Available online at www.sciencedirect.com

SCIENCE DIRECT®

Statistics & Probability Letters 76 (2006) 280–290

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Almost sure versions of the Darling–Erdős theorem ☆

István Berkes[a,*], Michel Weber[b]

[a]Department of Statistics, Technical University Graz, Steyrergasse 17/IV, A-8010 Graz, Austria
[b]Mathématique (I.R.M.A.), Université Louis-Pasteur et C.N.R.S., 7, rue René Descartes, 67084 Strasbourg, Cedex, France

Received 2 November 2004; received in revised form 27 June 2005
Available online 24 August 2005

Abstract

We prove a.s. limit theorems corresponding to the classical Darling–Erdős theorem for the maxima of normalized partial sums of i.i.d. random variables. Our results yield the analogue of the a.s. central limit theorem for the Darling–Erdős max functional and its variants. Unlike in standard a.s. central limit theory, our theorems involve nonlogarithmic averages.

MSC: primary 60F15; secondary 60F05

Keywords: Normalized maxima; Extremal distribution; Almost sure limit theorem; Weighted averages

1. Introduction

Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with $EX_1 = 0, EX_1^2 = 1$ and let $S_n = X_1 + \cdots + X_n$. By the classical theorem of Darling and Erdős (1956) we have, under slight additional moment conditions,

$$\alpha_k \left(\max_{1 \leqslant j \leqslant k} \frac{S_j}{\sqrt{j}} - \beta_k \right) \xrightarrow{\mathscr{D}} H, \tag{1}$$

☆Research supported by Hungarian National Foundation for Scientific Research, Grants T 043037 and T 037886.
*Corresponding author.
E-mail addresses: berkes@tugraz.at (I. Berkes), weber@math.u-strasbg.fr (M. Weber).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.08.007

where

$$\alpha_k = (2\log\log k)^{1/2}, \quad \beta_k = (2\log\log k)^{1/2} + \frac{\log\log\log k - \log 4\pi}{2(2\log\log k)^{1/2}} \quad (k \geqslant 3) \tag{2}$$

and H is the distribution with distribution function $\mathrm{e}^{-\mathrm{e}^{-x}}$. Darling and Erdős (1956) assumed $E|X_1|^3 < \infty$; Oodaira (1976) and Shorack (1979) weakened this to $E|X_1|^{2+\delta} < \infty$. An optimal condition was given by Einmahl (1989).

The purpose of this paper is to study the asymptotic behavior of the more general functional

$$M_k^{(f)} = \max_{k/f(k) \leqslant j \leqslant k} \frac{S_j}{\sqrt{j}},$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a nondecreasing function with $1 \leqslant f(x) \leqslant x$. For $f(x) \to \infty$ the limit distribution of $M_k^{(f)}$ is again the extremal distribution $\mathrm{e}^{-\mathrm{e}^{-x}}$. Indeed, relation (30) at the end of our paper shows that

$$a_k^{(f)}(M_k^{(f)} - b_k^{(f)}) \xrightarrow{\mathscr{D}} H, \tag{3}$$

where

$$a_k^{(f)} = (2\log\log f(k))^{1/2}$$
$$b_k^{(f)} = (2\log\log f(k))^{1/2} + \frac{\log\log\log f(k) - \log 4\pi}{2(2\log\log f(k))^{1/2}} \tag{4}$$

with the same H as in (1). In the case $f(x) = c \ \ (1 \leqslant c < \infty)$ we get a different behavior: in this case $M_k^{(f)}$ remains bounded in probability and it follows easily from Donsker's theorem that

$$M_k^{(f)} \xrightarrow{\mathscr{D}} G_c, \tag{5}$$

where G_c is the distribution of $\sup_{1/c \leqslant t \leqslant 1} W(t)/\sqrt{t}$, where W is a Wiener process. To see the connection between (3) and (5), let us write (5) in the equivalent form

$$\alpha_c(M_k^{(f)} - \beta_c) \xrightarrow{\mathscr{D}} H_c \quad \text{as } k \to \infty,$$

where $H_c(x) = G_c(\beta_c + x/\alpha_c)$. The last relation shows that (3) remains valid in the extreme case $f(x) = c$, but the limit H should be replaced by H_c. Using scaling and the Darling–Erdős theorem for the Wiener process, it is easy to see that $\lim_{c\to\infty} H_c(x) = H(x)$ for all x. Thus, relation (3) is formally the limit of (5) as $c \to \infty$.

