ON TAIL PROBABILITIES FOR MARTINGALES

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 (\cite{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 (\cite{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 It could also be used to prove Dvoretzky’s general central limit theorem (unpublished) for dependent summands: I hope to explore this elsewhere.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 (\cite{16} 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 \((\Omega, \mathcal{F}, P)\) be a probability triple. Let \(\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \ldots\) be an

\footnotesize

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-$\sigma$-fields of $\mathcal{F}$. Say $\tau$ is a stopping time if $\tau$ is a function on $\Omega$, which takes the values $0, 1, 2, \ldots, \infty$, such that $\{\tau = n\} \in \mathcal{F}_n$ for $n = 0, 1, 2, \ldots$; but $P(\tau = \infty) > 0$ is allowed. Let $X_1, X_2, \ldots$ be random variables on $(\Omega, \mathcal{F}, P)$, such that $X_n$ is $\mathcal{F}_n$-measurable. Let $V_n = \text{Var} \{X_n | \mathcal{F}_{n-1}\}$. Until further notice, assume

$$(1.1) \quad |X_n| \leq 1 \quad \text{and} \quad E[X_n | \mathcal{F}_{n-1}] = 0.$$ 

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

(1.2) \textbf{Definition.} $\lambda$ is a positive number.

(a) $e(\lambda) = e^\lambda - 1 - \lambda,$

(b) $f(\lambda) = e^{-\lambda} - 1 + \lambda,$

(c) $\exp x = e^x,$

(d) $Q(x, y) = \exp \{\lambda y - e(\lambda)v\},$

(e) $R(x, y) = \exp \{\lambda h - f(\lambda)v\},$

(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 $Q$ is superharmonic and $R$ subharmonic under (1.1). More exactly, let $y$ be real and $v$ nonnegative. Let $X$ be a random variable with $|X| \leq 1$ and $E(X) = 0$ and $E(X^2) = V$. Then

(1.3) \textbf{Lemma.}

(a) $E[\exp(\lambda X)] \leq \exp[V e(\lambda)],$

(b) $E[Q(x, y + X)] \leq Q(x, y),$

(c) $E[\exp(\lambda X)] \geq \exp[V f(\lambda)],$

(d) $E[R(x, y + X)] \geq R(x, y).$

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

(1.4) \textbf{Corollary.}

(a) $\{Q(T_n, S_n), \mathcal{F}_n: n = 0, 1, \ldots\}$ is an expectation-decreasing martingale;

(b) $\{R(T_n, S_n), \mathcal{F}_n: n = 0, 1, \ldots\}$ is an expectation-increasing martingale.

(1.5) \textbf{Main Inequality.} If $\sigma$ is a uniformly bounded stopping time, then

(a) $E[Q(T_\sigma, S_\sigma)] \leq 1$ and

(b) $E[R(T_\sigma, S_\sigma)] \geq 1.$
By appropriate limiting arguments, (1.5) can be extended to very general $\sigma$. All the inequalities in this paper come from (1.5), by proper choice of $\lambda$ and $\sigma$

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 $b$,

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

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

$$\lim \sup \frac{S_n}{(2T_n \log \log T_n)^{\frac{1}{4}}} \leq 1 \quad \text{a.e. on } \{\sum V_i = \infty\};$$

see (6.1).

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

(1.7)  **Definition.** Fix $a > 0$. Let $\tau_a$ be the least $n$ if any with $S_n \geq a$, 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 $\{W_a < b\} = \{S_n \geq a \text{ and } T_n < b \text{ for some } n\} \cup \{\sup_n T_n < b\}$. 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.** $E[\exp[-f(\lambda)W_a]] \geq \exp[-\lambda(a+1)]$ for $a > 0$.

To get this from (1.5b), use $\tau_a$ for $\sigma$. 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 $a = o(b^4)$.

(1.9)  **Corollary.** $P(W_a \geq b) < 5(a+1)/b^4$ 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 $(2\log \log b)^{\frac{1}{4}}$.

(1.10)  **Corollary.** Suppose $\delta, a, b$ 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

$$P(W_a < b) > \frac{1}{2} \exp[-(\frac{1}{3} + 2\delta)a^2/b].$$

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

$$\lim \sup \frac{S_n}{(2T_n \log \log T_n)^{\frac{1}{4}}} \geq 1 \quad \text{a.e. on } \{\sum V_i = \infty\}.$$

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

\begin{equation}
\Pr\{\sum_{i=1}^{n} V_i = \infty\} = 1.
\end{equation}

Then \(\Pr\{\tau_a < \infty\} = 1\), because (1.9) makes \(\Pr\{W_a < \infty\} = 1\). 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 \(E[\exp(-e(\lambda)W_a)] \leq \exp(-\lambda a)\) for \(a > 0\).

