A universal result in almost sure central limit theory

István Berkes*, Endre Csáki

A. Rényi Institute of Mathematics, Hungarian Academy of Sciences
P.O.B. 127, H–1364 Budapest, Hungary

Abstract

The discovery of the almost sure central limit theorem (Brosamler, 1988; Schatte, 1988) revealed a new phenomenon in classical central limit theory and has led to an extensive literature in the past decade. In particular, a.s. central limit theorems and various related 'logarithmic' limit theorems have been obtained for several classes of independent and dependent random variables. In this paper we extend this theory and show that not only the central limit theorem, but *every* weak limit theorem for independent random variables, subject to minor technical conditions, has an analogous almost sure version. For many classical limit theorems this involves logarithmic averaging, as in the case of the CLT, but we need radically different averaging processes for 'more sensitive' limit theorems. Several examples of such a.s. limit theorems are discussed.

MSC: Primary 60F15; Secondary 60F05.

Keywords: Almost sure central limit theorem; logarithmic averages; summation methods.

* Corresponding author.

E-mail addresses: berkes@renyi.hu (I. Berkes), csaki@renyi.hu (E. Csáki).

1. Introduction

Starting with Brosamler (1988) and Schatte (1988), in the past decade several authors investigated the almost sure central limit theorem and related 'logarithmic' limit theorems for partial sums of independent random variables. The simplest form of the a.s. central limit theorem (Brosamler, 1988; Schatte, 1988; Lacey and Philipp, 1990; Fisher, 1989) states that if $X_1, X_2, \dots$ are i.i.d. random variables with mean 0, variance 1 and partial sums S_n then

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{S_k}{\sqrt{k}} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x \tag{1.1}$$

where I denotes indicator function. Relation (1.1) is a weighted strong law for the events $A_k = \{S_k/\sqrt{k} < x\}$; note that the ordinary strong law

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k\le N} I\left\{\frac{S_k}{\sqrt{k}} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x$$

fails even for $x = 0$ by the arc sine law. This means that the relative frequencies of the events $A_k = \{S_k/\sqrt{k} < x\}$ fluctuate without a limit, but an a.s. limit exists if we replace the counting measure with the logarithmic measure $\mu(A) = \sum_{k\in A} 1/k$, $A \subset N$. This remarkable property of the logarithmic measure has been studied intensively in recent years and many extensions and variants of (1.1) have been obtained. Several papers investigated the general ASCLT

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{S_k - a_k}{b_k} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x \tag{1.2}$$

for independent r.v.'s; see e.g. Atlagh (1993), Atlagh and Weber (1992, 1996), Berkes (1995), Berkes and Dehling (1993, 1994), Ibragimov (1996), Ibragimov and Lifshits (1998, 1999), Lifshits (2000a, 2000b), Móri (1993), Peligrad and Révész (1991), Rodzik and Rychlik (1994, 1996) and the references in the survey paper Berkes (1998). As it turned out, under mild moment conditions relation (1.2) follows from the the ordinary (weak) CLT

$$(S_n - a_n)/b_n \xrightarrow{\mathcal{D}} N(0,1) \tag{1.3}$$

but in general the validity of (1.2) is a delicate question of a totally different character as (1.3). (See e.g. the counterexample in Lifshits, 2000a.) Starting with Weigl (1989) and Csörgő and Horváth (1992), several papers dealt with the fine asymptotic properties of the sum

$$\sum_{k\le N} \frac{1}{k}\left(I\left\{\frac{S_k-a_k}{b_k}<x\right\}-\Phi(x)\right).$$

See Section 4 of Berkes (1998) for detailed references. For further results in the field (higher dimensions, local theorems, large deviations, weakly dependent sequences, etc.) we also refer to Berkes (1998).

All the above results are related to the central limit theorem, but there also exist a few results of the type (1.2) in connection with other classical weak limit theorems. Marcus and Rosen (1995), Csáki and Földes (1995) and Horváth and Khoshnevisan (1995) obtained analogues of (1.2) for local times and Fahrner and Stadtmüller (1998) and Cheng et al. (1998) proved a similar result for extreme order statistics. In fact, they showed that if $X_1, X_2, \dots$ are i.i.d.r.v.'s with $M_n = \max_{k\le n} X_k$ and for some numerical sequences (a_k), (b_k) we have

$$(M_k - a_k)/b_k \xrightarrow{\mathcal{D}} G$$

for a nondegenerate distribution G, then

$$\lim_{N\to\infty} \frac{1}{\log N}\sum_{k\le N}\frac{1}{k} I\left\{\frac{M_k-a_k}{b_k}<x\right\} = G(x) \qquad \text{a.s. for any } x. \tag{1.4}$$

Relation (1.4) is the exact analogue of the a.s. central limit theorem (1.2) for extremal statistics and raises the question which other weak limit theorems for independent r.v.'s have an almost sure 'logarithmic' version. The purpose of this paper is to prove the surprising fact that not only the CLT and limit theorems for local times and extrema, but *every* weak limit theorem for independent r.v.'s, subject to minor technical conditions, has an almost sure version. To formulate our result, we need a few preparatory remarks about the structure of weak limit theorems.

The generic form of a weak limit theorem for a sequence (X_n) of r.v.'s is

$$f_k(X_1, X_2, \dots) \xrightarrow{\mathcal{D}} G \tag{1.5}$$

where $f_k : R^\infty \to R$ are measurable functions and G is a distribution function. A few examples corresponding to well known limit theorems are:

$$(a)\qquad f_k(x_1, x_2, \dots) = (x_1 + \cdots + x_k)/\sqrt{k} \qquad \text{(CLT)}$$

$$(b)\qquad f_k(x_1, x_2, \dots) = a_k(\max_{i\le k} x_i - b_k) \qquad \text{(Limit theorems for extrema)}$$

$$(c)\qquad f_k(x_1, x_2, \dots) = \sqrt{k}\sup_t \left| k^{-1} \sum_{i\le k} I(x_i \le t) - F(t) \right| \qquad \text{(empirical d.f.'s)}$$

$$(d)\qquad f_k(x_1, x_2, \dots) = a_k \left(\sum_{i\le k} I\{x_1 + \cdots + x_i = 0\} - b_k \right) \qquad \text{(Local times)}$$

$$(e)\qquad f_k(x_1, x_2, \dots) = a_k \left(\sum_{1\le i_1 < \cdots i_m \le k} h(x_{i_1}, \dots, x_{i_m}) - b_k \right) \qquad \text{(U-statistics).}$$

In most cases of interest, the functions f_k depend only on finitely many of the x_i's and in this paper we will consider only limit theorems of this kind. In this case, (1.5) reduces to the form

$$f_k(X_1, X_2, \dots, X_{n_k}) \xrightarrow{\mathcal{D}} G$$

for some sequence (n_k) of positive integers. For notational simplicity, we first formulate our results for the case $n_k = k$ (satisfied in all the above examples); we shall return to the general case in the next section.

The functions f_k in a limit theorem

$$f_k(X_1, X_2, \dots, X_k) \xrightarrow{\mathcal{D}} G \tag{1.6}$$

must be measurable, but not all measurable f_k correspond to interesting limit theorems. For example, if $f_k(x_1, x_2, \dots) = x_1 + x_2$ for all $k \ge 1$ and G denotes the convolution of the distributions of X_1 and X_2, then (1.6) expresses a true statement, but it cannot be called a "limit theorem" because it involves only the variables X_1 and X_2. A true limit theorem must involve infinitely many of the variables X_i, and it must have the property that for each $k \ge 1$, the influence of the initial variables $X_1, X_2, \dots, X_k$ on $f_l(X_1, \dots, X_l)$ becomes eventually negligible as $l \to \infty$, so that the validity of (1.6) is not influenced by changing finitely many X_i's. We will formalize this condition by assuming that for each $l > k \ge 1$ there exists a measurable function $f_{k,l} : R^{l-k} \to R$ such that

$$E\left(|f_l(X_1, \dots, X_l) - f_{k,l}(X_{k+1}, \dots, X_l)| \wedge 1\right) \le c_k/c_l \tag{1.7}$$

for some nondecreasing sequence $c_k \to \infty$. For $c_k = k^\alpha$, $\alpha > 0$ (1.7) means that $f_l(X_1, \dots, X_l)$ is close in probability to $f_{k,l}(X_{k+1}, \dots, X_l)$ if k/l is small and thus f_l changes little in probability if we change its first εl variables, ε small. It is easy to verify (see Section 5) that all the limit theorems listed above satisfy this condition. In fact, as experience shows, most "usual" weak limit theorems have this proportionality property. For $c_k = (\log k)^\alpha$, $\alpha > 0$, the right hand side of (1.7) will be small if $\log k / \log l \le \varepsilon$, i.e. $k \le l^\varepsilon$ with ε small. That is, in this case (1.7) means that $f_l(x_1, \dots, x_l)$ depends negligibly on its first l^ε variables. Note that this segment is shorter than the segment of length εl obtained in the case $c_k = k^\alpha$: the dependence of f_l on its initial variables became more sensitive. For even more slowly increasing (c_k) the segment of the initial, "irrelevant" variables of f_l gets even shorter, indicating a further increase of sensitivity of f_l.

A simple example for a limit theorem

$$f_k(X_1, \dots, X_k) \xrightarrow{\mathcal{D}} G$$

where the dependence of f_k on its first εk variables is not negligible (i.e. (1.7) does not hold with $c_k = k^\alpha$) is the Darling–Erdős (1956) limit theorem for the maximum of normed partial sums of independent random variables. This theorem states that if (X_n) is a sequence of independent random variables with mean 0, variance 1 and uniformly bounded third absolute moments, then letting $S_n = X_1 + \cdots + X_n$ we have

$$a_n \left(\max_{i \le n} \frac{S_i}{\sqrt{i}} - b_n \right) \xrightarrow{\mathcal{D}} G$$

for suitable (a_n) and (b_n), where

$$G(x) = \exp(-e^{-x}).$$

Here $\max_{i \le n} \left(S_i / \sqrt{i} \right)$ behaves like the maximum of the Ornstein-Uhlenbeck process U in $[0, \log n]$ and thus the probability that the maximum of $S_i/\sqrt{i}$, $1 \le i \le n$ is attained for some $i \le \sqrt{n}$ is approximately 1/2. Hence changing the first $\sqrt{n}$ of the variables $X_1, \dots, X_n$ can change the value of $\max_{i \le n} \left(S_i / \sqrt{i} \right)$ radically on a set of probability near 1/2 and thus (1.7) fails with $c_k = k^\alpha$.

We can now formulate our first general result providing the a.s. version of the weak limit theorems.

Theorem 1. *Let $X_1, X_2, \dots$ be independent random variables satisfying the weak limit theorem*

$$f_k(X_1, X_2, \dots, X_k) \xrightarrow{\mathcal{D}} G \tag{1.8}$$

where $f_k : R^k \to R$ $(k = 1, 2 \dots)$ are measurable functions and G is a distribution function. Assume that for each $1 \le k < l$ there exists a measurable function $f_{k,l} : R^{l-k} \to R$ such that

$$E\left(|f_l(X_1, \dots, X_l) - f_{k,l}(X_{k+1}, \dots, X_l)| \wedge 1\right) \le A(c_k/c_l) \tag{1.9}$$

with a constant $A > 0$ and a positive, nondecreasing sequence (c_n) satisfying $c_n \to \infty$, $c_{n+1}/c_n = O(1)$. Put

$$d_k = \log(c_{k+1}/c_k), \quad D_n = \sum_{k \le n} d_k. \tag{1.10}$$

Then we have

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \le N} d_k I\left\{f_k(X_1, \dots, X_k) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G \tag{1.11}$$

where C_G denotes the set of continuity points of G. The result remains valid if we replace the weight sequence (d_k) by any (d_k^*) such that $0 \le d_k^* \le d_k$, $\sum d_k^* = \infty$.