The previous remarks describe the distributional behavior of $M_k^{(f)}$ completely. The purpose of the present paper is to prove almost sure versions of these results in the spirit of the a.s. central limit theory. In the case $f(x) = c$ an almost sure version of (5) was obtained by Antonini and Weber (2004), who showed that

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{ \max_{k/c \leqslant j \leqslant k} \frac{S_j}{\sqrt{j}} \leqslant x \right\} = G_c(x) \ \text{ a.s. for any } x \in \mathbb{R}. \tag{6}$$

On the other hand, in the case $f(x)=x$ it was proved by Berkes and Csáki (2001) that under the additional moment condition $E|X_1|^{2+\delta}<\infty$ $(\delta>0)$ we have

$$\lim_{N\to\infty}\frac{1}{\log\log N}\sum_{k=3}^{N}\frac{1}{k\log k}I\left\{\alpha_k\left(\max_{1\leqslant j\leqslant k}\frac{S_j}{\sqrt{j}}-\beta_k\right)\leqslant x\right\}=\mathrm{e}^{-\mathrm{e}^{-x}}\quad \text{a.s. for any } x\in\mathbb{R}. \tag{7}$$

Note that in (6) and (7) we have different averaging processes; this is due to the fact that the dependence of the sequence $M_k^{(f)}$ for $f(x)=x$ is much stronger than for $f(x)=c$ (as reflected by the covariance estimates in Antonini and Weber (2004) and Berkes and Csáki (2001)) and thus the indicators in (7) require a stronger averaging method to converge a.s. Since (3) connects the extremal cases (1) and (5), it is natural to expect that the a.s. version of (3) will involve averaging processes which change continuously from logarithmic to log log average as f changes between $f(x)=c$ and $f(x)=x$. This is indeed the case, as the main results of our paper will show. We will prove:

Theorem 1. *Let $X_1,X_2,\dots$ be i.i.d. random variables with $EX_1=0$, $EX_1^2=1$ and $EX_1^2(\log_+|X_1|)^\alpha<\infty$, $\alpha>1$. Let $S_n=X_1+\cdots+X_n$ and let $f:\mathbb{R}^+\to\mathbb{R}^+$ be a nondecreasing function with $1\leqslant f(x)\leqslant x^{1/4}$, $f(x)\to+\infty$. Let*

$$Z_k=a_k^{(f)}(M_k^{(f)}-b_k^{(f)}) \tag{8}$$

with $a_k^{(f)},b_k^{(f)}$ defined in (4) and let (c_n) be a positive nondecreasing sequence satisfying $c_{n+1}/c_n=\mathrm{O}(1)$ and

$$c_{n/f(n)^3}\geqslant Ac_n,\quad c_n/n^{1/6}\quad \text{is nonincreasing} \tag{9}$$

for some constant $0<A<1$. (Here, and in the sequel, c_j is meant as $c_{[j]}$ for nonintegral j.) Then for any $x\in\mathbb{R}$ we have

$$\lim_{N\to\infty}\frac{1}{D_N}\sum_{k=1}^{N}d_kI\{Z_k\leqslant x\}=\mathrm{e}^{-\mathrm{e}^{-x}}\qquad \text{a.s.} \tag{10}$$

where

$$d_n=\log(c_{n+1}/c_n),\quad D_n=\log c_{n+1}. \tag{11}$$

Theorem 2. *Let $X_1,X_2,\dots$ be independent r.v.'s with mean 0, variance 1 and uniformly bounded $(2+\delta)$th moments, where $\delta>0$. Let $S_n=X_1+\cdots+X_n$ and let $f:\mathbb{R}^+\to\mathbb{R}^+$ be a nondecreasing function with $1\leqslant f(x)\leqslant x$, $f(x)\to+\infty$. Then (10) holds with $d_k=1/(k\log k)$, $D_N=\log\log N$.*

Theorem 1 covers the case of "small" f, while Theorem 2 covers the remaining cases by showing that (10) always holds with log log averages. The novel feature of our theorems is the appearance of nonstandard (i.e. nonlogarithmic) weights in (10); these are given indirectly through the function f. In Section 2 we will give a detailed analysis of the weight sequences in our theorems.

