

Optimal Averaging Procedures in Almost Sure Central Limit Theory

Siegfried Hörmann¹

Abstract

Let X_1, X_2, \dots be i.i.d. random variables with $EX_1 = 0, EX_1^2 = 1, S_n = X_1 + \dots + X_n$ and let (d_k) be a positive numerical sequence. We investigate the a.s. convergence of the averages

$$\frac{1}{D_N} \sum_{k=1}^N d_k I\{S_k/\sqrt{k} \leq x\},$$

where $D_N = \sum_{k=1}^N d_k$. In the case of $d_k = 1/k$ we have logarithmic means and by the almost sure central limit theorem the above averages converge a.s. to $\Phi(x)$, the standard normal distribution function. It is also known that the analogous convergence relation fails for $d_k = 1$ (ordinary averages). In this paper we give a fairly complete solution of the problem for which weight sequences the above convergence relation and its refinements hold.

1 Introduction

1.1 The almost sure central limit theorem

One of the most frequently investigated topics in classical probability theory is the fluctuation of the partial sums S_k of i.i.d. random variables X_1, X_2, \dots . In case of $EX_1 = 0$ and $EX_1^2 = 1$ the central limit theorem (CLT) yields that

$$P\{k^{-1/2}S_k \leq x\} \longrightarrow \Phi(x) \quad (k \rightarrow \infty).$$

Hence one might expect that the average time the path of the process $(k^{-1/2}S_k)_{k \geq 1}$ spends below $x \in \mathbb{R}$ is asymptotically $\Phi(x)$, i.e.

$$\frac{1}{N} \sum_{1 \leq k \leq N} I\{k^{-1/2}S_k \leq x\} \rightarrow \Phi(x) \quad \text{a.s.} \quad (N \rightarrow \infty). \quad (1.1)$$

Of course, the random variables $I\{k^{-1/2}S_k \leq x\}$ are not independent and have also no other structural property which would allow the application of the law of large numbers

¹Department of Statistics, Steyrergasse 17/IV, University of Technology Graz, Austria; shoermann@tugraz.at

(LLN). And indeed (1.1) is not true! The easiest way to see this is to observe that by Lévy's arc-sine law we have

$$\frac{1}{N} \sum_{1 \leq k \leq N} I\{S_k \leq 0\} \xrightarrow{\mathcal{L}} T, \quad (1.2)$$

where T is the arc-sine distribution with density

$$a(x) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}} \quad 0 < x < 1.$$

Note that $a(x)$ takes its minimum at $x = 1/2$ and becomes infinite at 0 and 1. Thus for fixed large N the least likely value of the fraction of time the path $(k^{-1/2}S_k)_{k=1}^N$ spends on the positive half axis is $1/2$, and it is much more likely that the path remains most of the time either below or above the x -axis rather than to level off to an equilibrium.

A possible way to overcome the strong dependence of the random variables $I\{S_k \leq 0\}$ has been shown by Erdős and Hunt (1953). They proved that if X_1 has a continuous and symmetric distribution function then

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{k=1}^N \frac{1}{k} I\{S_k \leq 0\} = 1/2 \quad \text{a.s.} \quad (1.3)$$

The difference between (1.2) and (1.3) lies in the summation method. In the latter we used logarithmic means instead of the common Cesàro means. This result is a special case of the so called *almost sure central limit theorem* (ASCLT) proved first by Brosamler (1988) and Schatte (1988). (Actually, a very similar result has been already stated by Lévy in 1937 (cf. Lévy, 2003), but without a proof). In order to formulate this result, let (Ω, \mathcal{F}, P) be the probability space on which the sequence (X_n) is defined.

