq m$, and

$$p_m = P\{S_n > g(n) \quad \text{for some} \quad n \geq m\} = P\{T_m < \infty\}$$

admits the decomposition

$$p_m = P\{T_m = m\} + P\{m < T_m < \infty\} = P\{S_m > g(m)\} + P\{m < T_m < \infty\}. \tag{3.10}$$

The convergence rate for the first term is known from studying the WLLN; the second is new. For a simple lower bound we have

$$p_m \geq P\{S_m > g(m)\}. \tag{3.11}$$

Siegmund (1975) used the relation (3.10), together with the Bahadur—Ranga Rao estimate for the first term and his own analysis of the second, to solve the convergence rate problem in the SLLN case $g(t) = ct$. Strassen (1967) solved the problem for boundaries $g \in U$ not too far from the $U \backslash L$ interface (3.8)—roughly speaking, for $g(t) = o(t^{3/5})$ as $t \to \infty$. The main contribution of this paper is to complete the solution to the convergence rate problem by bridging the gap between Strassen's boundaries and Siegmund's; see Section 6.

We set the stage by reviewing, in the next two sections, the results of Strassen and Siegmund.

4. Strassen's result. We recall the omnibus restriction to the mgf case. Modulo a precise definition of the adjective "smooth", Strassen's result (1967, Theorem 1.4) can be stated as follows:

THEOREM 4.1. (Strassen). *If $g \in U$ has a smooth derivative, $0 < \Psi \uparrow$, and $g(t) \leq t^{3/5-\gamma}$ for some $\gamma > 0$, then*

$$p_m \sim J_m \equiv \int_m^\infty \frac{1}{\sqrt{2\pi}} \frac{g'(t)}{\sqrt{t}} e^{-\Psi^2(t)/2}\, dt. \square \tag{4.1}$$

One precise definition of "smooth" is that g be continuously differentiable and satisfy

$$g'(u) \sim g'(t) \quad \text{as} \quad u \sim t \to \infty.$$

REMARK 4.2. (a) Strassen used an intricate argument to show that for Brownian motion B

$$P\{B(t) > g(t) \quad \text{for some} \quad t \geq s\} \sim J_s \quad \text{as} \quad s \to \infty \tag{4.2}$$

and used this, along with approximation of S by B via Skorohod embedding (see Breiman, 1968), to deduce his invariance principle (4.1). In light of (2.9a) one might expect that the restriction on the growth of g could be eased to $g(t) = o(t^{2/3})$. This can in fact be done (Fill, 1980, Theorem 3.2.1).

(b) Theorem 4.3 in Strassen (1967), a lemma to (4.1) credited by Strassen to F. Jonas, is in error. As noted by Sawyer (1972), the Skorohod embedding time for X may not have finite mgf even though X does. Theorem 3.2.1 in Fill (1980) repairs the proof of (and yields a result somewhat better than) Theorem 4.1 by using the dyadic quantile-transformation approximation of S by B due to Komlós, Major, and Tusnády (1975; 1976) instead of Skorohod embedding.

(c) If we assume $g(t)/t^{3/5} \downarrow$, then

$$g'(t) \asymp \frac{\Psi(t)}{\sqrt{t}}.$$

(The relation $a(t) \asymp b(t)$ means that $a(t) = O(b(t))$ and $b(t) = O(a(t))$.) In fact,

$$\frac{1}{2} \leq g'(t) \bigg/ \left(\frac{\Psi(t)}{\sqrt{t}} \right) \leq \frac{3}{5}.$$

Thus

$$p_m \asymp \int_m^\infty \frac{\Psi(t)}{t} e^{-\Psi^2(t)/2}\, dt,$$

which is the tail integral in the KPEF test (3.7). □

EXAMPLE 4.3. Define $g \in U$ at the $U \backslash L$ interface according to (3.8), with $\rho > 3$ and $\beta > 0$. Then

$$P\{S_m > g(m)\} \sim [2\sqrt{\pi}(Lm)(L_2 m)^2 (L_3 m) \cdots (L_{\rho-2} m)(L_{\rho-1} m)^{1+\beta}]^{-1} \tag{4.3}$$

which is of *much* smaller order of magnitude than

$$p_m \sim [2\sqrt{\pi}\beta(L_{\rho-1} m)^\beta]^{-1}. \tag{4.4}$$

In contrast we shall see for Siegmund's boundaries and for those of Theorem 6.1 (and also for Strassen's when g is not too close to L) that

$$p_m \asymp P\{S_m > g(m)\}. \tag{4.5}$$

The extreme reluctance with which (4.4) tends to zero is a well-known phenomenon connected with the LIL. Were g only slightly smaller we would have $g \in L$ and hence $p_m = 1$ for every m. □

5. Siegmund's result. In stating Siegmund's solution to the SLLN case $g(t) = ct$ we assume that $c > 0$ is sufficiently small. The criterion of smallness, explained in Section 2 above and detailed in Fill (1980, Section 3.4), is the same as for the Bahadur-Ranga Rao WLLN result. We further assume that if F is a lattice distribution with span h, then c is a point in that lattice.

THEOREM 5.1. (Siegmund). *If $g(t) = ct$ with $c > 0$ as above, then there is a constant $\gamma > 0$ for which*

$$p_m \sim (1 + \gamma)P\{S_m > g(m)\}. \square \tag{5.1}$$

REMARK 5.2. (a) Siegmund determined the constant γ explicitly and remarked that

$$\gamma \to 1 \quad as \quad c \to 0. \tag{5.2}$$

Nevertheless, for fixed c the constant γ, like the constant β and the series λ in (2.10), depends heavily on the component distribution F. So Siegmund's result, unlike Strassen's, is far from an invariance principle.

(b) Siegmund utilized the decomposition (3.10). In analyzing the second term he used the fundamental identity of sequential analysis, the large-deviation result of Bahadur and Ranga Rao, and some renewal-theoretic calculations. The same kind of approach is used in proving Lemma 9.1 to Theorem 6.1.

(c) For a generalization of Siegmund's theorem to linear boundaries g with nonzero intercept, see Fill (1980, Theorem 3.4.1). □

6. Completion of the solution to the convergence rate problem. The solution to the convergence rate problem in the mgf case is completed by Theorem 6.1 below, which overlaps somewhat with Strassen's theorem.