The proofs of Theorems 1 and 2 will use an invariance argument: we will first show the statement for the Wiener process and then obtain the general i.i.d. case by strong approximation. Our main tool will be a general version of the a.s. central limit theorem for nonlinear functionals, proved in Berkes and Csáki (2001). A consequence of the invariance method used in the proof is that Theorems 1 and 2 remain valid for any (dependent) sequence (X_n) of random variables for which there exists a Wiener process $\{W(t),\ t\geqslant 0\}$ such that

$$\sum_{k=1}^{n} X_k = W(n) + \mathrm{O}(n^{1/2}(\log n)^{-\beta}) \quad \text{a.s.}$$

with some constant $\beta>\frac{1}{2}$. The class of such sequences (X_n) includes various types of mixing sequences, martingale difference sequences, Markov chains, Gaussian processes, etc; see Philipp and Stout (1975).

The results of our paper concern maxima of normalized partial sums of i.i.d. random variables and the limit in (10) is an extremal distribution. It is worth comparing our results with known a.s. limit theorems for extremal statistics of i.i.d. random variables. Fahrner and Stadtmüller (1998) and Cheng et al. (1998) proved independently that if $X_1, X_2, \dots$ are i.i.d. random variables, $M_n = \max(X_1, \dots, X_n)$ and

$$(M_n - a_n)/b_n \xrightarrow{\mathscr{D}} G$$

for some real sequences a_n, b_n and a nondegenerate distribution function G, then we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{\frac{M_k - a_k}{b_k} \leqslant x\right\} = G(x) \quad \text{a.s. for any } x \in \mathbb{R}.$$

For additional strong limit theorems for the partial sums

$$\sum_{k=1}^{n} \frac{1}{k} f\left(\frac{M_k - a_k}{b_k}\right)$$

see e.g. Berkes and Horváth (2001), Fahrner (2001). Despite their formal similarity, there is a fundamental difference between these results and the results of our paper: the Darling–Erdős type functional $M_k^{(f)}$ contains the maxima of the strongly dependent random variables $S_k/\sqrt{k}$ and this leads to a.s. limit theorems involving an essentially different, nonstandard averaging process described in Theorem 1.

2. Analysis of the averaging sequences in Theorems 1 and 2

Theorem 1 above provides a large class of sequences (c_n) and thus a large class of averaging methods for which the limit relation (10) holds. To clarify the meaning of the theorem, we recall a few facts from analysis. 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

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

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 $D_n^* = \mathrm{O}(D_n)$ 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 $D_n^\alpha \leqslant D_n^* \leqslant D_n^\beta$ 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 $D_n^* = \mathrm{O}(D_n^\varepsilon)$ for all $\varepsilon>0$ then the summation method defined by $\mathbf{D}^*$ is strictly stronger than the method defined by $\mathbf{D}$. For example, logarithmic summation defined by $d_k = 1/k$ is stronger than Cesàro (or $(C,1)$) summation defined by $d_k = 1$ and is weaker than log log averaging defined by $d_k = 1/(k \log k)$; on the other hand, all summation procedures defined by $d_k = (\log k)^\alpha/k,\ \ \alpha>-1$ are equivalent to logarithmic summation.

The above remarks show that the faster the sequence D_n in (10) grows, the stronger the limit theorem (10) becomes. Hence, given a function f satisfying the assumptions of Theorem 1, one should choose the sequence D_n in (10) as large as possible to optimize Theorem 1. In what follows, we will give a few examples for evaluating the weight sequence D_n in Theorem 1. Observe that $c_{n+1}/c_n = \mathrm{O}(1)$ implies $D_n - D_{n-1} = \mathrm{O}(1)$ and thus the first relation of (9) can be written equivalently as

$$D_n - D_{n/f(n)^3} = \mathrm{O}(1). \tag{12}$$

Examples. (a) If $f(x) = \mathrm{O}(\mathrm{e}^{(\log x)^\alpha}),\ 0<\alpha<1,$ then (10) holds with $D_n = \log n$, i.e. with log averages.

(b) If $f(x) = x^\beta,\ 0<\beta\leqslant\frac{1}{4}$, then (10) holds with $D_n = \log\log n$, and this is the optimal (i.e. largest) choice, since (12) implies $D_n = \mathrm{O}(\log\log n)$. Thus in this case Theorem 1 yields log log averages.

(c) If $f(x) = \mathrm{e}^{\log x/(\log\log x)^{2\omega(x)}}$ where $\omega(x)$ is nondecreasing and tends to $+\infty$ so slowly that $\omega'(x)\leqslant 1/(x\log x)$, then (10) holds with $D_n = (\log\log n)^{\omega(n)}$. The corresponding averaging process in (10) lies strictly between the log and log log averages.