To prove this, use (1.5a) with \(\tau_a\) for \(\sigma\). There is a detailed argument in (4.6). Since \(e(\lambda)\) and \(f(\lambda)\) are very close for small \(\lambda\), 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 \(\varepsilon\), there is a finite \(c = c(\varepsilon)\) such that: \(a > c\) and \(b/a^2 > c\) imply

\begin{equation}
(1 - \varepsilon)(2/\pi)^{\frac{1}{2}} \leq (b^2/a^2)\Pr\{W_a \geq b\} \leq (1 + \varepsilon)(2/\pi)^{\frac{1}{2}}.
\end{equation}

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. \(0 \leq \alpha \leq \beta < \infty\). Let \(\sigma\) be a stopping time with \(\Pr\{\alpha \leq T_\sigma \leq \beta\} = 1\). Then

\begin{equation}
\exp[\alpha f(\lambda)] \leq E[\exp(\lambda S_\sigma)] \leq \exp[\beta e(\lambda)].
\end{equation}

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

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

\begin{equation}
\alpha > 1/\delta \quad \text{and} \quad \Pr[\alpha < T_\sigma < (1 + \delta)\alpha] > 1 - \delta.
\end{equation}

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 X = \{X_t, \mathcal{F}_t\} satisfies (1.1), define X(t) by the requirements that X(T_n) = S_n and X(\cdot) is linearly interpolated. As will be seen, \( \sum X_i \) converges a.e. on \( \{ \sum V_i < \infty \} \); define X(t) = \( \sum X_i \) for \( t > \sum V_i \). Let \( X^*(t) = X(t) \) for \( 0 \leq t \leq 1 \), so \( X^* \) is a random element of C[0, 1].

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

\[
|E[\phi(X^*)] - \int \phi \, d\pi| < \varepsilon
\]

for any system \( \{X_i, \mathcal{F}_i\} \) satisfying (1.1), subject to the conditions \( P[\sum_{i=1}^{n} V_i < 1] < \delta \) 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 \( \mathcal{F}\)-measurable and \( E[X_n | \mathcal{F}_{n-1}] = 0 \); do not assume the \( X_i \) are bounded.

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

\[
P[S_n \geq a \text{ and } T_n \leq b \text{ for some } n \leq \tau] \leq \left[ \frac{b}{Ka + b} \right]^{K^2} \exp \left[ -\frac{a^2}{2(Ka + b)} \right].
\]

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

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

\[L(b) = \text{ess sup}_\omega \sup_{n \leq \sigma_b(\omega)} |X_n(\omega)|.\]

This \( L(b) \) is not random, \( 0 \leq L(b) \leq \infty \), 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 \( \phi \) is some function which increases to \( \infty \) slowly. To make this kind of condition clearer, check that \( L(b) \) is the smallest real number for which \( i \leq n \) and \( T_n(\omega) > b \geq T_{n-1}(\omega) \) imply \( |X_i(\omega)| \leq L(b) \), almost surely. So \( L \) is the smallest non-decreasing function such that \( |X_{n+1}(\omega)| \leq L[T_n(\omega)] \) for \( n = 0, 1, \ldots \) almost surely. And \( L = o(\phi) \) means: there is a
function $\varepsilon(b)$ of $b$, tending to 0 at $\infty$, but with $\varepsilon(b)\phi(b)$ non-decreasing, and $|X_{n+1}(\omega)| \leq \varepsilon(T_n(\omega))\phi[T_n(\omega)]$ for all $n$.

As is immediate from (1.9),

(2.3) Proposition. $P(W_a \geq b) < 5[a + L(b)]/b^4$ for positive $a$ and $b$.

As is immediate from (1.10),

(2.4) Proposition. Let $0 < \delta < 1/4$. 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

$$P(W_a < b) > \frac{1}{3} \exp\left[-\frac{1}{2}(a^2/b)(1 + 4\delta)\right].$$

As is immediate from (1.14),

(2.5) Proposition. Let $0 < \alpha < \beta < \infty$. Let $\tau$ be a stopping time with $P(\alpha < T_\tau < \beta) = 1$. Then

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

Arguing as for (4.5),

(2.6) Proposition. Suppose $L(b) = o(b^4)$ as $b \to \infty$. Then $\sup \bar{S}_n = \infty$ and $\inf \underline{S}_n = -\infty$, a.e. on $\{\sum V_i = \infty\}$.

Of course, $\bar{S}_n$ converges to a finite limit $\bar{S}_\infty$ 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 $L(b) = o(b^4)$, but the result is false when $L(b) = o(b^{4+\epsilon})$.

Similarly, (2.7) Proposition. Suppose $L(b) = o(b/\log \log b)^4$ as $b \to \infty$. Then

$$\limsup_{n \to \infty} S_n/(2T_n \log \log T_n)^4 = 1 \quad \text{a.e. on } \{\sum V_i = \infty\}.$$ 

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 $a$, with $a = (1 - r^{-1})\phi(r^k)$ 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 $L(b) = o(b/\log \log b)^4$ 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 $L(b) = o(b^4)$ as $b \to \infty$.

The condition $L(b) = o(b^4)$ 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 \( P(X_n = \pm n^{1/2}) = \frac{1}{2} n^{-1} \). So \( V_n = 1 \),
\( T_n = n \), and \( L(b) \sim b^1 \). But \( S_n/n \) is not a normal. Theorem (2.8)
is a little sharper than Lévy’s result ([16] page 246), since \( \tau \) 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 \( x^2 g'(x) = h(x) \), 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 \( x \neq 0 \). \( \square \)

(3.2) **Corollary.** \( \exp(\lambda x) \leq 1 + \lambda x + x^2 e(\lambda) \) for \( \lambda \geq 0 \) and \( x \leq 1 \).

**Proof.** \( g(\lambda x) \leq g(\lambda) \), by (3.1). \( \square \)

(3.3) **Proposition.** \( \int_{|x|<1} Q_{n}(T_n, S_n) dP \leq 1 \) for \( \lambda \geq 0 \) and any stopping time \( \sigma \), provided
(3.4) \( X_n \leq 1 \) a.e. and \( E[X_n | \mathcal{F}_{n-1}] \leq 0 \) a.e. for all \( n \).

No lower bound is assumed on \( X_n \).