Recall that

$$K(\xi) = \log(Ee^{\xi X})$$

denotes the cumulant generating function corresponding to F. The so-called Cramér series

$$\lambda(\xi) = \textstyle\sum_{k=0}^{\infty} \lambda_k \xi^k$$

for F is defined implicitly for ξ near 0 by

$$\xi^3\lambda(\xi) = K(z) - z\xi + \tfrac{1}{2}\xi^2, \quad z(\equiv z(\xi)) \quad \text{given by} \quad K'(z) = \xi. \tag{6.1}$$

Let g: $(0, \infty) \to (0, \infty)$ and write

$$g(t) = \sqrt{t}\Psi(t).$$

Define

$$p_m = P\{S_n > g(n) \quad \text{for some} \quad n \geq m\}. \tag{6.2}$$

THEOREM 6.1. *Suppose that as $t \uparrow \infty$*

$$\frac{g(t)}{t^{1/2+\delta}} \uparrow \tag{6.3}$$

for some $0 < \delta < 1/2$ and

$$\frac{g(t)}{t} \downarrow 0. \tag{6.4}$$

If g is continuously differentiable and if for some $0 < r < 1$

(6.5) $$g'(u) \sim g'(t) \text{ when } t, u \to \infty \text{ with } t \le u \le t[1 + 1/\Psi^{2r}(t)],$$

then $g \in U$ *and*

(6.6) $$p_m \sim I_m \equiv \frac{g'(m)}{\sqrt{m}\Psi'(m)} P\{S_m > g(m)\} \quad as \quad m \to \infty. \square$$

REMARK 6.2. (a) That g belongs to U is an easy consequence of the KPEF test.

(b) The theorem's method of proof requires that g be kept away from the $U \backslash L$ interface and from linearity; hence the growth conditions (6.3) and (6.4).

(c) The rate of convergence of the factor $P\{S_m > g(m)\}$ to zero is given by Cramér's theorem (2.7) and (2.8). Thus in the present case the rate of convergence depends on F, but only through a (typically) finite number of cumulants of F.

(d) One might expect that a result like (4.1), but with a Cramer-like correction to the exponential factor, would hold. Indeed, Fill (1980, Lemma 3.2.4) shows that (6.6) can be recast in the form

(6.6a) $$p_m = \tilde{J}_m \equiv \int_m^\infty \frac{1}{\sqrt{2\pi}} \frac{g'(t)}{\sqrt{t}} e^{-\Psi^2(t)/2} \exp\left[\Psi^2(t) \frac{\Psi(t)}{\sqrt{t}} \lambda\left(\frac{\Psi(t)}{\sqrt{t}}\right)\right] dt.$$

In particular, (4.1) holds for $g(t) = o(t^{2/3})$.

(e) As in Siegmund's case, g increases rapidly enough that (4.5) holds (cf. Lemma 7.1(k)). $\square$

EXAMPLE 6.3. The transitions from Theorems 4.1 to 6.1 to 5.1 are smooth. Let $g(t) = t^{1/2+\delta}$ with $0 < \delta < \frac{1}{2}$. Then

$$p_m \sim (1 + 1/(2\delta))P\{S_m > g(m)\}.$$

As δ tends to its lower limit 0, the factor $(1 + 1/(2\delta))$ tends to ∞. This is consistent with Example 4.3. As δ tends to its upper limit $\frac{1}{2}$, $1/(2\delta)$ tends to 1, which is consistent with (5.2) in Remark 5.2(a). Furthermore, we have seen in Section 2 that the form of $P\{S_m > g(m)\}$ varies smoothly from the normal approximation to Cramér's theorem to the Bahadur-Ranga Rao result. $\square$

In Sections 7–11 we prove Theorem 6.1, thereby completing the solution to the general problem of convergence rates related to the SLLN.

7. Some facts about the boundary. The present section is reserved for a list of elementary properties of g resulting from the assumptions (6.3–6.5). The proofs are very easy.

LEMMA 7.1. *Let* $g: (0, \infty) \to (0, \infty)$ *be continuously differentiable and satisfy* (6.3–6.5), *and let* $g(t) = t^{1/2}\Psi(t)$. *Then*

(a) $\dfrac{\Psi(t)}{t^\delta} \uparrow$ (whence $\Psi(t) \uparrow \infty$) and $\dfrac{\Psi(t)}{\sqrt{t}} \downarrow 0$;

(b) $\left(\dfrac{1}{2} + \delta\right) \dfrac{g(t)}{t} \le g'(t) \le \dfrac{g(t)}{t}$;

(c) $\Psi'(t) = t^{-1/2}\left(g'(t) - \dfrac{1}{2}\dfrac{g(t)}{t}\right)$;

(d) $\dfrac{\delta\, \Psi(t)}{t} \le \Psi'(t) \le \dfrac{1}{2}\dfrac{\Psi(t)}{t}$;

(e) $\frac{d}{dt}\left(\frac{\Psi(t)}{\sqrt{t}}\right) = -\frac{1}{2}\frac{\Psi(t)}{t^{3/2}} + \frac{\Psi'(t)}{\sqrt{t}}$;

(f) $0 \le -\frac{d}{dt}\left(\frac{\Psi(t)}{\sqrt{t}}\right) \le \left(\frac{1}{2} - \delta\right)\frac{\Psi(t)}{t^{3/2}}$;

(g) $0 \le -\Psi^2(t)\frac{d}{dt}\left(\frac{\Psi(t)}{\sqrt{t}}\right) \le \left(\frac{1}{2} - \delta\right)\frac{\Psi^3(t)}{t^{3/2}} = o\left(\frac{\Psi^2(t)}{t}\right)$ as $t \to \infty$;

(h) $\Psi(u) \sim \Psi(t)$ as $u \sim t \to \infty$

(i) $\Psi^2(u) - \Psi^2(t) \to 0$ when $t, u \to \infty$ with $t \le u \le t + 1$;

(j) $\Psi'(u) \sim \Psi'(t)$ when $t, u \to \infty$ as in (6.5);

(k) $2 \le \frac{g'(t)}{\sqrt{t}\Psi'(t)} \le 1 + \frac{1}{2\delta}$.

8. Restriction of g-crossing times to a finite interval. Throughout this section "fact $(\cdot)$" refers to part $(\cdot)$ of Lemma 7.1.

We wish to define $v_m > m$ in such a way that

$$p_{v_m} = o(p_m) \quad \text{as} \quad m \to \infty \tag{8.1}$$

but g is virtually linear over $[m, v_m)$, for then we will use Siegmund's method to approximate the probability

$$p_{m,v_m} = P\{S_n > g(n) \quad \text{for some} \quad m \le n < v_m\}.$$

As it turns out, an appropriate choice is

$$v_m = \lfloor m(1 + 1/\Psi^{2r}(m))\rfloor, \tag{8.2}$$

where we denote by $\lfloor x \rfloor$ the integer part, or "floor", of x. Fact (a) implies that

$$v_m \sim m \tag{8.3}$$

and (6.5) yields

$$g'(t) \sim g'(m) \quad \text{when} \quad m, t \to \infty \quad \text{with} \quad m \le t \le v_m. \tag{8.4}$$

