The Annals of Probability
1975, Vol. 3, No. 1, 100–118

ON TAIL PROBABILITIES FOR MARTINGALES[1]

By David A. Freedman

University of California, Berkeley

Watch a martingale with uniformly bounded increments until it first crosses the horizontal line of height a. The sum of the conditional variances of the increments given the past, up to the crossing, is an intrinsic measure of the crossing time. Simple and fairly sharp upper and lower bounds are given for the Laplace transform of this crossing time, which show that the distribution is virtually the same as that for the crossing time of Brownian motion, even in the tail. The argument can be adapted to extend inequalities of Bernstein and Kolmogorov to the dependent case, proving the law of the iterated logarithm for martingales. The argument can also be adapted to prove Lévy's central limit theorem for martingales. The results can be extended to martingales whose increments satisfy a growth condition.

1. Introduction. In 1937, Lévy ([16] Theorem 67, page 243) showed that Lindeberg's central limit theorem (with an error bound) could be extended to martingales. Using this, he generalized Kolmogorov's law of the iterated logarithm to martingales ([16] pages 258 ff). This paper presents another method for proving these results, by extending inequalities of Bernstein and Kolmogorov on tail probabilities for sums of independent variables to the dependent case; these inequalities are powerful enough to prove the law of the iterated logarithm. The method also gives a bound on the Laplace transform, sharp enough to prove the central limit theorem. It could also be used to prove Dvoretzky's general central limit theorem (unpublished) for dependent summands: I hope to explore this elsewhere.

There is a review of the literature in Godwin (1964) and in Karlin and Studden (1966). Also see Bennett (1962), Hoeffding (1963), Loève ((1963) Section 18), and Steiger (1969).

The general approach used here is to transform a given martingale into two processes, one being expectation-decreasing and the other expectation-increasing. The first gives upper bounds, and the second lower bounds, on the probabilities of interest. I learned the idea from Theorem (2.12.1) of Dubins and Savage (1965); but this paper only uses standard martingale arguments.

Let $(\Omega, \mathscr{F}, P)$ be a probability triple. Let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \cdots$ be an

Received December 27, 1972; April 19, 1974.

[1] Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant AFOSR-71-2100. The United States Government is authorized to reproduce and distribute reports for governmental purposes notwithstanding any copyright notation hereon.

AMS 1970 *subject classifications.* 60F10, 60F15, 60F05, 60G45, 60G40.

Key words and phrases. Martingale, crossing time, tail probability, law of the iterated logarithm, central limit theorem.

increasing sequence of sub-σ-fields of $\mathscr{F}$. Say τ is a *stopping time* if τ is a function on Ω, which takes the values $0, 1, 2, \cdots, \infty$, such that $\{\tau = n\} \in \mathscr{F}_n$ for $n = 0, 1, 2, \cdots$; but $P\{\tau = \infty\} > 0$ is allowed. Let $X_1, X_2, \cdots$ be random variables on $(\Omega, \mathscr{F}, P)$, such that X_n is $\mathscr{F}_n$-measurable. Let $V_n = \operatorname{Var}\{X_n \mid \mathscr{F}_{n-1}\}$. Until further notice, assume

$$|X_n| \leqq 1 \quad \text{and} \quad E\{X_n \mid \mathscr{F}_{n-1}\} = 0 . \tag{1.1}$$

The restrictive condition is the uniform boundedness, which will be relaxed to a growth condition in Section 2.

(1.2) DEFINITION. λ is a positive number.

(a) $e(\lambda) = e^{\lambda} - 1 - \lambda$,
(b) $f(\lambda) = e^{-\lambda} - 1 + \lambda$,
(c) $\exp x = e^{x}$,
(d) $Q_\lambda(v, y) = \exp\{\lambda y - e(\lambda)v\}$,
(e) $R_\lambda(v, y) = \exp\{\lambda h - f(\lambda)v\}$,
(f) $S_n = X_1 + \cdots + X_n$, so $S_0 = 0$,
(g) $T_n = V_1 + \cdots + V_n$, so $T_0 = 0$,
(h) $a \wedge b = \min\{a, b\}$.

Informally, T_n is an intrinsic measure of time, while n is a nominal measure of time: this is discussed in [3]. If S_n is considered as occurring at time T_n, rather than time n, the process $\{S_n\}$ is remarkably like Brownian motion, even when many tail probabilities are computed.

The results of this paper all follow from the observation that Q_λ is superharmonic and R_λ subharmonic under (1.1). More exactly, let y be real and v nonnegative. Let X be a random variable with $|X| \leqq 1$ and $E(X) = 0$ and $E(X^2) = V$. Then

(1.3) LEMMA.

(a) $E\{\exp(\lambda X)\} \leqq \exp[Ve(\lambda)]$,
(b) $E\{Q_\lambda(v + V, y + X)\} \leqq Q_\lambda(v, y)$,
(c) $E\{\exp(\lambda X)\} \geqq \exp[Vf(\lambda)]$,
(d) $E\{R_\lambda(v + V, y + X)\} \geqq R_\lambda(v, y)$.

Of these, (a) and (c) are proved in Section 2; inequalities (b) and (d) are immediate consequences.

(1.4) COROLLARY.

(a) $\{Q_\lambda(T_n, S_n), \mathscr{F}_n : n = 0, 1, \cdots\}$ *is an expectation-decreasing martingale*;
(b) $\{R_\lambda(T_n, S_n), \mathscr{F}_n : n = 0, 1, \cdots\}$ *is an expectation-increasing martingale.*

(1.5) MAIN INEQUALITY. If σ is a uniformly bounded stopping time, then

(a) $E\{Q_\lambda(T_\sigma, S_\sigma)\} \leqq 1$ and
(b) $E\{R_\lambda(T_\sigma, S_\sigma)\} \geqq 1$.

By appropriate limiting arguments, (1.5) can be extended to very general σ. All the inequalities in this paper come from (1.5), by proper choice of λ and σ.

The first consequence is the extension of the inequalities of Bernstein and Kolmogorov. It makes explicit a bound implied by (31) of [5]. This bound was discovered by Steiger (1969).

(1.6) THEOREM. *For any positive numbers a and b,*

$$P\{S_n \geqq a \text{ and } T_n \leqq b \text{ for some } n\} \leqq \left(\frac{b}{a+b}\right)^{a+b} e^a \leqq \exp\left[-\frac{a^2}{2(a+b)}\right].$$

To get this from (1.5a), let σ be the least n with $S_n \geqq a$. There is a detailed argument in (4.1). This inequality shows that

$$\limsup S_n/(2T_n \log\log T_n)^{\frac{1}{2}} \leqq 1 \qquad \text{a.e. on } \{\sum V_i = \infty\};$$

see (6.1).

To get the other half of the law of the iterated logarithm, make the

(1.7) DEFINITION. Fix $a > 0$. Let τ_a be the least n if any with $S_n \geqq a$, and let $\tau_a = \infty$ if there is no such n. Let $W_a = T_{\tau_a} = \sum_{i=1}^{\tau_a} V_i$.

So $\{W_a < b\} = \{S_n \geqq a \text{ and } T_n < b \text{ for some } n\} \cup \{\sup_n T_n < b\}$. Thus W_a is the total amount of conditional variance it takes for the partial sum process to cross the a-line, if it crosses; otherwise, W_a is the total amount of conditional variance.

(1.8) THEOREM. $E\{\exp[-f(\lambda)W_a]\} \geqq \exp[-\lambda(a+1)]$ *for* $a > 0$.

To get this from (1.5b), use τ_a for σ. There is a detailed argument in (4.2). This theorem prevents W_a from being too large. Here is a more precise statement, interesting when $a = o(b^{\frac{1}{2}})$.

(1.9) COROLLARY. $P\{W_a \geqq b\} < 5(a+1)/b^{\frac{1}{2}}$ *for positive a and b.*

This follows from (1.8) by Chebychev's inequality. There is a detailed argument in (4.4). Consequently, $\sup_n S_n = \infty$ a.e. on $\{\sum V_i = \infty\}$, as argued in (4.5).

The next result is interesting when a is around $(2b \log\log b)^{\frac{1}{2}}$.

(1.10) COROLLARY. *Suppose* δ, a, b *are positive, with* $\delta < \frac{1}{3}$ *and* $b/a > 9/\delta^2$ *and* $a^2/b > (16/\delta^2)\log(64/\delta^2)$. *Then*

$$P\{W_a < b\} > \tfrac{1}{2}\exp[-(\tfrac{1}{2} + 2\delta)a^2/b].$$

This follows from (1.8) by a complicated analytical argument, carried out in (4.31). As in (6.3), it shows that

$$\limsup S_n/(2T_n \log\log T_n)^{\frac{1}{2}} \geqq 1 \qquad \text{a.e. on } \{\sum V_i = \infty\}.$$

Nothing so far stops W_a from being small. For instance, all the V_i could vanish identically, so $X_i = 0$ and $\tau_a = \infty$ and $W_a = 0$ identically. Suppose,

however, that

$$P\{\textstyle\sum_{i=1}^{\infty} V_i = \infty\} = 1\,. \tag{1.11}$$

Then $P\{\tau_a < \infty\} = 1$, because (1.9) makes $P\{W_a < \infty\} = 1$. And the distribution of W_a is virtually the same, for all processes satisfying (1.1) and (1.11).