**Proof.** It is enough to show that \( \{Q_n(T_n, S_n), \mathcal{F}_n\} \) is an expectation-decreasing martingale. This follows from
(3.5) \( E[\exp(\lambda X)] \leq 1 + e(\lambda) \text{Var } X \leq \exp[e(\lambda) \text{Var } X] \) for \( \lambda \geq 0 \) and random variables \( X \) satisfying \( X \leq 1 \) and \( E(X) \leq 0 \).

It is enough to prove (3.5) when \( E(X^2) < \infty \). Let \( \mu \) be the distribution of \( X \), so
\( \mu = p\mu_0 + (1 - p)\mu_- \), where \( 0 \leq p \leq 1 \) and \( \mu_0 \) is a probability on \( (-\infty, 1] \) with
\( \int x \mu_0(dx) = 1 \) and \( \mu_- \) 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
\[
E[\exp(\lambda X)] = p \int \exp(\lambda x) \mu_0(dx) + (1 - p) \int \exp(\lambda x) \mu_-(dx)
\leq p[1 + e(\lambda)v_0] + (1 - p)[1 + e(\lambda)v_-]
\leq 1 + e(\lambda) \text{Var } X,
\]
using (3.2) on the first term in the first line, and Schwarz’s inequality to check
\( pv_0 + (1 - p)v_- \leq \text{Var } X \). \( \square \)

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

\[
\sup \frac{1}{\text{Var } X} [E[\exp(\lambda X)] - 1] = \sup \frac{1}{\text{Var } X} \log E[\exp(\lambda X)] = e(\lambda),
\]
where the sup is taken over all \( X \) with \( X \leq 1 \) and \( E(X) \leq 0 \). The extreme \( X \) take only the values \(-\varepsilon\) and 1, and have mean 0.

(3.6) **Proposition.** \( E[R_c(T_n, S_n)] \geq 1 \) for any \( \lambda \geq 0 \) and uniformly bounded stopping time \( \sigma \), assuming condition (1.1).

**Proof.** This follows from

\[ E[\exp(\lambda X)] \geq \exp\{f(\lambda) \text{ Var } X\} \]

for random variables \( X \) with \( X \geq -1 \) 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 \( E[X^n] = a \). Let \( \theta(\lambda) = E[\exp(\lambda X)] - \exp\{f(\lambda) \text{ Var } X\} \). Check that \( \theta(0) = 0 \) and \( \theta(\lambda) \geq 0 \).

Case 2, in which \( X \) takes only the two values \(-b \) and \( a \), with positive \( a, b \) and \( b \leq 1 \). There is a \( c \geq 1 \) with \( bc = 1 \). Let \( Y = cX \) and use Case 1 on \( Y \). So \( E[\exp(\lambda X)] = E[\exp(\lambda/c)Y] \geq \exp\{f(\lambda/c) \text{ Var } Y\} = \exp\{f(\lambda/c)c^2 \text{ Var } X\} = \exp\{f(\lambda/c) \cdot (\lambda/c)^2 \text{ Var } X\} \geq \exp\{f(\lambda) \cdot \lambda^2 \text{ Var } X\} = \exp\{f(\lambda) \text{ Var } X\} \).

The last inequality follows from (3.1), since \( f(x)/x^2 = g(-x) \) and \( -\lambda \leq -\lambda/c \).

The general case. Let \( \mu \) be the distribution of \( X \), so \( \mu = \int \mu_n \theta(\sigma a) \), where \( \mu_n \) is a probability on \([-1, \infty)\) residing on two points and having mean 0. Let \( v_n = \int x^2 \mu_n(dx) \), so \( \text{Var } X = \int v_n \theta(d\sigma) \). Let \( \phi_n(\lambda) = \int \exp(\lambda x) \mu_n(dx) \). Use Case 2 on \( \mu_n \), followed by Jensen’s inequality: \( E[\exp(\lambda X)] = \int \phi_n(\lambda) \theta(\sigma a) \geq \int \exp\{f(\lambda)v_n(\theta(\sigma a)) \geq \exp\{f(\lambda) \cdot \int v_n \theta(d\sigma a)\} = \exp\{f(\lambda) \text{ Var } X\} \).

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

\[ X_n \geq -1 \text{ a.e. and } E[X_n|\mathcal{F}_n] = 0 \text{ a.e. for all } n. \]

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

\[ |X_n| \leq 1 \text{ a.e. and } E[X_n|\mathcal{F}_{n-1}] \geq 0 \text{ a.e. for all } n, \]

even if \( V_n \) is replaced by the larger quantity \( E[X_n^2|\mathcal{F}_{n-1}] \). 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 \( b \),

\[ P[S_n \geq a \text{ and } T_n \leq b \text{ for some } n = 1, 2, \ldots] \]

\[ \leq \left( \frac{b}{a+b} \right)^{a+b} e^a \leq \exp\left[-\frac{a^2}{2(a+b)}\right]. \]

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

\[ 1 \geq \int A \exp[\lambda S_{\sigma} - c(\lambda)T_{\sigma}] dP \geq P(A) \cdot \exp[\lambda a - c(\lambda)b], \]
so \( P[A] \leq \exp[-\lambda a + e(\lambda)b] \). The minimizing \( \lambda \) 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 \( S_a \leq a + 1 \) for \( n \leq \tau_a \). Then use (3.6) with \( \tau_a \wedge n \) for \( a \):

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

Let \( n \to \infty \) and use dominated convergence. \( \square \)

(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 \( P[X_n = -1] \) for all \( n \) > 0, and \( P[W_a = \infty] > 0 \). This contradicts (4.2), by letting \( \lambda \) approach 0.