LEMMA 8.1. *If v_m is defined by* (8.2), *then* (8.1) *holds.*

PROOF. By subadditivity

$$p_m \le \textstyle\sum_{n \ge m} P\{S_n > g(n)\}; \tag{8.5}$$

we'll replace m by v_m in (8.5) to get an asymptotic upper bound on p_{v_m}. By Cramér's theorem,

$$\begin{aligned} P\{S_n > g(n)\} &\sim [\sqrt{2\pi}\Psi(n)]^{-1} \exp\left\{-\frac{1}{2}\Psi^2(n)\left[1 - 2\frac{\Psi(n)}{\sqrt{n}}\lambda\left(\frac{\Psi(n)}{\sqrt{n}}\right)\right]\right\} \\ &\sim [\sqrt{2\pi}\tilde{\Psi}(n)]^{-1} \exp\left\{-\frac{1}{2}\tilde{\Psi}^2(n)\right\}, \end{aligned} \tag{8.6}$$

where

$$\tilde{\Psi}(t) \equiv \left\{\Psi^2(t)\left[1 - 2\frac{\Psi(t)}{\sqrt{t}}\lambda\left(\frac{\Psi(t)}{\sqrt{t}}\right)\right]\right\}^{1/2}. \tag{8.7}$$

Note

(8.8) $$\tilde{\Psi}(t) \sim \Psi(t) \quad as \quad t \to \infty.$$

Also, as $t \to \infty$

$$2\tilde{\Psi}(t)\tilde{\Psi}'(t) = 2\Psi(t)\Psi'(t)\left[1 - 2\frac{\Psi(t)}{\sqrt{t}}\lambda\left(\frac{\Psi(t)}{\sqrt{t}}\right)\right]$$
$$- 2\Psi^2(t) \cdot \frac{d}{dt}\left(\frac{\Psi(t)}{\sqrt{t}}\right) \cdot \left[\lambda\left(\frac{\Psi(t)}{\sqrt{t}}\right) + \frac{\Psi(t)}{\sqrt{t}}\lambda'\left(\frac{\Psi(t)}{\sqrt{t}}\right)\right]$$
$$= (1 + o(1))2\Psi(t)\Psi'(t) - (1 + o(1))2\lambda_0\Psi^2(t) \cdot \frac{d}{dt}\left(\frac{\Psi(t)}{\sqrt{t}}\right)$$
$$= (1 + o(1))2\Psi(t)\Psi'(t)$$

by facts (d) and (g), so that

(8.9) $$\tilde{\Psi}'(t) \sim \Psi'(t).$$

Hence when $t, u \to \infty$ with $t \le u \le t + 1$

(8.10) $$0 \le \tilde{\Psi}^2(u) - \tilde{\Psi}^2(t) \le (1 + o(1))\frac{\Psi^2(t)}{t} = o(1)$$

(cf. proof of fact (i)). From (8.5–8.10) and facts (h) and (d) follows

$$\sqrt{2\pi}p_m \le (1 + o(1)) \textstyle\sum_{n \ge m}[\tilde{\Psi}(n)]^{-1}\exp\left[-\frac{1}{2}\tilde{\Psi}^2(n)\right]$$
$$= (1 + o(1)) \int_m^\infty [\tilde{\Psi}(t)]^{-1}\exp\left[-\frac{1}{2}\tilde{\Psi}^2(t)\right] dt$$
$$\le (1 + o(1))\, \delta^{-1} \int_m^\infty \frac{t}{\tilde{\Psi}^2(t)}\exp\left[-\frac{1}{2}\tilde{\Psi}^2(t)\right]\tilde{\Psi}'(t)\, dt$$
$$= O\left(\int_m^\infty \tilde{\Psi}^{1/\delta - 2}(t)\exp\left[-\frac{1}{2}\tilde{\Psi}^2(t)\right]\tilde{\Psi}'(t)\, dt\right).$$

But

$$\int_m^\infty \tilde{\Psi}^{1/\delta - 2}(t)\exp\left[-\frac{1}{2}\tilde{\Psi}^2(t)\right]\tilde{\Psi}'(t)\, dt = \int_{\tilde{\Psi}(m)}^\infty u^{1/\delta - 2}e^{-u^2/2}\, du$$
$$\sim \tilde{\Psi}^{1/\delta - 3}(m)\exp\left[-\frac{1}{2}\tilde{\Psi}^2(m)\right]$$

so

(8.11) $$p_m = O(\tilde{\Psi}^{1/\delta - 3}(m)\exp[-\tfrac{1}{2}\tilde{\Psi}^2(m)]).$$

Now by the mean value theorem, (8.8), (8.9), facts (j) and (d), and (8.2),

$$\tilde{\Psi}^2(v_m) - \tilde{\Psi}^2(m) \ge (1 + o(1))2\delta\frac{\tilde{\Psi}^2(m)}{m}(v_m - m) = (1 + o(1))2\delta\tilde{\Psi}^{2(1-r)}(m)$$

and thus (8.11)(with m replaced by v_m) easily yields

(8.12) $$p_{v_m} = o([\sqrt{2\pi}\tilde{\Psi}(m)]^{-1}\exp[-\tfrac{1}{2}\tilde{\Psi}^2(m)]) = o(P\{S_m > g(m)\}).$$

To complete the proof of Lemma 8.1 use the obvious bound

(8.13) $$p_m \ge P\{S_m > g(m)\}. \square$$

Define

(8.14) $$p_{m,v_m} = P\{S_n > g(n) \quad \text{for some} \quad m \le n < v_m\};$$

clearly,

(8.15) $$p_{m,v_m} \le p_m \le p_{m,v_m} + p_{v_m}.$$

We therefore immediately obtain

COROLLARY 8.2. *With p_m given by* (6.2) *and p_{m,v_m} by* (8.14),

(8.16) $$p_m \sim p_{m,v_m} \quad as \quad m \to \infty. \square$$

9. Linearization of the boundary. We define here straight lines $\underline{\ell}_m$ *and* $\bar{\ell}_m$, both passing through the point $(m, g(m))$, which well approximate g over the interval $[m, v_m]$. The definition of $\underline{\ell}_m$ (respectively, $\bar{\ell}_m$) together with the mean value theorem will imply that this line minorizes (majorizes) g on $[m, v_m]$.

The slope $\underline{\epsilon}_m$ (respectively, $\bar{\epsilon}_m$) of $\underline{\ell}_m$ ($\bar{\ell}_m$) is defined to be the minimum (maximum) value of g' over the interval $[m, v_m]$. We now treat both lines at once by writing, for example, ℓ_m indifferently for $\underline{\ell}_m$ or $\bar{\ell}_m$. By (8.4)

(9.1) $$\epsilon_m \sim g'(m) \quad \text{as} \quad m \to \infty.$$

The y-intercept of ℓ_m is

(9.2) $$\alpha_m = g(m) - m\epsilon_m.$$

By analogy with (6.2) and (8.14) define

(9.3) $$p_m(\ell_m) = P\{S_n > \ell_m(n) \quad \text{for some} \quad n \ge m\}$$

and

(9.4) $$p_{m,v_m}(\ell_m) = P\{S_n > \ell_m(n) \quad \text{for some} \quad m \le n < v_m\}.$$

The key to analyzing (9.3) is the following lemma, to be proved in the next section.