(1.12) THEOREM. *Suppose* (1.1) *and* (1.11). *Then* $E\{\exp[-e(\lambda)W_a]\} \leqq \exp(-\lambda a)$ *for* $a > 0$.

To prove this, use (1.5a) with τ_a for σ. There is a detailed argument in (4.6). Since $e(\lambda)$ and $f(\lambda)$ are very close for small λ, inequalities (1.8) and (1.12) effectively determine the tail of W_a. The argument is carried out in Section 5, but seems to involve a nonconstructive step, so $c(\varepsilon)$ in the next result is not explicitly determined:

(1.13) COROLLARY. *Suppose* (1.1) *and* (1.11). *For any positive* ε, *there is a finite* $c = c(\varepsilon)$ *such that*: $a > c$ *and* $b/a^2 > c$ *imply*

$$(1 - \varepsilon)(2/\pi)^{\frac{1}{2}} \leqq (b^{\frac{1}{2}}/a)P\{W_a \geqq b\} \leqq (1 + \varepsilon)(2/\pi)^{\frac{1}{2}}$$

Of course, as $a \to \infty$ the distribution of W_a/a^2 tends to the distribution of the time for ordinary Brownian motion to cross the level 1, by the invariance principle for martingales ([11] pages 89 ff). The result (1.13) is interesting because it says something about finite a. The amount of variance needed to escape from a strip (as opposed to crossing a line) is described in [3].

Inequalities (1.5) also prove a variant of the central limit theorem, as shown in Section 7.

(1.14) THEOREM. $0 \leqq \alpha \leqq \beta < \infty$. *Let* σ *be a stopping time with* $P\{\alpha \leqq T_\sigma \leqq \beta\} = 1$. *Then*

$$\exp[\alpha f(\lambda)] \leqq E\{\exp(\lambda S_\sigma)\} \leqq \exp[\beta e(\lambda)]\,.$$

Since $e(\lambda)$ and $f(\lambda)$ are essentially $\frac{1}{2}\lambda^2$ for small λ,

(1.15) COROLLARY. *For each* n, *let* $0 < \alpha_n < \beta_n$ *be real numbers, and let* σ_n *be a stopping time. As* $n \to \infty$, *suppose* $\alpha_n \to \infty$ *and* $\beta_n/\alpha_n \to 1$ *and* $P\{\alpha_n < T_{\sigma_n} < \beta_n\} \to 1$. *Then the distribution of* $S_{\sigma_n}/\alpha_n^{\frac{1}{2}}$ *weak* converges to normal, with mean* 0 *and variance* 1. *The convergence is uniform, in the following sense. Fix any weak* neighborhood* $\mathscr{N}$ *of the* $N(0, 1)$ *distribution. Then there is a small positive number* $\delta = \delta(\mathscr{N})$ *such that the distribution of* $S_\sigma/\alpha^{\frac{1}{2}}$ *falls in* $\mathscr{N}$ *for any system* $\{X_i, \mathscr{F}_i\}$ *satisfying* (1.1), *any stopping time* σ, *and any real number* α, *provided*

$$\alpha > 1/\delta \qquad \text{and} \qquad P\{\alpha < T_\sigma < (1 + \delta)\alpha\} > 1 - \delta\,.$$

This will be proved in Section 7. A similar application could be made to the invariance principle ([11], pages 89 ff). A variant of (1.15) will establish the convergence of the finite-dimensional joint distributions. A variant of (1.6) will establish tightness, using the lemma of Arzela-Ascoli. Donsker's original argument will then prove the following variant of his invariance principle. Let

$C[0, 1]$ be the space of continuous functions on $[0, 1]$, endowed with the uniform topology. Let the probability π on $C[0, 1]$ be the distribution of Brownian motion, with time parameter confined to $[0, 1]$. Let φ be a bounded, measurable function on $C[0, 1]$, which is continuous π—almost everywhere. If $\mathbf{X} = \{X_i, \mathscr{F}_i\}$ satisfies (1.1), define $\mathbf{X}(t)$ by the requirements that $\mathbf{X}(T_n) = S_n$ and $\mathbf{X}(\cdot)$ is linearly interpolated. As will be seen, $\sum X_i$ converges a.e. on $\{\sum V_i < \infty\}$; define $\mathbf{X}(t) = \sum X_i$ for $t > \sum V_i$. Let $X^*(t) = \mathbf{X}(t)$ for $0 \leqq t \leqq 1$, so X^* is a random element of $C[0, 1]$.

(1.16) COROLLARY. *For any $\varepsilon > 0$, there is a $\delta > 0$ such that*

$$|E\{\varphi(X^*)\} - \int \varphi \, d\pi| < \varepsilon$$

for any system $\{X_i, \mathscr{F}_i\}$ satisfying (1.1), *subject to the conditions $P\{\sum_1^\infty V_i < 1\} < \delta$ and $|X_n| < \delta$ for all n with $T_{n-1} < 1$.*

Similarly, Chover's argument [4] for Strassen's invariance principle can be adapted to martingales, although the Skorokhod embedding works too ([11] pages 89 ff).

2. The growth condition. Many of the inequalities can be extended to variables which are individually bounded, where the bound is not uniform but grows at a controlled rate. Certain unbounded variables can then be handled by truncation. As in Section 1, assume X_n is $\mathscr{F}$-measurable and $E\{X_n \mid \mathscr{F}_{n-1}\} = 0$; do not assume the X_n are bounded.

(2.1) PROPOSITION. *Let τ be a stopping time, and K a positive real number. Suppose $P\{|X_i| \leqq K \text{ for } i \leqq \tau\} = 1$. Then for all positive real numbers a and b,*

$$P\{S_n \geqq a \text{ and } T_n \leqq b \text{ for some } n \leqq \tau\} \leqq \left[\left(\frac{b}{Ka + b}\right)^{Ka+b} e^{Ka}\right]^{1/K^2}$$

$$\leqq \exp\left[-\frac{a^2}{2(Ka + b)}\right].$$

This follows from (1.6): replace X_i by X_i/K for $i \leqq \tau$, and by 0 for $i > \tau$. To continue, it is helpful to make the

(2.2) DEFINITION. Let σ_b be the least n if any with $T_n > b$, and $\sigma_b = \infty$ if none. Let

$$L(b) = \text{ess sup}_\omega \sup_{n \leqq \sigma_b(\omega)} |X_n(\omega)| .$$

This $L(b)$ is not random, $0 \leqq L(b) \leqq \infty$, and L is non-decreasing. Previous results can be extended to processes which satisfy a growth condition of the form $L(b) = O[\phi(b)]$, where ϕ is some function which increases to ∞ slowly. To make this kind of condition clearer, check that $L(b)$ is the smallest real number for which $i \leqq n$ and $T_n(\omega) > b \geqq T_{n-1}(\omega)$ imply $|X_i(\omega)| \leqq L(b)$, almost surely. So L is the smallest non-decreasing function such that $|X_{n+1}(\omega)| \leqq L[T_n(\omega)]$ for $n = 0, 1, \cdots$ almost surely. And $L = o(\phi)$ means: there is a

function ε(b) of b, tending to 0 at ∞, but with $\varepsilon(b)\phi(b)$ non-decreasing, and $|X_{n+1}(\omega)| \leqq \varepsilon[T_n(\omega)]\phi[T_n(\omega)]$ for all n.

As is immediate from (1.9),

(2.3) PROPOSITION. *$P\{W_a \geqq b\} < 5[a + L(b)]/b^{\frac{1}{2}}$ for positive a and b.*

As is immediate from (1.10),

(2.4) PROPOSITION. *Let $0 < \delta < \frac{1}{3}$. Let a and b be positive, with $L(b)$ finite, $b/a > 9L(b)/\delta^2$, and $a^2/b > (16/\delta^2)\log(64/\delta^2)$. Then*

$$P\{W_a < b\} > \tfrac{1}{2}\exp[-\tfrac{1}{2}(a^2/b)(1 + 4\delta)]\,.$$

As is immediate from (1.14),

(2.5) PROPOSITION. *Let $0 < \alpha < \beta < \infty$. Let τ be a stopping time with $P\{\alpha < T_\tau < \beta\} = 1$. Then*

$$\exp\left\{\frac{\alpha}{L(\beta)^2} f[\lambda L(\beta)]\right\} \leqq E\{\exp(\lambda S_\tau)\} \leqq \exp\left\{\frac{\beta}{L(\beta)^2} e[\lambda L(\beta)]\right\}.$$

Arguing as for (4.5),

(2.6) PROPOSITION. *Suppose $L(b) = o(b^{\frac{1}{2}})$ as $b \to \infty$. Then $\sup_n S_n = \infty$ and $\inf_n S_n = -\infty$, a.e. on $\{\sum V_i = \infty\}$.*

Of course, S_n converges to a finite limit S_∞ as $n \to \infty$, a.e. on $\{\sum V_i < \infty\}$, without any growth condition.

