The Annals of Probability
1993, Vol. 21, No. 3, 1640–1670

SOME LIMIT THEOREMS IN LOG DENSITY

BY I. BERKES[1] AND H. DEHLING

Hungarian Academy of Sciences and Groningen University

Motivated by recent results on pathwise central limit theorems, we study in a systematic way log-average versions of classical limit theorems. For partial sums S_k of independent r.v.'s we prove under mild technical conditions that $(1/\log N)\Sigma_{k\le N}(1/k)I\{S_k/a_k \in \cdot\} \to G(\cdot)$ (a.s.) if and only if $(1/\log N)\Sigma_{k\le N}(1/k)P(S_k/a_k \in \cdot) \to G(\cdot)$. A functional version of this result also holds. For partial sums of i.i.d. r.v.'s attracted to a stable law, we obtain a pathwise version of the stable limit theorem as well as a strong approximation by a stable process on log dense sets of integers. We also give necessary and sufficient conditions for the law of large numbers in log density.

1. Introduction. In recent years many interesting extensions of classical probability limit theorems involving log average and log density have been obtained. The basic result and starting point of these investigations is the a.s. central limit theorem discovered by Brosamler [5] and Schatte [23] for i.i.d. r.v.'s having finite $(2+\delta)$th moments and later proved by Fisher [13] and Lacey and Philipp [15] to hold assuming only finite variances:

THEOREM. *Let $X_1, X_2 \dots$ be i.i.d. r.v's with $EX_1 = 0$, $EX_1^2 = 1$ and set $S_n = X_1 + \cdots + X_n$. Then*

$$(1.1)\qquad \lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{ \frac{S_k}{\sqrt{k}} \in A \right\} = (2\pi)^{-1/2} \int_A e^{-t^2/2}\, dt \quad a.s.,$$

for any Borel-set $A \subset \mathbf{R}$ with $\lambda(\partial A) = 0$. Moreover, the exceptional set of probability zero can be chosen to be independent of A. Here I denotes the indicator function.

Certain special cases of (1.1) have been known for almost 40 years; for example, Erdős and Hunt [10] proved that

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\{S_k > 0\} = \frac{1}{2} \quad \text{a.s.}$$

holds for any i.i.d. sequence (X_n) having a continuous symmetric distribution function. (For a related local limit theorem assuming only $EX_1 = 0$, see Chung

[1]Research supported by Hungarian National Foundation for Scientific Research Grants 1808 and 1905.

AMS 1991 *subject classifications*. 60F05, 60F17, 60F15.

Key words and phrases. Pathwise central limit theorem, log-averaging methods, stable convergence, strong approximation, law of large numbers.

and Erdős [6].) The limiting behaviour of the two-parameter process

$$\Delta_{N,t} = \frac{1}{(\log N)^{1/2}} \sum_{k \le Nt} \frac{1}{k}\left(I\left\{ \frac{S_k}{\sqrt{k}} < x \right\} - \Phi(x) \right), \quad -\infty < x < +\infty, 0 \le t \le 1)$$

was investigated by Csörgő and Horváth [7]. In particular, their results yield the precise rate of convergence, both in the a.s. and distributional sense, in (1.1) as well as functional limit theorems related to (1.1). (See also Weigl [27] and Peligrad and Révész [20].)

An interesting "log" version of Strassen's a.s. invariance principle was proved by Fisher [13] who proved that if $X_1, X_2 \dots$ are i.i.d. r.v.'s with $EX_1 = 0$, $EX_1^2 = 1$, then $\{X_n\}_{n=1}^{\infty}$ can be redefined on a suitable probability space together with a Wiener process W such that with probability 1 the relation

$$\sum_{i=1}^{n} X_i - W(n) = o(\sqrt{n}) \quad \text{as } n \to \infty \tag{1.2}$$

holds for "almost all" n in the sense that for a.e. ω there exists a set $N_\omega \subset N$ of integers of logarithmic density 0 such that at ω, (1.2) holds for $N \notin N_\omega$. [For a set $A \subset N$ of positive integers, the log density $\mu_L(A)$ of A is defined by

$$\mu_L(A) = \lim_{N \to \infty} \frac{1}{\log N} \sum_{k \le N, k \in A} \frac{1}{k}$$

provided that the limit exists.] Recall that Strassen's classical a.s. invariance principle (see [26]) says that for all n we have the approximation

$$\sum_{i=1}^{n} X_i - W(n) = o(n \log\log n)^{1/2} \quad \text{a.s. as } n \to \infty \tag{1.3}$$

(with a suitable construction of W). Moreover, as Major [17] proved, assuming only finite second moments the remainder term $o(n \log\log n)^{1/2}$ in (1.3) is the best possible.

The purpose of this paper is to study, in a systematic way, "log" analogues of further classical limit theorems of probability theory. In Section 2 we prove a general version of the a.s. central limit theorem and its functional version for independent, not necessarily identically distributed random variables. In fact, under mild growth conditions on the partial sums S_n of an independent sequence (X_n) such as

$$E(\log\log|S_n/a_n|)^{1+\varepsilon} \le (\log\log n)^{\delta}, \qquad 0 < \delta < \varepsilon, \tag{1.4}$$

we give necessary and sufficient criteria for the generalized a.s. central limit theorem

$$\lim_{N \to \infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{ \frac{S_k}{a_k} < x \right\} = G(x) \quad \text{a.s.} \tag{1.5}$$

and its functional version in terms of "weak" characteristics of (X_n). An immediate consequence of these criteria is the fact that under (1.4) the a.s.

central limit theorem (1.5) is a consequence of the weak limit relation

$$S_n/a_n \to_{\mathscr{D}} G. \tag{1.6}$$

Thus we see that, despite its pointwise character, the a.s. central limit theorem is [under (1.4)] a weaker statement than the distributional result (1.6). In [2] we constructed examples showing the crucial difference between (1.5) and (1.6). We found, for example, i.i.d. sequences (X_n) satisfying (1.5) with a simple G (e.g., stable or mixed stable distribution) such that (1.6) fails badly; in fact, under (1.5) it is possible that S_n/a_n has many different limit distributions along different subsequences which do not even resemble G.

In Section 3 we investigate the "log" analogues of the strong and weak laws of large numbers. Noting that the validity of the a.s. central limit theorem (1.5) is not affected by the behaviour of S_n/a_n on a set of n's with log density zero, it is natural to define the logarithmic version of laws of large numbers by admitting exceptional sets of log density zero in the usual definitions

$$S_n/a_n \to 0 \quad \text{a.s.} \quad \text{or} \quad S_n/a_n \to_P 0. \tag{1.7}$$

[Actually, in the strong law of large numbers we shall admit an exceptional set depending on ω as in Fisher's invariance principle (1.2).] In Section 3 we shall prove the surprising result that in the presence of such exceptional sets the weak and strong law of large numbers are *equivalent*. This result shows again that despite its "strong" character, a.s. convergence on a set of log density 1 is much weaker than ordinary a.s. convergence and in the context of normed sums of independent r.v.'s it is even weaker than convergence in probability. The equivalence of the log versions of the weak and strong law of large numbers also enables us to give simple necessary and sufficient criteria for them in terms of individual r.v.'s X_n.

Finally, in Section 4 we extend Fisher's a.s. invariance principle (1.2) to arbitrary i.i.d. sequences (X_n) with $EX_1^2 = +\infty$ belonging to the domain of attraction of the normal law or an α-stable law, $0 < \alpha < 2$. That is, we show the analogue of Fisher's result for any i.i.d. sequence (X_n) such that for some numerical sequences $a_n > 0$ and b_n,

$$\frac{1}{a_n}\left(\sum_{i=1}^{n} X_i - b_n\right) \to_{\mathscr{D}} G \tag{1.8}$$

with a nondegenerate distribution G. Again, permitting exceptional n-sets of log density zero changes the nature of the problem essentially, leading to the simple and all-covering "log" a.s. approximation results in Section 4, in contrast to the highly complicated and implicit ordinary a.s. invariance principles existing under $EX_1^2 = +\infty$ (see e.g., [14], [18] and [19] for such results). In fact, we shall see that under (1.8) the "log" a.s. approximation of $\sum_{i=1}^{n} X_i$ with Wiener or stable processes holds with the same remainder term as ordinary approximation in probability, another example of the phenomenon observed in Section 3.

2. The a.s. central limit theorem.

THEOREM 1. *Let $X_1, X_2 \dots$ be independent r.v.'s and $a_n > 0$, b_n numerical sequences such that setting $S_n = X_1 + \cdots + X_n$ we have*

$$Ef\left(\left|\frac{S_n - b_n}{a_n}\right|\right) \le (\log\log n)^{-1-\varepsilon} f\left(e^{(\log n)^{1-\varepsilon}}\right), \qquad n \ge n_0 \tag{2.1}$$

for some $\varepsilon > 0$ where $f \ge 0$ is a Borel measurable function on $(0, \infty)$ such that both $f(x)$ and $x/f(x)$ are eventually nondecreasing and the right-hand side of (2.1) *is nondecreasing for $n \ge n_0$. Assume also that*

$$a_l/a_k \ge C(l/k)^{\gamma}, \qquad l \ge k \tag{2.2}$$

for some constants $C > 0$, $\gamma > 0$. Then for any distribution function G the following statements are equivalent:

(a) *For any Borel set $A \subset \mathbf{R}$ with $G(\partial A) = 0$ we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{\frac{S_k - b_k}{a_k} \in A\right\} = G(A) \quad a.s., \tag{2.3}$$

where the exceptional set of probability zero is independent of A.

(b) *For any Borel set $A \subset \mathbf{R}$ with $G(\partial A) = 0$ we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} P\left\{\frac{S_k - b_k}{a_k} \in A\right\} = G(A). \tag{2.4}$$

COROLLARY 1. *Let $X_1, X_2, \dots$ be independent r.v.'s and $a_n > 0$, b_n numerical sequences such that setting $S_n = X_1 + \cdots + X_n$,* (2.1) *and* (2.2) *hold. Then for any distribution G the a.s. central limit theorem* (2.3) *holds provided*

$$(S_n - b_n)/a_n \to_{\mathscr{D}} G. \tag{2.5}$$

To clarify the meaning of (2.1) let us write out its special cases for $f(x) = x^p$, $f(x) = (\log x)^{\gamma}$, $f(x) = (\log\log x)^{1+\delta}$, respectively:

$$E\left|\frac{S_n - b_n}{a_n}\right|^p \le e^{(\log n)^{1-\varepsilon}}, \qquad 0 < p \le 1, 0 < \varepsilon < 1, \tag{2.1a}$$

$$E\left(\log\left|\frac{S_n - b_n}{a_n}\right|\right)^{\gamma} \le (\log n)^{\gamma'}, \qquad 0 < \gamma' < \gamma, \tag{2.1b}$$

$$E\left(\log\log\left|\frac{S_n - b_n}{a_n}\right|\right)^{1+\delta} \le (\log\log n)^{\delta'}, \qquad 0 < \delta' < \delta. \tag{2.1c}$$

[Here, and in the sequel, $\log x$ and $\log\log x$ are meant as $\log(x \vee e)$ and $\log\log(x \vee e^e)$, respectively.] Clearly, the slower the function f in (2.1) grows, the weaker type generalized moments of $(S_n - b_n)/a_n$ appear on the left-hand side of (2.1), but the more restrictive growth rates are required for these generalized moments. The choices $f(x) = x^p$ and $f(x) = (\log\log x)^{1+\delta}$ represent the limits for reasonable choices of f; for $f(x) = \log\log x$ the right-hand side of (2.1) tends to zero and thus (2.1) implies $(S_n - b_n)/a_n \to_P 0$. We do not know if Theorem 1 holds under

$$E\left(\log\log\left|\frac{S_n - b_n}{a_n}\right|\right) = O(1)$$

or even without (2.1). At any rate, (2.1c) is a rather mild restriction that holds in most standard situations; in fact, even if we replace the right-hand side by $O(1)$, condition (2.1c) is not much stronger than the stochastic boundedness of $(S_n - b_n)/a_n$, a natural assumption in problems concerning the limit distributional behaviour of $(S_n - b_n)/a_n$. [With the right-hand side tending to $+\infty$, (2.1c) does not even imply the stochastic boundedness of $(S_n - b_n)/a_n$.] Condition (2.2) is satisfied, for example, if

$$(2.6) \qquad n^{-\gamma}a_n \text{ is nondecreasing for some } \gamma > 0$$

or

$$(2.7) \qquad a_n = n^{\rho}L(n) \quad \text{where } \rho > 0 \text{ and } L \text{ is slowly varying.}$$

[The last statement follows from the fact that by the representation theorem for slowly varying functions, $L(l)/L(k) \geq C(l/k)^{-\varepsilon}$ for any $\varepsilon > 0$ and $l \geq k \geq k_0(\varepsilon)$.] As the examples at the end of this section will show, (2.6) cannot be essentially weakened; neither $n^{-\gamma}a_n \to \infty$ nor $(\log n)^{-\gamma}a_n \uparrow \infty$, $\gamma > 0$, suffice in Theorem 1.