Theorem 1 shows that if a weak limit theorem (1.8) satisfies condition (1.9) then it has an a.s. weighted version with weights depending on the sequence c_k in (1.9). In the case $c_k = k^\alpha$ (1.10) gives

$$d_k \sim \text{const} \cdot 1/k$$

and thus (1.11) reduces to

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{f_k(X_1, \dots, X_k) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G. \tag{1.12}$$

Thus in this case the a.s. limit theorem corresponding to our weak limit theorem will involve logarithmic averages. This covers a very large class of limit theorems; several examples will be given in Section 5. Among others, we shall prove there analogues of the pointwise CLT for partial sums and partial maxima, extremal order statistics, empirical distribution functions, U–statistics, local times, return times, etc. If $c_k = (\log k)^\alpha$, then (1.10) gives

$$d_k \sim \text{const} \cdot \frac{1}{k \log k}$$

and (1.11) reduces to

$$\lim_{N\to\infty} \frac{1}{\log\log N} \sum_{k\le N} \frac{1}{k\log k} I\{f_k(X_1,\dots,X_k)<x\} = G(x) \qquad \text{a.s. for any } x\in C_G.$$

Such loglog averaging will appear, e.g., in the a.s. version of the Darling-Erdős limit theorem, see Section 5. Clearly, the slower the sequence (c_n) tends to infinity, the more "sensitive" limit theorems are permitted by (1.9), and the more the weights d_k differ from $1/k$.

Any sequence $\mathbf{D}=(d_1,d_2,\dots)$ of positive numbers with $\sum d_n=\infty$ defines a linear summation method (Riesz summation of order 1) as follows. Given a real sequence (x_n), put

$$\sigma_n^{(\mathbf{D})} = D_n^{-1}\sum_{k\le n} d_k x_k \qquad \text{where} \qquad D_n=\sum_{k\le n} d_k.$$

We say that (x_n) is $\mathbf{D}$-summable if $\sigma_n^{(\mathbf{D})}$ has a finite limit. By a classical theorem of Hardy (see e.g. Chandrasekharan and Minakshisundaram, 1952, p. 35; see also pp. 37-38 for a more general version due to Hirst), if two sequences $\mathbf{D}=(d_n)$ and $\mathbf{D}^*=(d_n^*)$ with partial sums D_n and D_n^* satisfy $D_n^*=O(D_n)$ then, under mild regularity conditions, the summation procedure defined by $\mathbf{D}^*$ is stronger (i.e. more effective) than the procedure defined by $\mathbf{D}$ in the sense that if a sequence (x_n) is $\mathbf{D}$-summable then it is also $\mathbf{D}^*$-summable and to the same limit. Moreover, if $D_n^\alpha \le D_n^* \le D_n^\beta$ for some $0<\alpha<\beta$ and sufficiently large n, then by a theorem of Zygmund (see also Chandrasekharan and Minakshisundaram, 1952, p. 35) the summation procedures defined by $\mathbf{D}$ and $\mathbf{D}^*$ are equivalent, i.e., $\sigma_n^{(\mathbf{D})}$ converges for some (x_n) iff $\sigma_n^{(\mathbf{D}^*)}$ does. Finally, if $D_n^*=O(D_n^\gamma)$ for all $\gamma>0$ then the summation method defined by $\mathbf{D}^*$ is strictly stronger than the method defined by $\mathbf{D}$. (For further results see Chandrasekharan and Minakshisundaram, 1952, Chapter 2; we refer also to Bingham and Rogers, 1991 for various connections between summation methods and probability theory.) For example, logarithmic summation defined by $d_k=1/k$ is stronger than Cesàro (or $(C,1)$) summation defined by $d_k=1$; on the other hand, all summation procedures defined by $d_k=(\log k)^\alpha/k,\ \alpha>-1$ are equivalent to logarithmic summation. Using the terminology of summation procedures, Theorem 1 means that the more "sensitive" the functional f_k in the weak limit theorem

$$f_k(X_1,\dots,X_k) \xrightarrow{\mathcal{D}} G$$

is (i.e. the slower the sequence (c_n) in (1.9) tends to infinity), the more effective summation procedure has to be used in the corresponding strong limit theorem

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{f_k(X_1,\dots,X_k) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G.$$

While the essential new information provided by Theorem 1 concerns nonlinear functionals f_k, the theorem sheds also new light on a curious phenomenon observed earlier in connection with the original a.s. central limit theorem. If X_n are independent random variables satisfying the Lindeberg condition (and thus the CLT), then the pointwise CLT can still fail (see Berkes and Dehling, 1993; Ibragimov and Lifshits, 1999), but as Atlagh (1993) showed, with properly chosen weights the pointwise CLT is always valid. More generally, Ibragimov and Lifshits (1999) showed that if (X_n) is a sequence of independent random variables satisfying (1.3) with some $b_n \uparrow \infty$ and the left hand side of (1.3) has uniformly bounded p-th moments for some $p > 0$, then setting

$$d_k = (b_k - b_{k-1})/b_k, \quad D_n = \sum_{k\le n} d_k$$

we have

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{\frac{S_k - a_k}{b_k} < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G.$$

Theorem 1 explains this phenomenon: if an independent sequence (X_n) with partial sums S_n satisfies $(S_n - a_n)/b_n \xrightarrow{\mathcal{D}} G$, then removing $X_1,\dots,X_k$ from S_n for some $k \le n$ will result of a change of $(S_n - a_n)/b_n$ in the order of magnitude b_k/b_n and thus Theorem 1 applies with $c_n = b_n$. If b_n grows like n^γ for some $\gamma > 0$, then $b_l/b_k \sim l/k$ and thus by Theorem 1 a pointwise CLT with weights $1/k$ is valid. However, if b_n grows, e.g., like $(\log n)^\gamma$, $\gamma > 0$, then we have a more sensitive functional and Theorem 1 yields nonstandard weights. As the examples in Berkes and Dehling (1993), Ibragimov and Lifshits (1999) show, the use of weights different from $1/k$ is really necessary in this case.

Theorem 1 is the simplest one of the 'universal' results proved in our paper. In Section 2 we will formulate several extensions of Theorem 1 and in Section 3 we will discuss the weight sequences occuring in a.s. versions of weak limit theorems. The proofs of our theorems will be given in Section 4 and in Section 5 we will give examples and applications of our results.

2. Further results

Theorem 1 covers a very large class of limit theorems and condition (1.9) will be verified easily in all applications considered in Section 5. It should be noted, however, that condition (1.9) can be substantially weakened; the following result yields an essentially optimal condition under which the weak limit theorem

$$f_k(X_1, \dots, X_k) \xrightarrow{\mathcal{D}} G \tag{2.1}$$

implies the a.s. result (1.11). Let $\log_+ x = \log x$ if $x \geq 1$ and 0 otherwise.

Theorem 2. *Let $X_1, X_2, \dots$ be independent random variables, $f_k : R^k \to R$ $(k = 1, 2 \dots)$ measurable functions and assume that for each $1 \leq k < l$ there exists a measurable function $f_{k,l} : R^{l-k} \to R$ such that*

$$E\left(|f_l(X_1, \dots, X_l) - f_{k,l}(X_{k+1}, \dots, X_l)| \wedge 1\right) \leq C\left(\log_+ \log_+ (c_l/c_k)\right)^{-(1+\varepsilon)} \tag{2.2}$$

for some constants $C > 0$, $\varepsilon > 0$ and a positive, nondecreasing sequence (c_n) satisfying $c_n \to \infty$, $c_{n+1}/c_n = O(1)$. Put

$$d_k = \log(c_{k+1}/c_k), \quad D_n = \sum_{k \leq n} d_k. \tag{2.3}$$

Then for any distribution function G the relations

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\leq N} d_k I\left\{f_k(X_1, \dots, X_k) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G \tag{2.4}$$

and

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\leq N} d_k P\left\{f_k(X_1, \dots, X_k) < x\right\} = G(x) \qquad \text{for any } x \in C_G \tag{2.5}$$

are equivalent. The result remains valid if we replace the weight sequence (d_k) by any (d_k^*) such that $0 \leq d_k^* \leq d_k$, $\sum d_k^* = \infty$.

Note that the $\log_+ \log_+$ in (2.2) equals 0 for $1 \leq c_l/c_k \leq e$; in this case the right hand side of (2.2) is meant to be $+\infty$ (and thus (2.2) is automatically satisfied in this case). Actually, the proof of Theorem 2 will use condition (2.2) only for $c_l/c_k \geq 3$, when $\log_+ \log_+$ can be replaced by $\log \log$.

Assumption (2.2) of Theorem 2 is weaker than condition (1.9) in Theorem 1 and is actually sharp: Theorem 2 becomes false if we assume (2.2) only for $\varepsilon = 0$. This follows from a recent counterexample of Lifshits (2000a) related to the pointwise CLT, see Section 5. Observe also that in the theorem we state not only that the weak limit theorem (2.1) implies the strong result (2.4), but that the strong result (2.4) is actually equivalent to the weighted weak result (2.5). In addition to the case of weakly converging $f_k(X_1, \dots, X_k)$, relation (2.5) covers also situations where the distribution of $f_k(X_1, \dots, X_k)$ fluctuates without a limit. In Berkes and Dehling (1994) and Berkes et al. (1991) several examples are constructed (in the case of normalized partial sums) where (2.4) holds, but (2.1) fails. A more natural example for this phenomenon is the St. Petersburg game, see Berkes et al. (1999).

Theorem 1 permits various further generalizations. In what follows, we will investigate more general limit theorems of the type

$$f_k(X_1, X_2, \dots, X_{n_k}) \xrightarrow{\mathcal{D}} G$$

with arbitrary n_k and also a.s. versions of weak limit theorems defined along subsequences. Finally, we will investigate the case when the discrete parameter sequence X_n is replaced by a continouos parameter process.

Schatte (1988) and Atlagh and Weber (1992) proved that if (X_n) is a sequence of i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$ then setting $S_k = \sum_{i \leq k} X_i$ we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k \leq N} I\left\{ \frac{S_{2^k}}{\sqrt{2^k}} < x \right\} = \Phi(x) \qquad \text{a.s. for any } x.$$

(Actually, Schatte assumed also $E|X_1|^3 < +\infty$.) In other words, if we consider $S_n/\sqrt{n}$ only along the subsequence 2^k, then the logarithmic averages in the pointwise central limit theorem can be replaced by ordinary (Cesàro) averages. The following theorem shows how to choose the weights in our general a.s. limit theorem (1.11) when we consider the functional $f_k(X_1, \dots, X_k)$ only along a given subsequence (n_k) of integers.

Theorem 3. *Let $X_1, X_2, \dots$ be independent random variables, $f_k : R^k \to R$ $(k = 1, 2 \dots)$ measurable functions and assume that for each $1 \leq k < l$ there exists a measurable function*

$f_{k,l} : R^{l-k} \to R$ such that

$$E\left(|f_l(X_1,\dots,X_l) - f_{k,l}(X_{k+1},\dots,X_l)| \wedge 1\right) \le C\left(\log_+\log_+(l/k)\right)^{-(1+\varepsilon)} \tag{2.6}$$

for some $\varepsilon > 0$. Let (n_k) be an increasing sequence of positive integers satisfying $n_{k+1}/n_k = O(1)$. Set

$$d_k = \log(n_{k+1}/n_k), \quad D_n = \sum_{k\le n} d_k. \tag{2.7}$$

Then for any distribution function G the relations

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{f_{n_k}(X_1,\dots,X_{n_k}) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G$$

and

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k P\left\{f_{n_k}(X_1,\dots,X_{n_k}) < x\right\} = G(x) \qquad \text{for any } x \in C_G$$

are equivalent. The result remains valid if we replace the weight sequence (d_k) by any (d_k^*) such that $0 \le d_k^* \le d_k$, $\sum d_k^* = \infty$.