The above examples show that the transition from log to log log averages in Theorem 1 takes place in a very narrow strip, namely for functions $f(x) = \mathrm{e}^{\log x/\psi(x)}$, where $\psi(x)$ tends to $+\infty$ very slowly. For the functions in (a) we still have log averages, while in (b) we have log log averages; (c) describes an intermediate situation. Actually, (b) describes the worst possible case: as Theorem 2 shows, relation (10) is always valid with log log averages. (Accordingly, for any function f in Theorem 1, relation (12) is satisfied with $D_n = \log\log n$.)

To verify the examples, we first show that in case (a) relation (9) holds with $c_n = \exp((\log n)^{1-\alpha})$, or, what is the same, relation (12) holds with $D_n = (\log(n+1))^{1-\alpha}$. (Actually, we can work with $D_n = (\log n)^{1-\alpha}$, since by Zygmund's theorem quoted above, this means no difference in (10).) By the mean value theorem we get for large n

$$D_n - D_{n/f(n)^3} \leqslant (1-\alpha)\left(\log\frac{n}{f(n)^3}\right)^{-\alpha} 3\log f(n) \leqslant (1-\alpha)(\log\sqrt{n})^{-\alpha} 3(\log n)^\alpha = \mathrm{O}(1).$$

Thus Theorem 1 holds with $D_n = (\log n)^{1-\alpha}$ in (10) and hence by Zygmund's theorem, relation (10) holds also with log averages. In case (b), (12) becomes

$$D_n - D_{n^\gamma} = \mathrm{O}(1)$$

with $\gamma = 1 - 3\beta$ and by induction it is easy to see that this implies $D_n \leqslant C \log\log n$, provided C is large enough. Finally to get (c) let us note that for the function f in (c) and $D_n = (\log\log n)^{\omega(n)}$, the left side of (12) can be bounded by $I_1 + I_2$ where

$$I_1 = (\log\log n)^{\omega(n)} - \left(\log\log\frac{n}{f(n)^3}\right)^{\omega(n)},$$

$$I_2 = \left(\log\log\frac{n}{f(n)^3}\right)^{\omega(n)} - \left(\log\log\frac{n}{f(n)^3}\right)^{\omega(n/f(n)^3)}.$$

Clearly

$$\left|\log\log n - \log\log\frac{n}{f(n)^3}\right| = \left|\log\left(1 - \frac{3\log f(n)}{\log n}\right)\right| = \mathrm{O}\left(\frac{\log f(n)}{\log n}\right) \tag{13}$$

and thus by the mean value theorem and the definition of f we get

$$I_1 = \mathrm{O}\left(\frac{\log f(n)}{\log n}\omega(n)(\log\log n)^{\omega(n)-1}\right) = \mathrm{O}(1).$$

Another application of the mean value theorem yields

$$I_2 \leqslant \left|\omega(n) - \omega(n/f(n)^3)\right|(\log\log n)^{\omega(n)}\log\log\log n$$

and here we have, using $\omega'(x) \leqslant 1/(x\log x)$ and (13),

$$\left|\omega(n) - \omega(n/f(n)^3)\right| \leqslant \int_{n/f(n)^3}^{n} \frac{1}{x\log x}\,\mathrm{d}x = \log\log n - \log\log\frac{n}{f(n)^3} = \mathrm{O}\left(\frac{\log f(n)}{\log n}\right).$$

Thus

$$\begin{aligned} I_2 &\leqslant \mathrm{O}\left(\frac{\log f(n)}{\log n}(\log\log n)^{\omega(n)}\log\log\log n\right) \\ &= \mathrm{O}\left(\frac{1}{(\log\log n)^{2\omega(n)}}(\log\log n)^{\omega(n)}\log\log\log n\right) \\ &= \mathrm{O}(1) \end{aligned}$$

completing the proof of Example (c).

3. Proof of the theorems

We will deduce Theorems 1 and 2 from a general version of the a.s. central limit theorem for nonlinear functionals ("universal ASCLT"), proved in Berkes and Csáki (2001). We formulate here a special case of Theorem 5 of Berkes and Csáki (2001) for the Wiener process, which will suffice for our purposes.