In the first version of this paper we proved Theorem 1 under the simpler but more restrictive moment condition

$$E\left|\frac{S_n - b_n}{a_n}\right|^p = O(1), \qquad p > 0$$

instead of (2.1). The observation that the theorem remains valid, with only minor changes in the proof, under generalized moment assumptions of the type (2.1) is due to T. F. Móri who also constructed the examples at the end of this section.

Corollary 1 shows, as we already pointed out in the introduction, that the a.s. central limit theorem (2.3) is [under (2.1)] a weaker statement than the distributional result (2.5). In [2] examples are constructed to show that even in the i.i.d. case (2.3) can hold while (2.5) fails badly. The examples of [2] also show that while for i.i.d. r.v.'s (X_n) the limit distribution G in (2.5) is necessarily stable, this is not valid in the case of (2.3).

Our next theorem is the functional version of Theorem 1.

THEOREM 2. *Let $X_1, X_2, \ldots$ be independent r.v.'s and $a_n > 0$, b_n numerical sequences such that setting $S_n = X_1 + \cdots + X_n$,* (2.1) *and* (2.2) *hold. Let, for each $n \geq 1$, $0 = t_{n0} < t_{n1} < \cdots < t_{nn} = 1$ be a division of $[0,1]$ and* *define the normed partial sum process* $\{s_n(t),\ 0 \le t \le 1\}$ *by*

$$s_n(t) = \begin{cases} (S_j - b_j)/a_n, & \text{for } t = t_{nj},\ 0 \le j \le n, \\ \text{linear in between}. \end{cases}$$

Let finally μ *be a probability measure on* $D[0,1]$. *Then the following statements are equivalent:*

(a*) *For any Borel subset* A *of* $D[0,1]$ *with* $\mu(\partial A) = 0$ *we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\{s_k(\cdot) \in A\} = \mu(A) \quad a.s., \tag{2.8}$$

and the exceptional set of probability zero does not depend on A.

(b*) *For any Borel subset* A *of* $D[0,1]$ *with* $\mu(\partial A) = 0$ *we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} P\{s_k(\cdot) \in A\} = \mu(A). \tag{2.9}$$

Similarly to Corollary 1, a sufficient condition for the a.s. functional central limit theorem (2.8) is

$$\lim_{N\to\infty} P\{s_N(\cdot) \in A\} = \mu(A), \qquad A \subset D[0,1],\ \mu(\partial A) = 0, \tag{2.10}$$

that is, the ordinary functional central limit theorem. It is also worth noting that criterion (2.9) is equivalent to $\mu_N \to \mu$ weakly where μ_N denotes the probability measure on $D[0,1]$ defined by

$$\mu_N(A) = \left(\sum_{k \le N} \frac{1}{k} \right)^{-1} \sum_{k \le N} \frac{1}{k} P\{s_k(\cdot) \in A\}, \qquad A \subset D[0,1].$$

Thus (2.9) holds if and only if it is valid for finite dimensional sets A and $\{\mu_N\}$ is tight which is true, for example, if $\{s_N\}$ is tight.

We mention one more consequence of Theorem 2 concerning the i.i.d. case.

COROLLARY 2. *Let* (X_n) *be an i.i.d. sequence and* $a_n > 0$, b_n *numerical sequences such that setting* $S_n = X_1 + \cdots + X_n$ *we have*

$$(S_n - b_n)/a_n \to_{\mathscr{D}} G \tag{2.11}$$

with G *nondegenerate (necessarily* α*-stable for some* $0 < \alpha \le 2$). *Let* $\{\xi(t),\ 0 \le t \le 1\}$ *be the uniquely determined process with independent, stationary* α*-stable increments such that* $\xi(0) = 0$ *and* $\xi(1) =_{\mathscr{D}} G$. *Then* (X_n) *satisfies the a.s. functional central limit theorem* (2.8) *with* $t_{nk} = k/n$, $1 \le k \le n$, *where* μ *is the measure on* $D[0,1]$ *corresponding to* ξ.

In fact, (2.11) and the results of [8] easily imply that the ordinary functional limit theorem (2.10) is valid.

We turn now to the proof of Theorem 2; the proof of Theorem 1 is very similar (in fact simpler). Replacing X_n by $X_n - c_n$ where $c_n = b_n - b_{n-1}$, we can assume $b_n = 0$. Let $x_0 \geq 0$ denote a number such that $f(x)$ and $x/f(x)$ are nondecreasing for $x \geq x_0$; clearly $f(x)$ is continuous on $[x_0, \infty)$ and $f(\infty) = \infty$. Moreover, we can assume, without loss of generality, that $x_0 = 0$ and $f(0) = 0$. In fact, observe that

$$\hat{f}(x) = \begin{cases} \dfrac{x}{x_0} f(x_0), & \text{for } 0 \leq x < x_0, \\ f(x), & \text{for } x \geq x_0, \end{cases}$$

satisfies these stronger properties as well as (2.1) with an additional term $+O(1)$ on the right-hand side, which can be absorbed into

$$(\log\log n)^{-1-\varepsilon} f\left(\exp(\log n)^{1-\varepsilon}\right)$$

by replacing ε with a smaller number. [Recall that the right-hand side of (2.1) was assumed to be nondecreasing.] We next prove the following simple lemma.

LEMMA 1. *Under the assumptions of Theorem* 2 *we have, setting* $M_N = \max_{1 \leq k \leq N} |S_k|$,

$$Ef(M_N/a_N) \leq C^* \max_{1 \leq k \leq N} Ef(|S_k|/a_k), \qquad N \geq N_0 \tag{2.12}$$

for some constant $C^* > 0$.

PROOF. Set $\rho(N) = \max_{1 \leq k \leq N} Ef(|S_k|/a_k)$ and choose $\lambda > 0$ so that $f(\lambda) = 4\rho(N)/C$ where C is the constant in (2.2). Clearly we can assume without loss of generality that $C \leq 1$ and thus observing that $a_N \geq Ca_k$ for any $1 \leq k \leq N$ by (2.2), we get

$$\begin{aligned} P(|S_k| \geq \lambda a_N) &\leq P(|S_k| \geq C\lambda a_k) \leq f(C\lambda)^{-1} Ef(|S_k|/a_k) \\ &\leq 1/4, \qquad 1 \leq k \leq N, \end{aligned}$$

by the choice of λ and the fact that $f(C\lambda) \geq Cf(\lambda)$ by the assumptions made on f. Thus

$$P(|S_N - S_k| \geq 2\lambda a_N) \leq 1/2, \qquad 1 \leq k \leq N,$$

whence by a maximal inequality of Skorohod (see, e.g., Breiman [4], page 45) we get

$$P(M_N > xa_N) \leq 2P(|S_N| > xa_N/2) \quad \text{for } x \geq 4\lambda$$

that is,

$$P(\xi > x) \leq 2P(\eta > x), \qquad x > 0,$$

where

$$\xi = M_N/a_N \cdot I\{M_N/a_N > 4\lambda\}, \qquad \eta = 2|S_N|/a_N \cdot I\{2|S_N|/a_N > 4\lambda\}.$$

Thus an integration by parts yields $Ef(\xi) \leq 2Ef(\eta)$ whence we get, observing that $f(ax) \le af(x)$ for any $a \ge 1$, $x > 0$ by the assumptions made on f,

$$\begin{aligned} Ef(M_N/a_N) &\le Ef(\xi) + f(4\lambda) \le 2Ef(\eta) + 4f(\lambda) \\ &\le 4Ef(|S_N|/a_N) + 16C^{-1}\rho(N) \le C^*\rho(N) \end{aligned}$$

where $C^* = 4 + 16/C$, proving Lemma 1. □

Let $BL(D[0,1])$ denote the set of functions g: $D[0,1] \to \mathbf{R}$ such that for some $K > 0$,

$$|g(x) - g(y)| \le Kd(x,y), \qquad |g(x)| \le K \quad \text{for all } x, y \in D[0,1], \tag{2.13}$$

where d is the Skorohod metric (see, e.g., [3] page 111). By Theorem (8.3) of Dudley [9], statements (a*) and (b*) in Theorem 2 are equivalent, respectively, to the following:

(a′) For all $g \in BL(D[0,1])$,

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} g(s_k(\cdot)) = \int g\, d\mu \quad \text{a.s.} \tag{2.14}$$

(b′) For all $g \in BL(D[0,1])$,

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} Eg(s_k(\cdot)) = \int g\, d\mu.$$

[Note that by the separability argument in the proof of [9], Theorem (8.3), in statements (a′) and (b′) it suffices to consider g's from a suitable countable subset of $BL(D[0,1])$, whence it follows easily that if (a′) holds then the exceptional zero set in (2.14) can be chosen to be independent of g.] Hence to prove Theorem 2, it suffices to show that for any $g \in BL(D[0,1])$ we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} \xi_k = 0 \quad \text{a.s.} \tag{2.15}$$

where

$$\xi_k = g(s_k(\cdot)) - Eg(s_k(\cdot)).$$

To prove (2.15) we first show that

$$|E(\xi_k \xi_l)| \le 4K^2 \frac{Ef(M_k/a_k)}{f(a_l/a_k)}, \qquad 1 \le k < l, \tag{2.16}$$

where K denotes a constant such that (2.13) holds. Define, to this end, the function $s_{k,l}^*$: $[0,1] \to \mathbf{R}$ by

$$s_{k,l}^*(t) = \begin{cases} 0, & \text{if } 0 \le t \le t_{l,k}, \\ (S_j - S_k)/a_l, & \text{if } t_{l,j} \le t < t_{l,j+1},\ k \le j \le l-1. \end{cases}$$

Clearly

$$d(s_l, s_{k,l}^*) \le \|s_l - s_{k,l}^*\|_\infty \le M_k/a_l$$

and thus by (2.13),

$$|g(s_l) - g(s^*_{k,l})| \le K\frac{M_k}{a_l} \wedge 2K.$$

Also, $s^*_{k,l}$ is independent of s_k and thus setting $\lambda = a_l/a_k$ and using $x/y \le f(x)/f(y)$ for $x \le y$, we get

$$\begin{aligned}|E(\xi_k\xi_l)| &= |\mathrm{Cov}(g(s_k), g(s_l))| = |\mathrm{Cov}(g(s_k), g(s_l) - g(s^*_{k,l}))|\\ &\le 2K^2E\left(\frac{M_k}{a_l} \wedge 2\right) = 4K^2E\left(\frac{1}{\lambda}\left(\frac{M_k}{2a_k} \wedge \lambda\right)\right)\\ &\le \frac{4K^2}{f(\lambda)}Ef\left(\frac{M_k}{2a_k} \wedge \lambda\right) \le \frac{4K^2}{f(\lambda)}Ef\left(\frac{M_k}{a_k}\right),\end{aligned}$$

proving (2.16).