Condition (2.6) covers 'proportional' limit theorems, i.e. limit theorems of the type (2.1) where f_k depends weakly on its first $o(k)$ variables. (As we noted in Section 1, most "usual" limit theorems satisfy this condition; see the examples in Section 5.) It is worth writing out the theorem in detail in a special case, e.g. in the case of the CLT, when (X_n) is an i.i.d. sequence with $EX_1 = 0$, $EX_1^2 = 1$ and

$$f_l(x_1,\dots,x_l) = (x_1 + \dots + x_l)/\sqrt{l}$$

(see example (a) in Section 1). In this case Theorem 3 states that if (n_k) is an increasing sequence of positive integers with $n_{k+1}/n_k = O(1)$ and d_k is defined by (2.7), then

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{\frac{S_{n_k}}{\sqrt{n_k}} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x.$$

If $n_k = [k^\alpha]$ $(\alpha \ge 1)$ then $d_k \sim \alpha/k$, i.e. the weights are still logarithmic. If $n_k = [e^{k^\alpha}]$ $(0 < \alpha \le 1)$, then $d_k \sim \alpha/k^{1-\alpha}$, which corresponds to a weaker summation procedure. The case $n_k = 2^k$ covers the results in Atlagh and Weber (1992), Schatte (1988) mentioned above. Clearly, the faster (n_k) grows, the weaker the summation procedure in Theorem 3 becomes. For example, in the case $n_k = 2^k$ one can use a weaker summation procedure (namely Cesàro summation) to get the a.s. convergence of indicators than in the case $n_k = k$, when logarithmic summation is needed. This effect is exactly the opposite as the effect of the functional f_k getting more and more sensitive: this latter leads to the need of using stronger averaging procedures. The same interpretation holds for Theorem 3 in case of general functionals f_k.

Our next theorem is a common generalization of Theorems 1–3:

Theorem 4. *Let $X_1, X_2, \dots$ be independent random variables, (n_k) an increasing sequence of positive integers, $f_k : R^{n_k} \to R$ $(k = 1, 2 \dots)$ measurable functions and assume that for each $1 \le k < l$ there exists a measurable function $f_{k,l} : R^{n_l - n_k} \to R$ such that*

$$E\left(|f_l(X_1, \dots, X_{n_l}) - f_{k,l}(X_{n_k+1}, \dots, X_{n_l})| \wedge 1\right) \le C\left(\log_+ \log_+ (c_l/c_k)\right)^{-(1+\varepsilon)} \tag{2.8}$$

for some constants $C > 0$, $\varepsilon > 0$ and a positive, nondecreasing sequence (c_n) with $c_n \to \infty$, $c_{n+1}/c_n = O(1)$. Put

$$d_k = \log(c_{k+1}/c_k), \quad D_n = \sum_{k \le n} d_k.$$

Then for any distribution function G the relations

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \le N} d_k I\left\{f_k(X_1, \dots, X_{n_k}) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G \tag{2.9}$$

and

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \le N} d_k P\left\{f_k(X_1, \dots, X_{n_k}) < x\right\} = G(x) \qquad \text{for any } x \in C_G \tag{2.10}$$

are equivalent. The result remains valid if we replace the weight sequence (d_k) by any (d_k^*) such that $0 \le d_k^* \le d_k$, $\sum d_k^* = \infty$.

In addition to the situations considered in Theorems 1, 2 and 3, Theorem 4 covers limit theorems of the type

$$f_k(X_1, \dots, X_{n_k}) \xrightarrow{\mathcal{D}} G \tag{2.11}$$

with arbitrary (n_k).

In conclusion we extend our theorems to the case when the functionals f_k depend not on an independent sequence (X_n), but on a process $\{X(t),\, t \geq 0\}$ with independent increments:

Theorem 5. *Let $\{X(t),\, t \geq 0\}$ be a process with $X(0) = 0$ and independent increments and let $\xi_1, \xi_2, \ldots$ be random variables such that ξ_k is measurable with respect to $\sigma\{X(t),\, 0 \leq t \leq k\}$. Assume that for each $1 \leq k < l$ there exists a random variable $\xi_{k,l}$ measurable with respect to $\sigma\{X(t') - X(t) : k \leq t \leq t' \leq l\}$ such that*

$$E\left(|\xi_l - \xi_{k,l}| \wedge 1\right) \leq C \left(\log_+ \log_+ (c_l/c_k)\right)^{-(1+\varepsilon)}$$

for some constants $C > 0$, $\varepsilon > 0$ and a positive, nondecreasing sequence (c_n) with $c_n \to \infty$, $c_{n+1}/c_n = O(1)$. Put

$$d_k = \log(c_{k+1}/c_k), \quad D_n = \sum_{k \leq n} d_k.$$

Then for any distribution function G the relations

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \leq N} d_k I\{\xi_k < x\} = G(x) \qquad \text{a.s. for any } x \in C_G \tag{2.12}$$

and

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \leq N} d_k P\{\xi_k < x\} = G(x) \qquad \text{for any } x \in C_G \tag{2.13}$$

are equivalent.

This theorem is the exact analogue of Theorem 2 for processes; the analogues of Theorems 1, 3 and 4 can be formulated similarly. A typical application of Theorem 5 is the Darling-Erdős limit theorem for $\sup_{1\leq t\leq n} \left(W(t)/\sqrt{t}\right)$ where W is a Wiener process (see Section 5.)

We note that the following result is also true, where the sums in (2.12) and (2.13) are replaced by integrals.

Theorem 6. *Let $\{X(t),\, t \geq 0\}$ be a process with $X(0) = 0$ and independent increments and let $\{\xi(t),\, t \geq 0\}$ be a process such that $\xi(t)$ is measurable with respect to $\sigma\{X(u),\, 0 \leq$*

$u \le t\}$. Assume that for each $1 \le s < t$ there exists a random variable $\xi(s,t)$ measurable with respect to $\sigma\{X(u') - X(u) : s \le u \le u' \le t\}$ such that

$$E\left(|\xi(t) - \xi(s,t)| \wedge 1\right) \le C\left(\log_+ \log_+ (c(t)/c(s))\right)^{-(1+\varepsilon)}$$

for some constants $C > 0$, $\varepsilon > 0$ and a positive, nondecreasing, continuous function $(c(t),\ t \ge 1)$ with $\lim_{t\to\infty} c(t) = +\infty$. Put $D(t) = \log c(t)$. Then for any distribution function G the relations

$$\lim_{T\to\infty} \frac{1}{D(T)} \int_1^T I\{\xi(t) < x\}\, dD(t) = G(x) \qquad \text{a.s. for any } x \in C_G \tag{2.14}$$

and

$$\lim_{T\to\infty} \frac{1}{D(T)} \int_1^T P\{\xi(t) < x\}\, dD(t) = G(x) \qquad \text{for any } x \in C_G \tag{2.15}$$

are equivalent.

3. Notes on weight sequences

Most pointwise central limit theorems in the literature use the weights $d_k = 1/k$ and it might appear that these weights are the 'natural' ones. In many situations, the weights $1/k$ are very convenient to work with: for example, in the case of i.i.d. normal random variables, the use of an exponential time transformation together with the ergodic theorem yields a very elegant proof of the pointwise central limit theorem. However, Theorem 1 shows that even in these cases the weight sequence $1/k$ is only one in a very large class of weight sequences that work equally well: for example, in the case of i.i.d. random variables (X_n) with mean 0 and variance 1 the relation

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{\frac{S_k}{\sqrt{k}} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x \tag{3.1}$$

holds for all weight sequences $0 \le d_k \le 1/k$, $\sum d_k = \infty$, and it is easy to see that (3.1) remains valid also for several sequences $d_k \ge 1/k$. For example, (3.1) holds with

$$d_k = (\log k)^\alpha / k, \qquad (\alpha > -1) \tag{3.2}$$

(in fact, by Zygmund's theorem, the summation method belonging to the sequence in (3.2) is equivalent to logarithmic summation), and we shall see that (3.1) holds even with

$$d_k = \exp((\log k)^\alpha)/k \tag{3.3}$$

provided $0 \le \alpha < 1/2$. On the other hand, Schatte (1988) proved that (3.1) does not hold with $d_k = 1$ (Cesàro summation). Clearly, the larger the weight sequence (d_k) is, the stronger is the result (3.1) (see our earlier remarks on summation methods) and it would be of considerable interest to determine the optimal weights. As (3.3) gives logarithmic averaging for $\alpha = 0$ and Cesàro averaging for $\alpha = 1$, we see that in the pointwise CLT (3.1) we can go at least "halfways" from logarithmic to Cesàro averaging. Whether (3.3) works also for some $1/2 \le \alpha < 1$ remains open. We note, however, that for every $1/2 \le \alpha < 1$ relation (3.1) holds at least with convergence in probability.

Similarly to the above remarks, the weights in Theorems 2-5 are far from being the only possible (or optimal) ones. For example, Zygmund's theorem shows that Theorems 2-5 remain valid with

$$d_k = \log(c_{k+1}/c_k)(\log c_k)^\alpha, \qquad (\alpha > -1)$$

and if the right side of (2.2) is sharpened to $C(c_l/c_k)^{-\gamma}$ for some $\gamma > 0$ then we will prove that one can even choose

$$d_k = \log(c_{k+1}/c_k)\exp((\log c_k)^\alpha), \qquad 0 \le \alpha < 1/2.$$

On the other hand, the conclusion of the theorem fails generally for

$$d_k = \log(c_{k+1}/c_k)c_k^\alpha, \qquad \alpha > 0.$$

Again, the optimal weight sequence remains unknown.

In conclusion we note that by the assumption $c_{k+1}/c_k = O(1)$ made in Theorems 1, 2, 4 and 5 the weight sequences (d_k) in all of our theorems are bounded. This condition can be easily removed: the proofs our theorems remain valid, with obvious changes, if c_{k+1}/c_k grows, e.g., with polynomial speed. However, it must be pointed out that the case $c_{k+1}/c_k \to \infty$ covers only relatively uninteresting situations. For example, under $c_{k+1}/c_k \to \infty$ condition (2.8) of Theorem 4 implies that

$$f_l(X_1, \ldots, X_{n_l}) - f_{l-1,l}(X_{n_{l-1}+1}, \ldots, X_{n_l}) \xrightarrow{P} 0$$

and thus setting $f_k^* = f_{k-1,k}$, the limit theorem (2.11) reduces to the relation

$$f_k^*(X_{n_{k-1}+1}, \ldots, X_{n_k}) \xrightarrow{\mathcal{D}} G$$

where for different k's the left hand side contains disjoint sets of the X_i's. From the law of the iterated logarithm it will follow easily that in this case the conclusion of the theorem holds with any positive weight sequence (d_n) satisfying

$$E_n^2 := \sum_{k \le n} d_k^2 \to \infty \tag{3.4}$$

and

$$d_n = o\left(E_n/(\log\log E_n)^{1/2}\right). \tag{3.5}$$

Moreover, this result is optimal, i.e. replacing the o in (3.5) by O the result becomes false. Condition (3.5) permits almost exponential increase of the weights d_n; for example, the conclusion of the theorem holds if $d_n = e^{n/(\log n)^\gamma}$, $\gamma > 1$, and we will see that it generally fails if $d_n = e^{n/\log n}$. That our theorems cannot hold with exponential weights (d_n) (except in trivial cases) is obvious from the fact that for exponential (d_n) the summability procedure belonging to (d_n) is equivalent to convergence (see Chandrasekharan and Minakshisundaram, 1952, p. 13), and thus (2.9) means almost sure convergence of $I\{f_k(X_1, \ldots, X_{n_k}) < x\}$ to $G(x)$ for all $x \in C_G$, which is impossible except if G is concentrated in a single point c and $f_k(X_1, \ldots, X_{n_k}) \to c$ a.s.