(4.4) **The Proof of (1.9).** Use Chebychev's inequality on (1.8) to see

\[
P[W_a \geq b] = P[1 - \exp[-f(\lambda)W_a] \geq 1 - \exp[-f(\lambda)b]]
\]

\[
\leq \frac{1 - \exp[-\lambda(a + 1)]}{1 - \exp[-f(\lambda)b]}.
\]

Putting \( \lambda = b^{-1} \) gives

\[
1 - \exp[-\lambda(a + 1)] < \lambda(a + 1) = (a + 1)/b^3
\]

and

\[
1 - \exp[-f(\lambda)b] > f(\lambda)b - \frac{1}{2}f(\lambda)^3b^3
\]

\[
> \left( \frac{\lambda^3}{2} - \frac{\lambda^5}{6} \right) b - \frac{\lambda^7b^2}{8}
\]

\[
= \frac{1}{2} - \frac{1}{6} - \frac{1}{8}b^{-4}
\]

\[
> \frac{5}{8}, \text{ for } b > 1.
\]

If \( b \leq 1 \), the bound is at least 5, dwarfing \( P[W_a \geq b] \). \( \square \)

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

\begin{enumerate}
  \item \( \sup_n S_n = \infty \) and \( \inf_n S_n = -\infty \), a.e. on \( \{\sum V_i = \infty\} \)
  \item \( S_n \) converges to a finite limit \( S_\infty \) as \( n \to \infty \), a.e. on \( \{\sum V_i < \infty\} \).
\end{enumerate}

**Proof.** Claim (a). By symmetry, only the sup need be argued. (4.2) implies, \( P[W_a < \infty] = 1 \). So \( \{\tau_a = \infty \} \) 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. \( \square \)

(4.6) **The Proof of (1.12).** Relation (3.3) shows

\[
\frac{1}{\tau_a < c} \exp[\lambda S_{\tau_a} - e(\lambda)W_{\tau_a}] dP \leq 1.
\]

But \( S_{\tau_a} \geq a \) on \( \{\tau_a < c\} \). So

\[
\frac{1}{\tau_a < c} \exp[-e(\lambda)W_{\tau_a}] dP \leq \exp(-\lambda a).
\]
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. \( \square \)

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 \((b \log \log b)^\delta \). 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. \( \square \)

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 \( \lambda \geq 0 \),

(4.11a) \( E[\exp[-e(\lambda)W_a]] \leq \exp(-\lambda a) \)

(4.11b) \( E[\exp[-f(\lambda)W_a]] \geq \exp(-\lambda(a + 1)) \).

Let \( \delta, a, b \) be positive, with

(4.12a) \( \delta < \frac{1}{2} \)

(4.12b) \( \frac{b}{a} > \frac{9}{\delta^2} \)

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

Then

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

**Proof.** Let

(4.14) \( \phi(\lambda) = \frac{\lambda^2}{2} > \frac{\lambda^2}{6} \) and \( \lambda = (1 + \delta) \frac{a}{b} \) and \( k = \frac{a^2}{b} \) and \( N = \frac{2}{\delta^2} \).

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

(4.15) \( \phi(\lambda) \int_0^a P[W_a < x] \exp[-\phi(\lambda)x] dx > \exp[-\lambda(a + 1)] \).

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

\( I_1 = [0, Na] \) and \( I_2 = (Na, (1 - 2\delta)b] \) and \( I_3 = (b, 2b] \),

\( I_4 = (2b, \infty) \) and \( I_5 = ((1 - 2\delta)b, b] \).

Let

\( \gamma_i = \phi(\lambda) \int_{I_i} P[W_a < x] \exp[-\phi(\lambda)x] dx \).
Of these, the terms with \( i \leq 4 \) are errors. As will be seen,

\[
\gamma_i < \left( \frac{1}{k} - \frac{1}{2} \right) \exp[-(1 + \delta)k] \\
\gamma_i < \frac{1}{2} \exp[-(1 + \delta)k] \quad \text{for } i = 2, 3, 4.
\]

The main term is \( \gamma_5 \). As will be seen,

\[
\gamma_5 < P[W_a < b] \cdot \exp[-(\frac{1}{2} + 2\delta^3)k].
\]

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

\[
\exp[-\lambda(a + 1)] = \exp\left[-\left(1 + \delta\right)\frac{a}{b}\right] \cdot \exp[-(1 + \delta)k] \quad \text{by (4.14)}
\]

\[
> \exp\left(-\frac{1}{\delta_0}\right) \cdot \exp[-(1 + \delta)k] \quad \text{by (4.12)}
\]

\[
> (1 - \delta_0) \exp[-(1 + \delta)k].
\]

Using (4.15) and (4.16: 1–4):

\[
\gamma_5 > \exp[-\lambda(a + 1)] - \gamma_1 - \gamma_2 - \gamma_3 - \gamma_4 > \frac{1}{2} \exp[-(1 + \delta)k].
\]

Using (4.16: 5)

\[
P[W_a < b] > \exp\left[(\frac{1}{2} - 2\delta^3)k\right] \cdot \gamma_5
\]

\[
> \frac{1}{2} \exp[-(\frac{1}{2} + \delta + 2\delta^3)k]
\]

\[
> \frac{1}{2} \exp[-(\frac{1}{2} + 2\delta)k] \quad \text{by (4.12) a}.
\]

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

\[
\exp(-\delta^3 k/8) < 1/(8k),
\]