Theorem A *Let X_1, X_2, \dots be a sequence of i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$ and define $S_k = X_1 + \dots + X_k$. Then there is a set $N \in \mathcal{F}$ with $P(N) = 0$ such that for any Borel set $A \subset \mathbb{R}$ with $\lambda(\partial A) = 0$*

$$\frac{1}{\log N} \sum_{k=1}^N \frac{1}{k} I\{k^{-1/2}S_k \in A\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_A e^{-t^2/2} dt \quad \text{on } \Omega \setminus N. \quad (1.4)$$

Brosamler and Schatte proved Theorem A assuming $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$; for a proof assuming only finite second moments, see Lacey and Philipp (1990). In the past 15 years, a wide literature dealt with various extensions of the ASCLT; for a survey we refer to Berkes (1998).

The crucial new feature of the a.s. central limit theorem (1.4) is its pathwise character: the limit of the left hand side of (1.4) can be computed for any fixed $\omega \in \Omega$ knowing only the values of X_n , $n = 1, 2, \dots$ at ω . Hence the ASCLT is often called *pathwise central limit theorem*.

The following lemma will be the basis for a heuristic proof of Theorem A.

Lemma 1 *Let $(\mathbb{W}(t))_{t \geq 0}$ be a standard Brownian motion process and let f be a continuous function. Then*

$$\frac{1}{\log T} \int_1^T \frac{1}{t} f(t^{-1/2} \mathbb{W}(t)) dt \rightarrow Ef(\mathbb{W}(1)) \quad \text{a.s.} \tag{1.5}$$

Proof. By parameter transformation we get

$$\frac{1}{\log T} \int_1^T \frac{1}{t} f(t^{-1/2} \mathbb{W}(t)) dt = \frac{1}{\log T} \int_0^{\log T} f(Z(t)) dt,$$

where $Z(t) = e^{-t/2} \mathbb{W}(e^t)$ is (the stationary) *Ornstein-Uhlenbeck process*. Hence the result follows from the ergodic theorem. □

Using Lemma 1, Theorem A can be deduced now by an invariance argument. Assume that f is uniformly continuous. Replacing

$$t^{-1} f(t^{-1/2} \mathbb{W}(t)) \quad \text{by} \quad k^{-1} f(k^{-1/2} \mathbb{W}(k)) \quad \text{if } t \in [k, k + 1)$$

does not harm the limiting behavior in (1.5). Now assuming slightly more than two moments for X_1 , by an almost sure invariance principle (cf. Csörgő and Révész, 1981: 108) we can redefine \mathbb{W} and X_1, X_2, \dots jointly on a new probability space in such a way that

$$k^{-1/2} \mathbb{W}(k) - k^{-1/2} S_k = o(1) \quad \text{a.s.}$$

Since f is uniformly continuous, it follows that

$$f(k^{-1/2} \mathbb{W}(k)) - f(k^{-1/2} S_k) \rightarrow 0 \quad \text{a.s.} \quad (k \rightarrow \infty).$$

Finally, it is not hard to see that by an approximation argument the result of Lemma 1 extends to indicator functions $f = I_A$ where $A \subset \mathbb{R}$ is a Borel set with $\lambda(\partial A) = 0$.

1.2 Linear summation methods

The argument above shows why logarithmic summation in the ASCLT is natural from an ergodic point of view. Note that logarithmic summation is stronger than Cesàro summation, i.e. if for a numerical sequence $(x_k)_{k \geq 1}$ we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k = x \quad (\text{say}),$$

then

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{k=1}^N \frac{1}{k} x_k = x.$$

That the converse is false can be seen by defining

$$x_k = \begin{cases} 0 & \text{if } k \in \{2^{2^i}, \dots, 2^{2^{i+1}} - 1\} \\ 1 & \text{if } k \in \{2^{2^{i+1}}, \dots, 2^{2^{(i+1)}} - 1\} \end{cases} \quad (i \geq 0).$$

In some sense, this sequence (x_k) exhibits a similar behavior as $(I\{k^{-1/2}S_k \leq x\})$. Long periods of 0's are followed by even longer periods of 1's and conversely. Let $x = 0$. By the CLT the probability that $k^{-1/2}S_k$ lies outside the interval $(-\epsilon, \epsilon)$ is close to 1 for small $\epsilon > 0$. If we assume that $|X_k| \leq 1$ for all k , then it will take at least $n > \epsilon\sqrt{k}$ steps until $I\{k^{-1/2}S_k \leq 0\}$ will change from 0 to 1 or the other way round. In order to compensate this persistence in the same state we have to weigh down the random variables $I\{k^{-1/2}S_k \leq x\}$. Thus, the explanation that logarithmic means are a proper tool in this theory might also result from the point of view of summability theory. By this argument, one should also expect to get stronger results if we use larger weights.