Now

$$E\left(\sum_{k\le N}\frac{1}{k}\xi_k\right)^2 \le 2\sum_{1\le k\le l\le N}\frac{1}{kl}|E(\xi_k\xi_l)|. \tag{2.17}$$

By (2.1), (2.2), (2.16) and Lemma 1, the contribution of those terms in the sum of the right-hand side of (2.17) where $l/k \ge \exp((\log N)^{1-\varepsilon/2})$, is

$$\begin{aligned}&\le 8K^2C^*(\log\log N)^{-1-\varepsilon}\frac{f(e^{(\log N)^{1-\varepsilon}})}{f(Ce^{\gamma(\log N)^{1-\varepsilon/2}})}\sum_{1\le k\le l\le N}\frac{1}{kl}\\ &\le 16K^2C^*(\log\log N)^{-1-\varepsilon}(\log N)^2.\end{aligned}$$

On the other hand, the trivial estimate $|E(\xi_k\xi_l)| \le 4K^2$ shows that the contribution of those terms where $l/k \le \exp((\log N)^{1-\varepsilon/2})$, is

$$\begin{aligned}&\le 8K^2\sum_{k=1}^{N}\frac{1}{k}\sum_{l=k}^{k\exp(\log N)^{1-\varepsilon/2}}\frac{1}{l} \le \text{const.}\sum_{k=1}^{N}\frac{1}{k}(\log N)^{1-\varepsilon/2}\\ &\le \text{const.}(\log N)^{2-\varepsilon/2}.\end{aligned}$$

Hence setting $T_N = (\log N)^{-1}\Sigma_{k\le N}k^{-1}\xi_k$, we get

$$ET_N^2 \le \text{const.}(\log\log N)^{-1-\varepsilon},$$

whence for $N_k = \exp(\exp(k^{1-\varepsilon/2}))$ we have

$$ET_{N_k}^2 \le \text{const.}\, k^{-1-\rho}$$

for some $\rho > 0$. By the Beppo Levi theorem we have $\Sigma_{k=1}^{\infty}|T_{N_k}| < +\infty$ a.s., implying $T_{N_k} \to 0$ a.s. Now for $N_k \le N \le N_{k+1}$ we have

$$|T_N| \le |T_{N_k}| + \frac{1}{\log N}\sum_{i=N_k+1}^{N}\frac{1}{i} \le |T_{N_k}| + \left(1 - \frac{\log N_k}{\log N}\right).$$

Since $\log N_{k+1}/\log N_k \to 1$, it follows that $T_N \to 0$ a.s., completing the proof of Theorem 2. □

REMARK. For $C \geq 1$, condition (2.2) of Theorem 1 requires that $n^{-\gamma}a_n$ be nondecreasing. The following two examples, due to T. F. Móri, show that upon replacing this condition by either $n^{-\gamma}a_n \to \infty$ or $(\log n)^{-\gamma}a_n \uparrow \infty$, $\gamma > 0$, Theorem 1 becomes false.

EXAMPLE 1. Let $X_1, X_2, \dots$ be independent, normally distributed r.v.'s with $EX_i = 0$, $EX_i^2 = 1/i$, $i = 1, 2, \dots$ and set $S_n = X_1 + \cdots + X_n$, $a_n = (\log n)^{1/2}$. Then $S_n/a_n \to_{\mathscr{D}} N(0,1)$ and thus (2.4) holds with $b_n = 0$ and $G = N(0,1)$. Further $E|S_n/a_n|^2 = O(1)$ and thus (2.1) is satisfied with $f(x) = x^p$, $0 < p \leq 2$. It is also easily seen that (X_n) obeys the Lindeberg condition and thus it satisfies the arc sine law (see, e.g., [22]), that is, $P(\eta_N < x) \to 2\pi^{-1} \arcsin\sqrt{x}$, $0 \leq x \leq 1$, where

$$\eta_N = \left(\sum_{k \leq N} \frac{1}{k} \right)^{-1} \sum_{k \leq N} \frac{1}{k} I\{S_k > 0\}.$$

Thus $\eta_N \to 1/2$ a.s. is not valid, that is, (2.3) does not hold.

EXAMPLE 2. Let $X_1, X_2, \dots$ be independent r.v.'s such that $X_{2^{2^n}}$ is $N(0, 2^{2^{n+1}})$ distributed, $n = 1, 2, \dots$ and $X_k = 0$ for all other indices. Let $S_N = X_1 + \cdots + X_N$ and $a_k = 2^{2^n}$ for $2^{2^n} \leq k < 2^{2^{n+1}}$; clearly $\sqrt{k} \leq a_k \leq k$, $S_N/a_N \to_{\mathscr{D}} N(0,1)$ and also $E|S_N/a_N|^2 = O(1)$ so that (2.1) and (2.4) hold with $f(x) = x^p$, $0 < p \leq 2$, $b_n = 0$ and $G = N(0,1)$. If relation (2.3) were also valid, then

$$\eta_N = (\log N)^{-1} \sum_{k \leq N} k^{-1} g(S_k/a_k)$$

would converge a.s. to a constant for any bounded Lipschitz-1 function g on the real line. But for $N = 2^{2^n} - 1$ we have

$$\eta_N = \sum_{i \leq n} g\left(S_{2^{2^i}}/2^{2^i}\right)\lambda_{i,N},$$

where

$$\lambda_{i,N} = (\log N)^{-1} \sum_{j=2^{2^i}}^{2^{2^{i+1}}} j^{-1} = 2^{i-n} + O(2^{-n});$$

moreover, $S_{2^{2^i}}/2^{2^i} = X_{2^{2^i}}/2^{2^i} + \varepsilon_i$ where $E(\varepsilon_i^2) = O(2^{-2^i})$. These remarks imply easily that the distribution of η_N converges, as N runs through the above special values, to the distribution of the r.v. $\eta = \sum_{i=1}^{\infty} 2^{-i} g(\xi_i)$, where ξ_i are i.i.d. $N(0,1)$ r.v.'s. Choosing a g such that the distribution of η is nondegenerate [e.g., $g(x) = \phi(x)$ will do], we get a contradiction.

3. The law of large numbers.

DEFINITION. Let $\xi_1, \xi_2, \dots$ and ξ be r.v.'s. We say that

$$\xi_n \to_P \xi \quad (\log),$$

if there exists a set $H \subset \mathbf{N}$ of log density 1 such that $\xi_n \to_P \xi$ as $n \to \infty$, $n \in H$. We say that

$$\xi_n \to \xi \quad \text{a.s. } (\log),$$

if for a.e. ω there exists a set $H_\omega \subset \mathbf{N}$ of log density 1 such that $\xi_n(\omega) \to \xi(\omega)$ as $n \to \infty$, $n \in H_\omega$.

We can now formulate the main result of this section.

THEOREM 3. *Let $X_1, X_2, \dots$ be independent r.v.'s with partial sums $S_n = X_1 + \cdots + X_n$ and let (a_n) be a positive numerical sequence such that* (2.1) *and* (2.2) *hold with $b_n = 0$. Then the following statements are equivalent*:

(a$_1$) $S_n/a_n \to 0$ *a.s.* (log)
(b$_1$) $S_n/a_n \to_P 0$ (log).

Moreover, if $X_n/a_n \to_P 0$ also holds then a third equivalent condition is:

(c$_1$) *We have*

$$\frac{1}{\log N} \sum_{k \le N} \frac{1}{k} \{G_k(a_k) \wedge 1\} \to 0 \quad \text{as } N \to \infty$$

and

$$\frac{1}{\log N} \sum_{k \le N} \frac{1}{k} \{H_k(a_k) \wedge 1\} \to 0 \quad \text{as } N \to \infty,$$

where

$$G_k(\lambda) = \sum_{j=1}^{k} \left[\int_{|x| \ge \lambda} dF_j(x) + \lambda^{-2} \left\{ \int_{|x| < \lambda} x^2 \, dF_j(x) - \left(\int_{|x| < \lambda} x \, dF_j(x) \right)^2 \right\} \right],$$

$$H_k(\lambda) = \frac{1}{\lambda} \left| \sum_{j=1}^{k} \int_{|x| < \lambda} x \, dF_j(x) \right|$$

and F_j is the distribution function of X_j.

Applying Theorem 3 to $X_n' = X_n - c_n$ where $c_n = b_n - b_{n-1}$, we get the analogous result for $(S_n - b_n)/a_n$. In particular, it follows that under the assumptions of Theorem 1 the relations

$$(S_n - b_n)/a_n \to 0 \quad \text{a.s. } (\log)$$

and

$$(S_n - b_n)/a_n \to_P 0 \quad (\log)$$

are equivalent.

COROLLARY 3. *Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with distribution function F satisfying either*

$$\liminf_{x\to\infty} x(1 - F(x) + F(-x)) > 0, \tag{3.1}$$

or

$$x(1 - F(x) + F(-x)) \le \exp\bigl((\log x)^{1-\varepsilon}\bigr), \qquad x \ge x_0 \tag{3.2}$$

for some $\varepsilon > 0$. Then setting $S_n = X_1 + \cdots + X_n$ we have

$$S_n/n \to 0 \quad a.s. \ (\log) \tag{3.3}$$

if and only if the following three conditions hold:

$$\begin{aligned} &\frac{1}{\log N}\sum_{k\le N}\left\{\int_{|x|\ge k} dF(x) \wedge \frac{1}{k}\right\} \to 0,\\ &\frac{1}{\log N}\sum_{k\le N}\left\{\frac{1}{k^2}\left[\int_{|x|<k} x^2\,dF(x) - \left(\int_{|x|<k} x\,dF(x)\right)^2\right] \wedge \frac{1}{k}\right\} \to 0,\\ &\frac{1}{\log N}\sum_{k\le N}\left\{\frac{1}{k}\left|\int_{|x|<k} x\,dF(x)\right| \wedge \frac{1}{k}\right\} \to 0. \end{aligned} \tag{3.4}$$

The assumptions of Corollary 3 require that the tails $1 - F(x) + F(-x)$ be either big [cf. (3.1)] or small [cf. (3.2)]. The only case excluded is when $x(1 - F(x) + F(-x))$ fluctuates irregularly between small and large values. It seems likely that Corollary 3 holds without any restrictions on F, but we cannot prove this.

For the proof of Theorem 3 and Corollary 3 we shall need the following lemma.