LEMMA 9.1. *Let ℓ_m denote the straight line*

(9.5) $$\ell_m(t) = \alpha_m + \epsilon_m t$$

and define $p_m(\ell_m)$ by (9.3). *If $\epsilon_m > 0$ satisfies*

(9.6) $$\epsilon_m \to 0$$

and

(9.7) $$\sqrt{m}\epsilon_m \to \infty,$$

and if

(9.8) $$\limsup_{m\to\infty} \frac{|\alpha_m|}{\epsilon_m m} < 1,$$

then

(9.9) $$p_m(\ell_m) \sim \frac{2\epsilon_m}{\epsilon_m - \dfrac{\alpha_m}{m}} \cdot P\{S_m > \ell_m(m)\} \quad \text{as} \quad m \to \infty. \square$$

REMARK 9.2. The conditions (9.6–9.7) demand that ϵ_m tend to 0, but not too quickly. Assumption (9.8) requires, loosely speaking, that the limiting proportional contribution of the constant term to the value at $t = m$ of either $\ell_m(t) = \alpha_m + \epsilon_m t$ or $\tilde{\ell}_m(t) = -\alpha_m + \epsilon_m t$ is less than ½. It follows from (9.8–9.9) that

$$p_m(\ell_m) \asymp P\{S_m > \ell_m(m)\}. \square$$

According to (6.3–6.4), (9.1–9.2), and Lemma 7.1(b), $\ell_m = \underline{\ell}_m$ or $\bar{\ell}_m$ satisfies the assumptions of the lemma. Recall that

$$\ell_m(m) = g(m). \tag{9.10}$$

Furthermore, by (9.2), (9.1), and Lemma 7.1(b)–(c)

$$\begin{aligned}\epsilon_m - \frac{\alpha_m}{m} &= 2\left[\epsilon_m - \frac{1}{2}\frac{g(m)}{m}\right] = 2\left[(1 + o(1))g'(m) - \frac{1}{2}\frac{g(m)}{m}\right] \\ &= (1 + o(1))2\sqrt{m}\Psi'(m).\end{aligned} \tag{9.11}$$

Combining (9.1) and (9.9–9.11) we get

$$p_m(\ell_m) \sim \frac{g'(m)}{\sqrt{m}\Psi'(m)} P\{S_m > g(m)\} = I_m \quad \text{as} \quad m \to \infty, \tag{9.12}$$

completing the analysis of (9.3).

In Lemma 9.3 below we will prove the analogue

$$p_{v_m}(\ell_m) \equiv P\{S_n > \ell_m(n) \quad \text{for some} \quad n \geq v_m\} = o(p_m(\ell_m)) \tag{9.13}$$

to (8.1). As an immediate corollary (cf. Corollary 8.2),

$$p_{m,v_m}(\ell_m) \sim p_m(\ell_m) \quad \text{as} \quad m \to \infty. \tag{9.14}$$

We combine (9.14) and (9.12) to obtain the rate of convergence for (9.4):

$$p_{m,v_m}(\ell_m) \sim I_m \quad \text{as} \quad m \to \infty. \tag{9.15}$$

Moreover, since $g(n)$ is trapped between $\underline{\ell}_m(n)$ and $\bar{\ell}_m(n)$ we have

$$p_{m,v_m}(\bar{\ell}_m) \leq p_{m,v_m} \leq p_{m,v_m}(\underline{\ell}_m). \tag{9.16}$$

The main result (6.6) follows from (8.16) and (9.15–9.16).

The remainder of this section is devoted to the following corollary to Lemma 9.1.

LEMMA 9.3. *With $\ell_m = \underline{\ell}_m$ or $\bar{\ell}_m$, (9.13) holds.*

PROOF. To begin the proof of (9.13), we apply Lemma 9.1 to the left side to yield

$$p_{v_m}(\ell_m) \sim \frac{2\epsilon_m}{\epsilon_m - \dfrac{\alpha_m}{v_m}} P\{S_{v_m} > \ell_m(v_m)\}, \tag{9.17}$$

recalling (8.3) to verify the hypotheses (9.7–9.8). In light of (8.3) and (9.8), the first factor on the right in (9.17) asymptotes to the first factor on the right in (9.9), or, in the present context, to the first factor in I_m (recall the proof of (9.12)). Moreover, by Cramér's result

$$P\{S_{v_m} > \ell_m(v_m)\} \sim [\sqrt{2\pi}\Psi_m(v_m)]^{-1}\exp\left\{-\frac{1}{2}\Psi_m^2(v_m)\left[1 - 2\frac{\Psi_m(v_m)}{\sqrt{v_m}}\lambda\left(\frac{\Psi_m(v_m)}{\sqrt{v_m}}\right)\right]\right\}; \tag{9.18}$$

here we have written

$$\ell_m(t) = \sqrt{t}\Psi_m(t) \tag{9.19}$$

so that $\Psi_m(m) = \Psi(m)$ and

$$\begin{aligned}\Psi_m(v_m) &= v_m^{-1/2}\ell_m(v_m) = v_m^{-1/2}(\alpha_m + \epsilon_m v_m) \\ &= (1 + o(1))m^{-1/2}(\alpha_m + \epsilon_m m) = (1 + o(1))\Psi(m).\end{aligned} \tag{9.20}$$

Put

$$\tilde{\Psi}_m(t) = \left\{\Psi_m^2(t)\left[1 - 2\frac{\Psi_m(t)}{\sqrt{t}}\lambda\left(\frac{\Psi_m(t)}{\sqrt{t}}\right)\right]\right\}^{1/2}; \tag{9.21}$$

then

$$P\{S_{v_m} > \ell_m(v_m)\} \sim [\sqrt{2\pi}\Psi(m)]^{-1}\exp\left[-\frac{1}{2}\tilde{\Psi}_m^2(v_m)\right]. \tag{9.22}$$

But with

$$f(\xi) = \xi^2[1 - 2\xi\lambda(\xi)] \sim \xi^2 \quad \text{as} \quad \xi \to 0, \tag{9.23}$$

so that

$$f'(\xi) = 2\xi[1 - 2\xi\lambda(\xi)] - 2\xi^2[\lambda(\xi) + \xi\lambda'(\xi)] \sim 2\xi \quad \text{as} \quad \xi \to 0, \tag{9.24}$$