4. Proof of the theorems

We shall give the proof of Theorem 4; the proofs of Theorems 5 and 6 are similar. Despite the large generality of our theorems, the proofs will be rather simple; we will use essentially the same second order argument that was used in the proof of the original versions of the ASCLT (see e.g. Lacey and Philipp, 1990; Schatte, 1988). Let

$$0 \le d_k \le \log(c_{k+1}/c_k), \qquad D_n = \sum_{k \le n} d_k \to \infty. \tag{4.1}$$

By a well known principle in the theory of the pointwise central limit theorem (see e.g. Lacey and Philipp, 1990), it suffices to prove that for any bounded Lipschitz 1 function $g: R \to R$ we have

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \le N} d_k \left[g\left(f_k(X_1, \ldots, X_{n_k})\right) - Eg\left(f_k(X_1, \ldots, X_{n_k})\right)\right] = 0 \qquad \text{a.s.} \tag{4.2}$$

Put

$$\xi_k = g\left(f_k(X_1, \ldots, X_{n_k})\right) - Eg\left(f_k(X_1, \ldots, X_{n_k})\right)$$

and let $K \geq 1$ denote a constant such that

$$|g(x)| \leq K \quad \text{and} \quad |g(x) - g(y)| \leq K|x - y| \quad \text{for any } x, y \in R. \tag{4.3}$$

Then for any $1 \leq k < l$ with $c_l/c_k \geq 3$ we get, using (2.8), (4.3) and the fact that $f_k(X_1, \ldots, X_{n_k})$ and $f_{k,l}(X_{n_k+1}, \ldots, X_{n_l})$ are independent,

$$\begin{aligned}
&|E(\xi_k \xi_l)| \\
&= \left|\mathrm{Cov}\left(g\left(f_k(X_1, \ldots, X_{n_k})\right), g\left(f_l(X_1, \ldots, X_{n_l})\right)\right)\right| \\
&= \left|\mathrm{Cov}\left(g\left(f_k(X_1, \ldots, X_{n_k})\right), g\left(f_l(X_1, \ldots, X_{n_l})\right) - g\left(f_{k,l}(X_{n_k+1}, \ldots, X_{n_l})\right)\right)\right| \\
&\leq 2K \cdot E\left|g\left(f_l(X_1, \ldots, X_{n_l})\right) - g\left(f_{k,l}(X_{n_k+1}, \ldots, X_{n_l})\right)\right| \\
&\leq 2K \cdot E\left(K|f_l(X_1, \ldots, X_{n_l}) - f_{k,l}(X_{n_k+1}, \ldots, X_{n_l})| \wedge 2K\right) \\
&\leq 4CK^2 \left(\log\log(c_l/c_k)\right)^{-(1+\varepsilon)}.
\end{aligned} \tag{4.4}$$

Now

$$E\left(\sum_{k \leq N} d_k \xi_k\right)^2 \leq 2 \sum_{1 \leq k \leq l \leq N} d_k d_l |E(\xi_k \xi_l)|. \tag{4.5}$$

Let N be so large that $D_N \geq 4$. By (4.4), the contribution of those terms in the sum on the right side of (4.5) where

$$c_l/c_k \geq \exp\left(D_N^{1/2}\right)$$

is at most

$$C^*(\log D_N)^{-1-\varepsilon} \sum_{1 \leq k \leq l \leq N} d_k d_l \leq C^*(\log D_N)^{-1-\varepsilon} D_N^2.$$

On the other hand, letting $M = \sup_{n \geq 1}(c_{n+1}/c_n)$, the relation $c_l/c_k \leq \exp(D_N^{1/2})$ implies

$$\log c_{l+1} - \log c_k \leq \log M + D_N^{1/2}$$

and thus (4.1) and the trivial estimate $|E(\xi_k \xi_l)| \leq 4K^2$ show that the contribution of those terms on the right hand side of (4.5) where $c_l/c_k < \exp\left(D_N^{1/2}\right)$ is

$$\begin{aligned}
&\leq 8K^2 \sum_{k=1}^N d_k \sum_{\{l \geq k:\, c_l \leq c_k \exp(D_N^{1/2})\}} d_l \leq \mathrm{const} \cdot \sum_{k=1}^N d_k \sum_{\{l \geq k:\, c_l \leq c_k \exp(D_N^{1/2})\}} (\log c_{l+1} - \log c_l) \\
&\leq \mathrm{const} \cdot \sum_{k=1}^N d_k \left(\log M + D_N^{1/2}\right) \leq \mathrm{const} \cdot D_N^{3/2}.
\end{aligned}$$

Hence setting

$$T_N = \frac{1}{D_N} \sum_{k \le N} d_k \xi_k$$

we get

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

Let $\eta > 0$ be so small that $(1+\varepsilon)(1-\eta) > 1$. Since (d_n) is bounded by (4.1) and $c_{n+1}/c_n = O(1)$, we have $D_{n+1}/D_n \to 1$ and thus we can choose a nondecreasing sequence (N_k) of positive integers such that

$$D_{N_k} \sim e^{k^{1-\eta}} \tag{4.6}$$

and consequently

$$ET_{N_k}^2 \le \text{const} \cdot k^{-1-\varrho}$$

for some $\varrho > 0$. Hence we have $\sum_{k=1}^{\infty} |T_{N_k}|^2 < +\infty$ a.s., implying $T_{N_k} \to 0$ a.s. Now for $N_k < N \le N_{k+1}$ we have

$$|T_N| \le |T_{N_k}| + \frac{2K}{D_N} \sum_{i=N_k+1}^{N} d_i = |T_{N_k}| + 2K \left(1 - \frac{D_{N_k}}{D_N}\right).$$

Since $D_{N_{k+1}}/D_{N_k} \to 1$ by (4.6), it follows that $T_N \to 0$ a.s., completing the proof of (4.2).

If condition (2.8) of Theorem 4 is assumed only for $l > k \ge A$ with some constant $A > 0$, then the theorem remains valid with the summations in (2.9) and (2.10) extended for $A \le k \le N$. (The proof requires only trivial changes.) In this case $f_l, f_{k,l}, c_k$ need not even be defined if k or l is $< A$. A similar remark applies in our other theorems.

It is easily seen that if the right hand side of (2.8) is replaced by $C(c_l/c_k)^{-\gamma}$ for some $\gamma > 0$, then in Theorem 4 we can choose

$$d_k = \log(c_{k+1}/c_k) \, \exp\left((\log c_k)^{\alpha}\right) \qquad 0 \le \alpha < 1/2.$$

In this case instead of (4.4) we get $|E(\xi_k \xi_l)| \le \text{const} \cdot (c_l/c_k)^{-\gamma}$ and thus the contribution of those terms in (4.5) where $c_l/c_k \ge (\log D_N)^{2/\gamma}$ is at most $D_N^2 (\log D_N)^{-2}$. On the other hand, in the present case we get by elementary calculations

$$D_n \sim \text{const} \cdot (\log c_n)^{1-\alpha} \exp\left((\log c_n)^{\alpha}\right)$$

and consequently

$$\exp\left((\log c_n)^\alpha\right) \sim \text{const} \cdot D_n/(\log D_n)^{(1-\alpha)/\alpha}.$$

Thus the contribution of those terms in (4.5) where $c_l/c_k < (\log D_N)^{2/\gamma}$ is

$$\begin{aligned}
&\le 8K^2 \cdot \sum_{k=1}^N d_k \sum_{\{l\ge k:\, c_l \le c_k (\log D_N)^{2/\gamma}\}} (\log c_{l+1} - \log c_l)\exp((\log c_l)^\alpha) \\
&\le \text{const} \cdot \exp((\log c_N)^\alpha) \sum_{k=1}^N d_k \log\log D_N = \text{const} \cdot \exp((\log c_N)^\alpha)\, D_N \log\log D_N \\
&\le \text{const} \cdot D_N^2 \log\log D_N/(\log D_N)^{(1-\alpha)/\alpha} \le \text{const} \cdot D_N^2/(\log D_N)^{-(1+\varepsilon)}
\end{aligned}$$

for sufficiently large N; here $\varepsilon > 0$ by $\alpha < 1/2$. The rest of the proof is the same as above.

The previous argument breaks down for $1/2 \le \alpha < 1$, but it is worth noting that even in this case we have $T_N \xrightarrow{P} 0$ and thus relation (2.10) implies (2.9) at least in probability. Whether we can have a.s. convergence in this case remains open.

We finally prove the claim, made in Section 3, that if in Theorem 4 the function $f_k(x_1, \dots, x_{n_k})$ depends only on $x_{n_{k-1}+1}, \dots, x_{n_k}$ (i.e., if different f_k's depend on disjoint segments of the sequence $x_1, x_2, \dots$), then the conclusion of the theorem holds for any weight sequence (d_n) satisfying (3.4), (3.5) and this becomes false if we replace the o in (3.5) by O. The direct part is easy: letting g denote a bounded Lip 1 function on R, we have to verify, just as in the proof above,

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k \xi_k = 0 \qquad \text{a.s.} \tag{4.7}$$

where

$$\xi_k = g(f_k(X_{n_{k-1}+1}, \dots, X_{n_k})) - Eg(f_k(X_{n_{k-1}+1}, \dots, X_{n_k})).$$

Clearly, (3.5) implies

$$\max_{k\le n} d_k = o\left(E_n/(\log\log E_n)^{1/2}\right)$$

and thus

$$E_n^2 \le \left(\max_{k\le n} d_k\right) D_n = o\left(D_n E_n/(\log\log E_n)^{1/2}\right)$$

whence

$$E_n = o\left(D_n/(\log\log E_n)^{1/2}\right). \tag{4.8}$$

Now (ξ_n) is a uniformly bounded sequence of independent, zero mean random variables and if $E\xi_n^2 \geq c > 0$ for some constant c then the variance of $\sum_{k\leq n} d_k\xi_k$ lies between positive constant multiples of E_n and thus (3.5) shows that Kolmogorov's LIL applies to the sequence $(d_n\xi_n)$. Hence we get, using also (4.8), that

$$\sum_{k\leq n} d_k\xi_k = O\left(E_n(\log\log E_n)^{1/2}\right) = o(D_n) \qquad \text{a.s.}$$

and thus (4.7) is valid. If $\inf E\xi_n^2 = 0$ then we get the same conclusion by replacing ξ_n by $\xi_n^* = \xi_n + \zeta_n$, where ζ_n are independent r.v.'s, independent also of the ξ_n's, such that $P(\zeta_n = 1) = P(\zeta_n = -1) = 1/2$.