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

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

Using Chebychev's inequality on (4.11 a),

\[
P[W_a < x] = P[\exp[-e(\lambda)W_a] > \exp[-e(\lambda)x]]
\]

\[
< \exp[e(\lambda)x] \cdot E[\exp[-e(\lambda)W_a]]
\]

\[
< \exp[-\lambda a + e(\lambda)x].
\]

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

The Proof of (4.16: 1). If \( x \leq Na \), then

\[
P[W_a < x] < \theta = \left( \frac{1}{k} - \frac{1}{2} \right) \exp[-(1 + \delta)].
\]

This follows from (4.18): if \( x \leq Na \), then \( a^2/(a + x) \geq a/(N + 1) \). Next, \( N + 1 < 3/\delta^3 \), so \( a^2/(a + x) > (\delta^3/3)a = (\delta^3/3)(b/a)k > 3k > (1 + \delta + \frac{1}{8})k \); and \( k/6 > 5 \), so \( \exp(-k/6) < 2^{-5} < 1/8 - \delta_0 \). This settles (4.19). To complete the proof of (4.16: 1),

\[
\gamma_i < \theta \cdot \phi(\lambda) \cdot \int_0^{x} \exp[-(\phi(\lambda)x) \, dx < \theta \cdot \phi(\lambda) \cdot \int_0^{x} \exp[-\phi(\lambda)x] \, dx = \theta.
\]
The proof of (4.16:2). Confirm that \( Na < (1 - 2\delta)b \), so \( I_2 \) is proper. If \( x \geq Na \), then \( a^2/(a + x) \geq a^2/(1 + N^{-1}) = (N/N + 1)a^2/x \). Use (4.18) and (4.14):

\[
\eta_2 < \exp[\frac{1}{2}(1 - 2\delta)b] \cdot \phi(\lambda) \cdot \gamma,
\]

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

\[
\theta(x) = \frac{x^2}{2} + \frac{N}{N + 1} \cdot \frac{a^2}{2x}.
\]

By calculation,

\[
\theta(x) \geq (1 + \delta)^2 (1 - 2\delta) + \frac{N}{N + 1} \cdot \frac{1}{1 - 2\delta},
\]

(4.23)

\[
\mu > 1 + \delta + \frac{1}{4}\delta^2.
\]

By calculation,

\[
\theta(x) > (1 + \delta + \frac{1}{4}\delta^2)k \quad \text{for} \quad x \in I_2.
\]

(4.24)

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

\[
\gamma < (1 - 2\delta)b \exp(-\frac{1}{4}\delta^2) \exp[-(1 + \delta)k].
\]

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

\[
\phi(\lambda)\gamma < \frac{1}{4}k \exp(-\frac{1}{4}\delta^2) \exp[-(1 + \delta)k].
\]

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

\[
\exp(\frac{1}{4}\lambda^2(1 - 2\delta)b) \leq \exp(\frac{1}{4}\delta^2k).
\]

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

The proof of (4.16:3). As for (4.20),

\[
\eta_3 < \exp(\frac{1}{8}x^2 \cdot 2b \cdot \phi(\lambda) \cdot \gamma),
\]

where \( \gamma = \int_{I_2} \exp[-\theta(x)] \, dx \), and \( \theta \) is still defined by (4.21). By (4.22), the minimum of \( \theta \) on \( I_3 \) occurs at the left endpoint \( b \), and is \( \mu k \), where \( \mu = \frac{1}{4}(1 + \delta)^2 + N/(N + 1) \). By computation, \( \mu > 1 + \delta + \frac{1}{4}\delta^2 \). Consequently, \( \gamma < b \cdot \exp[-(1 + \delta + \frac{1}{4}\delta^2)k] \). But \( \phi(\lambda)b < \frac{1}{2}(1 + \delta)^2k < k \); so

\[
\phi(\lambda)\gamma < k \exp[-(1 + \delta + \frac{1}{4}\delta^2)k].
\]
Next
\begin{equation}
\exp\left(\frac{4}{3} \delta^3 \cdot 2b \right) = \exp\left[ \frac{4}{3}(1 + \delta)^3 \frac{a}{b} k \right] < \exp\left[ \frac{4}{3}(1 + \delta)^3 \frac{\delta^3}{9} k \right] < \exp\left( \frac{4}{3} \delta^3 k \right).
\end{equation}
So \( \eta_3 < \alpha \exp[-(1 + \delta)k] \), where \( \alpha = k \cdot \exp\left( \frac{4}{3} \delta^3 k \right) \cdot \exp\left( -\frac{4}{3} \delta^3 k \right) < \frac{8}{9} \) by (4.17). This proves (4.16:3).

**The Proof of (4.16:4).** Confirm that \( \eta_4 \leq \phi(\lambda) \int_{\mathbb{R}} \exp\left[ -\phi(\lambda) x \right] dx = \exp\left[ -2b\phi(\lambda) \right] = \exp\left( \frac{4}{3} \lambda^3 \cdot 2b \right) \cdot \exp\left[ -(1 + \delta)^3 k \right] \). The first factor in the last expression was shown to be less than \( \exp\left( \frac{4}{3} \delta^3 k \right) \) in (4.30). Confirm that \( \frac{4}{3} \delta^3 - (1 + \delta)^3 < -k^{-1} \log 3 - (1 + \delta)k \), completing the proof of (4.16:4).