LEMMA 2. *Let x_n be a numerical sequence. Then the following statements are equivalent:*

(i) *There exists a subset $H \subset \mathbf{N}$ of log density 0 such that $x_n \to 0$ as $n \to \infty$, $n \notin H$.*

(ii) *For all $\varepsilon > 0$, the set $A(\varepsilon) = \{n: |x_n| > \varepsilon\}$ has log density 0, that is,*

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{k\le N}\frac{1}{k}I\{|x_k| > \varepsilon\} = 0.$$

Moreover, if x_n is bounded then (i) *and* (ii) *are equivalent to*

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{k\le N}\frac{1}{k}|x_k| = 0. \tag{iii}$$

For the equivalence of (i) and (ii) see Lemma (4.9) of Fisher [13]; the equivalence of (ii) and (iii) is obvious from the inequality $\varepsilon I\{|x_k| > \varepsilon\} \le |x_k| \le \varepsilon + CI\{|x_k| > \varepsilon\}$ where $C = \sup_k |x_k|$.

PROOF OF THEOREM 3. We first show the equivalence of (a_1) and (b_1). Clearly, in the case when $b_n = 0$ and G is the distribution concentrated at the origin, statements (a) and (b) in Theorem 1 reduce to

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\{|S_k/a_k| > \varepsilon\} = 0 \quad \text{a.s. for any } \varepsilon > 0 \tag{3.5}$$

and

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} P\{|S_k/a_k| > \varepsilon\} = 0 \quad \text{for any } \varepsilon > 0, \tag{3.6}$$

respectively. By Lemma 2, (3.5) is equivalent to (a_1) while (3.6) can be written equivalently as

$$\mu_L\{n\colon P(|S_n/a_n| > \varepsilon) > \delta\} = 0 \quad \text{for any } \varepsilon > 0,\ \delta > 0. \tag{3.7}$$

Setting

$$x_n = \inf\{\rho > 0\colon P(|S_n/a_n| > \rho) \le \rho\},$$

(3.7) implies

$$\mu_L\{n\colon x_n > \varepsilon\} \le \mu_L\{n\colon P(|S_n/a_n| > \varepsilon) > \varepsilon\} = 0 \quad \text{for all } \varepsilon > 0$$

whence we get, using Lemma 2, that $x_n \to 0$ along a sequence $H \subset \mathbf{N}$ of log density 1. But then by a well-known property of convergence in probability we get $S_n/a_n \to_P 0$ as $n \to \infty$, $n \in H$, that is, (b_1) holds. Conversely, (b_1) trivially implies (3.7) and thus (3.6).

To prove the equivalence of (b_1) and (c_1) we first note that by (2.2) we have $a_N \ge Ca_k$ for any $1 \le k \le N$ and thus $X_n/a_n \to_P 0$ implies

$$\max_{1\le k\le N} P(|X_k| \ge \varepsilon a_N) \to 0 \quad \text{for any } \varepsilon > 0.$$

Hence by the standard degenerate convergence criterion (see, e.g., Loève [16], page 317) for a fixed set $H \subset \mathbf{N}$ of integers we have

$$(X_1 + \cdots + X_n)/a_n \to_P 0 \quad \text{as } n \to \infty,\ n \in H \tag{3.8}$$

if and only if

$$x_n := \sum_{k=1}^{n} \int_{|x|\ge a_n} dF_k(x) \to 0 \quad \text{as } n \to \infty,\ n \in H, \tag{3.9}$$

$$y_n := a_n^{-2} \sum_{k=1}^{n} \left[\int_{|x|<a_n} x^2\, dF_k(x) - \left(\int_{|x|<a_n} x\, dF_k(x) \right)^2 \right] \to 0, \quad \text{as } n \to \infty,\ n \in H, \tag{3.10}$$

$$z_n := a_n^{-1} \sum_{k=1}^{n} \int_{|x|<a_n} x\, dF_k(x) \to 0 \quad \text{as } n \to \infty,\ n \in H. \tag{3.11}$$

[Note that in [16] one has εa_n instead of a_n in (3.9) but since (3.9), (3.10) and (3.11) together trivially imply (3.8) by truncation and the Chebyshev inequality, (3.9), (3.10) and (3.11) together are equivalent to (3.8).] Thus we see that statement (b_1) holds if and only if

$$x_n \to 0, \quad y_n \to 0, \quad z_n \to 0 \quad (\log), \tag{3.12}$$

where for a numerical sequence c_n the relation $c_n \to 0$ (log) means that $c_n \to 0$ on a set of log density 1. By $x_n \geq 0$, $y_n \geq 0$, (3.12) is equivalent to

$$(x_n + y_n) \wedge 1 \to 0 \quad (\log) \quad \text{and} \quad |z_n| \wedge 1 \to 0 \quad (\log). \tag{3.13}$$

Applying Lemma 2 it follows that (3.13) is equivalent to (c_1). □

Proof of Corollary 3. Assume first that (3.2) holds for some $\varepsilon > 0$; set $\psi(x) = \exp((\log x)^{1-\varepsilon})$ and $c_n = n\exp((\log n)^{1-\varepsilon/2})$. It is easy to see that

$$n\int_{|x|\geq c_n} dF(x) \to 0 \quad \text{as } n \to \infty, \tag{3.14}$$

$$nc_n^{-1}\int_{|x|<c_n} x\,dF(x) \to 0 \quad \text{as } n \to \infty, \tag{3.15}$$

$$nc_n^{-2}\int_{|x|<c_n} x^2\,dF(x) \to 0 \quad \text{as } n \to \infty. \tag{3.16}$$

In fact, by (3.2) the left-hand side of (3.14) is $\leq nc_n^{-1}\psi(c_n) \leq nc_n^{-1}\psi(n^2) = o(1)$; on the other hand, an integration by parts shows that the absolute value of the left-hand side of (3.15) is bounded by

$$\begin{aligned} nc_n^{-1}\left(6 + \int_3^{c_n}\psi(x)/x\,dx\right) + n\psi(c_n)c_n^{-1} &\leq 3nc_n^{-1}\psi(c_n)\log c_n \\ &\leq 3nc_n^{-1}\psi(n^2)\log n^2 = o(1). \end{aligned}$$

The proof of (3.16) is the same. Relations (3.14)–(3.16) evidently imply (by truncating the r.v.'s X_i, $1 \leq i \leq n$ at c_n and applying the Chebyshev inequality) that $S_n/c_n \to_P 0$ and since

$$c_{mn}/c_n \leq m^{1+\tau(n)},$$

where $\tau(n) = \text{const.}(\log n)^{-\varepsilon/2} \to 0$, the proof of Theorem (6.1) of [1] shows that $E|S_n/c_n|^p$ remains bounded for any $0 < p < 1$ and consequently

$$E|S_n/n|^{1/2} \leq K\exp\{(\log n)^{1-\varepsilon/2}\} \tag{3.17}$$

for some constant $K > 0$. Since $X_n/n \to_P 0$ trivially, Theorem 3 applies to the sequence (X_n) [see (2.1a) in Section 2], proving the equivalence of (3.3) and (3.4). [We replace the two relations in (c_1) again with three relations; see (3.9)–(3.11) in the proof of $(b_1) \Leftrightarrow (c_1)$ in Theorem 3.]

To complete the proof of Corollary 3 it suffices to show that if (3.1) holds then (3.3) and the first relation of (3.4) fail. To this effect, let us observe that (3.5) trivially implies (3.6) by the bounded convergence theorem and thus the implication $(a_1) \Rightarrow (b_1)$ in Theorem 3 holds without any assumption on (X_n) and (a_n). Similarly, the proof of Theorem 3 shows that the equivalence of (b_1) and (c_1) holds without (2.1), that is, assuming only (2.2) and $X_n/a_n \to_P 0$ for the independent sequence (X_n). Now (2.2) trivially holds for $a_n = n$ and $X_n/n \to_P 0$ holds for any i.i.d. sequence (X_n), hence in Corollary 3 relation (3.3) implies (3.4) for any i.i.d. sequence (X_n). Since under (3.1) the first relation of (3.4) trivially fails, it follows that under (3.1) both (3.3) and (3.4) are false. □

We formulate one more corollary of the proof of Theorems 2 and 3 which we shall need in Section 4.

COROLLARY 4. *Under the conditions of Theorem* 3 *the following two statements are equivalent*:

(d) $$a_n^{-1} \max_{1 \le k \le n} |S_k| \to 0 \quad a.s. \,(\log)$$

(e) $$a_n^{-1} \max_{1 \le k \le n} |S_k| \to_P 0 \quad (\log).$$

Moreover, (d) *and* (e) *follow from* $S_n/a_n \to_P 0$.

PROOF. By Lemma 2 and an argument used in the proof of $(a_1) \Leftrightarrow (b_1)$ in Theorem 3, (d) and (e) are equivalent to

$$\frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\Big\{ \max_{1 \le j \le k} |S_j|/a_k > \varepsilon \Big\} \to 0 \quad \text{a.s. for all } \varepsilon > 0 \tag{3.18}$$

and

$$\frac{1}{\log N} \sum_{k \le N} \frac{1}{k} P\Big\{ \max_{1 \le j \le k} |S_j|/a_k > \varepsilon \Big\} \to 0 \quad \text{for all } \varepsilon > 0, \tag{3.19}$$

respectively. Let μ_N denote the probability measure on $D[0,1]$ with mass $k^{-1}L(N)^{-1}$ at the points $s_k(\cdot)$, $1 \le k \le N$, where $L(N) = \sum_{k \le N} k^{-1} \sim \log N$. Then (3.18) is equivalent to the statement that $\mu_N(A_\varepsilon) \to 0$ a.s. for all $\varepsilon > 0$ where $A_\varepsilon = \{x \in D[0,1]\colon \|x\|_\infty \ge \varepsilon\}$ and hence to

$$\mu_N \to_w \delta_0 \quad \text{a.s.}, \tag{3.20}$$

where $\to_w$ denotes weak convergence. (Note that the sup functional is continuous in the Skorohod topology.) By Theorem (8.3) of Dudley [9] and the separability argument in its proof, (3.20) is equivalent to

$$\int g(x)\, d\mu_N(x) \to g(0) \quad \text{a.s. for all } g \in BL(D[0,1]),$$

that is,

$$\frac{1}{L(N)} \sum_{k \le N} \frac{1}{k} g(s_k(\cdot)) \to g(0) \quad \text{a.s. for all } g \in BL(D[0,1]). \tag{3.21}$$

In a completely similar fashion, (3.19) is equivalent to

$$\frac{1}{L(N)} \sum_{k \le N} \frac{1}{k} Eg(s_k(\cdot)) \to g(0) \quad \text{for all } g \in BL(D[0,1]). \tag{3.22}$$

Now relation (2.15) in the proof of Theorem 2 implies the equivalence of (3.21) and (3.22) and hence that of (d) and (e).