The result (2.6) is due to Lévy ([16], page 248 footnote). I do not know what happens when $L(b) = o(b^{\frac{1}{2}})$, but the result is false when $L(b) = o(b^{\frac{1}{2}+\varepsilon})$.

Similarly,

(2.7) PROPOSITION. *Suppose $L(b) = o(b/\log\log b)^{\frac{1}{2}}$ as $b \to \infty$. Then*

$$\limsup_{n\to\infty} S_n/(2T_n \log\log T_n)^{\frac{1}{2}} = 1 \qquad \text{a.e. on} \quad \{\textstyle\sum V_i = \infty\}\,.$$

This can be argued like (6.1) and (6.3). There is one new difficulty: in estimating $P\{W_a < b\}$ for the process shifted by σ_k, with $a = (1 - r^{-\frac{1}{2}})\phi(r^k)$ and $b = r^{k+1} - r^k$, check that $L(b)$ for the shifted process is at most $L(r^{k+1})$ for the old process, so $b/(L(b)a) \to \infty$ as $k \to \infty$.

The growth condition $L(b) = o(b/\log\log b)^{\frac{1}{2}}$ is the one used by Kolmogorov (1929) for the independent case. In a sense, it is best possible, as shown by an example of Marcinkiewicz and Zygmund (1937). For a discussion of other growth conditions in the independent case, see Feller, (1943). The iterated logarithm for martingales with this growth condition was established by Stout (1970), in the main special case where $\sum V_i = \infty$ a.e.

(2.8) PROPOSITION. *The central limit theorem* (1.15) *and the invariance principle* (1.16) *hold with the growth condition $L(b) = o(b^{\frac{1}{2}})$ as $b \to \infty$.*

The condition $L(b) = o(b^{\frac{1}{2}})$ is best possible, even in the case of independence:

because the Lindeberg condition is necessary (Feller (1966) Section XV. 6). Suppose, for instance, that the X_n are independent, and X_n takes the three values 0 and $\pm n^{-\frac{1}{2}}$, with $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = \pm n^{\frac{1}{2}}) = \frac{1}{2}n^{-1}$. So $V_n = 1$, $T_n = n$, and $L(b) \sim b^{\frac{1}{2}}$. But $S_n/n^{\frac{1}{2}}$ does not converge to normal. Theorem (2.8) is a little sharper than Lévy's result ([16] page 246), since τ is a more general stopping time than Lévy's T.

3. The transformation. This section proves the main technical estimates corresponding to (1.3)–(1.5).

(3.1) LEMMA. *Let* $g(0) = \frac{1}{2}$ *and* $g(x) = (e^x - 1 - x)/x^2$ *for* $x \neq 0$. *Then* g *is increasing.*

PROOF. Check that $g'(0) = 0$ and $x^3g'(x) = h(x)$, where $h(x) = xe^x - 2e^x + x + 2$. So $h''(x) = xe^x$, and h' is increasing on $(0, \infty)$, decreasing on $(-\infty, 0)$. But $h'(0) = 0$, so h' is nonnegative, and h is increasing. But $h(0) = 0$, so $h(x)$ has the same sign as x, and $g'(x) > 0$ for $x \neq 0$. □

(3.2) COROLLARY. $\exp(\lambda x) \leqq 1 + \lambda x + x^2e(\lambda)$ *for* $\lambda \geqq 0$ *and* $x \leqq 1$.

PROOF. $g(\lambda x) \leqq g(\lambda)$, by (3.1). □

(3.3) PROPOSITION. $\int_{[\sigma<\infty]} Q_\lambda(T_\sigma, S_\sigma)\, dP \leqq 1$ *for* $\lambda \geqq 0$ *and any stopping time* σ, *provided*

(3.4) $X_n \leqq 1$ a.e. *and* $E\{X_n \mid \mathscr{F}_{n-1}\} \leqq 0$ a.e. *for all* n.

No lower bound is assumed on X_n.

PROOF. It is enough to show that $\{Q_\lambda\{T_n, S_n\}, \mathscr{F}_n\}$ is an expectation-decreasing martingale. This follows from

(3.5) $E\{\exp(\lambda X)\} \leqq 1 + e(\lambda) \operatorname{Var} X \leqq \exp[e(\lambda) \operatorname{Var} X]$ for $\lambda \geqq 0$ and random variables X satisfying $X \leqq 1$ and $E(X) \leqq 0$.

It is enough to prove (3.5) when $E(X^2) < \infty$. Let μ be the distribution of X, so $\mu = p\mu_0 + (1 - p)\mu_-$, where $0 \leqq p \leqq 1$ and μ_0 is a probability on $(-\infty, 1]$ with $\int x\mu_0(dx) = 1$ and μ_- is a probability on $(-\infty, 0)$. Let $v_0 = \int x^2\mu_0(dx)$ and $v_- = \int x^2\mu_-(dx) - (\int x\mu_-(dx))^2$. Then

$$\begin{aligned} E\{\exp(\lambda X)\} &= p \int \exp(\lambda x)\mu_0(dx) + (1 - p) \int \exp(\lambda x)\mu_-(dx) \\ &\leqq p[1 + e(\lambda)v_0] + (1 - p)[1 + e(\lambda)v_-] \\ &\leqq 1 + e(\lambda) \operatorname{Var} X, \end{aligned}$$

using (3.2) on the first term in the first line, and Schwarz's inequality to check $pv_0 + (1 - p)v_- \leqq \operatorname{Var} X$. □

In a sense, $e(\lambda)$ is best possible:

$$\sup \frac{1}{\operatorname{Var} X} [E\{\exp(\lambda X)\} - 1] = \sup \frac{1}{\operatorname{Var} X} \log E\{\exp(\lambda X)\} = e(\lambda),$$

where the sup is taken over all X with $X \leqq 1$ and $E(X) \leqq 0$. The extreme X take only the values $-\varepsilon$ and 1, and have mean 0.

(3.6) PROPOSITION. *$E\{R_\lambda(T_\sigma, S_\sigma)\} \geqq 1$ for any $\lambda \geqq 0$ and uniformly bounded stopping time σ, assuming condition* (1.1).

PROOF. This follows from

(3.7) $E\{\exp(\lambda X)\} \geqq \exp\{f(\lambda)\ \mathrm{Var}\ X\}$, for random variables X with $X \geqq -1$ and $E(X) = 0$.

Inequality (3.7) can be proved by cases, as follows.

Case 1, *in which X takes only the two values -1 and $a > 0$.* Then $P\{X = -1\} = a/(1 + a)$ and $P\{X = a\} = 1/(1 + a)$, so $E\{X^2\} = a$. Let $\theta(\lambda) = E\{\exp(\lambda X)\} - \exp\{f(\lambda)\ \mathrm{Var}\ X\}$. Check that $\theta(0) = 0$ and $\theta'(\lambda) \geqq 0$.

Case 2, *in which X takes only the two values $-b$ and a, with positive a, b and $b \leqq 1$.* There is a $c \geqq 1$ with $bc = 1$. Let $Y = cX$ and use Case 1 on Y. So $E\{\exp(\lambda X)\} = E\{\exp[(\lambda/c)Y]\} \geqq \exp[f(\lambda/c)\ \mathrm{Var}\ Y] = \exp[f(\lambda/c)c^2\ \mathrm{Var}\ X] = \exp[\lambda^2 \cdot f(\lambda/c) \cdot (\lambda/c)^{-2}\ \mathrm{Var}\ X] \geqq \exp[\lambda^2 \cdot f(\lambda) \cdot \lambda^{-2}\ \mathrm{Var}\ X] = \exp[f(\lambda)\ \mathrm{Var}\ X]$. The last inequality follows from (3.1), since $f(x)/x^2 = g(-x)$ and $-\lambda \leqq -\lambda/c$.

The general case. Let μ be the distribution of X, so $\mu = \int \mu_\alpha \theta(d\alpha)$, where μ_α is a probability on $[-1, \infty)$ residing on two points and having mean 0. Let $v_\alpha = \int x^2 \mu_\alpha(dx)$, so $\mathrm{Var}\ X = \int v_\alpha \theta(d\alpha)$. Let $\phi_\alpha(\lambda) = \int \exp(\lambda x)\mu_\alpha(dx)$. Use Case 2 on μ_α, followed by Jensen's inequality: $E\{\exp(\lambda X)\} = \int \phi_\alpha(\lambda)\theta(d\alpha) \geqq \int \exp[f(\lambda)v_\alpha]\theta(d\alpha) \geqq \exp[f(\lambda) \cdot \int v_\alpha \theta(d\alpha)] = \exp[f(\lambda)\ \mathrm{Var}\ X]$. □

Again, $f(\lambda)$ is best possible. Of course, (3.6) holds if $\{X_n\}$ satisfies a weaker condition than (1.1), namely:

(3.8) $X_n \geqq -1$ a.e. and $E\{X_n \mid \mathscr{F}_n\} = 0$ a.e. for all n.