**The Proof of (4.16:5).** Since \( P(W_a < x) \leq P(W_a < b) \) for \( x \leq b \),
\begin{equation}
\eta_5 \leq P(W_a < b) \cdot \int_{-\infty}^{\frac{4}{3} \delta^3 \cdot 2b} \phi(\lambda) \exp\left[ -\phi(\lambda) x \right] dx < P(W_a < b) \cdot \int_{\mathbb{R}} \phi(\lambda) \exp\left[ -\phi(\lambda) x \right] dx = P(W_a < b) \cdot \exp\left[ \frac{4}{3} \lambda^3 (1 - 2\delta) b \right] \cdot \exp\left[ \frac{4}{3}(1 + \delta)^3 (1 - 2\delta) k \right].
\end{equation}

In the last line, the second factor is less than \( \exp\left( \frac{4}{3} \delta^3 k \right) \) by (4.27). And \( -\frac{4}{3}(1 + \delta)^3 (1 - 2\delta) < -\frac{4}{3} + 2\delta^3 - \frac{4}{3} \delta^3 \). This settles (3.16:5) and the theorem. \[ \]

(3.31) **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 \( W_a^* \geq W_a \), 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 \( Y_1, Y_2, \ldots \) which are independent and \( \pm 1 \) with probability \( \frac{1}{2} \) each. Make all the \( Y_n \)'s independent of all the \( \mathcal{F}_n \)'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 \( S_{\infty} \) on \( \{ \sum V_i < \infty \} \). So
\begin{equation}
\sum X_i \text{ converges to a finite limit } S_{\infty} \text{ and } T_{\infty} = \sum V_i < \infty, \text{ a.e. on } \{ \tau_a = \infty \}.
\end{equation}

If \( \tau_a = \infty \), let \( \tau_a^* \) be the least \( n = 0, 1, \ldots \) with \( S_{\infty} + Y_1 + \cdots + Y_n \geq a \). Let \( W_a^* = W_a \) when \( \tau_a < \infty \), and let \( W_a^* = T_{\infty} + \tau_a^* \) when \( \tau_a = \infty \). So \( W_a^* \geq W_a \).

I will now argue that \( W_a^* \) satisfies (4.11a). Use (3.3) with \( \tau_a \wedge n \) for \( \sigma \) to see that \( \int Q_1(T_{\tau_a \wedge n}, S_{\tau_a \wedge n}) \, dP \leq 1 \). Let \( n \to \infty \) and use (4.32):
\begin{equation}
\int_{\{ \tau_a < \infty \}} Q_1(W_a, S_{\tau_a}) \, dP + \int_{\{ \tau_a = \infty \}} Q_1(T_{\infty}, S_{\infty}) \, dP \leq 1.
\end{equation}
Given \( \{ \tau_a = \infty \} \), the process \( Q_1(T_{\infty} + n, S_{\infty} + Y_1 + \cdots + Y_n) \) is an expectation-decreasing martingale relative to the \( \sigma \)-fields \( \mathcal{F}_n^* \), where \( \mathcal{F}_n^* \) is the \( \sigma \)-field
generated by all the $\mathcal{F}_t$ and by $Y_1, \ldots, Y_n$, as follows from (3.5). So
\begin{equation}
\sum_{t_\alpha = \omega}^\infty Q_\alpha (T_\infty + \tau_\alpha, S_\omega + \sum_i i^* Y_i) \, dP \leq \sum_{t_\alpha = \omega}^\infty Q_\alpha (T_\omega, S_\omega) \, dP.
\end{equation}

Let $S^* = S_\omega$ when $\tau_\omega < \infty$, and $S^* = S_\omega + \sum_i i^* Y_i$ when $\tau_\omega = \infty$. So (4.33) and (4.34) show that $\sum Q_\alpha (W_\alpha, S^*) \, dP \leq 1$. But $S^* \geq a$, and this proves (4.11a) for $W_\alpha^*$.

The argument for (4.11 b) is similar, using (3.6) and (3.7): since $S_n \leq a + 1$ for $n \leq \tau_\alpha^*$, 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 $P[W_\alpha \geq b]$ when $a = o(b^k)$. 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 $\int_0^\infty e^{\lambda x} h_n(x) \, dx \to (2/\pi\lambda)^{1/2}$ as $n \to \infty$, for all positive $\lambda$. Then $h_n(x) \to [2/(\pi x)]^{1/2}$ 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 $n$th process to cross $a_n$, in the sense of (1.7). Let $Z_n = W_{a_n}/b_n$, and $h_n(x) = (b_n^{1/2} a_n) P(Z_n \geq x)$. I will argue that $h_n(x) \to (2/\pi x)^{1/2}$ for all positive $x$; putting $x = 1$ gives the theorem.

Let $\theta$ be the inverse function to $e(\cdot)$, and let $\phi$ be the inverse function to $f$, so $e^{\theta(\lambda)} - 1 - \theta(\lambda) = e^{-\phi(\lambda)} - 1 + \phi(\lambda) = \lambda$ for $\lambda > 0$. Check
\begin{equation}
\theta(\lambda)/(2\lambda)^{1/2} \text{ and } \phi(\lambda)/(2\lambda)^{1/2} \text{ tends to 1 as } \lambda \text{ tends to 0}.
\end{equation}

Let
\begin{align*}
\theta_n(\lambda) &= (2\lambda)^{-1/2} (b_n^{1/2} a_n) \left[ 1 - \exp \left[ -\theta(\lambda/b_n) a_n \right] \right], \\
\phi_n(\lambda) &= (2\lambda)^{-1/2} (b_n^{1/2} a_n) \left[ 1 - \exp \left[ -\phi(\lambda/b_n) (a_n + 1) \right] \right].
\end{align*}

Using (5.3),
\begin{equation}
\theta_n(\lambda) \to 1 \text{ and } \phi_n(\lambda) \to 1 \text{ as } n \to \infty, \text{ for all positive } \lambda.
\end{equation}