Assume now $S_n/a_n \to_P 0$; observing that $a_n \ge Ca_k$ for $1 \le k \le n$ by (2.2), it follows that

$$P(|S_k| \ge \varepsilon a_n) \le 1/4 \quad \text{for } 1 \le k \le n,\ n \ge n_0(\varepsilon),$$

that is,

$$P(|S_n - S_k| \ge 2\varepsilon a_n) \le 1/2 \quad \text{for } 1 \le k \le n,\ n \ge n_0(\varepsilon).$$

By Skorohod's maximal inequality (see, e.g., Breiman [4], page 45) the last relation implies

$$P\Big(\max_{1 \le k \le n} |S_k| \ge 4\varepsilon a_n\Big) \le 2P(|S_n| \ge 2\varepsilon a_n) \to 0 \quad \text{as } n \to \infty,$$

that is, (e) holds. □

EXAMPLE 3. To conclude this section, we construct a sequence (X_n) of i.i.d. symmetric r.v.'s such that setting $S_n = X_1 + \cdots + X_n$ we have $S_n/n \to 0$ a.s. (log) but $S_n/n \not\to_P 0$. To this end, let Y be a symmetric r.v. taking the values $\pm 3, \pm 4, \dots$ such that

$$P(|Y| = i) = p_i = Ci^{-2}(\log i)^{-1}, \qquad i = 3, 4, \dots,$$

for some (uniquely determined) constant $C > 0$. Let G denote the distribution function of Y. Clearly

$$P(|Y| \ge n) \sim Cn^{-1}(\log n)^{-1} \quad \text{as } n \to \infty, \tag{3.23}$$

$$\int_{|x| \le n} x^2\, dG(x) \sim Cn(\log n)^{-1} \quad \text{as } n \to \infty. \tag{3.24}$$

Let μ be the atomic probability measure on $\mathbf{N}$ defined by $\mu(\{n\}) = p_n$, $n = 3, 4, \dots$ and construct a new probability measure μ' from μ by concentrating, for each $k \ge 1$, the total mass of μ on the interval $I_k = [2^{k^4}, 2^{k^4+k^2}]$ onto the single point $t_k = 2^{k^4+[k^2/2]}$. Let X be a symmetric r.v. such that the distribution of $|X|$ is μ' and let F denote the distribution function of X. Set also

$$H = \bigcup_{k=1}^{\infty} I_k.$$

Letting $k = k(N) \sim \text{const.}(\log N)^{1/4}$ denote the largest integer such that $I_k \subset [0, N]$, we have

$$\sum_{i \le N,\, i \in H} i^{-1} \le \sum_{r \le k+1} \sum_{i \in I_r} i^{-1} \le \sum_{r \le k+1} 2r^2 \log 2 = O\big((\log N)^{3/4}\big),$$

and thus H is of log density zero. We also claim that

$$nP(|X| \geq n) \to 0 \quad \text{as } n \to \infty,\ n \notin H, \tag{3.25}$$

$$n^{-1} \int_{|x| \leq n} x^2\, dF(x) \to 0 \quad \text{as } n \to \infty,\ n \notin H, \tag{3.26}$$

$$\limsup_{n \to \infty} nP(|X| \geq n) = +\infty, \tag{3.27}$$

$$E\left(|X| e^{-\sqrt{\log |X|}}\right) < +\infty. \tag{3.28}$$

Equation (3.25) follows from (3.23) since by the definition of μ' we have $P(|X| \geq n) = P(|Y| \geq n)$ for $n \notin H$. To show (3.26) observe that

$$\int_{I_k} x^2\, dF(x) = t_k^2 \sum_{i \in I_k} \frac{1}{2} p_i = \frac{C \cdot 2^{k^4 + k^2 + O(1)}}{2k^4 \log 2},$$

$$\int_{I_k} x^2\, dG(x) = \sum_{i \in I_k} \frac{1}{2} i^2 p_i \sim \frac{C \cdot 2^{k^4 + k^2}}{2k^4 \log 2},$$

whence by the definition of μ' and H we get

$$\int_{|x| \leq n} x^2\, dF(x) \approx \int_{|x| \leq n} x^2\, dG(x) \quad \text{as } n \to \infty,\ n \notin H,$$

where the symbol $\approx$ means that the ratio of the two sides remains between positive constants. Hence (3.26) follows from (3.24). To prove (3.27) it suffices to observe that

$$t_k P(|X| \geq t_k) = t_k P\left(|Y| \geq 2^{k^4}\right) \sim \frac{C t_k}{2^{k^4} k^4 \log 2} \to +\infty.$$

Finally, letting $q_i = P(|X| = i)$ and $r_k = \sum_{i \in I_k} q_i \sim C \cdot 2^{-k^4} k^{-4} (\log 2)^{-1}$ we get that the expectation in (3.28) equals

$$\sum_{i \notin H} i e^{-\sqrt{\log i}} C i^{-2} (\log i)^{-1} + \sum_{k=1}^{\infty} t_k e^{-\sqrt{\log t_k}} r_k < \infty$$

since the above asymptotics for r_k show that the kth term of the second sum is less than or equal to $\exp(-\text{const. } k^2)$. Let now $X_1, X_2, \ldots$ be i.i.d. r.v.'s with distribution function F and set $S_n = X_1 + \cdots + X_n$. Truncating the r.v.'s X_i, $1 \leq i \leq n$ at n and using the Chebyshev inequality, (3.25) and (3.26) imply

$$S_n / n \to_P 0 \quad \text{as } n \to \infty,\ n \notin H. \tag{3.29}$$

Also, letting $\psi(x) = x \exp(-\sqrt{\log x})$ and $c_n = n \exp(2\sqrt{\log n})$, we have $\sum_{k \geq n} c_k^{-2} = O(n c_n^{-2})$ by Lemma 3 in Section 4 since $\exp(-\sqrt{\log x})$ is slowly varying; further $\psi(c_n) \geq n$ for $n \geq n_0$ and thus using $E\psi(|X|) < +\infty$ we get

$$\sum_{n \geq 1} P(|X| \geq c_n) \leq \sum_{n \geq 1} P(\psi(|X|) \geq n) < +\infty.$$

Hence applying Theorem 18 in Petrov [21], Chapter 9, Section 3 we get

$$S_n/c_n \to 0 \quad \text{a.s.}$$

whence, as in the proof of Corollary 3, it follows that

$$E|S_n/n|^{1/2} \le K \exp\left(\sqrt{\log n}\,\right).$$

Hence Theorem 3 applies to the sequence (X_n) [cf. (2.1a)] and thus (3.29) entails $S_n/n \to 0$ a.s. (log). Finally, using (3.27) and the classical degenerate convergence criterion (see, e.g., [16] page 317) it follows that $S_n/n \not\to_P 0$.

4. A.s. invariance principles. Before we formulate our results, we recall a few classical facts from probability theory. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with distribution function F. By the classical theory (see, e.g., Feller [12], pages 540–547 and also pages 302–306) there exist numerical sequences $a_n > 0$ and b_n such that (1.8) holds with a nondegenerate G if and only if there exists a slowly varying function $L(x)$ such that

$$\int_{-x}^{x} t^2\, dF(t) \sim x^{2-\alpha} L(x) \quad \text{as } x \to \infty \tag{4.1}$$

with $0 < \alpha \le 2$ and when $\alpha < 2$ then also

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \to p, \qquad \frac{F(-x)}{1 - F(x) + F(-x)} \to q \quad \text{as } x \to \infty \tag{4.2}$$

for some $p \ge 0$, $q \ge 0$. (Clearly $p + q = 1$.) If $\alpha = 2$ then the limit distribution G is normal; for $0 < \alpha < 2$ the above necessary and sufficient condition is equivalent to

$$1 - F(x) \sim p\frac{2-\alpha}{\alpha} x^{-\alpha} L(x), \quad F(-x) \sim q\frac{2-\alpha}{\alpha} x^{-\alpha} L(x), \quad x \to \infty. \tag{4.3}$$

In this case the limit distribution G is nonnormal stable and satisfies

$$1 - G(x) \sim p\frac{2-\alpha}{\alpha} x^{-\alpha}, \qquad G(-x) \sim q\frac{2-\alpha}{\alpha} x^{-\alpha}, \qquad x \to \infty. \tag{4.4}$$

In all cases the norming factor a_n in (1.8) satisfies

$$a_n^{\alpha} \sim nL(a_n) \quad \text{as } n \to \infty. \tag{4.5}$$

Moreover, the sequence $L(a_n)$ is also slowly varying, that is, a_n is regularly varying with exponent $1/\alpha$. Further, if for $\alpha > 1$ we center F to 0 expectation [note that for $\alpha > 1$, F has a finite expectation by (4.1)], then in (1.8) we can take $b_n = 0$ for all $\alpha \ne 1$. Finally, (4.1) is equivalent to

$$\lim_{x \to \infty} \frac{x^2(1 - F(x) + F(-x))}{\int_{-x}^{x} t^2\, dF(t)} = \frac{2-\alpha}{\alpha}. \tag{4.6}$$

We can now formulate our results.

THEOREM 4. *Let $X_1, X_2, \dots$ be i.i.d. r.v.'s such that* (1.8) *holds where G is the standard normal distribution. Assume, without loss of generality, that $EX_1 = 0$. Then the sequence $\{X_n\}$ can be redefined on a suitable probability space together with a Wiener process W such that*

$$\max_{1\le k\le n}\left|\sum_{i=1}^{k} X_i - W(a_k^2)\right| = o(a_n) \quad a.s.\ (\log). \tag{4.7}$$

THEOREM 5. *Let $X_1, X_2, \dots$ be i.i.d. r.v.'s such that* (1.8) *holds where G is an α-stable distribution, $0 < \alpha < 2$. In the case $\alpha > 1$ assume also, without loss of generality, that $EX_1 = 0$. Then after suitably enlarging the probability space there exist i.i.d. r.v.'s $Y_1, Y_2, \dots$ with distribution G such that setting $\lambda_i = L(a_i)^{1/\alpha}$ [L is the slowly varying function in* (4.3)], *we have for $\alpha \ne 1$*

$$\max_{1\le k\le n}\left|\sum_{i=1}^{k} (X_i - \lambda_i Y_i)\right| = o(a_n) \quad a.s.\ (\log). \tag{4.8}$$

The result remains valid also for $\alpha = 1$ except that in this case a centering constant c_k should be subtracted from the sum on the left-hand side of (4.8).

As the example at the end of this section will show, the centering c_k is really necessary in (4.8) in the case $\alpha = 1$.

We formulated Theorem 4 and Theorem 5 separately only to give Theorem 4 a more traditional form. As the simple calculation at the end of the proof of Theorem 4 shows, (4.7) is equivalent to (4.8) with $Y_1, Y_2, \dots$ i.i.d. $N(0,1)$ r.v.'s and $\lambda_i = L(a_i)^{1/\alpha}$ where $\alpha = 2$ and $L(x) = \int_{-x}^{x} t^2\, dF(t)$. Hence defining L universally for all $0 < \alpha \le 2$ by (4.1), Theorem 5 holds in a unified form for all nondegenerate limits G (stable or normal) in (1.8).

We now turn to the proof of Theorems 4 and 5. In what follows, we shall make repeated use of the following well known property of regularly varying functions (see, e.g., [12], pages 272–274):

LEMMA 3. *Let the function $Z > 0$ be regularly varying with exponent γ. Then*

$$\int_1^x Z(t)\,dt \sim \sum_{k\le x} Z(k) \sim \frac{1}{\gamma+1} x Z(x) \quad \text{as } x \to \infty \text{ if } \gamma > -1,$$

$$\int_x^\infty Z(t)\,dt \sim \sum_{k\ge x} Z(k) \sim \frac{1}{|\gamma+1|} x Z(x) \quad \text{as } x \to \infty \text{ if } \gamma < -1.$$

PROOF OF THEOREM 4. Let W be a Wiener process and $a_k > 0$ a numerical sequence. By the functional a.s. central limit theorem for the Wiener process, in the case $a_k = \sqrt{k}$ we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I_A\{W^{(a_k)}\} = \lambda_w(A) \quad \text{a.s.} \tag{4.9}$$

for any Borel set $A \subset C(0,\infty)$ with $\lambda_w(\partial A) = 0$; here λ_w is the Wiener measure and $W^{(a_k)}$ is the scaled process defined by

$$W^{(a_k)}(t) = a_k^{-1} W(a_k^2 t), \qquad t \geq 0. \tag{4.10}$$

The simple proof of this fact given in [15] shows that (4.9) remains valid for arbitrary a_k with $a_k/\sqrt{k} \to +\infty$. [In fact, the faster a_k tends to infinity, the better the covariance estimates used in the proof of (4.9) become.] In the sequel, we will use (4.9) in the case when a_k is the norming factor in Theorem 4; clearly then $a_k = \sqrt{k}\, L(k)$, where L is slowly varying and $\lim_{k\to\infty} L(k) = +\infty$.

Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with distribution function F such that for some numerical sequences $a_n > 0$ and b_n, (1.8) holds where G is the standard normal distribution. Then X_1 has a finite expectation and thus without loss of generality we may assume $EX_1 = 0$. Let

$$X_k' = X_k I(|X_k| < a_k), \qquad X_k'' = X_k I(|X_k| \geq a_k).$$

We first note that

$$\max_{1\leq k\leq n} \left| \sum_{i=1}^{k} X_i'' \right| = o(a_n) \quad \text{a.s. (log)}. \tag{4.11}$$

Indeed, by the classical facts summarized above the function $L(x) = \int_{-x}^{x} t^2\, dF(t)$ is slowly varying at infinity and setting $U(x) = 1 - F(x) + F(-x)$, we have

$$x^2 U(x)/L(x) \to 0 \quad \text{a.s. } x \to +\infty. \tag{4.12}$$

Now using Lemma 3 and (4.12) we get

$$E|X_k''| = \int_{|x|\geq a_k} |x|\, dF(x) = \int_{a_k}^{\infty} U(x)\, dx + a_k U(a_k)$$
$$= o\left(\int_{a_k}^{\infty} x^{-2} L(x)\, dx + a_k^{-1} L(a_k) \right) = o\left(a_k^{-1} L(a_k)\right).$$

Since $a_k^{-1} L(a_k)$ varies regularly with exponent $-1/2$, Lemma 3 and (4.5) yield

$$\sum_{k=1}^{n} a_k^{-1} L(a_k) \sim 2n a_n^{-1} L(a_n) \sim 2a_n$$

and thus

$$E\left(\frac{1}{a_n} \sum_{k=1}^{n} |X_k''| \right) \to 0. \tag{4.13}$$

Equation (4.13) and the implication (2.7) $\Rightarrow$ (2.2) observed in Section 2 show that the sequence $|X_k''|$ satisfies the assumptions of Theorem 1 (and thus of Theorems 2 and 3) with $f(x) = x$, $b_n = 0$; further, (4.13) implies

$a_n^{-1}\Sigma_{k=1}^n|X_k''| \to_P 0$ and thus Theorem 3 yields

$$\sum_{k=1}^{n} |X_k''| = o(a_n) \quad \text{a.s. (log)}$$

which clearly implies (4.11).

By the Skorohod embedding theorem the sequence $\{X_n\}$ can be redefined on a suitable probability space together with a Wiener process W such that

$$\sum_{k=1}^{n} (X_k' - EX_k') = W(T_1 + \cdots + T_n), \qquad n \geq 1, \tag{4.14}$$

where $T_1, T_2, \ldots$ are independent r.v.'s with

$$ET_n = E(X_n' - EX_n')^2 = \int_{-a_n}^{a_n} x^2\,dF(x) - \left(\int_{-a_n}^{a_n} x\,dF(x)\right)^2$$

and

$$ET_n^2 \leq 4E(X_n' - EX_n')^4 \leq 64E(X_n')^4 = 64\int_{-a_n}^{a_n} x^4\,dF(x).$$

Now for any $\varepsilon > 0$ there is $x_0 = x_0(\varepsilon)$ such that $\int_{x_0}^{\infty} dF(x) \leq \varepsilon$ and thus using $\int_{-\infty}^{+\infty} x^2\,dF(x) = +\infty$ we have for $n \geq n_0$,

$$\begin{aligned}\left|\int_{-a_n}^{a_n} x\,dF(x)\right| &\leq \left|\int_{-x_0}^{x_0} x\,dF(x)\right| \\ &\quad + \left(\int_{x_0 \leq |x| \leq a_n} x^2\,dF(x)\right)^{1/2} \left(\int_{|x| \geq x_0} dF(x)\right)^{1/2} \\ &\leq 2\varepsilon^{1/2}\left(\int_{|x| \leq a_n} x^2\,dF(x)\right)^{1/2}\end{aligned}$$

and thus

$$ET_n \sim \int_{-a_n}^{a_n} x^2\,dF(x) = L(a_n).$$

Moreover, by (4.12), (4.5) and Lemma 3 we get

$$\begin{aligned}\int_{-a_n}^{a_n} x^4\,dF(x) &= 4\int_0^{a_n} U(x)x^3\,dx + a_n^4 U(a_n) \\ &= o\left(\int_0^{a_n} xL(x)\,dx + a_n^2 L(a_n)\right) = o\big(a_n^2 L(a_n)\big).\end{aligned}$$

Set $D_n = \Sigma_{i=1}^n T_i$. Since $a_n^2 L(a_n)$ is regularly varying with exponent 1, using the above estimates, (4.5) and Lemma 3 we get

$$ED_n \sim \sum_{k=1}^{n} L(a_k) \sim nL(a_n) \sim a_n^2$$

and

$$E|D_n - ED_n|^2 \le \sum_{k=1}^{n} ET_k^2 = o\left(\sum_{k=1}^{n} a_k^2 L(a_k)\right)$$
$$= o\big(na_n^2 L(a_n)\big) = o\big(a_n^4\big),$$

that is,

$$E|D_n/a_n^2 - 1| \to 0. \tag{4.15}$$

Since a_n^2 is regularly varying with exponent 1, (4.15) and the implication (2.7) $\Rightarrow$ (2.2) observed in Section 2 show that the independent sequence (X_n^*) whose nth partial sum is $D_n - a_n^2$ satisfies the assumptions of Theorems 1–3 with $f(x) = x$, $b_n = 0$ and a_n^2 in place of a_n. Also, $D_n/a_n^2 \to_P 1$ by (4.15) and thus Corollary 4 yields

$$\max_{1 \le k \le n} |D_k - a_k^2| = o\big(a_n^2\big) \quad \text{a.s. (log)}. \tag{4.16}$$

Further, (4.13) and $EX_1 = 0$ imply $\sum_{k=1}^{n} |EX_k'| = \sum_{k=1}^{n} |EX_k''| = o(a_n)$ which, together with (4.14) and (4.11), yields

$$\max_{1 \le k \le n} \left| \sum_{i=1}^{k} X_i - W(D_k) \right| = o(a_n) \quad \text{a.s. (log)}.$$

Hence to prove Theorem 4 it remains to show

$$\max_{1 \le k \le n} |W(D_k) - W\big(a_k^2\big)| = o(a_n) \quad \text{a.s. (log)},$$

which by Lemma 2 is equivalent to

$$\limsup_{N \to \infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{ \max_{1 \le j \le k} |W(D_j) - W\big(a_j^2\big)| \ge \varepsilon a_k \right\} = 0 \quad \text{a.s.} \tag{4.17}$$

for every $\varepsilon > 0$.

For any $\delta > 0$, the indicator function in (4.17) is bounded by

$$I\left\{ \max_{1 \le j \le k} |D_j - a_j^2| \ge \delta a_k^2 \right\} + I\left\{ \sup_{\substack{0 \le t \le t' \le 2a_k^2 \\ |t - t'| \le \delta a_k^2}} |W(t) - W(t')| \ge \varepsilon a_k \right\}.$$

(Note that $\max_{1 \le k \le n} a_k \sim a_n$ by a well-known property of regularly varying functions; see, e.g., Seneta [24], pages 19–20, Property 4°). By (4.16) we have

$$\lim_{N \to \infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{ \max_{1 \le j \le k} |D_j - a_j^2| \ge \delta a_k^2 \right\} = 0 \quad \text{a.s.,}$$

and thus the lim sup in (4.17) is bounded by

$$\limsup_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{ \sup_{\substack{0\le t\le t'\le 2a_k^2 \\ |t-t'|\le \delta a_k^2}} |W(t) - W(t')| \ge \varepsilon a_k \right\} \tag{4.18}$$

$$= \limsup_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left(W^{(a_k)} \in A_\delta\right)$$

for any $\delta > 0$, where $W^{(a_k)}$ is defined by (4.10) and

$$A_\delta = \left\{ x \in C(0,\infty) \colon \sup_{\substack{0\le t\le t'\le 2 \\ |t-t'|\le \delta}} |x(t) - x(t')| \ge \varepsilon \right\}.$$

It is easily seen that the boundary of A_δ has Wiener measure 0 and thus by (4.9) the lim sup in (4.18) equals $\lambda_w(A_\delta)$. But $\lambda_w(A_\delta) \to 0$ as $\delta \to 0$ by the continuity of the sample paths of W. Thus the lim sup in (4.17) equals 0 for each $\varepsilon > 0$ and the proof of Theorem 4 is complete. □

Let $d_n = \sum_{k=1}^{n} L(a_k)$. By Lemma 3 and (4.5) we have $d_n \sim nL(a_n) \sim a_n^2$ and thus the above argument yields

$$\max_{1\le k\le n} |W(d_k) - W(a_k^2)| = o(a_n) \quad \text{a.s. (log)}.$$

Now letting $\lambda_i = L(a_i)^{1/2}$, $Y_i = (d_i - d_{i-1})^{-1/2}(W(d_i) - W(d_{i-1}))$, $i = 1, 2, \dots$, we have $W(d_n) = \sum_{i=1}^{n} \lambda_i Y_i$, and thus (4.7) can be equivalently written as (4.8) with $Y_1, Y_2, \dots$ i.i.d. $N(0,1)$ r.v.'s.

Proof of Theorem 5. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with distribution function F satisfying (4.3) where $0 < \alpha < 2$, $p \ge 0$, $q \ge 0$, $p + q = 1$ and L is a slowly varying function. Enlarge the probability space to carry an i.i.d. sequence $\{Z_k\}$ independent of $\{X_k\}$, with distribution function Φ. Since a_n in (1.8) satisfies $a_n/n^\gamma \to +\infty$ for any $1/2 < \gamma < 1/\alpha$, the law of the iterated logarithm shows that it suffices to prove (4.8) for the sequence $X_k^* = X_k + Z_k$. Hence without loss of generality we may assume that F is continuous. We show that the conclusion of the theorem holds with Y_k defined by the standard quantile transform, that is,

$$Y_k = G^{-1}(F(X_k)), \qquad k = 1, 2, \dots, \tag{4.19}$$

where G^{-1} is defined by

$$G^{-1}(x) = \inf\{t \colon G(t) \ge x\}.$$

Since F is continuous, $F(X_k)$ is uniform over $(0,1)$ and $G^{-1}(F(X_k))$ has distribution G. We show that (4.8) holds with $\lambda_k = L(a_k)^{1/\alpha}$ for any $\alpha \ne 1$ and in the case $\alpha = 1$ with an extra centering constant c_k on the left-hand side of (4.8).