Conversely, let (X_n) be an i.i.d. sequence with $P(X_1 = 1) = P(X_1 = -1) = 1/2$ and let $n_k = k$, $f_k(x_1, \ldots, x_{n_k}) = f_k(x_k) = x_k$, $d_k = e^{k/\log k}$. Simple calculations show that

$$D_n = \sum_{k\leq n} d_k = \sum_{k\leq n} e^{k/\log k} \sim e^{n/\log n}\log n \tag{4.9}$$

$$E_n^2 = \sum_{k\leq n} d_k^2 = \sum_{k\leq n} e^{2k/\log k} \sim \frac{1}{2}e^{2n/\log n}\log n \tag{4.10}$$

and thus

$$d_n = O\left(E_n/(\log\log E_n)^{1/2}\right).$$

Let $G(x)$ denote the distribution function having jump $1/2$ at $x = -1$ and $x = 1$. Clearly $P\{f_k(X_k) < x\} = G(x)$ for all x and thus

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\leq N} d_k P\{f_k(X_k) < x\} = G(x) \qquad \text{for any } x \in C_G$$

but we will show that

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\leq N} d_k I\{f_k(X_k) < x\} = G(x) \qquad \text{a.s. for any } x \in C_G$$

is not valid. Specifically, we prove that setting $\eta_k = I\{f_k(X_k) < 0\} - P\{f_k(X_k) < 0\}$, the relation

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k\leq n} d_k\eta_k = 0 \qquad \text{a.s.} \tag{4.11}$$

is not valid. To see this, let i_n be the smallest integer such that $D_{i_n} \geq e^n$. By the relation $D_{n+1}/D_n \to 1$ we have $D_{i_n} \sim e^n$ and it is also easily seen that $i_n \sim n \log n$. Relations (4.9), (4.10) imply

$$\sum_{i=i_n+1}^{i_{n+1}} d_i^2 \sim \text{const} \cdot \frac{e^{2n}}{\log n}. \tag{4.12}$$

Now η_n are i.i.d. symmetric two-valued random variables and simple calculations show that the finite sequence $\{d_i \eta_i;\ i_n + 1 \leq i \leq i_{n+1}\}$ satisfies the conditions of Feller's large deviation theorem (see Feller, 1943, Theorem 1). It follows that if $x_n = c_0 \sqrt{\log n}$ $(n = 1, 2, \dots)$ with a sufficiently small absolute constant c_0 then

$$P\left\{ \sum_{i=i_n+1}^{i_{n+1}} d_i \eta_i \geq x_n \left(\sum_{i=i_n+1}^{i_{n+1}} d_i^2 \right)^{1/2} \right\} \geq \text{const} \cdot \frac{1}{\sqrt{n}} \qquad \text{as } n \to \infty.$$

In view of (4.12) this implies that for a sufficiently small positive constant c_1 we have

$$P\left\{ \sum_{i=i_n+1}^{i_{n+1}} d_i \eta_i \geq c_1 e^n \right\} \geq \text{const} \cdot \frac{1}{\sqrt{n}}$$

and thus

$$P\left\{ \sum_{i=i_n+1}^{i_{n+1}} d_i \eta_i \geq c_1 e^n \quad \text{i.o.} \right\} = 1. \tag{4.13}$$

Now if (4.11) were true then by the definition of i_n and $D_{i_n} \sim e^n$ it would follow that

$$\sum_{i=i_n+1}^{i_{n+1}} d_i \eta_i = o(e^n) \qquad \text{a.s.}$$

which contradicts to (4.13).

5. Applications

In this section we shall give several applications of our theorems.

1. *Partial sums.* Let $X_1, X_2, \ldots$ be independent random variables with partial sums $S_n = \sum_{k \le n} X_k$ and assume that $(S_n - a_n)/b_n \xrightarrow{\mathcal{D}} G$ for some distribution function G and numerical sequences (a_n), (b_n) satisfying $b_n \uparrow \infty$, $b_{n+1}/b_n = O(1)$. Assume also that

$$E\left(\log_+ \log_+ \left|\frac{S_n - a_n}{b_n}\right|\right)^{1+\delta} \le K \qquad (n = 1, 2, \ldots) \tag{5.1}$$

for some constants $\delta > 0$, $K > 0$. Then the assumptions of Theorem 2 are satisfied with $c_k = b_k$ and

$$f_l(x_1, \ldots, x_l) = \left(\sum_{i=1}^{l} x_i - a_l\right)/b_l, \quad f_{k,l}(x_{k+1}, \ldots, x_l) = \left(\sum_{i=k+1}^{l} x_i - (a_l - a_k)\right)/b_l.$$

Indeed, letting $g(x) = 1 + (\log_+ \log_+ x)^{1+\delta}$, the function $x/g(x)$ is continuous for $x \ge 0$ and increasing for $x \ge x_0$ and thus there exists a number $a_0 > 0$ such that $x/y \le g(x)/g(y)$ for $0 \le x \le y$, $y \ge a_0$. Hence letting $\lambda = b_l/b_k$ and assuming $\lambda \ge a_0$ we get, using (5.1),

$$\begin{aligned} &E\left(|f_l(X_1, \ldots, X_l) - f_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1\right) \\ &= E\left(\left|\frac{S_k - a_k}{b_l}\right| \wedge 1\right) = E\left(\frac{1}{\lambda}\left\{\left|\frac{S_k - a_k}{b_k}\right| \wedge \lambda\right\}\right) \le \frac{1}{g(\lambda)} Eg\left(\left|\frac{S_k - a_k}{b_k}\right| \wedge \lambda\right) \\ &\le \frac{1}{g(\lambda)} Eg\left(\left|\frac{S_k - a_k}{b_k}\right|\right) \le \text{const} \cdot \left(\log_+ \log_+ \frac{b_l}{b_k}\right)^{-(1+\delta)}. \end{aligned} \tag{5.2}$$

Increasing the constant if necessary, the last expression in (5.2) will exceed 1 for $1 \le b_l/b_k \le a_0$, and thus the relation

$$E\left(|f_l(X_1, \ldots, X_l) - f_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1\right) \le \text{const} \cdot \left(\log_+ \log_+ \frac{b_l}{b_k}\right)^{-(1+\delta)}$$

holds for all $l \ge k \ge 1$, showing that Theorem 2 applies with $c_k = b_k$. If we assume also

$$b_l/b_k \ge C(l/k)^\gamma \qquad (l \ge k) \tag{5.3}$$

for some constants $C > 0$, $\gamma > 0$, then the last expression in (5.2) is bounded by const $\cdot$ $(\log_+ \log_+(l/k))^{-(1+\delta)}$ and thus Theorem 2 applies with $c_k = k$, in which case (2.3) gives $d_k \sim \text{const}/k$. Hence we obtain

Theorem A. *Let $X_1, X_2, \dots$ be independent random variables with partial sums S_n and assume that $(S_n - a_n)/b_n \xrightarrow{\mathcal{D}} G$ for some distribution function G and numerical sequences (a_n), (b_n) satisfying $b_n \uparrow \infty$, $b_{n+1}/b_n = O(1)$. Assume also that (5.1) holds for some constants $\delta > 0$, $K > 0$. Then*

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{\frac{S_k - a_k}{b_k} < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G \tag{5.4}$$

where d_k and D_k are defined by

$$d_k = \log(b_{k+1}/b_k), \quad D_n = \sum_{k\le n} d_k. \tag{5.5}$$

If the norming factors b_n satisfy also (5.3), then (5.4) holds with $d_k = 1/k$, $D_N = \log N$.

In the case $d_k = 1/k$ this was proved by Berkes and Dehling (1993); the general case extends also a result of Ibragimov and Lifshits (1999) who proved the same conclusion under the stronger moment condition

$$E\left|\frac{S_n - a_n}{b_n}\right|^p \le K \qquad (n = 1, 2, \dots, p > 0)$$

instead of (5.1).

Recently Lifshits (2000a) constructed a sequence (X_n) of independent r.v.'s with mean 0 and finite variances whose partial sums S_n satisfy $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} N(0,1)$ and

$$E\left(\log_+ \log_+ \left|\frac{S_n}{\sqrt{n}}\right|\right) \le K \qquad (n = 1, 2, \dots)$$

for some constant $K > 0$ but the a.s. central limit theorem

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{S_k}{\sqrt{k}} < x\right\} = \Phi(x) \qquad \text{a.s. for any } x$$

is not valid. This shows that condition (5.1) of Theorem A is sharp even in the case $a_n = 0$, $b_n = \sqrt{n}$. In view of the estimate (5.2), this implies that assumption (2.2) of Theorem 2 is also sharp, and thus the same holds for the analogous conditions in Theorems 3, 4, 5.

2. *Extremes.* Let $X_1, X_2, \ldots$ be i.i.d. random variables and (a_n), (b_n) numerical sequences. Then the assumptions of Theorem 1 are satisfied with $c_k = k$ and

$$f_l(x_1, \ldots, x_l) = a_l \left(\max_{i \le l} x_i - b_l \right), \qquad f_{k,l}(x_{k+1}, \ldots, x_l) = a_l \left(\max_{k+1 \le i \le l} x_i - b_l \right).$$

Indeed, in this case $f_l(X_1, \ldots, X_l) - f_{k,l}(X_{k+1}, \ldots, X_l)$ differs from zero only if

$$\max_{1 \le i \le k} X_i > \max_{k+1 \le i \le l} X_i$$

and we will see below that the probability of this event is $\le k/l$. Therefore

$$E\left(|f_l(X_1, \ldots, X_l) - f_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1 \right) \le k/l.$$

Thus we obtain the following result (see Cheng et al., 1998; Fahrner and Stadtmüller, 1998):

Theorem B. *Let $X_1, X_2, \ldots$ be i.i.d. random variables such that setting $M_k = \max_{i \le k} X_i$ we have*

$$a_k(M_k - b_k) \xrightarrow{\mathcal{D}} G$$

for some numerical sequences (a_n), (b_n) and a distribution function G. Then

$$\lim_{N \to \infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{ a_k(M_k - b_k) < x \right\} = G(x) \qquad \text{a.s. for any } x \in C_G.$$

We note that, as observed in Cheng et al. (1998), Fahrner and Stadtmüller (1998), Theorem B becomes false if we replace logarithmic averages by ordinary (Cesàro) averages.

It remains to prove that if $1 \le k < l$ and $X_1, \ldots, X_l$ are i.i.d. random variables, then

$$P(\max_{1 \le i \le k} X_i > \max_{k+1 \le i \le l} X_i) \le k/l. \tag{5.6}$$

Letting F denote the (left continuous) distribution function of X_1, the distribution functions of the two random variables in (5.6) are $H_1(x) = F(x)^k$ and $H_2(x) = F(x)^{l-k}$ and thus the probability in (5.6) equals

$$\int_{-\infty}^{\infty} H_2(x) dH_1(x) = \int_{-\infty}^{\infty} F(x)^{l-k} k F(x)^{k-1} dF(x) \le \int_0^1 k t^{l-1} dt = k/l$$

where in the second step we used the fact that

$$\int_{-\infty}^{\infty} \psi(F(x))dF(x) \leq \int_0^1 \psi(t)dt \tag{5.7}$$

for any nondecreasing function ψ on $[0,1]$. To verify (5.7), let $F^{-1}(t) = \sup\{x : F(x) \leq t\}$ and let U be a r.v. uniformly distributed on $(0,1)$. Then $F(F^{-1}(t)) \leq t$ for all $t \in (0,1)$ and the r.v. $Y = F^{-1}(U)$ has distribution function F. Thus the left hand side of (5.7) equals

$$E\psi(F(Y)) = E\psi(F(F^{-1}(U))) \leq E\psi(U) = \int_0^1 \psi(t)dt,$$

as claimed.

Theorem B extends easily to independent, not identically distributed random variables. Note that the identical distribution of the X_i was used only to obtain (5.6) and the proof remains valid if (5.6) holds with the right hand side replaced by $C(k/l)$ for a constant C. This modified inequality holds, in turn, if for any $l \geq 1$ and any permutation $\{i_1, \dots, i_l\}$ of $\{1, 2, \dots, l\}$ we have

$$P\left(X_{i_1} \geq \max\left(X_{i_2}, \dots, X_{i_l}\right)\right) \leq C/l. \tag{5.8}$$

Condition (5.8) is satisfied, e.g., if the distributions of the X_i are continuous and resemble each other in the sense that there exists a continuous distribution function F and positive constants γ_1, γ_2 (necessarily $\gamma_1 \leq 1$) such that for all $i \geq 1$

$$\gamma_1(1 - F(t)) \leq P(X_i \geq t) \leq \gamma_2(1 - F(t)) \qquad \text{for all } t \in R.$$

Indeed, letting F_i denote the distribution function of X_i and using the postulated bounds for $P(X_i \geq t)$ we get that the probability in (5.8) equals

$$\begin{aligned}
&\int_{-\infty}^{\infty} (1 - F_{i_1}(x))d\left(\prod_{r=2}^{l} F_{i_r}(x)\right) \\
&\leq \gamma_2 \int_{-\infty}^{\infty} (1 - F(x))d\left(\prod_{r=2}^{l} F_{i_r}(x)\right) = \gamma_2 \int_{-\infty}^{\infty} \left(\prod_{r=2}^{l} F_{i_r}(x)\right) dF(x) \\
&\leq \gamma_2 \int_{-\infty}^{\infty} (1 - \gamma_1(1 - F(x)))^{l-1}\, dF(x) \leq \gamma_2 \int_0^1 (1 - \gamma_1(1 - u))^{l-1}\, du \leq \frac{\gamma_2}{\gamma_1} l^{-1}
\end{aligned}$$

where we used (5.7) again.