A similar argument will prove (3.6) under the alternative condition

(3.9) $|X_n| \leqq 1$ a.e. and $E\{X_n \mid \mathscr{F}_{n-1}\} \geqq 0$ a.e. for all n,

even if V_n is replaced by the larger quantity $E\{X_n^2 \mid \mathscr{F}_{n-1}\}$. This can be used to sharpen some later inequalities, but not in any very interesting way.

4. Inequalities. This section proves results (1.6) to (1.12).

(4.1) THEOREM. *Suppose* (3.4). *For any positive numbers a and b,*

$$P\{S_n \geqq a \text{ and } T_n \leqq b \text{ for some } n = 1, 2, \cdots\} \leqq \left(\frac{b}{a+b}\right)^{a+b} e^a \leqq \exp\left[-\frac{a^2}{2(a+b)}\right].$$

Let σ be the least n if any with $S_n \geqq a$, and $\sigma = \infty$ if none. Let $A = \{S_n \geqq a$ and $T_n \leqq b$ for some $n\}$. Then $\sigma < \infty$ and $S_\sigma \geqq a$ and $T_\sigma \leqq b$ on A. Use (3.3)

$$1 \geqq \int_A \exp[\lambda S_\sigma - e(\lambda)T_\sigma]\, dP \geqq P\{A\} \cdot \exp[\lambda a - e(\lambda)b],$$

so $P\{A\} \leqq \exp[-\lambda a + e(\lambda)b]$. The minimizing λ is $\log[(a+b)/b]$, which gives the first bound, and the second follows by calculus.

(4.2) THE PROOF OF (1.8). Confirm that $S_n \leqq a+1$ for $n \leqq \tau_a$. Then use (3.6) with $\tau_a \wedge n$ for σ:

$$1 \leqq E\{\exp[\lambda S_{\tau_a \wedge n} - f(\lambda)T_{\tau_a \wedge n}]\} \leqq \exp[\lambda(a+1)] \cdot \{\exp[-f(\lambda)T_{\tau_a \wedge n}]\}.$$

Let $n \to \infty$ and use dominated convergence. □

(4.3) EXAMPLE. Theorem (4.2) is false just assuming (3.8). For instance, make the X_n independent, X_n being -1 or n^2 and having mean 0. Then $P\{X_n = n^2\} = 1/(n^2+1)$, so $P\{X_n = -1 \text{ for all } n\} > 0$, and $P\{W_a = \infty\} > 0$. This contradicts (4.2), by letting λ approach 0.

(4.4) THE PROOF OF (1.9). Use Chebychev's inequality on (1.8) to see

$$P\{W_a \geqq b\} = P\{1 - \exp[-f(\lambda)W_a] \geqq 1 - \exp[-f(\lambda)b]\}$$
$$\leqq \frac{1 - \exp[-\lambda(a+1)]}{1 - \exp[-f(\lambda)b]}.$$

Putting $\lambda = b^{-\frac{1}{2}}$ gives

$$1 - \exp[-\lambda(a+1)] < \lambda(a+1) = (a+1)/b^{\frac{1}{2}}$$

and

$$1 - \exp[-f(\lambda)b] > f(\lambda)b - \tfrac{1}{2}f(\lambda)^2 b^2$$
$$> \left(\frac{\lambda^2}{2} - \frac{\lambda^3}{6}\right)b - \frac{\lambda^4 b^2}{8}$$
$$= \tfrac{1}{2} - \tfrac{1}{8} - \tfrac{1}{6}b^{-\frac{1}{2}}$$
$$> \tfrac{5}{24} \quad \text{for} \quad b > 1.$$

If $b \leqq 1$, the bound is at least 5, dwarfing $P\{W_a \geqq b\}$. □

(4.5) COROLLARY. *This proves the following result of Lévy* ([16] Theorem 68 on page 247). *Suppose* (1.1).

(a) $\sup_n S_n = \infty$ *and* $\inf_n S_n = -\infty$, a.e. *on* $\{\sum V_i = \infty\}$
(b) S_n *converges to a finite limit* S_∞ *as* $n \to \infty$, a.e. *on* $\{\sum V_i < \infty\}$.

PROOF. *Claim* (a). By symmetry, only the sup need be argued. (4.2) implies, $P\{W_a < \infty\} = 1$. So $\{\tau_a = \infty \text{ and } \sum V_i = \infty\} \subset \{W_a = \infty\}$ has probability 0. That is, $P\{S_n < a \text{ for all } n \text{ and } \sum V_i = \infty\} = 0$. Claim (b) holds even for unbounded X_i by Kolmogorov's inequality. □

(4.6) THE PROOF OF (1.12). Relation (3.3) shows

$$\int_{\{\tau_a < \infty\}} \exp[\lambda S_{\tau_a} - e(\lambda)W_a]\, dP \leqq 1. \tag{4.7}$$

But $S_{\tau_a} \geqq a$ on $\{\tau_a < \infty\}$. So

$$\int_{\{\tau_a < \infty\}} \exp[-e(\lambda)W_a]\, dP \leqq \exp(-\lambda a). \tag{4.8}$$

And $W_a = \infty$ a.e. on $\{\tau_a = \infty\}$, because $\sum V_i = \infty$ a.e. So $\exp[-e(\lambda)W_a] = 0$ a.e. on $\{\tau_a = \infty\}$, and the integral can be extended over the whole space. □

This result is false if, for instance, all the V_i vanish. The next main step is proving (1.10), which estimates $P\{W_a < b\}$ from below when a is the order $(b \log\log b)^{\frac{1}{2}}$. The proof is disappointingly hard. The main analytical difficulty is isolated in (4.10): here is a preliminary

(4.9) LEMMA. *If $\alpha > 0$ and $x > 2\log\alpha$, then $e^x > \alpha x$.*

PROOF. By calculus, $\alpha - 2\log\alpha$ has its minimum at $\alpha = 2$ and is positive there. By more calculus, $e^x - \alpha x$ increases with x for $x > \log\alpha$; but this function is positive at $x = 2\log\alpha$ by the previous remark. □

The proof of the next result is hard, and can be skipped without much loss.

(4.10) PROPOSITION. *For each $a > 0$, let W_a be a nonnegative random variable. Suppose that for all $\lambda \geqq 0$,*

$$E\{\exp[-e(\lambda)W_a]\} \leqq \exp(-\lambda a) \tag{4.11a}$$

$$E\{\exp[-f(\lambda)W_a]\} \geqq \exp(-\lambda(a+1)) . \tag{4.11b}$$

Let δ, a, b be positive, with

$$\delta < \tfrac{1}{3} \tag{4.12a}$$

$$\frac{b}{a} > \frac{9}{\delta^2} \tag{4.12b}$$

$$\frac{a^2}{b} > \frac{16}{\delta^2}\log\frac{64}{\delta^2} . \tag{4.12c}$$

Then

$$P\{W_a < b\} > \tfrac{1}{2}\exp[-(\tfrac{1}{2} + 2\delta)a^2/b] . \tag{4.13}$$

PROOF. Let

$$\phi(\lambda) = \frac{\lambda^2}{2} > \frac{\lambda^3}{6} \quad \text{and} \quad \lambda = (1+\delta)\frac{a}{b} \quad \text{and} \tag{4.14}$$

$$k = \frac{a^2}{b} \quad \text{and} \quad N = \frac{2}{\delta^2} .$$

Now $0 < \phi < f$, so $E\{\exp[-\phi(\lambda)W_a]\} > \exp[-\lambda(a+1)]$, by (4.11b). Integrating by parts,

$$\phi(\lambda)\int_0^\infty P\{W_a < x\}\exp[-\phi(\lambda)x]\,dx > \exp[-\lambda(a+1)] . \tag{4.15}$$

The interval of integration $[0, \infty)$ in (4.15) splits into five subintervals:

$$I_1 = [0, Na] \quad \text{and} \quad I_2 = (Na, (1-2\delta)b] \quad \text{and} \quad I_3 = (b, 2b] ,$$
$$I_4 = (2b, \infty) \quad \text{and} \quad I_5 = ((1-2\delta)b, b]) .$$