From the standard representation formula for slowly varying functions (see, e.g., [12], page 274) it follows that if $\varepsilon_k \downarrow 0$ sufficiently slowly then

$$\lim_{k \to \infty} \sup_{\varepsilon_k \le x \le \varepsilon_k^{-1}} \left| \frac{L(a_k x)}{L(a_k)} - 1 \right| = 0. \tag{4.20}$$

Fix such a sequence ε_k; we can also assume that ε_k is slowly varying. We shall prove the following lemma:

LEMMA 4. *There exist numerical sequences* d_n, e_n, f_n *such that*

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \lambda_k Y_k) I(|X_k| \le \varepsilon_k a_k) - d_n \to_P 0, \tag{4.21}$$

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \lambda_k Y_k) I(|X_k| \ge \varepsilon_k^{-1} a_k) - e_n \to_P 0, \tag{4.22}$$

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \lambda_k Y_k) I(\varepsilon_k a_k < |X_k| < \varepsilon_k^{-1} a_k) - f_n \to_P 0. \tag{4.23}$$

Moreover, for $\alpha \ne 1$ *we have* $d_n + e_n + f_n \to 0$.

Clearly, Lemma 4 implies

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \lambda_k Y_k) - c_n \to_P 0, \tag{4.24}$$

where c_n is a centering sequence vanishing for $\alpha \ne 1$. Since $X_k - \lambda_k Y_k$ are independent by (4.19) and a_n is regularly varying with exponent $1/\alpha$, (4.24) and Corollary 4 imply the statement of Theorem 5 if we show that for some $p > 0$,

$$E \left| \frac{1}{a_n} \left(\sum_{k=1}^{n} (X_k - \lambda_k Y_k \right) - c_n \right|^p = O(1). \tag{4.25}$$

By Theorem (6.1) of [1] we have

$$E \left| \frac{1}{a_n} \left(\sum_{k=1}^{n} X_k - b_n \right) \right|^p = O(1) \quad \text{for any } 0 < p < \alpha,$$

where b_n is the centering constant in (1.8). On the other hand, using the canonical form of the characteristic functions of stable distributions (see, e.g., [12], pages 540–543) it follows that $a_n^{-1}(\sum_{k=1}^{n} \lambda_k Y_k - \hat{b}_n) =_{\mathscr{D}} Y_1$ and thus

$$E \left| \frac{1}{a_n} \left(\sum_{k=1}^{n} \lambda_k Y_k - \hat{b}_n \right) \right|^p = O(1) \quad \text{for any } 0 < p < \alpha$$

for some other sequence $\hat{b}_n$. (In the case $\alpha \ne 1$, $b_n = \hat{b}_n = 0$.) The last two relations show that c_n in (4.24) must satisfy $c_n = (b_n - \hat{b}_n)/a_n + O(1)$, but then (4.25) is also valid.

To prove Lemma 4 we first observe that (4.4) implies

$$G^{-1}(x) \sim \left(p\frac{2-\alpha}{\alpha}\right)^{1/\alpha}(1-x)^{-1/\alpha} \quad \text{as } x \uparrow 1$$

and

$$G^{-1}(x) \sim -\left(q\frac{2-\alpha}{\alpha}\right)^{1/\alpha} x^{-1/\alpha} \quad \text{as } x \downarrow 0,$$

whence using (4.3) we get

$$G^{-1}(F(x)) \sim \frac{x}{L(|x|)^{1/\alpha}} \quad \text{as } |x| \to \infty. \tag{4.26}$$

We also note that if $L_1 > 0$ is slowly varying at infinity and $L_1(x) = 1$ in some interval $0 \le x < x_0$, then for any $\gamma > 0$ we have

$$\int_{-x}^{x} |t|^\gamma L_1(|t|)\, dF(t) \le CL_1(x) \int_{-x}^{x} |t|^\gamma\, dF(t) \tag{4.27}$$

for some constant $C > 0$. The proof of this relation is similar to those in Lemma 3.

Now using (4.1), (4.19), (4.26), (4.27), $\lambda_k = L(a_k)^{1/\alpha}$ and the inequality $(x+y)^2 \le 2(x^2+y^2)$ we get that the square integral of the kth term of the sum in (4.21) equals

$$\begin{aligned} &\int_{-\varepsilon_k a_k}^{\varepsilon_k a_k} \left(x - \left(\frac{L(a_k)}{L(|x|)}\right)^{1/\alpha} x(1+o(1))\right)^2 dF(x) \\ &\ll \int_{-\varepsilon_k a_k}^{\varepsilon_k a_k} x^2\, dF(x) + L(a_k)^{2/\alpha} \int_{-\varepsilon_k a_k}^{\varepsilon_k a_k} x^2 L(|x|)^{-2/\alpha}\, dF(x) \qquad (4.28) \\ &\ll \int_{-\varepsilon_k a_k}^{\varepsilon_k a_k} x^2\, dF(x) \sim (\varepsilon_k a_k)^{2-\alpha} L(\varepsilon_k a_k), \end{aligned}$$

where $\ll$ means the same as the big O notation. [Note that since the validity of (4.3) depends only on the values of $L(x)$ for large x, we can assume that $L(x) = 1$ for $0 < x \le 1$.] Now $a_k^{2-\alpha} L(a_k)$ is regularly varying with exponent $(2-\alpha)/\alpha$ and thus Lemma 3, $L(\varepsilon_k a_k) \sim L(a_k)$ [cf. (4.20)] and (4.5) imply

$$\sum_{k=1}^{n} (\varepsilon_k a_k)^{2-\alpha} L(\varepsilon_k a_k) = o\left(\sum_{k=1}^{n} a_k^{2-\alpha} L(a_k)\right) = o\left(n a_n^{2-\alpha} L(a_n)\right) = o\left(a_n^2\right).$$

Thus the variance of the normalized sum in (4.21) tends to 0 and consequently (4.21) holds with d_n denoting the expectation of the normed sum.

To prove (4.22) let us write the sum as $S_1 + S_2$ where

$$S_1 = \sum_{k=1}^{n} (X_k - \lambda_k Y_k) I(|X_k| \ge \varepsilon_k^{-1} a_k) I(|X_k| \ge a_n),$$

$$S_2 = \sum_{k=1}^{n} (X_k - \lambda_k Y_k) I(|X_k| \ge \varepsilon_k^{-1} a_k) I(|X_k| < a_n).$$

By (4.3) and (4.5) we have

$$\begin{aligned} P(|X_k| \ge \varepsilon_k^{-1} a_k) &\ll (\varepsilon_k^{-1} a_k)^{-\alpha} L(\varepsilon_k^{-1} a_k) \\ &\sim (\varepsilon_k^{-1} a_k)^{-\alpha} L(a_k) \sim \varepsilon_k^{\alpha}/k \end{aligned}$$

and thus for any $\varepsilon > 0$ we have

$$\begin{aligned} P(S_1 \ne 0) &\le \sum_{k=1}^{n} P\{|X_k| \ge \varepsilon_k^{-1} a_k \vee a_n\} \\ &\ll \sum_{k \le \varepsilon n} P\{|X_k| \ge a_n\} + \sum_{\varepsilon n < k \le n} \varepsilon_k^{\alpha}/k \\ &\ll \varepsilon n a_n^{-\alpha} L(a_n) + (\varepsilon n)^{-1} n \sup_{\varepsilon n < k \le n} \varepsilon_k^{\alpha} \\ &\ll \varepsilon + o(1) \ll \varepsilon \end{aligned}$$

for $n \ge n_0(\varepsilon)$. Here (and in the sequel) the constants implied by $\ll$ do not depend on ε. In the sum S_2 all terms with $k > \varepsilon n$ vanish [provided $n \ge n_0(\varepsilon)$], since the regular variation of a_k with exponent $1/\alpha$ and $\varepsilon_k \to 0$ implies that for such terms we have $\varepsilon_k^{-1} a_k > a_n$. On the other hand, for $k \le \varepsilon n$ the absolute value of the kth term of the sum is less than or equal to $|X_k - \lambda_k Y_k| I(|X_k| < a_n)$, whose square integral can be estimated as in (4.28) except that the integral should be taken on the interval $(-a_n, a_n)$. Thus using (4.1), (4.5), (4.27) and Lemma 3 we get

$$\begin{aligned} \operatorname{Var} S_2 &\le \sum_{k \le \varepsilon n} \left(a_n^{2-\alpha} L(a_n) + \frac{L(a_k)^{2/\alpha}}{L(a_n)^{2/\alpha}} a_n^{2-\alpha} L(a_n) \right) \\ &\ll \varepsilon n a_n^{2-\alpha} L(a_n) + L(a_n)^{-2/\alpha} a_n^{2-\alpha} L(a_n) \varepsilon n L(a_{[\varepsilon n]})^{2/\alpha} \\ &\ll \varepsilon n a_n^{2-\alpha} L(a_n) \ll \varepsilon a_n^2 \end{aligned}$$

for $n \ge n_0(\varepsilon)$, completing the proof of (4.22) with $e_n = a_n^{-1} E(S_2)$.

Finally, to prove (4.23) let us note that by (4.19) the sum in (4.23) can be written as $\sum_{k=1}^{n} X_k \psi_k(X_k)$ where the function $\psi_k(x)$ is defined by

$$x \psi_k(x) = x - \lambda_k G^{-1}(F(x)) \quad \text{for } \varepsilon_k a_k < |x| < \varepsilon_k^{-1} a_k$$

and by $\psi_k(x) = 0$ otherwise. By (4.26) we have

$$\psi_k(x) = 1 - \left(\frac{L(a_k)}{L(|x|)}\right)^{1/\alpha}(1 + o(1))$$

uniformly for $\varepsilon_k a_k < |x| < \varepsilon_k^{-1} a_k$ which implies, in view of (4.20), that $\sup_{\varepsilon_k a_k < |x| < \varepsilon_k^{-1} a_k} |\psi_k(x)| \to 0$ and consequently $\sup_{x \in R} |\psi_k(x)| \to 0$. Choose a sequence $\delta_k \to 0$ such that $|\psi_k(x)| \le \delta_k$; clearly we can assume that $\delta_k \ge \varepsilon_k$ and that δ_k is slowly varying. Write the sum $\sum_{k=1}^n X_k \psi_k(X_k)$ as $T_1 + T_2$ where

$$T_1 = \sum_{k=1}^{n} X_k \psi_k(X_k) I(|X_k| \ge \delta_n^{-1} a_n),$$

$$T_2 = \sum_{k=1}^{n} X_k \psi_k(X_k) I(|X_k| < \delta_n^{-1} a_n).$$

By $\delta_n \ge \varepsilon_n$ and (4.20) we have $L(\delta_n^{-1} a_n) \sim L(a_n)$ and thus using Lemma 3, (4.1), (4.3) and (4.5) we get

$$\begin{aligned} P(T_1 \neq 0) &\le \sum_{k=1}^{n} P(|X_k| \ge \delta_n^{-1} a_n) \ll \sum_{k=1}^{n} (\delta_n^{-1} a_n)^{-\alpha} L(\delta_n^{-1} a_n) \\ &\sim n \delta_n^{\alpha} a_n^{-\alpha} L(a_n) = o(1) \end{aligned}$$

and

$$\begin{aligned} \operatorname{Var} T_2 &\le \sum_{k=1}^{n} \int_{|x| \le \delta_n^{-1} a_n} x^2 \psi_k^2(x)\, dF(x) \\ &\ll \sum_{k=1}^{n} \delta_k^2 (\delta_n^{-1} a_n)^{2-\alpha} L(\delta_n^{-1} a_n) \sim (\delta_n^{-1} a_n)^{2-\alpha} L(a_n) \sum_{k=1}^{n} \delta_k^2 \\ &\sim \delta_n^{\alpha-2} a_n^{2-\alpha} L(a_n) n \delta_n^2 = o(a_n^2), \end{aligned}$$

completing the proof of (4.23) with $f_n = a_n^{-1} E(T_2)$.