Obviously, one cannot hope that (5.6) holds for independent X_i with radically different distributions. An instructive example is given by the case when the distribution function of X_k is $F(x)^{c_k-c_{k-1}}$ where F is a fixed distribution function and (c_k) is an increasing sequence with $c_0=0$ and $c_1\ge 1$. In this case the distribution functions $H_1(x)$ and $H_2(x)$ of $\max_{1\le i\le k} X_i$ and $\max_{k+1\le i\le l} X_i$ are $F(x)^{c_k}$ and $F(x)^{c_l-c_k}$, respectively, and thus the probability in (5.6) is

$$\int_{-\infty}^{\infty} H_2(x)dH_1(x)=\int_{-\infty}^{\infty} F(x)^{c_l-c_k}c_kF(x)^{c_k-1}dF(x)\le c_k/c_l$$

where we used (5.7) in the last step. Thus (5.6) can fail, but Theorem 1 still applies in this case and we get the following theorem extending Theorem B:

Theorem C. *Let $X_1, X_2, \dots$ be independent random variables such that the distribution of X_k is $F(x)^{c_k-c_{k-1}}$ where (c_k) is an increasing sequence satisfying $c_0=0$, $c_1\ge 1$, $c_n\to\infty$ and $c_{n+1}/c_n=O(1)$. Let $M_k=\max_{i\le k} X_i$ and assume that*

$$a_k(M_k-b_k)\xrightarrow{\mathcal{D}} G$$

for some numerical sequences (a_n), (b_n) and a distribution function G. Then letting $d_k=\log(c_{k+1}/c_k)$ and $D_n=\sum_{k\le n} d_k$, we have

$$\lim_{N\to\infty}\frac{1}{D_N}\sum_{k\le N} d_kI\left\{a_k(M_k-b_k)<x\right\}=G(x)\qquad \text{a.s. for any } x\in C_G.$$

3. *Maxima of partial sums.* Let $X_1, X_2, \dots$ be independent random variables with partial sums $S_n=\sum_{k\le n} X_k$ and let $S_n^*=\max_{k\le n} S_k$. Assume that for some positive numerical sequence (b_n) we have

$$E\left(\log_+\log_+\left|\frac{S_n}{b_n}\right|\right)^{1+\delta}\le K\qquad (n=1,2,\dots)\tag{5.9}$$

for some $K>0$, $\delta>0$ and the analogous relation for S_n^* is also valid. Then the assumptions of Theorem 2 are satisfied with $c_k=b_k$ and

$$f_l(x_1,\dots,x_l)=\frac{1}{b_l}\max_{i\le l}(x_1+\dots+x_i)$$

$$f_{k,l}(x_{k+1},\dots,x_l)=\begin{cases}\dfrac{1}{b_l}\max\limits_{k+1\le i\le l}(x_{k+1}+\dots+x_i) & \text{if } i_0>k\\ 0 \text{ if } i_0\le k\end{cases}$$

where $1 \le i_0 \le l$ is the smallest integer where $s_i = x_1 + \ldots + x_i$, $1 \le i \le l$, reaches its maximum. Indeed, letting $s_j^* = \max_{i \le j} s_i$, observe that the difference

$$\Delta = f_l(x_1, \ldots, x_l) - f_{k,l}(x_{k+1}, \ldots, x_l)$$

equals s_k^*/b_l if $i_0 \le k$, while for $i_0 > k$ the sum $x_{k+1} + \ldots + x_i$, $k+1 \le i \le l$ reaches its maximum also at $i = i_0$ and thus $\Delta = s_k/b_l$. Therefore

$$E(|f_l(X_1, \ldots, X_l) - f_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1) \le E\left(\left|\frac{S_k}{b_l}\right| \wedge 1\right) + E\left(\left|\frac{S_k^*}{b_l}\right| \wedge 1\right)$$

and thus estimate (5.2) applies with trivial changes. Hence we obtain

Theorem D. *Let $X_1, X_2, \ldots$ be independent r.v.'s with partial sums S_n and let $S_n^* = \max_{k \le n} S_k$. Let (b_n) be a positive numerical sequence satisfying $b_n \uparrow \infty$, $b_{n+1}/b_n = O(1)$ and assume that*

$$S_n^*/b_n \xrightarrow{\mathcal{D}} G$$

for some distribution function G and also that (5.9) and its analogue for S_n^* hold. Then

$$\lim_{N \to \infty} \frac{1}{D_N} \sum_{k \le N} d_k I\left\{\frac{S_k^*}{b_k} < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G$$

where d_k and D_k are defined by (5.5). If the norming factors b_n satisfy also (5.3), then the last convergence relation holds with $d_k = 1/k$, $D_N = \log N$.

An analogous result holds for the absolute maxima $S_n^{**} = \max_{k \le n} |S_k|$; the proof is similar.

4. *Empirical distribution functions.* Let $X_1, X_2, \ldots$ be i.i.d. random variables with continuous distribution function F and let

$$F_n(x) = \frac{1}{n} \sum_{k \le n} I(X_k < x) \tag{5.10}$$

be the empirical distribution function of the sample $(X_1, \ldots, X_n)$. Then the assumptions of Theorem 1 are satisfied with $c_k = k$ and

$$f_l(x_1, \ldots, x_l) = \frac{1}{\sqrt{l}} \sup_x \left|\sum_{i \le l} (I(x_i < x) - F(x))\right|,$$

$$f_{k,l}(x_{k+1},\dots,x_l)=\frac{1}{\sqrt{l}}\sup_x\left|\sum_{k+1\le i\le l}(I(x_i<x)-F(x))\right|.$$

Indeed, letting $\lambda=\sqrt{l/k}$ and

$$T_k=\sup_x\left|\sum_{i\le k}(I(X_i<x)-F(x))\right|$$

we get

$$E|f_l(X_1,\dots,X_l)-f_{k,l}(X_{k+1},\dots,X_l)|\le E|T_k/\sqrt{l}|\le \text{const}\cdot(k/l)^{1/2}$$

since $E|T_k/\sqrt{k}|$ is bounded (see Dvoretzky et al., 1956, Lemma 2). Also, the classical theorem of Kolmogorov-Smirnov implies that

$$f_k(X_1,\dots,X_k)\xrightarrow{\mathcal{D}}G$$

where

$$G(x)=\sum_{j=-\infty}^{\infty}(-1)^je^{-2j^2x^2}.$$

Thus Theorem 1 implies

Theorem E. *Let $X_1,X_2,\dots$ be i.i.d. random variables with continuous distribution function F, let F_n be the empirical distibution function defined by (5.10) and let*

$$D_n=\sup_x|F_n(x)-F(x)|$$

be the Kolmogorov statistics. Then

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{k\le N}\frac{1}{k}I\left\{\sqrt{k}D_k<x\right\}=\sum_{j=-\infty}^{\infty}(-1)^je^{-2j^2x^2}\qquad\text{a.s. for any } x.$$

A similar argument yields the one-sided result

$$\lim_{N\to\infty}\frac{1}{\log N}\sum_{k\le N}\frac{1}{k}I\left\{\sqrt{k}D_k^+<x\right\}=1-e^{-2x^2}\qquad\text{a.s. for any } x$$

where $D_n^+ = \sup_x (F_n(x) - F(x))$.

5. *U-statistics.* Let $X_1, X_2, \dots$ be an i.i.d. sequence, $m \ge 1$ an integer and $h(x_1, \dots, x_m)$ a symmetric measurable function satisfying

$$Eh^2(X_1, \dots X_m) < \infty.$$

Let

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \dots < i_m \le n} h(X_{i_1}, \dots, X_{i_m})$$

be the corresponding U-statistic. Set $\theta = Eh(X_1, \dots, X_m)$, and put for $1 \le j \le m$

$$h_j(x_1, \dots, x_j) = Eh(x_1, \dots, x_j, X_{j+1}, \dots, X_m), \qquad \zeta_j = \operatorname{Var} h_j(X_1, \dots, X_j).$$

It is known (see e.g. Serfling, 1980, p. 182) that

$$0 = \zeta_0 \le \zeta_1 \cdots \le \zeta_m = \operatorname{Var} h(X_1, \dots, X_m) < \infty.$$

Let $c \ge 1$ denote the smallest integer such that $\zeta_c > 0$; we call c the critical parameter of the statistic U_n. It is known that $n^{c/2}(U_n - \theta)$ has a nondegenerate limit distribution. (See e.g. Koroljuk and Borovskich, 1994; we refer also to Denker, 1985; Giné and Zinn, 1994 and the references there for various further related results.) Put $\tilde{h}(X_1, \dots, X_m) = h(X_1, \dots, X_m) - \theta$ and set, for any $1 \le k < l$,

$$f_l(x_1, \dots, x_l) = \frac{l^{c/2}}{\binom{l}{m}} \sum_{1 \le i_1 < \dots < i_m \le l} \tilde{h}(x_{i_1}, \dots, x_{i_m}),$$

$$f_{k,l}(x_{k+1}, \dots, x_l) = \frac{l^{c/2}}{\binom{l}{m}} \sum_{k+1 \le i_1 < \dots < i_m \le l} \tilde{h}(x_{i_1}, \dots, x_{i_m}).$$

We claim that

$$E(|f_l(X_1, \dots, X_l) - f_{k,l}(X_{k+1}, \dots, X_l)|)^2 \le C(k/l) \qquad (k \le l) \tag{5.11}$$

for some positive constant C; this will show that Theorem 1 applies in our case. Relation (5.11) can be equivalently written as

$$E(l^{c/2} U_{k,l})^2 \le C(k/l) \tag{5.12}$$

where

$$U_{k,l} = \frac{1}{\binom{l}{m}} \sum\nolimits^{*}_{1\le i_1<\ldots<i_m\le l} \tilde{h}(X_{i_1},\ldots,X_{i_m}).$$

and the * means that at least one of $i_1,\ldots,i_m$ lies in the interval $[1,k]$. To prove (5.12) we note that if $\{a_1,\ldots,a_m\}$ and $\{b_1,\ldots,b_m\}$ are two sets of distinct positive integers with j common elements, then

$$E\tilde{h}(X_{a_1},\ldots,X_{a_m})\tilde{h}(X_{b_1},\ldots,X_{b_m}) = \zeta_j. \tag{5.13}$$

(see Serfling, 1980, p. 183). Now expanding $U_{k,l}^2$ and using (5.13) and $\zeta_j = 0$ for $j < c$, we get nonzero terms belonging only to such sets $\{a_1,\ldots,a_m\}$ and $\{b_1,\ldots,b_m\}$ which have at least c common elements. Thus among $a_1,\ldots,a_m,b_1,\ldots,b_m$ there are at most $2m-c$ different ones, moreover, at least one of these elements must be in the interval $[1,k]$. Therefore, the number of choices for the union set $\{a_1,\ldots,a_m,b_1,\ldots,b_m\}$ is at most kl^{2m-c-1}. Once this union set is fixed, each of the sets $\{a_1,\ldots,a_m\}$ and $\{b_1,\ldots,b_m\}$ can be chosen at most in C_m ways, where $C_m = \binom{2m}{m}$. Thus, the number of nonzero terms in $E(U_{k,l})^2$ is at most $C_m^2 kl^{2m-c-1}$. By the Cauchy-Schwarz inequality, the left hand side of (5.13) is at most $\gamma_0 = \operatorname{Var} h(X_1,\ldots,X_m)$ and since we have $\binom{l}{m} \sim l^m$, (5.12) follows. Hence we proved the following theorem:

Theorem F. *Let $X_1, X_2, \ldots$ be i.i.d. random variables, $m \ge 1$ a fixed integer and*

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1\le i_1<\ldots<i_m\le n} h(X_{i_1},\ldots,X_{i_m})$$

where $h(x_1,\ldots,x_m)$ a symmetric measurable function satisfying

$$Eh^2(X_1,\ldots X_m) < \infty.$$

Let $\theta = Eh(X_1,\ldots,X_m)$ and let c be the critical parameter of the statistic U_n. Then

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{k^{c/2}(U_k-\theta) < x\right\} = F(x) \qquad \text{a.s. for any } x \in C_F$$

where F is the limit distribution of $n^{c/2}(U_n-\theta)$.