Inequalities (1.8) and (1.12) imply
\begin{equation}
\exp \left[ -\phi(\lambda/b_n) (a_n + 1) \right] \leq E[\exp (\lambda Z_n)] \leq \exp \left[ -\theta(\lambda/b_n) a_n \right].
\end{equation}

Integrating by parts, $E[\exp (\lambda Z_n)] = 1 - \lambda \int_0^\infty e^{-\lambda x} h_n(x) \, dx$. So
\begin{equation}
(2/\lambda)^{1/2} \theta_n(\lambda) \leq \int_0^\infty e^{-\lambda x} h_n(x) \, dx \leq (2/\lambda)^{1/2} \phi_n(\lambda).
\end{equation}

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 $\phi(x) = (2x \log \log x)^{1/2}$ for $x \geq e^e$ and $\phi(x) = 1$ for $x < e^e$. 
(6.1) Theorem. Suppose condition (3.4). Then
\[ \limsup_{n \to \infty} S_n / \phi(T_n) \leq 1 \quad \text{a.e. on } \{ \sum V_i = \infty \}. \]

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 \( S_n > r^k \phi(T_n) \) for infinitely many \( n \). Let \( \sigma_k \) be the sup of \( n \) with \( T_n \leq r^k \). If \( \sum V_i = \infty \), then \( \sigma_k < \infty \) and \( \sigma_k \) increases to \( \infty \) with \( k \). So \( A \subset \limsup A_n \), where
\[ A_k = \{ \sum V_i = \infty \text{ and } S_n > r^k \phi(T_n) \text{ for some } n \text{ with } \sigma_k < n \leq \sigma_{k+1} \}. \]
But \( r^n < T_n \leq r^{k+1} \) for \( \sigma_k < n \leq \sigma_{k+1} \). So \( A_k \subset B_k \), where
\[ B_k = \{ S_n > r^k \phi(r^k) \text{ for some } n \text{ with } T_n \leq r^{k+1} \}. \]
Now use (4.1) to estimate \( P[B_k] \), with \( r^k \phi(r^k) \) for \( a \) and \( r^{k+1} \) for \( b \). Keep \( k \) so large that \( a + b \leq r^{k+3} \), which is feasible because \( a = o(r^k) \): then \( P[B_k] \leq \exp[-a^2/2(a + b)] \leq \exp[-\log \log r^k] \leq 1/k^2 \), 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 \( \tau \) be a stopping time. Then \( \tau + n \) is a stopping time. Let \( \mathcal{I}_\tau \) be the \( \sigma \)-field of events \( A \) such that \( A \cap \{ \tau = n \} \in \mathcal{I}_n\) for all \( n \geq 0 \). For \( n \geq 0 \), let \( \mathcal{I}_n^* = \mathcal{I}_{\tau+n} \). For \( n \geq 1 \), 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 \( \mathcal{I}_n^* \)-measurable and \( V_n^* \) is \( \mathcal{I}_n^* \)-measurable, for \( n \geq 1 \).
(b) \( E[X_n^* | \mathcal{I}_{n-1}] = 0 \) and \( E[X_n^* | \mathcal{I}_{n-1}] = V_n^* \) a.e., for \( n \geq 1 \).

Let \( P_\sigma(\omega, A) \) be a regular conditional probability on \( (\Omega, \mathcal{I}), \) given \( \mathcal{I}_\tau \).

(c) For \( P \)-almost all \( \omega \), the starred system satisfies (1.1), relative to the conditional probability triple \( (\Omega, \mathcal{I}, P_\sigma(\omega, \cdot)) \).

(6.3) Theorem. Suppose conditions (1.1). Then
\[ \limsup_{n \to \infty} S_n / \phi(T_n) \geq 1 \quad \text{a.e. on } \{ \sum V_i = \infty \}. \]

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 \( S_n \geq (1 - 3r^{-1}) \phi(T_n) \) for only finitely many \( n \). As before, let \( \sigma_k \) be the sup of \( n \) with \( T_n \leq r^k \), so \( \sigma_k \) is a stopping time, because \( V_n \) is \( \mathcal{I}_{n-1} \)-measurable. Let \( A_k \) be the event that \( \sigma_k < \infty \) and
\[ S_n - S_{a_k} \geq (1 - r^{-1}) \phi(r^{k+1}) \text{ for some } n > a_k \text{ with } T_n - T_{a_k} < r^{k+1} - r^k. \]

Let

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

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

\[ P[B_k \mid \mathcal{F}_{a_k}] > \frac{1}{2} \exp(-\theta \log \log r^{k+1}) \quad \text{on } \{ a_k < \infty \}. \]

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

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

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

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

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

\[ S_{a_k} > -2\phi(T_n) > -2\phi(r^k) > -2r^{-i}\phi(r^{k+1}). \]

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

\section{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).

\begin{equation}
\text{Lemma. Suppose (1.1). Let } S_{a_n} = \sum_{i=1}^{a_n} X_i, \text{ provided the sum converges to a finite limit. Let } \sigma \text{ be a stopping time, with } T_\sigma < \infty \text{ a.e. Let } S_\sigma = S_n \text{ when } \sigma = n, \text{ and } S_\sigma = S_{a_n} \text{ when } \sigma = \infty. \text{ Then } S_\sigma \text{ is defined and finite a.e.}
\end{equation}

\begin{equation}
\text{Theorem. Suppose (1.1). Let } \lambda \geq 0 \text{ and } 0 \leq \alpha < \beta < \infty. \text{ Let } \sigma \text{ be a stopping time.}
\end{equation}

(a) If \( P[T_\sigma \leq \beta] = 1 \), then \( E[\exp(\lambda S_\sigma)] \leq \exp[\beta e(\lambda)]. \)

(b) If \( P[\alpha \leq T_\sigma \leq \beta] = 1 \), then \( E[\exp(\lambda S_\sigma)] \geq \exp[\alpha f(\lambda)]. \) These bounds are sharp.