Finally we prove that for $\alpha \neq 1$ we have $d_n + e_n + f_n \to 0$ in Lemma 4. In view of the above arguments, this will follow if we show that for $\alpha \neq 1$ we have

$$\begin{aligned} &\sum_{k=1}^{n} E((X_k - \lambda_k Y_k) I(|X_k| \le \varepsilon_k a_k)) \\ &\quad + \sum_{k=1}^{n} E(X_k \psi_k(X_k) I(|X_k| < \delta_n^{-1} a_n)) = o(a_n) \end{aligned} \tag{4.29}$$

and

$$\sum_{k=1}^{n} E((X_k - \lambda_k Y_k) I(\varepsilon_k^{-1} a_k \le |X_k| < a_n)) = o(a_n). \tag{4.30}$$

Assume first $\alpha < 1$. Similarly as in (4.28) it follows that the absolute value of the k-th term of the first sum in (4.29) is bounded by

$$\int_{|x|\le\varepsilon_k a_k}\left(|x| + \frac{L(a_k)^{1/\alpha}}{L(|x|)^{1/\alpha}}|x|(1+o(1))\right) dF(x)$$

$$\ll \int_{|x|\le\varepsilon_k a_k}|x|\,dF(x) \le \int_0^{\varepsilon_k a_k}U(x)\,dx + \varepsilon_k a_k U(\varepsilon_k a_k) \tag{4.31}$$

$$\ll 2 + \int_1^{\varepsilon_k a_k}x^{-\alpha}L(x)\,dx + L(\varepsilon_k a_k)(\varepsilon_k a_k)^{1-\alpha}$$

$$\ll L(\varepsilon_k a_k)(\varepsilon_k a_k)^{1-\alpha} \ll L(a_k)(\varepsilon_k a_k)^{1-\alpha},$$

where $U(x) = 1 - F(x) + F(-x)$. Since $L(a_k)(\varepsilon_k a_k)^{1-\alpha}$ is regularly varying with exponent $(1-\alpha)/\alpha$, using Lemma 3 it follows that the first sum in (4.29) is

$$\ll \sum_{k=1}^{n} L(a_k)(\varepsilon_k a_k)^{1-\alpha} \ll nL(a_n)(\varepsilon_n a_n)^{1-\alpha} = o(a_n).$$

The second sum in (4.29) and the sum in (4.30) can be estimated similarly, using $|\psi_k| \le \delta_k$ in (4.29) and estimating

$$\sum_{k\le\varepsilon n} E(|X_k - \lambda_k Y_k| I(|X_k| < a_n))$$

in (4.30). [Note that for $n \ge n_0(\varepsilon)$ all terms of (4.30) with $k > \varepsilon n$ vanish.]

In the case $\alpha > 1$, F is assumed to be centered to 0 expectation; moreover, we have $b_n = 0$ in (1.8) whence it follows, using Theorem (6.1) of [1], that $EY_1 = 0$. By the definition of ψ_k the left-hand side of (4.29) equals

$$\sum_{k=1}^{n} E\big((X_k - \lambda_k Y_k) I(|X_k| < \varepsilon_k^{-1}a_k \wedge \delta_n^{-1}a_n\big).$$

Since $E(X_k - \lambda_k Y_k) = 0$, the absolute value of the last sum equals

$$\left|\sum_{k=1}^{n} E((X_k - \lambda_k Y_k) I(|X_k| \ge \varepsilon_k^{-1}a_k \wedge \delta_n^{-1}a_n)\right|$$

$$\le \sum_{k=1}^{n} E(|X_k - \lambda_k Y_k| I(|X_k| \ge \varepsilon_k^{-1}a_k))$$

$$+ \sum_{k=1}^{n} E(|X_k - \lambda_k Y_k| I(|X_k| \ge \delta_n^{-1}a_n)),$$

which can be estimated exactly as in (4.31), replacing the interval of integration by $|x| \ge \varepsilon_k^{-1}a_k$, respectively, $|x| \ge \delta_n^{-1}a_n$ and using the "outside" version of (4.27). The estimation of the sum in (4.30) is the same, noting again that for $n \ge n_0(\varepsilon)$ all terms with $k > \varepsilon n$ vanish and bounding the kth term by $E(|X_k - \lambda_k Y_k| I(|X_k| \ge \varepsilon_k^{-1}a_k))$. This completes the proofs of Lemma 4 and Theorem 5. □

EXAMPLE 4. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with distribution function F satisfying

$$1 - F(x) = \frac{1}{2x} + \frac{1}{x \log x}, \qquad F(-x) = \frac{1}{2x}$$

for $x \geq x_0$. Then (4.3), (4.5) hold with $\alpha = 1$, $p = q = 1/2$, $L(x) = 1$, that is, (1.8) is valid with $a_n = n$ and some b_n. Actually, $\int_{-x}^{x} t\, dF(t) \sim \log\log x$, $x \to \infty$, and thus the classical theory gives (see, e.g., [12], page 305) that $b_n = n^2 a_n^{-1} \int_{-a_n}^{a_n} t\, dF(t) \sim n \log\log n$. The limit distribution G satisfies (4.4), that is, it is symmetric Cauchy. Let $Y_1, Y_2, \dots$ be any i.i.d. sequence of r.v.'s with distribution G defined on the same space as the X_n's. Given $\varepsilon > 0$ choose $c > 0$ so large that $G(c) - G(-c) \geq 1 - \varepsilon$. Corollary 1 implies that with probability 1 we have

$$\frac{b_n}{n} - c < \frac{1}{n} \sum_{i=1}^{n} X_i < \frac{b_n}{n} + c$$

on a set of n's with log density $\geq 1 - \varepsilon$. Since $n^{-1} \sum_{i=1}^{n} Y_i =_{\mathscr{D}} G$ for all n, another application of Corollary 1 yields that with probability 1,

$$-c < \frac{1}{n} \sum_{i=1}^{n} Y_i < c$$

on a set of n's with log density $\geq 1 - \varepsilon$. From these relations it follows that for any joint construction of $\{X_n\}$ and $\{Y_n\}$ on the same space the inequality

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i) > \frac{1}{2} \log\log n$$

holds on a set of n's with log density 1.

REMARK. The proof of Theorem 5 depended on the weak approximation result (4.24) where $\lambda_k = L(a_k)^{1/\alpha}$ and c_n is a centering sequence vanishing for $\alpha \neq 1$. In fact, using the equivalence of the log versions of the weak and strong law of large numbers and the fortunate fact that by the construction (4.19) the r.v.'s $X_k - \lambda_k Y_k$ are independent, (4.24) directly implies the a.s. conclusion of Theorem 5. It should be noted, however, that even (4.24) is new; it is the weighted version of a result of Simons and Stout (Theorem 3 of [25]) which is not applicable for our purposes. In the case of Theorem 4 the situation is different: in this case the corresponding weak approximation result

$$\sum_{i=1}^{n} X_i - W(a_n^2) = o(a_n) \quad \text{in probability} \tag{4.32}$$

is known (see the last remark on page 7 in Mijnheer [18]) but the a.s. result (4.7) is not a consequence of (4.32) since the process on the left-hand side does not have independent increments. Still, as our proof shows, (4.7) is valid and with the same remainder term as (4.32); the key to the proof is again Theorem 3 used several times in the Skorohod embedding argument.

Acknowledgments. We are indebted to Professor T. F. Móri for fruitful conversations leading to a more general formulation of Theorems 1 and 2; he also constructed Examples 1 and 2 in Section 2. We are also grateful to the referee for many valuable remarks and suggestions leading to the improvement of our results.

REFERENCES

[1] DEACOSTA, A. and GINÉ, E. (1979). Convergence of moments and related functionals in the general central limit theorem in Banach spaces. *Z. Wahrsch. Verw. Gebiete* **48** 213–231.
[2] BERKES, I., DEHLING, H. and MÓRI, T. F. (1991). Counterexamples related to the a.s. central limit theorem. *Studia Sci. Math. Hung.* **26** 153–164.
[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
[4] BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, MA.
[5] BROSAMLER, G. (1988). An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.* **104** 561–574.
[6] CHUNG, K. L. and ERDŐS, P. (1951). Probability limit theorems assuming only the first moment *I. Mem. Amer. Math. Soc.* **6**.
[7] CSÖRGŐ, M. and HORVÁTH, L. (1992). Invariance principles for logarithmic averages. *Math. Proc. Cambridge Philos. Soc.* **112** 195–205.
[8] DABROWSKI, A., DEHLING, H. and PHILIPP, W. (1984). An almost sure invariance principle for triangular arrays of Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **65** 483–491.
[9] DUDLEY, R. M. (1976). *Probabilities and Metrics. Lecture Notes Series* **45**. Aarhus Univ.
[10] ERDŐS, P. and HUNT, G. A. (1953). Changes of signs of sums of random variables. *Pacific J. Math.* **3** 673–687.
[11] ESSEEN, C. G. (1968). On the concentration function of a sum of independent random variables. *Z. Wahrsch. Verw. Gebiete* **9** 290–308.
[12] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.
[13] FISHER, A. (1989). A pathwise central limit theorem for random walks. Preprint.
[14] FISHER, E. (1984). An almost sure invariance principle for random variables in the domain of attraction of a stable law. *Z. Wahrsch. Verw. Gebiete* **67** 461–471.
[15] LACEY, M. T. and PHILIPP, W. (1990). A note on the almost everywhere central limit theorem. *Statist. Probab. Lett.* **9** 201–205.
[16] LOÈVE, M. (1955). *Probability Theory.* Van Nostrand, New York.
[17] MAJOR, P. (1976). Approximation of partial sums of i.i.d. r.v.'s when the summands have only two moments. *Z. Wahrsch. Verw. Gebiete* **35** 221–229.
[18] MIJNHEER, J. (1980). A strong approximation of partial sums of i.i.d. random variables with infinite variance. *Z. Wahrsch. Verw. Gebiete* **52** 1–7.
[19] MIJNHEER, J. (1983). Strong approximations of partial sums of i.i.d. random variables in the domain of attraction of a symmetric stable distribution. In *Trans. Ninth Prague Conference on Information Theory, Statist. Decision Functions and Random Processes.* Academia, Prague.
[20] PELIGRAD, M. and RÉVÉSZ, P. (1991). On the almost sure central limit theorem. In *Almost Everywhere Convergence II* 209–225. Academic, New York.
[21] PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.
[22] PROHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** 157–214.
[23] SCHATTE, P. (1988). On strong versions of the central limit theorem. *Math. Nachr.* **137** 249–256.
[24] SENETA, E. (1976). *Regularly Varying Functions. Lecture Notes in Math.* **508**. Springer, Berlin.

[25] SIMONS, G. and STOUT, W. F. (1978). A weak invariance principle with applications to domains of attraction. *Ann. Probab.* **6** 294–315.

[26] STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

[27] WEIGL, A. (1989). Zwei Sätze über die Belegungszeit beim Random Walk. Diplomarbeit. Technical Univ., Vienna.

MATHEMATICAL INSTITUTE OF THE
HUNGARIAN ACADEMY OF SCIENCES
REÁLTANODA U. 13-15
H-1053 BUDAPEST
HUNGARY

DEPARTMENT OF MATHEMATICS
GRONINGEN UNIVERSITY
P.O. BOX 800
NL-9700 AV GRONINGEN
THE NETHERLANDS