6. *Local times.* Let $X_1, X_2, \ldots$ be i.i.d. integer valued random variables with $EX_1 = 0$. Put $\psi(v) = E(e^{ivX_1})$ and assume that $\psi(2\pi t) = 1$ if and only if t is an integer and ψ satisfies

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\psi(v)}{1-\lambda\psi(v)}dv \sim \frac{a}{(1-\lambda)^\beta}, \qquad \text{as} \quad \lambda \to 1$$

where $0 < \beta \le 1/2$. Putting $S_n = X_1 + \ldots + X_n$, it follows that the random walk $\{S_n,\, n \ge 1\}$ is aperiodic and X_1 is in the domain of normal attraction of a stable law of order $\alpha = 1/(1-\beta)$. (Note that $1 < \alpha \le 2$.) Define the local time $\xi(z,n)$ by

$$\xi(z,n) = \sum_{i=1}^{n} I\{S_i = z\}, \qquad z \in Z. \tag{5.14}$$

It was shown by Darling and Kac (1957) that

$$\lim_{n\to\infty} P\left\{\frac{\xi(0,n)}{an^\beta} < x\right\} = F_\beta(x) = \frac{1}{\pi\beta}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}}{j!j}\sin(\pi\beta j)\Gamma(1+\beta j)x^j. \tag{5.15}$$

Kesten and Spitzer (1979) showed that for some constant c we have

$$E\xi(0,n) \le cn^\beta, \qquad E\left(\xi(z,n)-\xi(0,n)\right)^2 \le c|z|^{\beta/(1-\beta)}n^\beta.$$

Thus choosing

$$f_l(x_1,\ldots,x_l) = \frac{1}{al^\beta}\sum_{i=1}^{l} I\{x_1+\ldots+x_i = 0\}$$
$$f_{k,l}(x_{k+1},\ldots,x_l) = \frac{1}{al^\beta}\sum_{i=k+1}^{l} I\{x_{k+1}+\ldots+x_i = 0\}$$

we have

$$\begin{aligned}
&E|f_l(X_1,\ldots,X_l) - f_{k,l}(X_{k+1},\ldots,X_l)| \\
&= E\left|\frac{\xi(0,l)}{al^\beta} - \frac{\xi(S_k,l)-\xi(S_k,k)}{al^\beta}\right| \\
&= E\left|\frac{\xi(0,k)}{al^\beta} - \frac{\xi(S_k,l)-\xi(S_k,k)-(\xi(0,l)-\xi(0,k))}{al^\beta}\right| \\
&\le \frac{C}{a}\left((k/l)^\beta + l^{-\beta}E|S_k|^{\beta/2(1-\beta)}l^{\beta/2}\right) \le C'\left((k/l)^\beta + (k/l)^{\beta/2}\right)
\end{aligned}$$

Thus using Theorem 1 we get the following

Theorem G. *Let the random walk $\{S_n,\ n \ge 1\}$ satisfy the above conditions and let $\xi(z,n)$ be its local time defined by (5.14). Then*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{\xi(0,k)}{ak^\beta} < x\right\} = F_\beta(x) \qquad \text{a.s. for any } x$$

where F_β is the distribution function defined by (5.15).

7. *Return times.* Let $0 = \tau_0 < \tau_1 < \dots$ be the successive times of return to the origin of a two dimensional simple symmetric random walk and put $X_n = \tau_n - \tau_{n-1}$ $(n = 1, 2, \dots)$. Clearly, (X_n) is an i.i.d. sequence; it is known (see Dvoretzky and Erdős, 1951) that

$$P(X_1 > t) \sim \frac{c}{\log t} \qquad \text{as } t \to \infty. \tag{5.16}$$

Setting $M_k = \max_{i\le k} X_i$, relation (5.16) implies

$$\frac{1}{k}\log M_k \xrightarrow{\mathcal{D}} H$$

where

$$H(x) = \begin{cases} e^{-c/x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases} \tag{5.17}$$

Now Theorem 1 applies to the sequence (X_n) with $c_k = k$ and

$$f_l(x_1,\dots,x_l) = \frac{1}{l}\log\max_{i\le l} x_i, \qquad f_{k,l}(x_{k+1},\dots,x_l) = \frac{1}{l}\log\max_{k+1\le i\le l} x_i.$$

(This can be verified exactly as in the case of Theorem B.) Hence we get

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{1}{k}\log M_k < x\right\} = H(x) \qquad \text{a.s. for any } x.$$

But $M_k \le \tau_k \le kM_k$ and thus $(\log\tau_k - \log M_k)/k \to 0$ a.s. Thus we proved

Theorem H. *Let $0 = \tau_0 < \tau_1 < \dots$ be the successive times of return to the origin of a two dimensional simple symmetric random walk. Then we have*

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{1}{k}\log \tau_k < x\right\} = H(x) \qquad \text{a.s. for any } x$$

where H is the distribution function defined by (5.17).

Actually, all we used about the i.i.d. sequence (X_n) was its positivity and (5.16), hence the same argument shows that if (X_n) is a positive i.i.d. sequence satisfying (5.16) then setting $S_k = \sum_{i\le k} X_i$ we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k\le N} \frac{1}{k} I\left\{\frac{1}{k}\log S_k < x\right\} = H(x) \qquad \text{a.s. for any } x.$$

8. *Darling-Erdős type limit theorems.* Let (X_n) be a sequence of independent random variables with mean 0, variance 1 and uniformly bounded third absolute moments. Put $S_k = X_1 + \dots + X_k$. By a well known theorem of Darling and Erdős (1956) we have

$$a_n \left(\max_{k\le n} \frac{S_k}{\sqrt{k}} - b_n\right) \xrightarrow{\mathcal{D}} G$$

where

$$a_n = (2\log\log n)^{1/2}, \quad b_n = (2\log\log n)^{1/2} + \frac{\log\log\log n - \log 4\pi}{2(2\log\log n)^{1/2}} \qquad (n\ge 3) \qquad (5.18)$$

and

$$G(x) = \exp(-e^{-x}). \qquad (5.19)$$

(Actually, the assumption on the third moments can be weakened, see Einmahl, 1989; Oodaira, 1976; Shorack, 1979.) An analogous result holds for the Wiener process W, in fact we have

$$a_n \left(\sup_{1\le t\le n} \frac{W(t)}{\sqrt{t}} - b_n\right) \xrightarrow{\mathcal{D}} G \qquad (5.20)$$

where (a_n), (b_n) and G are the same as above. We now show that

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{a_k \left(\sup_{1\le t\le k} \frac{W(t)}{\sqrt{t}} - b_k\right) < x\right\} = G(x) \qquad \text{a.s. for any } x$$

for a suitable (d_k), where $D_N = \sum_{k \le N} d_k$. To this end, let

$$c_n = \exp(\sqrt{\log\log n}), \quad A_n = \exp\left(\log n / \exp(\sqrt{\log\log n})\right) \qquad (n \ge 3),$$

$$\xi_l = a_l \left(\sup_{1 \le t \le l} \frac{W(t)}{\sqrt{t}} - b_l \right) \qquad (l \ge 3),$$

and put for $l \ge k \ge 3$

$$\xi_{k,l} = \begin{cases} a_l \left(\sup\limits_{A_l^2 \le t \le l} \dfrac{W(t) - W(k)}{\sqrt{t}} - b_l \right) & \text{if} \quad k \le A_l \\ 0 & \text{if} \quad k > A_l. \end{cases}$$

Clearly $\xi_{k,l}$ is measurable with respect to $\sigma\{X(t') - X(t) : k \le t \le t' \le l\}$. We claim that

$$E\left(|\xi_l - \xi_{k,l}| \wedge 1\right) \le 4(c_k/c_l)^{1/2} \qquad (3 \le k < l, \ l \ge l_0) \tag{5.21}$$

where l_0 is an absolute constant. In the case $k > A_l$ relation (5.21) is valid since the left hand side is at most 1, while the right hand side exceeds 1 since

$$c_k \ge \exp\left(\sqrt{\log\log A_l}\right) = \exp\left((\log\log l - \sqrt{\log\log l})^{1/2}\right) \ge \exp\left(\sqrt{\log\log l} - 1\right) \ge c_l/4.$$

To prove (5.21) for $k \le A_l$ we first note that the stationarity and Markov property for the Ornstein-Uhlenbeck process imply easily that

$$P\left\{ \sup_{1 \le t \le T} \frac{W(t)}{\sqrt{t}} = \sup_{1 \le t \le T'} \frac{W(t)}{\sqrt{t}} \right\} = \frac{\log T}{\log T'} \qquad \text{for any} \quad T' \ge T > 1 \tag{5.22}$$

and thus

$$P\left\{ \sup_{1 \le t \le l} \frac{W(t)}{\sqrt{t}} \ne \sup_{A_l^2 \le t \le l} \frac{W(t)}{\sqrt{t}} \right\} = \frac{\log A_l^2}{\log l} = \frac{2}{c_l}. \tag{5.23}$$

Now setting

$$\xi_l^* = a_l \left(\sup_{A_l^2 \le t \le l} \frac{W(t)}{\sqrt{t}} - b_l \right)$$

we have by (5.23)

$$P\left(\xi_l \ne \xi_l^*\right) = 2/c_l$$

and thus

$$E(|\xi_l - \xi_l^*| \wedge 1) \le 2/c_l \le 2c_k/c_l \le 2(c_k/c_l)^{1/2}. \tag{5.24}$$

Hence to prove (5.21) it suffices to show that

$$E(|\xi_l^* - \xi_{k,l}| \wedge 1) \le (c_k/c_l)^{1/2}. \tag{5.25}$$

Now by $k \le A_l$ we have

$$|\xi_l^* - \xi_{k,l}| \le \frac{|W(k)|}{A_l} a_l$$

and thus

$$E|\xi_l^* - \xi_{k,l}|^2 \le \frac{k}{A_l^2} a_l^2 \le \frac{a_l^2}{A_l} \le \exp\left(-\sqrt{\log l}\right) \le \frac{1}{c_l} \le \frac{c_k}{c_l}$$

which implies (5.25) by the Cauchy-Schwarz inequality. Since the number of pairs (k, l) with $3 \le k < l < l_0$ is finite, (5.21) implies that

$$E(|\xi_l - \xi_{k,l}| \wedge 1) \le C(c_k/c_l)^{1/2} \le C'(\log_+ \log_+(c_l/c_k))^{-2} \qquad (3 \le k < l)$$

for some constants $C > 0$, $C' > 0$. Hence using Theorem 5 we get that

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=3}^{N} d_k I\left\{a_k \left(\sup_{1\le t\le k} \frac{W(t)}{\sqrt{t}} - b_k\right) < x\right\} = G(x) \qquad \text{a.s. for any } x \tag{5.26}$$

where

$$d_k \sim \frac{1}{2k \log k \sqrt{\log\log k}}, \qquad D_N \sim \sqrt{\log\log N} \tag{5.27}$$

by $d_k = \log(c_{k+1}/c_k)$ and simple calculations. (Note that $\xi_l, \xi_{k,l}, c_k$ are defined only for $l \ge k \ge 3$, but this does not cause any difficulty, as we observed in Section 4.) By the theorem of Zygmund mentioned after Theorem 1 the summation procedure belonging to the weights in (5.27) is equivalent to the summation procedure belonging to $d_k^* = 1/k \log k$, $D_N^* \sim \log\log N$. Hence we proved the following

Theorem J. *Let W be a Wiener process. Then*

$$\lim_{N\to\infty} \frac{1}{\log\log N} \sum_{k=3}^{N} \frac{1}{k \log k} I\left\{a_k \left(\sup_{1\le t\le k} \frac{W(t)}{\sqrt{t}} - b_k\right) < x\right\} = G(x) \quad \text{a.s. for any } x \tag{5.28}$$

where (a_n), (b_n) and $G(x)$ are defined by (5.18) and (5.19), respectively.