we have

$$\begin{aligned}
\tilde{\Psi}_m^2(v_m) - \tilde{\Psi}_m^2(m) &= v_m f\left(\frac{\Psi_m(v_m)}{\sqrt{v_m}}\right) - mf\left(\frac{\Psi_m(m)}{\sqrt{m}}\right) \\
&= -v_m\left[f\left(\frac{\Psi_m(m)}{\sqrt{m}}\right) - f\left(\frac{\Psi_m(v_m)}{\sqrt{v_m}}\right)\right] + (v_m - m)f\left(\frac{\Psi_m(m)}{\sqrt{m}}\right) \\
&= -(1 + o(1))m \cdot 2\frac{\Psi(m)}{\sqrt{m}}\left(\frac{\Psi_m(m)}{\sqrt{m}} - \frac{\Psi_m(v_m)}{\sqrt{v_m}}\right) \\
&\quad + (1 + o(1))\frac{m}{\Psi^{2r}(m)}\frac{\Psi^2(m)}{m} \\
&= -(1 + o(1)) \cdot 2\sqrt{m}\Psi(m)\left[\left(\epsilon_m + \frac{\alpha_m}{m}\right) - \left(\epsilon_m + \frac{\alpha_m}{v_m}\right)\right] \\
&\quad + (1 + o(1))\Psi^{2(1-r)}(m) \\
&= -(1 + o(1)) \cdot 2\sqrt{m}\Psi(m)\frac{\alpha_m}{m}\left(1 - \frac{m}{v_m}\right) + (1 + o(1))\Psi^{2(1-r)}(m) \\
&= -(1 + o(1)) \cdot 2\sqrt{m}\Psi^{1-2r}(m) \cdot \frac{\alpha_m}{m} + (1 + o(1))\Psi^{2(1-r)}(m) \\
&= -(1 + o(1))\sqrt{m}\Psi^{1-2r}(m) \cdot 2\frac{\alpha_m}{m} \\
&\quad + (1 + o(1))\sqrt{m}\Psi^{1-2r}(m) \cdot \left(\epsilon_m + \frac{\alpha_m}{m}\right) \\
&= (1 + o(1))\sqrt{m}\Psi^{1-2r}(m) \cdot \left(\epsilon_m - \frac{\alpha_m}{m}\right) \\
&= (1 + o(1)) \cdot 2m\Psi'(m)\Psi^{1-2r}(m) \quad \text{(recall (9.11))} \\
&\geq (1 + o(1)) \cdot 2\delta\Psi^{2(1-r)}(m) \quad \text{(Lemma 7.1(d))}.
\end{aligned}$$

So

$$\frac{p_{v_m}(\ell_m)}{p_m(\ell_m)} = (1 + o(1))\frac{P\{S_{v_m} > \ell_m(v_m)\}}{P\{S_m > g(m)\}}$$
$$= (1 + o(1)) \cdot \exp\left\{ -\frac{1}{2} [\tilde{\Psi}_m^2(v_m) - \tilde{\Psi}_m^2(m)] \right\} = o(1),$$

completing the proof of (9.13). □

10. Line-crossing probabilities. This section is devoted to the following proof.

PROOF OF LEMMA 9.1. We first need to establish some notation. Define the distribution function F_m by

$$F_m(x) = F(\epsilon_m + x); \tag{10.1}$$

in other words, if X is distributed according to F_m, then $X - \epsilon_m$ has distribution function F. The effect of F_m will be to replace the linear boundary ℓ_m by a horizontal one: in obvious notation,

$$p_m(\ell_m) = P_{F_m}\{S_n > \alpha_m \quad \text{for some} \quad n \geq m\}.$$

Let K_m denote the cumulant generating function corresponding to F_m, so that

$$K_m(\xi) = K(\xi) - \epsilon_m \xi. \tag{10.2}$$

Recall the assumption that K is finite in an open interval I containing 0; K_m is also finite in this interval.

Since F has zero mean and unit variance, it is well-known that

$$K(\xi) \sim \frac{1}{2}\xi^2, \; K'(\xi) \sim \xi, \quad K''(\xi) \to 1 \quad \text{as} \quad \xi \to 0, \tag{10.3}$$

and that

$$K(0) = 0, \quad K'(0) = 0, \quad K''(\xi) > 0 \quad \text{for all} \quad \xi \in I. \tag{10.4}$$

From (10.2) and (10.4)

$$K_m(0) = 0, \quad K'_m(0) = -\epsilon_m, \quad K''_m(\xi) > 0 \quad \text{for all} \quad \xi \in I. \tag{10.5}$$

Hence there exists at most one nonzero value $\xi_1(m)$, necessarily positive, for which $K_m(\xi_1(m)) = 0$. We now show that $\xi_1(m)$ exists for sufficiently large m and that

$$\xi_1(m) \sim 2\epsilon_m \quad \text{as} \quad m \to \infty. \tag{10.6}$$

Indeed, let $\xi_m \to 0$ be an arbitrary real sequence. By (10.2–10.3)

$$K_m(\xi_m) = (1 + o(1)) \cdot \frac{1}{2}\xi_m^2 - \epsilon_m \xi_m \quad \text{as} \quad m \to \infty. \tag{10.7}$$

In particular, with $\xi_m \equiv C \cdot 2\epsilon_m$ $(C > 0)$

$$K_m(\xi_m) = 2C\epsilon_m^2[(1 + o(1))C - 1]. \tag{10.8}$$

If $C < 1$, then $K_m(\xi_m) \sim -2C(1 - C)\epsilon_m^2 < 0$; if $C > 1$, then $K_m(\xi_m) \sim 2C(C - 1)\epsilon_m^2 > 0$. This shows that $\xi_1(m)$ exists for sufficiently large m and that, for any pair of constants $0 < C < 1 < C'$, $C < \xi_1(m)/(2\epsilon_m) < C'$ for large m; (10.6) follows.

Similarly, for large m there exists exactly one point $\xi_0(m) \in (0, \xi_1(m))$ at which K'_m vanishes, and

$$\xi_0(m) \sim \epsilon_m \quad \text{as} \quad m \to \infty. \tag{10.9}$$

This follows from the fact that if $\xi_m \to 0$, then

$$K'_m(\xi_m) = (1 + o(1))\xi_m - \epsilon_m \quad \text{as} \quad m \to \infty. \tag{10.10}$$

Next, let P_m denote the probability under which $X, X_1, X_2, \cdots$ are i.i.d. with

$$P_m\{X \in dx\} = \exp[\xi_0(m) \cdot x - K_m(\xi_0(m))]F_m(dx). \tag{10.11}$$

P_m has cgf

$$\begin{aligned} \phi_m(\theta) &= K_m(\xi_0(m) + \theta) - K_m(\xi_0(m)) \\ &= K(\xi_0(m) + \theta) - K(\xi_0(m)) - \epsilon_m\theta; \end{aligned} \tag{10.12}$$

note that

$$\phi_m(0) = \phi'_m(0) = 0, \qquad \phi''_m(0) = K''(\xi_0(m)). \tag{10.13}$$

Put

$$\theta_0(m) = -\xi_0(m), \qquad \theta_1(m) = \xi_1(m) - \xi_0(m). \tag{10.14}$$

Then

$$\theta_0(m) \sim -\epsilon_m, \qquad \theta_1(m) \sim \epsilon_m \quad \text{as} \quad m \to \infty \tag{10.15}$$

and

$$\begin{aligned} \phi_m(\theta_0(m)) &= \phi_m(\theta_1(m)) = -K_m(\xi_0(m)) = -(K(\xi_0(m)) + \epsilon_m\theta_0(m)) \\ &\sim \frac{1}{2}\epsilon_m^2 \quad \text{as} \quad m \to \infty. \end{aligned} \tag{10.16}$$