Theorem 3. *Let W be a Wiener process and let $\xi_1, \xi_2, \ldots$ be random variables such that ξ_k is measurable with respect to $\sigma\{W(t), 0 \leqslant t \leqslant k\}$. Assume that $\xi_k \xrightarrow{\mathscr{D}} G$ for some distribution function G. Assume further that for each $1 \leqslant k < l$ there exists a random variable $\xi_{k,l}$ measurable with respect* *to* $\sigma\{W(t')-W(t): k \leqslant t \leqslant t' \leqslant l\}$ *such that*

$$E(|\xi_l - \xi_{k,l}| \wedge 1) \leqslant C(c_k/c_l)^\beta \tag{14}$$

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

$$d_k = \log(c_{k+1}/c_k), \quad D_n = \sum_{k \leqslant n} d_k. \tag{15}$$

Then

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k \leqslant N} d_k I\{\xi_k \leqslant x\} = G(x) \quad \text{a.s. for any } x \in C_G, \tag{16}$$

where C_G *denotes the set of continuity points of* G.

We turn now to the proof of the theorems. As a first step, we prove the results in the Wiener case, i.e. when the definition of $M_k^{(f)}$ is replaced by

$$M_k^{(f)} = \sup_{k/f(k) \leqslant t \leqslant k} \frac{W(t)}{\sqrt{t}},$$

where W is a Wiener process.

Assume first that f and c_n satisfy the assumptions of Theorem 1 and define

$$Z_l^* = a_l^{(f)}\left(\sup_{l/f(l) \leqslant t \leqslant l} \frac{W(t)}{\sqrt{t}} - b_l^{(f)}\right),$$

$$Z_{k,l}^* = \begin{cases} a_l^{(f)}\left(\sup_{l/f(l) \leqslant t \leqslant l} \frac{W(t)-W(k)}{\sqrt{t}} - b_l^{(f)}\right) & \text{if } k \leqslant l/f(l) \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

We claim that

$$E(|Z_l^* - Z_{k,l}^*| \wedge 1) \leqslant B(c_k/c_l) \quad (k \leqslant l) \tag{18}$$

with some constant B. Indeed, if $k \leqslant l/f(l)^3$, then

$$E|Z_l^* - Z_{k,l}^*| \leqslant a_l^{(f)} \frac{E|W(k)|}{\sqrt{l/f(l)}} \leqslant a_l^{(f)} \sqrt{f(l)}\sqrt{k/l} \leqslant f(l)\sqrt{k/l} \leqslant (k/l)^{1/6} \leqslant c_k/c_l$$

for $l \geqslant l_0$ by (4) and the second relation of (9). If $k > l/f(l)^3$, then by the first relation of (9) we have

$$\frac{c_k}{c_l} \geqslant \frac{c_{l/f(l)^3}}{c_l} \geqslant A$$

and therefore (18) will hold in this case, too. Thus (18) is verified for $l \geqslant l_0$; increasing the constant B if necessary, it will hold for all $l \geqslant 1$.

Let T be a r.v. with distribution function $\mathrm{e}^{-\mathrm{e}^{-x}}$. The Darling–Erdős theorem for the Wiener process states that

$$\alpha_n\left(\sup_{1\leqslant t\leqslant n}\frac{W(t)}{\sqrt{t}}-\beta_n\right)\xrightarrow{\mathscr{D}}T, \tag{19}$$

where α_n, β_n are defined in (2). By the scaling property of the Wiener process we have

$$\sup_{k/f(k)\leqslant t\leqslant k}\frac{W(t)}{\sqrt{t}}\overset{\mathscr{D}}{=}\sup_{1\leqslant t\leqslant f(k)}\frac{W(t)}{\sqrt{t}}$$

and since $\alpha_{f(n)}=a_n^{(f)}$, $\beta_{f(n)}=b_n^{(f)}$, relation (19) implies $Z_n^*\xrightarrow{\mathscr{D}}T$. Hence applying Theorem 3 with $\xi_k=Z_k^*$, $\xi_{k,l}=Z_{k,l}^*$ we get Theorem 1 for the Wiener process.