Let

$$\eta_i = \phi(\lambda)\int_{I_i} P\{W_a < x\}\exp[-\phi(\lambda)x]\,dx .$$

Of these, the terms with $i \leqq 4$ are errors. As will be seen,

$$\eta_1 < (\tfrac{1}{8} - \tfrac{1}{50})\exp[-(1+\delta)k] \tag{4.16:1}$$

$$\eta_i < \tfrac{1}{8}\exp[-(1+\delta)k] \qquad \text{for } i = 2, 3, 4\,. \tag{4.16:i}$$

The main term is η_5. As will be seen,

$$\eta_5 < P\{W_a < b\} \cdot \exp[(-\tfrac{1}{2} + 2\delta^2)k]\,. \tag{4.16:5}$$

For now, suppose (4.16). Look at the right side of (4.15). Check

$$\begin{aligned}\exp[-\lambda(a+1)] &= \exp\left[-\left(1+\delta\right)\frac{a}{b}\right] \cdot \exp[-(1+\delta)k] && \text{by (4.14)}\\ &> \exp(-\tfrac{1}{50}) \cdot \exp(-(1+\delta)k] && \text{by (4.12)}\\ &> (1 - \tfrac{1}{50})\exp[-(1+\delta)k]\,.\end{aligned}$$

Using (4.15) and (4.16:1–4):

$$\eta_5 > \exp[-\lambda(a+1)] - \eta_1 - \eta_2 - \eta_3 - \eta_4 > \tfrac{1}{2}\exp[-(1+\delta)k]\,.$$

Using (4.16:5)

$$\begin{aligned}P\{W_a < b\} &> \exp[(\tfrac{1}{2} - 2\delta^2)k] \cdot \eta_5\\ &> \tfrac{1}{2}\exp[-(\tfrac{1}{2} + \delta + 2\delta^2)k]\\ &> \tfrac{1}{2}\exp[-(\tfrac{1}{2} + 2\delta)k] && \text{by (4.12a)}\,.\end{aligned}$$

This reduces the task of proving (4.13) to the task of proving (4.16). There are two facts which will help. First,

$$\exp(-\delta^2 k/8) < 1/(8k)\,, \tag{4.17}$$

—which is a special case of (4.9). Second,

$$P\{W_a < x\} < \exp\left[-\frac{a^2}{2(a+x)}\right]. \tag{4.18}$$

Using Chebychev's inequality on (4.11a),

$$\begin{aligned}P\{W_a < x\} &= P\{\exp[-e(\lambda)W_a] > \exp[-e(\lambda)x]\}\\ &< \exp[e(\lambda)x] \cdot E\{\exp[-e(\lambda)W_a]\}\\ &< \exp[-\lambda a + e(\lambda)x]\,.\end{aligned}$$

The argument for (4.18) is completed as in (4.1).

THE PROOF OF (4.16:1). If $x \leqq Na$, then

$$P\{W_a < x\} < \theta = (\tfrac{1}{8} - \tfrac{1}{50})\exp[-(1+\delta)]\,. \tag{4.19}$$

This follows from (4.18): if $x \leqq Na$, then $a^2/(a+x) \geqq a/(N+1)$. Next, $N + 1 < 3/\delta^2$, so $a^2/(a+x) > (\delta^2/3)a = (\delta^2/3)(b/a)k > 3k > (1 + \delta + \frac{1}{8})k$; and $k/6 > 5$, so $\exp(-k/6) < 2^{-5} < \frac{1}{8} - \frac{1}{50}$. This settles (4.19). To complete the proof of (4.16:1),

$$\eta_1 < \theta \cdot \phi(\lambda) \cdot \textstyle\int_0^{Na} \exp[-\phi(\lambda)x]\,dx < \theta \cdot \phi(\lambda) \cdot \int_0^{\infty} \exp[-\phi(\lambda)x]\,dx = \theta\,.$$

THE PROOF OF (4.16:2). Confirm that $Na < (1 - 2\delta)b$, so I_2 is proper. If $x \geqq Na$, then $a^2/(a + x) \geqq a^2/x(1 + N^{-1}) = (N/N + 1)a^2/x$. Use (4.18) and (4.14):

$$\eta_2 < \exp[\tfrac{1}{6}\lambda^3(1 - 2\delta)b] \cdot \phi(\lambda) \cdot \gamma\,, \tag{4.20}$$

where $\gamma = \int_{I_2} \exp[-\theta(x)]\,dx$, and

$$\theta(x) = \frac{\lambda^2}{2}x + \frac{N}{N+1}\frac{a^2}{2x}\,. \tag{4.21}$$

By calculation,

(4.22) θ is strictly convex on $(0, \infty)$, and takes its minimum at $x_m = (N/N + 1)^{\frac{1}{2}}a/\lambda = (N/N + 1)^{\frac{1}{2}}b/(1 + \delta)$, and $(1 - \delta)b < x_m < b$.

This shows that the minimum of θ on I_2 occurs at the right endpoint $(1 - 2\delta)b$, and is μk, where

$$\mu = \frac{1}{2}\left[(1 + \delta)^2(1 - 2\delta) + \frac{N}{N+1}\frac{1}{1 - 2\delta}\right].$$

By calculation,

$$\mu > 1 + \delta + \tfrac{1}{4}\delta^2\,. \tag{4.23}$$

So

$$\theta(x) > (1 + \delta + \tfrac{1}{4}\delta^2)k \qquad \text{for } x \in I_2\,. \tag{4.24}$$

Return to (4.20). The length of I_2 is less than $(1 - 2\delta)b$. So (4.24) shows

$$\gamma < (1 - 2\delta)b \exp(-\tfrac{1}{4}\delta^2 k) \exp[-(1 + \delta)k]\,. \tag{4.25}$$

But $\phi(\lambda)(1 - 2\delta)b < \frac{1}{2}\lambda^2(1 - 2\delta)b = \frac{1}{2}(1 + \delta)^2(1 - 2\delta)k < \frac{1}{2}k$. So

$$\phi(\lambda)\gamma < \tfrac{1}{2}k \exp(-\tfrac{1}{4}\delta^2 k) \exp[-(1 + \delta)k]\,. \tag{4.26}$$

To estimate the first factor in (4.20), check $\frac{1}{6}\lambda^3(1 - 2\delta)b = \frac{1}{6}(1 + \delta)^3(1 - 2\delta)(a/b)k < \frac{1}{6}(1 + \delta)(a/b)k < \frac{1}{6} \cdot \frac{4}{3} \cdot (\delta^2/9)k$. So

$$\exp[\tfrac{1}{6}\lambda^3(1 - 2\delta)b] < \exp(\tfrac{1}{8}\delta^2 k)\,. \tag{4.27}$$

Combining this with (4.26) shows $\eta_2 < \alpha \exp[-(1 + \delta)k]$, where $\alpha = \frac{1}{2}k \exp(\frac{1}{8}\delta^2 k)\exp(-\frac{1}{4}\delta^2 k) < \frac{1}{8}$ by (4.17), settling (4.16:2).

THE PROOF OF (4.16:3). As for (4.20),

$$\eta_3 < \exp(\tfrac{1}{6}\lambda^3 \cdot 2b) \cdot \phi(\lambda) \cdot \gamma\,, \tag{4.28}$$

where $\gamma = \int_{I_3} \exp[-\theta(x)]\,dx$, and θ is still defined by (4.21). By (4.22), the minimum of θ on I_3 occurs at the left endpoint b, and is μk, where $\mu = \frac{1}{2}[(1 + \delta)^2 + N/(N + 1)]$. By computation, $\mu > 1 + \delta + \frac{1}{4}\delta^2$. Consequently, $\gamma < b \cdot \exp[-(1 + \delta + \frac{1}{4}\delta^2)k]$. But $\phi(\lambda)b < \frac{1}{2}(1 + \delta)^2 k < k$; so

$$\phi(\lambda)\gamma < k \exp[-(1 + \delta + \tfrac{1}{4}\delta^2)k]\,. \tag{4.29}$$

Next

$$\exp(\tfrac{1}{6}\lambda^3 \cdot 2b) = \exp\left[\tfrac{1}{3}(1+\delta)^3 \frac{a}{b} k\right] \tag{4.30}$$
$$< \exp\left[\tfrac{1}{3}(1+\delta)^3 \frac{\delta^2}{9} k\right] < \exp(\tfrac{1}{8}\delta^2 k) .$$

So $\eta_3 < \alpha \exp[-(1+\delta)k]$, where $\alpha = k \cdot \exp(\frac{1}{8}\delta^2 k) \cdot \exp(-\frac{1}{4}\delta^2 k) < \frac{1}{8}$ by (4.17). This proves (4.16:3).

THE PROOF OF (4.16:4). Confirm that $\eta_4 \leqq \phi(\lambda) \int_{2b}^{\infty} \exp[-\phi(\lambda)x]\, dx = \exp[-2b\phi(\lambda)] = \exp(\frac{1}{6}\lambda^3 \cdot 2b) \cdot \exp[-(1+\delta)^2 k]$. The first factor in the last expression was shown to be less than $\exp(\frac{1}{8}\delta^2 k)$ in (4.30). Confirm that $\frac{1}{8}\delta^2 - (1+\delta)^2 < -k^{-1}\log 3 - (1+\delta)k$, completing the proof of (4.16:4).