\text{Proof. Claim (a). Inequality (3.3) shows } E[Q_2(T_{a,n}, S_{a,n})] \leq 1. \text{ By assumption, } T_\sigma < \infty \text{ a.e. So } S_{a,n} \rightarrow S_\sigma \text{ a.e. as } n \rightarrow \infty, \text{ by (7.1). Let } n \rightarrow \infty \text{ and use Fatou’s lemma:}

\[ \exp[-\beta e(\lambda)] \cdot E[\exp(\lambda S_\sigma)] \leq E[\exp(\lambda S_\sigma - e(\lambda) T_\sigma)] \leq 1. \]

Claim (b). Inequality (3.6) shows \( E[R_2(T_{a,n}, S_{a,n})] \geq 1. \) Again, (7.1) makes
$R_t(T_{\alpha,n}, S_{\alpha,n}) \rightarrow R_t(T_{\alpha}, S_{\alpha}) \text{ a.e. as } n \rightarrow \infty$. And this sequence of variables is uniformly integrable, because

$$E[R_t(T_{\alpha,n}, S_{\alpha,n})^3] \leq E[\exp(2\lambda S_{\alpha,n})] \leq \exp[\beta e(2\lambda)],$$

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

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

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

The next result is standard.

(7.3) **Lemma.** Let $(\Omega, \mathcal{F}, P)$ be a probability triple, and let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ be sub-$\sigma$-fields of $\mathcal{F}$. For each $n$, let $X_n$ be $\mathcal{F}_n$-measurable, and suppose $Y_n = E[X_n | \mathcal{F}_{n-1}]$ is defined a.e. Let $\tau$ be a stopping time. Let $X_n' = X_n$ and $Y_n' = Y_n$ for $n \leq \tau$, while $X_n' = Y_n' = 0$ for $n > \tau$. Then $X_n'$ is $\mathcal{F}_n$-measurable, $Y_n'$ is $\mathcal{F}_{n-1}$-measurable, and $Y_n' = E[X_n' | \mathcal{F}_{n-1}]$.

(7.4) **Lemma.** Suppose (1.1) and suppose $(\Omega, \mathcal{F}, P)$ supports a uniform random variable independent of all the $\mathcal{F}_n$. Let $0 < \delta < 1$ and $0 < \alpha < \beta < \infty$. Let $\tau$ be a uniformly bounded stopping time, with $P[\alpha < T_{\tau} < \beta] = 1 - \delta$. Then on the same probability triple there are variables $X_n^*$, $\sigma$-fields $\mathcal{F}_n^*$, and variances $V_n^*$, satisfying (1.1), and a stopping time $\tau^*$ relative to $\{\mathcal{F}_n^*\}$, with $P[S_n^* = S_n] < \delta$ 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 $\tau \leq k$. Let $\tau'$ be the max of $n \leq \tau$ with $T_n < \beta$. Then $\tau'$ is also a stopping time, because $V_n$ is $\mathcal{F}_{n-1}$-measurable. Clearly, $T_{\tau'} \leq \beta$ everywhere and $\tau' = \tau$ on $G$.

Choose a positive $\varepsilon$ 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 $\mathcal{F}_n$, each variable taking the values $\pm \varepsilon$ with chance $\frac{1}{2}$. Let $\mathcal{F}_n^* = \mathcal{F}_n$ for $n \leq k$, while $\mathcal{F}_n^*$ is the $\sigma$-field spanned by $\mathcal{F}_k$ and $Y_1, \cdots, Y_{n-k}$ for $n > k$. Let

$$X_n^* = X_n \quad \text{and} \quad V_n^* = V_n \quad \text{for } n \leq \tau'.$$

$$X_n^* = V_n^* = 0 \quad \text{for } \tau' < n \leq k.$$

$$X_n^* = Y_{n-k} \quad \text{and} \quad V_n^* = \varepsilon^2 \quad \text{for } n > k.$$

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

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

$$\tau^* = k \quad \text{and} \quad S_n^* = \sum_i X_i^* = \sum_i X_i = S_\varepsilon = S_\tau \text{ on } G. \quad \square$$

(7.5) **The proof of (1.15).** Let $\delta_n > 0$ with $\delta_n \rightarrow 0$. Let $0 < \alpha_n < \beta_n < \infty$, with $\alpha_n \rightarrow \infty$ and $\beta_n/\alpha_n \rightarrow 1$. For each $n$, consider a process satisfying (1.1), and a stopping time $\tau_n$. Let $S(n)$ be the sum of the variables in the $n$th process.
TAIL PROBABILITIES FOR MARTINGALES

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

\begin{equation}
P[\beta_n < T(n) < \beta_n] > 1 - \delta_n.
\end{equation}

The problem is to show that the distribution of $S(n)/\alpha_n$ 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 $\tau_n$ to be uniformly bounded, say by $k_n$. Indeed, $\tau_n$ 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 $n$th 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 $\alpha_n$ and $\beta_n$ everywhere. Finally, since $\alpha_n f(\lambda/\alpha_n^4)$ and $\beta_n e(\lambda/\alpha_n^4)$ both tend to $\lambda^2/2$, convergence to the normal distribution follows from (7.2).

REFERENCES