Assume now that f satisfies the conditions of Theorem 2 and introduce the quantities $\widehat{Z}_l$ and $\widehat{Z}_{k,l}$ defined similarly as Z_l^* and $Z_{k,l}^*$ in (17), just with $l/f(l)$ replaced everywhere by $(l/f(l))\vee A_l^2$, where

$$A_l=\exp(\log l/\exp(\sqrt{\log\log l}))\quad(l\geqslant 3).$$

We show that

$$E(|Z_l^*-\widehat{Z}_{k,l}|\wedge 1)\leqslant D(c_k/c_l)\quad(k\leqslant l) \tag{20}$$

with some constant D, where

$$c_n=\exp\left(\tfrac{1}{2}\sqrt{\log\log n}\right). \tag{21}$$

To prove (20) we first observe that the stationarity and Markov property of the Ornstein–Uhlenbeck process $\mathrm{e}^{-t/2}W(\mathrm{e}^t)$ imply

$$P\left\{\sup_{1\leqslant t\leqslant T}\frac{W(t)}{\sqrt{t}}=\sup_{1\leqslant t\leqslant T'}\frac{W(t)}{\sqrt{t}}\right\}=\frac{\log T}{\log T'}\qquad\text{for any } T'\geqslant T>1. \tag{22}$$

Now $Z_l^*\neq\widehat{Z}_l$ can hold only if $l/f(l)<A_l^2$ and the sup of $W(t)/\sqrt{t}$ over the interval $[l/f(l),l]$ is reached somewhere in $[l/f(l),A_l^2]$, which implies that the sup of $W(t)/\sqrt{t}$ over $[1,l]$ is reached somewhere in $[1,A_l^2]$. Thus by (22) we have

$$P(Z_l^*\neq\widehat{Z}_l)\leqslant\frac{\log A_l^2}{\log l}=\frac{2}{c_l^2}$$

and thus to prove (20) it suffices to show that

$$E(|\widehat{Z}_l-\widehat{Z}_{k,l}|\wedge 1)\leqslant c_k/c_l\quad(k\leqslant l). \tag{23}$$

Now if $k\leqslant A_l$, then $\widehat{Z}_{k,l}$ is defined by the upper line of the definition (concerning the case $k\leqslant l/f(l)\vee A_l^2$) and thus

$$\begin{aligned}E|\widehat{Z}_l-\widehat{Z}_{k,l}|&\leqslant a_l^{(f)}E|W(k)|/A_l\leqslant a_l^{(f)}\sqrt{k/A_l^2}\leqslant a_l^{(f)}/\sqrt{A_l}\\&\leqslant(2\log\log l)^{1/2}/\sqrt{A_l}\leqslant\exp(-\sqrt{\log l})\leqslant 1/c_l\leqslant c_k/c_l,\end{aligned}$$

for $l \geqslant l_0$, verifying (23) and thus (20). If $k > A_l$, then

$$\begin{aligned} c_k &\geqslant \exp\left(\tfrac{1}{2}\sqrt{\log\log A_l}\right) \\ &= \exp\left(\tfrac{1}{2}(\log\log l - \sqrt{\log\log l})^{1/2}\right) \geqslant \exp\left(\tfrac{1}{2}(\sqrt{\log\log l} - 1)\right) \geqslant \frac{c_l}{3} \end{aligned}$$

and therefore (20) holds in this case, too, provided $D \geqslant 3$. We thus verified (20) for $l \geqslant l_0$; as in the previous case, the inequality will hold for all $l \geqslant 1$ if we suitably increase D. Applying Theorem 3 with $\xi_k = Z_k^*$, $\xi_{k,l} = \widehat{Z}_{k,l}$ we get the validity of (10)–(11) in the Wiener case with the sequence c_n defined by (21), i.e. we showed that

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=3}^{N} d_k I\left\{ a_k^{(f)} \left(\sup_{k/f(k)\leqslant t\leqslant k} \frac{W(t)}{\sqrt{t}} - b_k^{(f)} \right) \leqslant x \right\} = \mathrm{e}^{-\mathrm{e}^{-x}} \quad \text{a.s. for all } x. \tag{24}$$

Here

$$D_n = \log c_{n+1} \sim \tfrac{1}{2}(\log\log n)^{1/2}$$

and thus Zygmund's theorem implies that this averaging procedure is equivalent to log log averaging. This completes the proof of Theorem 2 in the Wiener case.