THE PROOF OF (4.16:5). Since $P\{W_a < x\} \leqq P\{W_a < b\}$ for $x \leqq b$,

$$\begin{aligned}
\eta_5 &\leqq P\{W_a < b\} \cdot \int_{(1-2\delta)b}^{b} \phi(\lambda)\exp[-\phi(\lambda)x]\, dx \\
&< P\{W_a < b\} \cdot \int_{(1-2\delta)b}^{\infty} \phi(\lambda)\exp[-\phi(\lambda)x]\, dx \\
&= P\{W_a < b\} \cdot \exp[-\phi(\lambda)(1-2\delta)b] \\
&= P\{W_a < b\} \cdot \exp[\tfrac{1}{6}\lambda^3(1-2\delta)b] \cdot \exp[-\tfrac{1}{2}(1+\delta)^2(1-2\delta)k] .
\end{aligned}$$

In the last line, the second factor is less than $\exp(\frac{1}{8}\delta^2 k)$ by (4.27). And $-\frac{1}{2}(1+\delta)^2(1-2\delta) < -\frac{1}{2} + 2\delta^2 - \frac{1}{8}\delta^2$. This settles (3.16:5) and the theorem. □

(4.31) THE PROOF OF (1.10). First, suppose $\sum V_i = \infty$ a.e. Then (4.11a) holds by (4.6), and (4.11b) holds by (4.2). So (4.11) proves (1.10) in this special case. What happens in general? The trouble is, (4.11a) can fail. But it is possible to construct a new variable $W_a^* \geqq W_a$, which satisfies (4.11a-b). Then (4.11) can be used on W_a^*; and (1.10) holds *a fortiori*.

To construct W_a^*, introduce new variables $Y_1, Y_2, \cdots$ which are independent and ± 1 with probability $\frac{1}{2}$ each. Make all the Y_n's independent of all the $\mathscr{T}_n$'s, which may require an enlargement of the basic probability triple. As (4.5) shows, $\tau_a < \infty$ almost surely on $\{\sum V_i = \infty\}$, and $\sum X_i$ converges almost surely to a finite limit S_∞ on $\{\sum V_i < \infty\}$. So

(4.32) $\sum X_i$ converges to a finite limit S_∞ and $T_\infty = \sum V_i < \infty$, a.e. on $\{\tau_a = \infty\}$.

If $\tau_a = \infty$, let τ_a^* be the least $n = 0, 1, \cdots$ with $S_\infty + Y_1 + \cdots + Y_n \geqq a$. Let $W_a^* = W_a$ when $\tau_a < \infty$, and let $W_a^* = T_\infty + \tau_a^*$ when $\tau_a = \infty$. So $W_a^* \geqq W_a$.

I will now argue that W_a^* satisfies (4.11a). Use (3.3) with $\tau_a \wedge n$ for σ to see that $\int Q_\lambda(T_{\tau_a \wedge n}, S_{\tau_a \wedge n})\, dP \leqq 1$. Let $n \to \infty$ and use (4.32):

$$\int_{\{\tau_a < \infty\}} Q_\lambda(W_a, S_{\tau_a})\, dP + \int_{\{\tau_a = \infty\}} Q_\lambda(T_\infty, S_\infty)\, dP \leqq 1 . \tag{4.33}$$

Given $\{\tau_a = \infty\}$, the process $Q_\lambda(T_\infty + n, S_\infty + Y_1 + \cdots + Y_n)$ is an expectation-decreasing martingale relative to the σ-fields $\mathscr{T}_n^*$, where $\mathscr{T}_n^*$ is the σ-field

generated by all the $\mathscr{T}_i$ and by $Y_1, \cdots, Y_n$, as follows from (3.5). So

$$(4.34) \qquad \int_{\{\tau_a=\infty\}} Q_\lambda(T_\infty + \tau_a^*, S_\infty + \sum_1^{\tau_a^*} Y_i)\, dP \leqq \int_{\{\tau_a=\infty\}} Q_\lambda(T_\infty, S_\infty)\, dP .$$

Let $S^* = S_{\tau_a}$ when $\tau_a < \infty$, and $S^* = S_\infty + \sum_1^{\tau_a^*} Y_i$ when $\tau_a = \infty$. So (4.33) and (4.34) show that $\int Q_\lambda(W_a^*, S^*)\, dP \leqq 1$. But $S^* \geqq a$, and this proves (4.11a) for W_a^*.

The argument for (4.11b) is similar, using (3.6) and (3.7): since $S_n \leqq a + 1$ for $n \leqq \tau_a^*$, the passages to the limit can all be justified. □

5. The Tauberian argument. This section uses inequalities (1.8) and (1.12) to get a sharp estimate (1.13) of $P\{W_a \geqq b\}$ when $a = o(b^{\frac{1}{2}})$. The argument is a rearrangement of Feller's ((1966) Section XIII. 5) proof of the Tauberian theorem. The first lemma is standard.

(5.1) LEMMA. *For each n, let h_n be a non-increasing function on $(0, \infty)$. Suppose $\int_0^\infty e^{-\lambda x} h_n(x)\, dx \to (2/\lambda)^{\frac{1}{2}}$ as $n \to \infty$, for all positive λ. Then $h_n(x) \to [2/(\pi x)]^{\frac{1}{2}}$ as $n \to \infty$, for all positive x.*

(5.2) THE PROOF OF (1.13). Let $a_n \to \infty$. For each n consider a process which satisfies (1.1) and the additional condition that the sum of the conditional variances is almost surely infinite. Let W_{a_n} be the intrinsic time for the nth process to cross a_n, in the sense of (1.7). Let $Z_n = W_{a_n}/b_n$, and $h_n(x) = (b_n^{\frac{1}{2}}/a_n)P(Z_n \geqq x)$. I will argue that $h_n(x) \to (2/\pi x)^{\frac{1}{2}}$ for all positive x; putting $x = 1$ gives the theorem.

Let θ be the inverse function to $e(\cdot)$, and let ϕ be the inverse function to f, so $e^{\theta(\lambda)} - 1 - \theta(\lambda) = e^{-\phi(\lambda)} - 1 + \phi(\lambda) = \lambda$ for $\lambda > 0$. Check

$$(5.3) \qquad \theta(\lambda)/(2\lambda)^{\frac{1}{2}} \text{ and } \phi(\lambda)/(2\lambda)^{\frac{1}{2}} \text{ tends to 1 as } \lambda \text{ tends to 0.}$$

Let

$$\theta_n(\lambda) = (2\lambda)^{-\frac{1}{2}}(b_n^{\frac{1}{2}}/a_n)\{1 - \exp[-\theta(\lambda/b_n)a_n]\}$$
$$\phi_n(\lambda) = (2\lambda)^{-\frac{1}{2}}(b_n^{\frac{1}{2}}/a_n)\{1 - \exp[-\phi(\lambda/b_n)(a_n + 1)]\} .$$

Using (5.3),

$$(5.4) \qquad \theta_n(\lambda) \to 1 \text{ and } \phi_n(\lambda) \to 1 \text{ as } n \to \infty, \text{ for all positive } \lambda.$$

Inequalities (1.8) and (1.12) imply

$$(5.5) \qquad \exp[-\phi(\lambda/b_n)(a_n + 1)] \leqq E\{\exp(-\lambda Z_n)\} \leqq \exp[-\theta(\lambda/b_n)a_n] .$$

Integrating by parts, $E\{\exp(-\lambda Z_n)\} = 1 - \lambda \int_0^\infty e^{-\lambda x} h_n(x)\, dx$. So

$$(5.6) \qquad (2/\lambda)^{\frac{1}{2}}\theta_n(\lambda) \leqq \int_0^\infty e^{-\lambda x} h_n(x)\, dx \leqq (2/\lambda)^{\frac{1}{2}}\phi_n(\lambda) .$$

Let $n \to \infty$ and use (5.4) and (5.1). □

6. The law of the iterated logarithm. Inequalities (1.6) and (1.10) are strong enough to prove Levy's form of the iterated logarithm for martingales with uniformly bounded increments. Let $\phi(x) = (2x \log\log x)^{\frac{1}{2}}$ for $x \geqq e^e$ and $\phi(x) = 1$ for $x < e^e$.