We now introduce the distributions associated through "exponential tilting" with that of X under P_m. For each real θ for which $\phi_m(\theta) < \infty$ let $P_{m,\theta}$ be the probability under which $X, X_1, X_2, \cdots$ are i.i.d. with

$$\begin{aligned} P_{m,\theta}\{X \in dx\} &= \exp[\theta x - \phi_m(\theta)]P_m\{X \in dx\} \\ &= \exp[\theta x - \phi_m(\theta)]\exp[\xi_0(m) \cdot x - K_m(\xi_0(m))]F_m(dx). \end{aligned} \tag{10.17}$$

The corresponding cgf $\phi_{m,\theta}$ is given by

$$\begin{aligned} \phi_{m,\theta}(\eta) &= \phi_m(\theta + \eta) - \phi_m(\theta) = K_m(\xi_0(m) + \theta + \eta) - K_m(\xi_0(m) + \theta) \\ &= K(\xi_0(m) + \theta + \eta) - K(\xi_0(m) + \theta) - \epsilon_m\eta. \end{aligned} \tag{10.18}$$

In particular

$$E_{m,\theta}X = \phi'_{m,\theta}(0) = \phi'_m(\theta) <, =, \text{or} > 0 \quad \text{according as} \quad \theta <, =, \text{or} > 0 \tag{10.19}$$

by (10.13) and the strict convexity of ϕ_m.

As special cases of (10.17),

$$P_{m,0} = P_m, \quad P_{m,\theta_1(m)}\{X \in dx\} = \exp[\xi_1(m) \cdot x]F_m(dx),$$

and

$$\begin{aligned} P_{m,\theta_0(m)}\{X \in dx\} &= \exp\{[\xi_0(m) + \theta_0(m)] \cdot x - [\phi_m(\theta_0(m)) + K_m(\xi_0(m))]\}F_m(dx) \\ &= F_m(dx) = F(\epsilon_m + dx) \end{aligned} \tag{10.20}$$

(cf. (10.14), (10.16), and (10.1)). Hence by (9.3) and (9.5)

$$p_m(\ell_m) = P_{m,\theta_0(m)}\{S_n > \alpha_m \quad \text{for some} \quad n \ge m\}. \tag{10.21}$$

Without loss of generality we consider $P_{m,\theta}$ to *be* the distribution of $X_1, X_2, \cdots$, defined on the space of (infinite) sequences of real numbers. Accordingly, let $P_{m,\theta}^{(n)}$ denote the restriction of $P_{m,\theta}$ to the σ-algebra generated by the first n coordinates ($n = 1, 2, \cdots$). Then for any θ' and θ'', $P_{m,\theta'}^{(n)}$ and $P_{m,\theta''}^{(n)}$ are mutually absolutely continuous, and by (10.17)

$$\frac{dP_{m,\theta'}^{(n)}}{dP_{m,\theta''}^{(n)}} = \exp\{(\theta' - \theta'')S_n - n[\phi_m(\theta') - \phi_m(\theta'')]\}. \tag{10.22}$$

In particular, by (10.14), (10.16), and (10.22)

$$\frac{dP^{(n)}_{m,\theta_0(m)}}{dP^{(n)}_{m,\theta_1(m)}} = \exp[-\xi_1(m)S_n] \quad (n = 1, 2, \cdots). \tag{10.23}$$

Let

$$T_m = \inf\{n : n \geq m,\ S_n > \alpha_m\},$$

the inf of the empty set being $+\infty$. Then by (10.21)

$$p_m(\ell_m) = P_{m,\theta_0(m)}\{S_m > \alpha_m\} + P_{m,\theta_0(m)}\{m < T_m < \infty\}. \tag{10.24}$$

By (9.5) and (10.20), the first probability in (10.24) is $P\{S_m > \ell_m(m)\}$. To complete the proof of (9.9) we shall use the approach of Siegmund (1975) to show that

$$P_{m,\theta_0(m)}\{m < T_m < \infty\} \sim \frac{\epsilon_m + \dfrac{\alpha_m}{m}}{\epsilon_m - \dfrac{\alpha_m}{m}} \cdot P\{S_m > \ell_m(m)\} \quad \text{as} \quad m \to \infty. \tag{10.25}$$

We first apply the fundamental identity of sequential analysis, to wit: by (10.23)

$$\begin{aligned} P_{m,\theta_0(m)}\{m < T_m < \infty\} &= \textstyle\sum_{n=m+1}^{\infty} \displaystyle\int_{\{T_m=n\}} \exp[-\xi_1(m)S_n]\, dP_{m,\theta_1(m)} \\ &= \int_{\{m<T_m<\infty\}} \exp[-\xi_1(m)S_{T_m}]\, dP_{m,\theta_1(m)} \\ &= \exp[-\xi_1(m)\alpha_m] \cdot \int_{\{m<T_m<\infty\}} \exp[-\xi_1(m)(S_{T_m} - \alpha_m)]\, dP_{m,\theta_1(m)}. \end{aligned} \tag{10.26}$$

Recalling $\theta_1(m) > 0$, (10.19) implies $E_{m,\theta_1(m)}X > 0$; hence by the SLLN

$$\{m < T_m < \infty\} = \{m < T_m \leq \infty\} = \{S_m \leq \alpha_m\} \text{ a.s. } P_{m,\theta_1(m)},$$

and the final integral in (10.26) equals

$$\begin{aligned} &\int_{\{S_m \leq \alpha_m\}} \exp[-\xi_1(m)(S_{T_m} - \alpha_m)]\, dP_{m,\theta_1(m)} \\ &\quad = P_{m,\theta_1(m)}\{S_m \leq \alpha_m\} \\ &\quad \cdot \int_{[0,\infty)} E_{m,\theta_1(m)}(\exp[-\xi_1(m)(S_{T_m} - \alpha_m)] \mid S_m = \alpha_m - y) \\ &\quad \cdot P_{m,\theta_1(m)}\{S_m \in \alpha_m - dy \mid S_m \leq \alpha_m\}. \end{aligned} \tag{10.27}$$

The last conditional expectation in (10.27) is

$$E_{m,\theta_1(m)}(\exp[-\xi_1(m)(S_{\tau(y)} - y)]), \tag{10.28}$$

where $\tau(y) = \inf\{n : S_n > y\}$. We show below that the expression (10.28) tends uniformly in $0 \leq y < \infty$ to 1 as $m \to \infty$; thus

$$P_{m,\theta_0(m)}\{m < T_m < \infty\} \sim \exp[-\xi_1(m)\alpha_m] \cdot P_{m,\theta_1(m)}\{S_m \leq \alpha_m\}. \tag{10.29}$$

Moreover, mimicking the proof of Cramér's theorem we'll show in Section 11 that

$$P_{m,\theta_1(m)}\{S_m \leq \alpha_m\} \sim \exp[\xi_1(m)\alpha_m] \cdot \frac{\epsilon_m + \dfrac{\alpha_m}{m}}{\epsilon_m - \dfrac{\alpha_m}{m}} \cdot P\{S_m > \ell_m(m)\}. \tag{10.30}$$

Then (10.25) follows from (10.29) and (10.30).