Using Theorem 6 instead of Theorem 5, the same procedure leads to

$$\lim_{T\to\infty} \frac{1}{\log\log T}\int_3^T \frac{1}{t\log t} I\left\{a_t\left(\sup_{1\le u\le t}\frac{W(u)}{\sqrt{u}} - b_t\right) < x\right\} dt = G(x) \quad \text{a.s. for any } x.$$

Using an a.s. invariance principle, it is easy to extend Theorem J for partial sums. Let (X_n) be a sequence of independent random variables with mean 0, variance 1 and uniformly bounded $(2+\delta)$-th moments for some $\delta > 0$; put $S_k = X_1 + \cdots + X_k$. Then one can define the sequence (X_n), together with a Wiener process W, on a suitable probability space such that

$$S_n - W(n) = O(n^{1/2-\eta}) \qquad \text{a.s.} \tag{5.29}$$

for some constant $\eta > 0$. (See e.g. Strassen, 1967.) The last relation easily implies

$$a_k\left(\max_{(\log k)^3\le i\le k}\frac{S_i}{\sqrt{i}} - \sup_{(\log k)^3\le t\le k}\frac{W(t)}{\sqrt{t}}\right) \to 0 \qquad \text{a.s.} \tag{5.30}$$

where (a_k) is defined by (5.18). (Cf. also Oodaira, 1976; Shorack, 1979.) Note that i and t in (5.30) are restricted to the interval $[(\log k)^3, k]$, but since the LIL implies

$$\sup_{1\le t\le (\log k)^3}\left|\frac{W(t)}{\sqrt{t}}\right| = O(\log\log\log k)^{1/2}) \qquad \text{a.s.,}$$

it follows that

$$a_k\left(\sup_{1\le t\le (\log k)^3}\frac{W(t)}{\sqrt{t}} - b_k\right) \to -\infty \qquad \text{a.s.} \tag{5.31}$$

and thus (5.28) remains valid if we extend the sup only for $(\log k)^3 \le t \le k$. Since changing the random variable ξ_k by $o(1)$ in (2.12) does not affect the validity of (2.12) (see e.g. Lacey and Philipp, 1990), (5.30) implies that (5.28) holds if $\sup_{1\le t\le k}(W(t)/\sqrt{t})$ is replaced by $\max_{(\log k)^3\le i\le k}(S_i/\sqrt{i})$. Finally, the last maximum can be replaced by $\max_{i\le k}(S_i/\sqrt{i})$, as it follows from the the analogue of (5.31) for the (X_n). Thus we proved the following result:

Theorem K. *Let (X_n) be a sequence of independent random variables with mean 0, variance 1 and uniformly bounded $(2+\delta)$-th moments for some $\delta > 0$. Put $S_k = \sum_{i \le k} X_i$. Then*

$$\lim_{N\to\infty} \frac{1}{\log\log N} \sum_{k=3}^{N} \frac{1}{k\log k} I\left\{ a_k \left(\max_{i\le k} \frac{S_i}{\sqrt{i}} - b_k \right) < x \right\} = G(x) \qquad \text{a.s for any } x$$

where (a_n), (b_n) and $G(x)$ are defined by (5.18) and (5.19), respectively.

In the case when the random variables X_n are i.i.d., the moment conditions in Theorem K can be weakened: in this case the theorem holds under assuming only

$$EX_1 = 0, \qquad EX_1^2 = 1, \qquad E\left(X_1^2 \log_+ \log_+ |X_1| \right) < \infty. \tag{5.32}$$

The proof is similar to the above, just in this case instead of (5.29) we use the a.s. invariance principle

$$S_n - W(n) = o\left(\sqrt{n} (\log\log n)^{-1/2} \right) \qquad \text{a.s.}$$

valid under (5.32) (see Einmahl, 1987). Observing also that the a.s. invariance principle (5.29) is actually valid for a large class of weakly dependent sequences (X_n) (see Philipp and Stout, 1975), it follows that Theorem K also holds in many dependent situations.

Acknowledgements

The authors are indebted to A. Chuprunov, L. Horváth, D. Khoshnevisan and W. Philipp for valuable conversations. Research supported by Hungarian National Foundation for Scientific Research, Grant T 29621.

References

Atlagh, M., 1993. Théorème central limite presque sûr et loi du logarithme itéré pour des sommes de variables aléatoires indépendantes. C. R. Acad. Sci. Paris Sér. I. 316, 929–933.

Atlagh, M., Weber, M., 1992. Un théorème central limite presque sûr relatif à des sous-suites. C. R. Acad. Sci. Paris Sér. I 315, 203–206.

Atlagh, M., Weber, M., 1996. Une nouvelle loi forte des grandes nombres. In: Bergelson, V., March, P., Rosenblatt, J. (Eds.), Convergence in Ergodic Theory and Probability, W. de Gruyter, Berlin, pp. 41–62.

Berkes, I., 1995. On the almost sure central limit theorem and domains of attraction. Probab. Theory Related Fields 102, 1–18.

Berkes, I., 1998. Results and problems related to the pointwise central limit theorem. In: Szyszkowicz, B. (Ed.), Asymptotic results in Probability and Statistics (a volume in honour of Miklós Csörgő), Elsevier, Amsterdam, pp. 59–96.

Berkes, I., Csáki, E., Csörgő, S., 1999. Almost sure limit theorems for the St. Petersburg game. Statist. Probab. Letters 45, 23–30.

Berkes, I., Dehling, H., 1993. Some limit theorems in log density. Ann. Probab. 21, 1640–1670.

Berkes, I., Dehling, H., 1994. On the almost sure central limit theorem for random variables with infinite variance. J. Theoret. Probab. 7, 667–680.

Berkes, I., Dehling, H., Móri, T.F., 1991. Counterexamples related to the a.s. central limit theorem. Studia Sci. Math. Hungar. 26, 153-164.

Bingham, N.H., Rogers, L.C.G., 1991. Summability methods and almost sure convergence. In: Bellow, A., Jones, R. (Eds.), Almost Everywhere Convergence II. Academic Press, New York, pp. 69–83.

Brosamler, G., 1988. An almost everywhere central limit theorem. Math. Proc. Cambridge Phil. Soc. 104, 561–574.

Chandrasekharan, K., Minakshisundaram, S., 1952. Typical Means. Oxford University Press.

Cheng, S., Peng, L., Qi, Y., 1998. Almost sure convergence in extreme value theory. Math. Nachr. 190, 43–50.

Csáki, E., Földes, A., 1995. On the logarithmic average of additive functionals. Statist. Probab. Letters 22, 261–268.

Csörgő, M., Horváth, L., 1992. Invariance principles for logarithmic averages. Math. Proc. Cambridge Phil. Soc. 112, 195–205.

Darling, D., Erdős, P., 1956. A limit theorem for the maximum of normalized sums of independent random variables. Duke Math. J. 23, 143–155.

Darling, D., Kac, M., 1957. On occupation times for Markoff processes. Trans. Amer. Math. Soc. 84, 444–458.

Denker, M., 1985. Asymptotic Distribution Theory in Nonparametric Statistics. F. Vieweg & Sohn, Braunschweig and Wiesbaden.

Dvoretzky, A., Erdős, P., 1951. Some problems on random walk in space. Proc. Second Berkeley Symp. Math. Statist. Prob., Univ. of California Press, pp. 353–367.

Dvoretzky, A., Kiefer, J., Wolfowitz, J., 1956. Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. Ann. Math. Statist. 27, 642–649.

Einmahl, U., 1987. Strong invariance principles for partial sums of independent random vectors. Ann. Probab. 15, 1419–1440.

Einmahl, U., 1989. The Darling-Erdős theorem for sums of i.i.d. random variables. Probab. Theory Related Fields 82, 241–257.

Fahrner, I., Stadtmüller, U., 1998. On almost sure max-limit theorems. Statist. Probab. Letters 37, 229–236.

Feller, W., 1943. Generalization of a probability limit theorem of Cramér. Trans. Amer. Math. Soc. 54, 361–372.

Fisher, A., 1989. A pathwise central limit theorem for random walks. Preprint.

Giné, E., Zinn, J., 1994. A remark on convergence in distribution of U-statistics. Ann. Probab. 22, 117–125.

Horváth, L., Khoshnevisan, D., 1995. Weight functions and pathwise local central limit theorems. Stochastic Process. Appl. 59, 105–123.

Ibragimov, I.A., 1996. On almost sure versions of limit theorems. (In Russian.) Dokl. Akad. Nauk 350, 301–303.

Ibragimov, I.A., Lifshits, M., 1998. On the convergence of generalized moments in almost sure central limit theorem. Statist. Probab. Letters 40, 343–351.

Ibragimov, I.A., Lifshits, M., 1999. On almost sure limit theorems. Th. Probab. Appl. 44, 254–272.

Kesten, H., Spitzer, F., 1979. A limit theorem related to a new class of self similar processes. Z. Wahrsch. Verw. Gebiete 50, 5–25.

Koroljuk, V.S., Borovskich, Yu.V., 1994. Theory of U-statistics. Kluwer, Dordrecht.

Lacey, M., Philipp, W., 1990. A note on the almost everywhere central limit theorem. Statist. Probab. Letters 9, 201–205.

Lifshits, M., 2000a. On the difference between CLT and ASCLT. (In Russian) Zapiski Seminarov POMI 260, 186–201.

Lifshits, M., 2000b. An almost sure limit theorem for sums of random vectors. Preprint.

Marcus, M., Rosen, J., 1995. Logarithmic averages for the local time of recurrent random walks and Levy processes. Stochastic Process. Appl. 59, 175–184.

Móri, T.F., 1993. On the strong law of large numbers for logarithmically weighted sums. Ann. Univ. Sci. Budapest. Sect. Math. 36, 35–46.

Oodaira, H., 1976. Some limit theorems for the maximum of normalized sums of weakly dependent random variables. Proc. Third Japan-USSR Symposium on Probability Theory. Lecture Notes Math. No. 550, Springer, Berlin. pp. 467–474.

Peligrad, M., Révész, P., 1991. On the almost sure central limit theorem. In: Bellow, A., Jones, R. (Eds.), Almost Everywhere Convergence II. Academic Press, New York, pp. 209–225.

Philipp, W., Stout, W., 1975. Almost sure invariance principles for partial sums of weakly dependent random variables. Memoirs of the AMS, No. 161.

Rodzik, B., Rychlik, Z., 1994. An almost sure central limit theorem for independent random variables. Ann. Inst. H. Poincaré 30, 1–11.

Rodzik, B., Rychlik, Z., 1996. On the central limit theorem for independent random variables with almost sure convergence. Probab. Math. Statist. 12, 299–309.

Schatte, P., 1988. On strong versions of the central limit theorem. Math. Nachr. 137, 249–256.

Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Shorack, G., 1979. Extension of the Darling and Erdös theorem on the maximum of normalized sums. Ann. Probab. 7, 1092–1096.

Strassen, V., 1967. Almost sure behavior of sums of independent random variables and martingales. Proc. Fifth Berkeley Symp. Math. Statist. Probab., Vol II, Part I, Univ. of California Press, pp. 315–343.

Weigl, A., 1989. Zwei Sätze über die Belegungszeit beim Random Walk. Diplomarbeit, TU Wien.