(6.1) THEOREM. *Suppose condition* (3.4). *Then*

$$\limsup_{n\to\infty} S_n/\phi(T_n) \leqq 1 \qquad \text{a.e. on} \quad \{\sum V_i = \infty\}.$$

PROOF. Fix r just a little bigger than 1. The main step is to show that $P\{A\} = 0$, where A is the event that: $\sum V_i = \infty$, and $S_n > r^2\phi(T_n)$ for infinitely many n. Let σ_k be the sup of n with $T_n \leqq r^k$. If $\sum V_i = \infty$, then $\sigma_k < \infty$ and σ_k increases to ∞ with k. So $A \subset \limsup A_k$, where

$$A_k = \{\sum V_i = \infty \text{ and } S_n > r^2\phi(T_n) \text{ for some } n \text{ with } \sigma_k < n \leqq \sigma_{k+1}\}.$$

But $r^k < T_n \leqq r^{k+1}$ for $\sigma_k < n \leqq \sigma_{k+1}$. So $A_k \subset B_k$, where

$$B_k = \{S_n > r^2\phi(r^k) \text{ for some } n \text{ with } T_n \leqq r^{k+1}\}.$$

Now use (4.1) to estimate $P\{B_k\}$, with $r^2\phi(r^k)$ for a and r^{k+1} for b. Keep k so large that $a + b \leqq r^{k+3}$, which is feasible because $a = o(r^k)$: then $P\{B_k\} \leqq \exp[-a^2/2(a+b)] \leqq \exp[-r \log\log r^k] \leqq 1/k^r$, which sums on k because $r > 1$. So $P\{\limsup B_k\} = 0$. □

In one respect, this theorem is a bit stronger than classical results, even for independent variables: because there is no condition on the negative tail of the X_i. At first sight, large negative tails only reduce S_n, so allowing them is frivolous. But very small masses could be placed at faraway negative values, so as to realize them only finitely often. This does not affect the asymptotic behavior of S_n. Then compensating masses could be allowed at 1, to bring the expectation back up to 0. This tends to increase S_n, which is the point of generalization. However, it also increases V_n, in such a way that $S_n/\phi(T_n)$ decreases. For example, look at (4.3) with the signs reversed.

The next result is standard.

(6.2) LEMMA. *Suppose* (1.1). *Let τ be a stopping time. Then $\tau + n$ is a stopping time. Let $\mathscr{T}_\tau$ be the σ-field of events A such that $A \cap \{\tau = n\} \in \mathscr{T}_n$ for all $n \geqq 0$. For $n \geqq 0$, let $\mathscr{T}_n^* = \mathscr{T}_{\tau+n}$. For $n \geqq 1$, let $X_n^* = X_{\tau+n}$ and $V_n^* = V_{\tau+n}$ on $\{\tau < \infty\}$, while $X_n^* = V_n^* = 0$ on $\{\tau = \infty\}$.*

(a) *X_n^* is $\mathscr{T}_n^*$-measurable and V_n^* is $\mathscr{T}_{n-1}^*$-measurable, for $n \geqq 1$.*

(b) *$E\{X_n^* \mid \mathscr{T}_{n-1}^*\} = 0$ and $E\{X_n^{*2} \mid \mathscr{T}_{n-1}^*\} = V_n^*$ a.e., for $n \geqq 1$.*

Let $P_\tau(\omega, A)$ be a regular conditional probability on $(\Omega, \mathscr{T})$, given $\mathscr{T}_\tau$.

(c) *For P-almost all ω, the starred system satisfies* (1.1), *relative to the conditional probability triple $(\Omega, \mathscr{T}, P_\tau(\omega, \cdot))$.*

(6.3) THEOREM. *Suppose conditions* (1.1). *Then*

$$\limsup_{n\to\infty} S_n/\phi(T_n) \geqq 1 \qquad \text{a.e. on} \quad \{\sum V_i = \infty\}.$$

PROOF. Fix r much larger than 1. The point to argue is that $P\{A\} = 0$, where A is the event that: $\sum V_i = \infty$, and $S_n \geqq (1 - 3r^{-\frac{1}{2}})\phi(T_n)$ for only finitely many n. As before, let σ_k be the sup of n with $T_n \leqq r^k$, so σ_k is a stopping time, because V_n is $\mathscr{T}_{n-1}$-measurable. Let A_k be the event that $\sigma_k < \infty$ and

$S_n - S_{\sigma_k} \geqq (1 - r^{-\frac{1}{2}})\phi(r^{k+1})$ for some $n > \sigma_k$ with $T_n - T_{\sigma_k} < r^{k+1} - r^k$. Let

$$B_k = A_k \cup \{\sigma_k < \infty \text{ and } \sup_n T_n - T_{\sigma_k} < r^{k+1} - r^k\}.$$

Now B_k is the event $W_a < b$ for the shifted process $X_{\sigma_k+1}, X_{\sigma_k+2}, \cdots$ σ-fields $\mathscr{T}_{\sigma_k}, \mathscr{T}_{\sigma_k+1}, \cdots$ and variances $V_{\sigma_k+1}, V_{\sigma_k+2}, \cdots$ with $a = (1 - r^{-\frac{1}{2}})\phi(r^{k+1})$ and $b = r^{k+1} - r^k$, on $\{\sigma_k < \infty\}$. Choose a positive $\delta < \frac{1}{3}$ so small that $\theta = (1 + 4\delta)(1 - r^{-\frac{1}{2}})^2/(1 - r^{-1}) < 1$. Then keep k so large that conditions (4.12b-c) are satisfied. By (6.2), inequality (1.10) can be used on the shifted process to get

$$P\{B_k \mid \mathscr{T}_{\sigma_k}\} > \tfrac{1}{2}\exp(-\theta \log\log r^{k+1}) \qquad \text{on} \quad \{\sigma_k < \infty\}.$$

The bound is of order $1/k^\theta$: the sum on k diverges to ∞ because $\theta < 1$. And $\sum V_i = \infty$ implies $\sigma_k < \infty$ for all k. So Lévy's conditional form of the Borel-Cantelli lemma ([16] page 249; also see [5]) shows

$$P\{\textstyle\sum V_i = \infty \text{ and only finitely many } B_k \text{ occur}\} = 0.$$

Clearly, $A_k = B_k$ when $\sum V_i = \infty$. So

$$P\{\textstyle\sum V_i = \infty \text{ and only finitely many } A_k \text{ occur}\} = 0.$$

However, if $\sum V_i = \infty$ and infinitely many A_k occur, then $S_n > (1 - 3r^{-\frac{1}{2}})\phi(T_n)$ for infinitely many n. Indeed, suppose $\sigma_k < n$ and $S_n - S_{\sigma_k} \geqq (1 - r^{-\frac{1}{2}})\phi(r^{k+1})$ and $T_n - T_{\sigma_k} < r^{k+1} - r^k$. Then $T_{\sigma_k} \leqq r^k$, so $T_n < r^{k+1}$, so $\phi(T_n) < \phi(r^{k+1})$. Use (6.1) on $\{-X_i\}$. After discarding a null set, for all sufficiently large k,

$$S_{\sigma_k} > -2\phi(T_{\sigma_k}) > -2\phi(r^k) > -2r^{-\frac{1}{2}}\phi(r^{k+1}).$$

So $S_n > (1 - 3r^{-\frac{1}{2}})\phi(r^{k+1}) > (1 - 3r^{-\frac{1}{2}})\phi(T_n)$. □

7. The central limit theorem. This section presents a variant of Lévy's central limit theorem for martingales with uniformly bounded increments ([16] page 243). The first result is immediate from (4.5).

(7.1) LEMMA. *Suppose* (1.1). *Let* $S_\infty = \sum_{i=1}^\infty X_i$, *provided the sum converges to a finite limit. Let* σ *be a stopping time, with* $T_\sigma < \infty$ a.e. *Let* $S_\sigma = S_n$ *when* $\sigma = n$, *and* $S_\sigma = S_\infty$ *when* $\sigma = \infty$. *Then* S_σ *is defined and finite* a.e.

(7.2) THEOREM. *Suppose* (1.1). *Let* $\lambda \geqq 0$ *and* $0 \leqq \alpha < \beta < \infty$. *Let* σ *be a stopping time.*

(a) *If* $P\{T_\sigma \leqq \beta\} = 1$, *then* $E\{\exp(\lambda S_\sigma)\} \leqq \exp[\beta e(\lambda)]$.