The first of our two assertions is that

$$E_{m,\theta_1(m)}(\exp[-\xi_1(m)R_y]) \to 1 \quad \text{as} \quad m \to \infty, \quad \text{uniformly in} \quad 0 \le y < \infty, \tag{10.31}$$

where $R_y = S_{\tau(y)} - y$ is the excess over the horizontal boundary with level y. In light of the inequality $1 \ge e^{-\eta} \ge 1 - \eta$ for $\eta \ge 0$ and (10.16), it is enough to show that the nonnegative quantity

$$\sup\{E_{m,\theta_1(m)}R_y : 0 \le y < \infty\} = o\left(\frac{1}{\epsilon_m}\right) \quad \text{as} \quad m \to \infty. \tag{10.32}$$

Indeed, for $0 \le y < \infty$ we find

$$(E_{m,\theta_1(m)}R_y)^2 \le E_{m,\theta_1(m)}R_y^2 \le \frac{4}{3} E_{m,\theta_1(m)}(X^+)^3 / E_{m,\theta_1(m)}X$$

using Theorem 3 in Lorden (1970). An easy dominated convergence argument shows that $E_{m,\theta_1(m)}(X^+)^3 \to E(X^+)^3 < \infty$, and by (10.19), (10.12), (10.14), (10.10), and (10.6)

$$\begin{aligned} E_{m,\theta_1(m)}X &= \phi'_m(\theta_1(m)) = K'(\xi_1(m)) - \epsilon_m \\ &= (1 + o(1))\xi_1(m) - \epsilon_m = (1 + o(1))\epsilon_m. \end{aligned} \tag{10.33}$$

Therefore

$$\sup\{E_{m,\theta_1(m)}R_y : 0 \le y < \infty\} \le (1 + o(1))\left[\frac{4}{3} E(X^+)^3/\epsilon_m\right]^{1/2} = O(\epsilon_m^{-1/2}) = o(\epsilon_m^{-1}), \tag{10.34}$$

which proves (10.32).

The remaining assertion is (10.30). Standardizing to zero mean and unit variance,

$$P_{m,\theta_1(m)}\{S_m \le \alpha_m\} = P_{m,\theta_1(m)}\left\{\frac{S_m - mE_{m,\theta_1(m)}X}{(m \operatorname{Var}_{m,\theta_1(m)}X)^{1/2}} \le -u_m\right\}$$

with

$$u_m = -(\alpha_m - mE_{m,\theta_1(m)}X)/(m \operatorname{Var}_{m,\theta_1(m)}X)^{1/2}. \tag{10.35}$$

From (10.33) and

$$\operatorname{Var}_{m,\theta_1(m)}X = \phi''_m(\theta_1(m)) = K''_m(\xi_1(m)) = K''(\xi_1(m)) \to 1 \tag{10.36}$$

and (9.8) follows

$$u_m \sim \sqrt{m}\left(\epsilon_m - \frac{\alpha_m}{m}\right). \tag{10.37}$$

Let G_m be the standardized distribution of $(-X)$ under $P_{m,\theta_1(m)}$, and let λ_m be the associated Cramér series. Were it not for the dependence of G_m on m, direct application of Cramér's result (which holds as well when $>$ is changed to $\ge$ on the left in (2.7)) would give

$$P_{m,\theta_1(m)}\{S_m \le \alpha_m\} \sim P\{Z > u_m\}\exp\left[\frac{u_m^3}{\sqrt{m}}\lambda_m\left(\frac{u_m}{\sqrt{m}}\right)\right]. \tag{10.38}$$

In fact, (10.38) can be established by rehashing the proof of Cramér's theorem and can be used to deduce (10.30). We shall follow a somewhat shorter route and verify (10.30) directly, but our proof, like Cramér's, will be based on the standard large-deviation technique of exponential tilting.

11. A Cramér-like result. In this section we complete the proof of Lemma 9.1 by establishing the Cramér-like result (10.30).

For abbreviation we introduce the notation

$$\Psi(m) = \ell_m(m)/\sqrt{m} = \sqrt{m}\left(\epsilon_m + \frac{\alpha_m}{m}\right). \tag{11.1}$$

This should cause no confusion; after all, whenever Lemma 9.1 is applied to the original convergence rate problem the identification (11.1) is made. In addition, let

$$z_m = z\left(\frac{\Psi(m)}{\sqrt{m}}\right) \tag{11.2}$$

with z defined in (6.1) and put

$$\theta_2(m) = z_m + \theta_0(m); \tag{11.3}$$

we shall soon tilt from $P_{m,\theta_1(m)}$ to $P_{m,\theta_2(m)}$ to compute the left side of (10.30). Observe

$$z_m \sim \frac{\Psi(m)}{\sqrt{m}} \to 0, \tag{11.4}$$

$$\theta_2(m) = \frac{\alpha_m}{m} + o(\epsilon_m) \to 0, \tag{11.5}$$

$$\phi_m(\theta_2(m)) = K(z_m) - K(\xi_0(m)) - \epsilon_m \theta_2(m). \tag{11.6}$$

Also note that

$$E_{m,\theta_2(m)}X = \phi'_m(\theta_2(m)) = K'(z_m) - \epsilon_m = \frac{\Psi(m)}{\sqrt{m}} - \epsilon_m = \frac{\alpha_m}{m} \tag{11.7}$$

and that

$$\sigma_m^2 \equiv \mathrm{Var}_{m,\theta_2(m)}X = \phi''_m(\theta_2(m)) = K''(z_m) \to 1. \tag{11.8}$$

We are going to use the Berry-Esséen theorem, and so the third absolute moments

$$\rho_m = E_{m,\theta_2(m)} \mid X \mid^3, \quad \rho = E \mid X \|^3 < \infty \tag{11.9}$$

will arise; dominated convergence gives

$$\rho_m \to \rho. \tag{11.10}$$

Let

$$\pi_m = P_{m,\theta_1(m)}\{S_m \le \alpha_m\} \tag{11.11}$$

denote the left side of (10.30). Putting $\theta' = \theta_1(m)$, $\theta'' = \theta_2(m)$, and $n = m$ in (10.22) yields

$$\pi_m = \exp\{-m[\phi_m(\theta_1(m)) - \phi_m(\theta_2(m))]\} \cdot \int_{(-\infty,\alpha_m]} \exp[(\theta_1(m) - \theta_2(m))s] P_{m,\theta_2(m)}\{S_m \in ds\}. \tag{11.12}$$

Recalling (10.16) and (11.6) and simplifying,

$$\pi_m = \exp\{-m[\epsilon_m z_m - K(z_m)]\} \int_{(-\infty,\alpha_m]} \exp[(\theta_1(m) - \theta_0(m) - z_m)s] \cdot P_{m,\theta_2(m)}\{S_m \in ds\}. \tag{11.13}$$