Let now the sequence (X_n) and the function f satisfy the conditions of Theorem 1 or Theorem 2. Using Theorem 2 of Einmahl (1987) in the case of Theorem 1 and Theorem 4.4 of Strassen (1967) in the case of Theorem 2, it follows that one can define the sequence (X_n), together with a Wiener process W, on a suitable probability space such that

$$S_n - W(n) = \mathrm{O}(n^{1/2}(\log n)^{-\beta}) \quad \text{a.s.} \tag{25}$$

for some $\beta > \frac{1}{2}$. Letting $T_k = (k/f(k)) \vee \log f(k)$, the last relation clearly implies

$$a_k^{(f)} \left(\max_{T_k \leqslant i \leqslant k} \frac{S_i}{\sqrt{i}} - \max_{T_k \leqslant i \leqslant k} \frac{W(i)}{\sqrt{i}} \right) \to 0 \quad \text{a.s.} \tag{26}$$

From well-known properties of the Wiener process (see e.g. Csörgő and Révész, 1981) it follows that

$$\sup_{n\leqslant t\leqslant n+1} \left| \frac{W(t)}{\sqrt{t}} - \frac{W(n)}{\sqrt{n}} \right| = \mathrm{O}\left(\frac{\log n}{\sqrt{n}}\right) \quad \text{a.s.}$$

and thus from (26) we can infer

$$a_k^{(f)} \left(\max_{T_k \leqslant i \leqslant k} \frac{S_i}{\sqrt{i}} - \sup_{T_k \leqslant t \leqslant k} \frac{W(t)}{\sqrt{t}} \right) \to 0 \quad \text{a.s.} \tag{27}$$

On the other hand, the LIL implies

$$\sup_{1\leqslant t\leqslant \log f(k)} \left| \frac{W(t)}{\sqrt{t}} \right| = \mathrm{O}((\log\log\log f(k))^{1/2}) \quad \text{a.s.}$$

and thus

$$a_k^{(f)}\left(\sup_{1\leqslant t\leqslant \log f(k)} \frac{W(t)}{\sqrt{t}} - b_k^{(f)}\right) \to -\infty \quad \text{a.s.} \tag{28}$$

Hence (24) remains valid if we extend the sup only for $T_k \leqslant t \leqslant k$. Also, an elementary argument shows that changing the r.v. in the indicator in (24) by o(1) does not effect the validity of (24), and thus in view of (27) it follows that (24) holds if $\sup_{k/f(k)\leqslant t\leqslant k}(W(t)/\sqrt{t})$ is replaced by $\max_{T_k\leqslant i\leqslant k}(S_i/\sqrt{i})$. Finally, the last maximum can be replaced by

$$\max_{k/f(k)\leqslant i\leqslant k}(S_i/\sqrt{i}),$$

as it follows from the analogue of (28) for the (X_n). This completes the proof Theorems 1 and 2.

In conclusion we note that, as we proved above,

$$Z_k^* = a_k^{(f)}\left(\sup_{k/f(k)\leqslant t\leqslant k} \frac{W(t)}{\sqrt{t}} - b_k^{(f)}\right) \xrightarrow{\mathscr{D}} T, \tag{29}$$

where T is a r.v. with distribution function $\mathrm{e}^{-\mathrm{e}^{-x}}$. Repeating the invariance argument in the previous paragraph, we get from (29)

$$a_k^{(f)}\left(\max_{k/f(k)\leqslant i\leqslant k} \frac{S_i}{\sqrt{i}} - b_k^{(f)}\right) \xrightarrow{\mathscr{D}} T \tag{30}$$

as claimed in the Introduction.

References

Antonini, R.G., Weber, M., 2004. The intersective ASCLT. Stochastic Anal. Appl. 22, 1009–1025.

Berkes, I., Csáki, E., 2001. A universal result in almost sure central limit theory. Stoch. Proc. Appl. 94, 105–134.

Berkes, I., Horváth, L., 2001. The logarithmic average of sample extremes is asymptotically normal. Stoch. Proc. Appl. 91, 77–98.

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

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

Csörgő, M., Révész, P., 1981. Strong approximations in probability and statistics. Academic Press, New York.

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

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., 2001. A strong invariance principle for the logarithmic average of sample maxima. Stoch. Proc. Appl. 93, 317–337.

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

Oodaira, H., 1976. Some limit theorems for the maximum of normalized sums of weakly dependent random variables. Proceedings of Third Japan-USSR Symposium on Probability Theory. Lecture Notes in Mathematics, vol. 550, Springer, Berlin, pp. 467–474.

Philipp, W., Stout, W., 1975. Almost sure invariance principles for partial sums of weakly dependent random variables. Memoirs of the American Mathematical Society, vol. 161.

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. In: Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, vol. II, Part I, University of California Press, Berkeley, CA, pp. 315–343.