(b) *If* $P\{\alpha \leqq T_\sigma \leqq \beta\} = 1$, *then* $E\{\exp(\lambda S_\sigma)\} \geqq \exp[\alpha f(\lambda)]$. *These bounds are sharp.*

PROOF. *Claim* (a). Inequality (3.3) shows $E\{Q_\lambda(T_{\sigma\wedge n}, S_{\sigma\wedge n})\} \leqq 1$. By assumption, $T_\sigma < \infty$ a.e. So $S_{\sigma\wedge n} \to S_\sigma$ a.e. as $n \to \infty$, by (7.1). Let $n \to \infty$ and use Fatou's lemma:

$$\exp[-\beta e(\lambda)] \cdot E\{\exp(\lambda S_\sigma)\} \leqq E\{\exp \lambda S_\sigma - e(\lambda)T_\sigma\} \leqq 1.$$

Claim (b). Inequality (3.6) shows $E\{R_\lambda(T_{\sigma\wedge n}, S_{\sigma\wedge n})\} \geqq 1$. Again, (7.1) makes

$R_\lambda(T_{\sigma\wedge n}, S_{\sigma\wedge n}) \to R_\lambda(T_\sigma, S_\sigma)$ a.e. as $n \to \infty$. And this sequence of variables is uniformly integrable, because

$$E\{R_\lambda(T_{\sigma\wedge n}, S_{\sigma\wedge n})^2\} \leqq E\{\exp(2\lambda S_{\sigma\wedge n})\} \leqq \exp[\beta e(2\lambda)],$$

from claim (a) on the stopping time $\sigma \wedge n$. Let $n \to \infty$:

$$\exp[-\alpha f(\lambda)] \cdot E\{\exp(\lambda S_\sigma)\} \geqq E\{\exp[\lambda S_\sigma - f(\lambda)T_\sigma]\} \geqq 1.$$

Sharpness. For (a), let the X_i be independent and identically distributed, taking the two values $-\varepsilon$ and 1, and having mean 0. Let $\sigma = N$ a.e., with $N\varepsilon = \beta$. Then S_σ is essentially centered Poisson. For (b), just reverse the signs. □

The next result is standard.

(7.3) LEMMA. *Let $(\Omega, \mathscr{T}, P)$ be a probability triple, and let $\mathscr{T}_0 \subset \mathscr{T}_1 \subset \cdots$ be sub-σ-fields of $\mathscr{T}$. For each n, let X_n be $\mathscr{T}_n$-measurable, and suppose $Y_n = E\{X_n \mid \mathscr{T}_{n-1}\}$ is defined* a.e. *Let τ be a stopping time. Let $X_n' = X_n$ and $Y_n' = Y_n$ for $n \leqq \tau$, while $X_n' = Y_n' = 0$ for $n > \tau$. Then X_n' is $\mathscr{T}_n$-measurable, Y_n' is $\mathscr{T}_{n-1}$-measurable, and $Y_n' = E\{X_n' \mid \mathscr{T}_{n-1}\}$.*

(7.4) LEMMA. *Suppose* (1.1) *and suppose $(\Omega, \mathscr{T}, P)$ supports a uniform random variable independent of all the $\mathscr{T}_n$. Let $0 < \delta < 1$ and $0 < \alpha < \beta < \infty$. Let τ be a uniformly bounded stopping time, with $P\{\alpha < T_\tau < \beta\} > 1 - \delta$. Then on the same probability triple there are variables X_n^*, σ-fields $\mathscr{T}_n^*$, and variances V_n^*, satisfying* (1.1), *and a stopping time τ^* relative to $\{\mathscr{T}_n^*\}$, with $P\{S_{\tau^*}^* \neq S_\tau\} < \delta$ and $P\{\alpha < T_{\tau^*}^* < \beta\} = 1$, where $S_n^* = X_1^* + \cdots + X_n^*$ and $T_n^* = V_1^* + \cdots + V_n^*$.*

PROOF. Suppose $\tau \leqq k$. Let τ' be the max of $n \leqq \tau$ with $T_n < \beta$. Then τ' is also a stopping time, because V_n is $\mathscr{T}_{n-1}$-measurable. Clearly, $T_{\tau'} < \beta$ everywhere and $\tau' = \tau$ on G.

Choose a positive ε so small that $\varepsilon^2 < \beta - \alpha$. Construct random variables $Y_1, Y_2, \cdots$ which are independent of one another and of all the $\mathscr{T}_n$, each variable taking the values $\pm\varepsilon$ with chance $\frac{1}{2}$. Let $\mathscr{T}_n^* = \mathscr{T}_n$ for $n \leqq k$, while $\mathscr{T}_n^*$ is the σ-field spanned by $\mathscr{T}_k$ and $Y_1, \cdots, Y_{n-k}$ for $n > k$. Let

$$\begin{aligned} X_n^* &= X_n \quad \text{and} \quad V_n^* = V_n && \text{for } n \leqq \tau' \\ X_n^* &= V_n^* = 0 && \text{for } \tau' < n \leqq k \\ X_n^* &= Y_{n-k} \quad \text{and} \quad V_n^* = \varepsilon^2 && \text{for } n > k. \end{aligned}$$

With the help of (7.3) for $n \leqq k$, check that the starred system satisfies (1.1).

Let τ^* be the least $n \geqq k$ with $T_n^* > \alpha$, so τ^* is a stopping time relative to $\{\mathscr{T}_n^*\}$. Clearly, $T_{\tau^*}^* > \alpha$ everywhere. Check $T_{\tau^*}^* < \beta$ everywhere, and

$$\tau^* = k \quad \text{and} \quad S_{\tau^*}^* = \textstyle\sum_1^k X_n^* = \sum_1^{\tau'} X_n = S_{\tau'} = S_\tau \quad \text{on} \quad G. \qquad \square$$

(7.5) THE PROOF OF (1.15). Let $\delta_n > 0$ with $\delta_n \to 0$. Let $0 < \alpha_n < \beta_n < \infty$, with $\alpha_n \to \infty$ and $\beta_n/\alpha_n \to 1$. For each n, consider a process satisfying (1.1), and a stopping time τ_n. Let $S(n)$ be the sum of the variables in the nth process

up to τ_n, and let $T(n)$ be the sum of the conditional variances in the nth process up to τ_n. Suppose

$$P\{\beta_n < T(n) < \beta_n\} > 1 - \delta_n . \tag{7.6}$$

The problem is to show that the distribution of $S(n)/\alpha_n^{\frac{1}{2}}$ tends to $N(0, 1)$: the uniformity follows by a conventional *reductio ad absurdum*. Here, $S(n)$ is only partially defined: it is undefined when the stopping time is infinite and the sum of the variables fails to converge. However, as (4.5) shows, $S(n)$ is defined a.e. on $\{\alpha_n < T(n) < \beta_n\}$. There is no loss in assuming each τ_n to be uniformly bounded, say by k_n. Indeed, τ_n could be replaced by $\tau_n \wedge k_n$, with k_n so large that (7.6) still holds, and the distribution of $S(n)$ over $\{\alpha_n < T(n) < \beta_n\}$ is disturbed very little (weak *) by the truncation. Of course, $S(n)$ is now defined off $\{\alpha_n < T(n) < \beta_n\}$, but the set has probability less than $\delta_n \to 0$, so the overall change is weak * negligible.

There is no loss in assuming the existence of uniform variables independent of the processes, by enlarging the underlying probability triple. So (7.4) can be used to modify the nth process in such a way as to keep (1.1), to disturb the distribution of $S(n)$ very little, and to get the sum of the conditional variances strictly between α_n and β_n everywhere. Finally, since $\alpha_n f(\lambda/\alpha_n^{\frac{1}{2}})$ and $\beta_n e(\lambda/\alpha_n^{\frac{1}{2}})$ both tend to $\lambda^2/2$, convergence to the normal distribution follows from (7.2). □

REFERENCES

[1] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[2] BENNETT, G. (1963). On the probability of large deviations from the expectation for sums of bounded independent random variables. *Biometrika* **50** 528–535.

[3] BLACKWELL, D. and FREEDMAN, D. A. (1973). On the amount of variance needed to escape from a strip. *Ann. Probability* **1** 772–787.

[4] CHOVER, J. (1967). On Strassen's version of the log log law. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **8** 83–90.

[5] DUBINS, L. E. and FREEDMAN, D. A. (1965). A sharper form of the Borel–Cantelli lemma and the strong law. *Ann. Math. Statist.* **36** 800–807.

[6] DUBINS, L. E. and SAVAGE, L. J. (1965). *How to Gamble if You Must.* McGraw-Hill, New York.

[7] FELLER, W. (1943 a). Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.* **54** 361–372.

[8] FELLER, W. (1943 b). The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc.* **54** 373–402.

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

[10] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1** 3rd ed. Wiley, New York.

[11] FREEDMAN, D. A. (1971). *Brownian Motion and Diffusion.* Holden-Day, San Francisco.

[12] GODWIN, H. J. (1964). *Inequalities on Distribution Functions.* Griffin, New York.

[13] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 13–30.

[14] KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems.* Interscience, New York.

[15] KOLMOGOROV, A. N. (1929). Uber das Gesetz des iterierten logarithmus. *Math. Ann.* **101** 126–135.

[16] LÉVY, P. (1954). *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris. This was originally printed in 1937.

[17] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[18] MARCINKIEWICZ, J. and ZYGMUND, A. (1937). Remarque sur la loi du logarithme itéré. *Fund. Math.* **29** 215–222.

[19] STEIGER, W. (1969). A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Statist.* **40** 764–769.

[20] STOUT, W. (1970). A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **15** 279–290.

[21] STRASSEN, V. (1965). Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** Part 1, 315–343.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720