If the approximation $\hat{\pi}_m$ to π_m is obtained from the right side of (11.13) by replacing $P_{m,\theta_2(m)}\{S_m \in ds\}$ with the normal distribution $P\{\alpha_m + m^{1/2}\sigma_m Z \in ds\}$ with the same mean and variance, then (completing the square and rearranging)

$$
\begin{aligned}
\hat{\pi}_m = {} & \exp[(\theta_1(m) - \theta_0(m))\alpha_m]\exp\left[\frac{1}{2} m\sigma_m^2(\theta_1(m) - \theta_2(m))^2\right] \\
& \cdot P\{Z \le -\sqrt{m}\,\sigma_m(\theta_1(m) - \theta_2(m)\} \\
& \cdot \exp\left\{m\left[K(z_m) - z_m \cdot \frac{\Psi(m)}{\sqrt{m}}\right]\right\} \\
= {} & \exp[(\theta_1(m) - \theta_0(m))\alpha_m] \\
& \cdot \exp\left[-\frac{1}{2}\Psi^2(m) + h(\sqrt{m}\,\sigma_m(\theta_1(m) - \theta_2(m)))\right] \\
& \cdot \exp\left[\frac{\Psi^3(m)}{\sqrt{m}}\lambda\left(\frac{\Psi(m)}{\sqrt{m}}\right)\right],
\end{aligned}
\tag{11.14}
$$

where

$$h(t) = \log(e^{t^2/2} \cdot P\{Z > t\}). \tag{11.15}$$

Note

$$h'(t) = t - \frac{1}{\sqrt{2\pi}} e^{-t^2/2}/P\{Z > t\} \sim -1/t \quad \text{as} \quad t \to \infty. \tag{11.16}$$

So by the mean value theorem

$$
\begin{aligned}
& h(\sqrt{m}\,\sigma_m(\theta_1(m) - \theta_2(m))) - h\left(\sqrt{m}\left(\epsilon_m - \frac{\alpha_m}{m}\right)\right) \\
& = -(1 + o(1))\sqrt{m}\left[\sigma_m(\theta_1(m) - \theta_2(m)) - \left(\epsilon_m - \frac{\alpha_m}{m}\right)\right] \Big/ \left[\sqrt{m}\left(\epsilon_m - \frac{\alpha_m}{m}\right)\right] \\
& = o(1),
\end{aligned}
\tag{11.17}
$$

and hence the second of the three factors on the right side of (11.14) is $(1 + o(1))$ times

$$
\begin{aligned}
& \exp\left[-\frac{1}{2}\Psi^2(m) + h\left(\sqrt{m}\left(\epsilon_m - \frac{\alpha_m}{m}\right)\right)\right] \\
& = \exp\left[-\frac{1}{2}\Psi^2(m)\right]\exp\left[\frac{1}{2} m\left(\epsilon_m - \frac{\alpha_m}{m}\right)^2\right] P\left\{Z > \sqrt{m}\left(\epsilon_m - \frac{\alpha_m}{m}\right)\right\} \\
& = (1 + o(1))\left(\epsilon_m - \frac{\alpha_m}{m}\right)^{-1} (2\pi m)^{-1/2} \exp\left[-\frac{1}{2}\Psi^2(m)\right] \\
& = (1 + o(1)) \frac{\epsilon_m + \dfrac{\alpha_m}{m}}{\epsilon_m - \dfrac{\alpha_m}{m}} [\sqrt{2\pi}\Psi(m)]^{-1} \exp\left[-\frac{1}{2}\Psi^2(m)\right] \\
& = (1 + o(1)) \frac{\epsilon_m + \dfrac{\alpha_m}{m}}{\epsilon_m - \dfrac{\alpha_m}{m}} P\{Z > \Psi(m)\}.
\end{aligned}
$$

In light of (2.7) it is now clear that

$$\hat{\pi}_m \sim \text{right side of (10.30)}. \tag{11.18}$$

Moreover, integration by parts in (11.13) leads to the error estimate

$$| \pi_m - \hat{\pi}_m | \le 2C \cdot \frac{\rho_m}{\sigma_m^3 \sqrt{m}} \cdot \exp[(\theta_1(m) - \theta_0(m))\alpha_m] \cdot \exp\left\{ m \left[K(z_m) - z_m \cdot \frac{\Psi'(m)}{\sqrt{m}} \right] \right\} \tag{11.19}$$

where C is the universal constant appearing in the Berry-Esséen bound $C \cdot \rho/(\sigma^3 n^{1/2})$ on the error in the central limit theorem (see Feller, 1971, Theorem XVI, 5.1). Comparing with (11.14) and recalling (11.15), the right side of (11.19) is

$$2C \cdot \frac{\rho_m}{\sigma_m^3 \sqrt{m}} \cdot \hat{\pi}_m \exp[- h(\sqrt{m}\sigma_m(\theta_1(m) - \theta_2(m)))]. \tag{11.20}$$

But

$$\exp[- h(t)] \sim \sqrt{2\pi} t \quad \text{as} \quad t \to \infty, \tag{11.21}$$

so by (11.8), (11.10), and (11.21), (11.20) is

$$(1 + o(1)) \cdot 2C\rho \sqrt{2\pi} \left(\epsilon_m - \frac{\alpha_m}{m} \right) \hat{\pi}_m = o(\hat{\pi}_m).$$

Together $\pi_m \sim \hat{\pi}_m$ and (11.18) gives (10.30), completing the proof of Lemma 9.1. □

Acknowledgement. I wish to thank Professor Michael Wichura of the University of Chicago for his extensive guidance. I would also like to thank the referee for many helpful suggestions.

REFERENCES

[1] Bahadur, R. R. and Rao, R. Ranga (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.
[2] Breiman, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.
[3] Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Act. Sci. et Ind.* **736**.
[4] Feller, W. (1968). *An Introduction to Probability Theory and its Applications* **1**, 3rd ed. Wiley, New York.
[5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications* **2**, 2nd ed. Wiley, New York.
[6] Fill, J. A. (1980). Convergence rates related to the strong law of large numbers. Stanford University Technical Report No. 14.
[7] Jain, N. C., Jogdeo, K., and Stout, W. F. (1975). Upper and lower functions for martingales and mixing processes. *Ann. Probability* **3** 119–145.
[8] Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF.I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.
[9] Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's and the sample DF.II. *Z. Wahrsch. verw. Gebiete* **34** 33–58.
[10] Lorden, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527.
[11] Sawyer, S. (1972). A remark on the Skorohod representation. *Z. Wahrsch. verw. Gebiete* **23** 67–74.
[12] Siegmund, D. (1975). Large deviation probabilities in the strong law of large numbers. *Z. Wahrsch. verw. Gebiete* **31** 107–113.
[13] Strassen, V. (1967). Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **2** 315–343.

Department of Statistics
Stanford University
Sequoia Hall
Stanford, California 94305