The elegance of the above heuristic argument for Theorem A raises the question if there exist other, substantially different averaging methods which also work in the ASCLT. Let us note first that log averaging is not unique: it was observed by Peligrad and Révész (1991) that

$$\frac{\alpha + 1}{(\log N)^{\alpha+1}} \sum_{k=1}^N \frac{(\log k)^\alpha}{k} I\{k^{-1/2}S_k \in A\} \rightarrow \Phi(A) \quad \text{a.s.} \quad (1.6)$$

for any $\alpha > -1$. While the averaging in (1.6) is rather similar to log averages, Berkes and Csáki (2001) showed that

$$\frac{1}{D_N} \sum_{k=1}^N d_k I\{k^{-1/2}S_k \in A\} \rightarrow \Phi(A) \quad \text{a.s.} \quad (1.7)$$

for

$$d_k = \exp((\log k)^\alpha)k^{-1}, \quad D_k = \sum_{j=1}^k d_j \quad (0 \leq \alpha < 1/2). \quad (1.8)$$

(Also, (1.7) holds for all smaller weight sequences (d_k) as well). To understand the connection between these results and the standard ASCLT (1.4), we shall review some facts on linear summation methods.

Let $\mathbf{D} = (D_N)$ be a positive non-decreasing sequence with $\lim_{N \rightarrow \infty} D_N = \infty$ and set $d_k = D_k - D_{k-1}$. We say that $(x_k)_{k \geq 1}$ is \mathbf{D} -summable to x if

$$\lim_{N \rightarrow \infty} \frac{1}{D_N} \sum_{k=1}^N d_k x_k = x. \quad (1.9)$$

By a result of Hardy (see e.g. Chandrasekharan and Minakshisundaram, 1952: 35), if \mathbf{D} and \mathbf{D}^* are summation procedures with $D_N^* = O(D_N)$, then under minor technical assumptions, the summation \mathbf{D}^* is stronger than \mathbf{D} , i.e. if a sequence (x_n) is \mathbf{D} -summable to x , then it is also \mathbf{D}^* -summable to x . Also, if (d_k) grows exponentially or faster, then (1.9) is equivalent to convergence of the sequence (x_k) , and hence this is the weakest summation method. By a result of Zygmund (see also Chandrasekharan and Minakshisundaram, 1952: 35) if $D_N^\alpha \leq D_N^* \leq D_N^\beta$ ($N \geq N_0$) for some $\alpha > 0, \beta > 0$, then \mathbf{D} and \mathbf{D}^* are equivalent, and if $D_N^* = O(D_N^\varepsilon)$ for any $\varepsilon > 0$, then \mathbf{D}^* is strictly stronger than \mathbf{D} . These results show that the larger the norming sequence D_N in (1.9) is, the stronger the relation (1.9) becomes. In view of Zygmund's theorem, (1.6) is actually equivalent to Theorem A.

Note also that the larger weights

$$d_k = k^{-\alpha}, \quad D_k = \sum_{j=1}^k d_j \quad (\alpha < 1)$$

define a summation equivalent to ordinary averaging, and hence do not work in the AS-CLT. On the other hand, the result of Berkes and Csáki provides summation methods in the ASCLT which are strictly weaker than logarithmic summation and thus yield a sharpening of the ordinary ASCLT. The summation procedures defined by (1.8) lie strictly between log and Cesàro averaging and are pairwise nonequivalent for different α 's. Note that for $\alpha = 0$, resp. $\alpha = 1$ (1.8) reduces to logarithmic, resp. ordinary averaging and thus the result of Berkes and Csáki shows that the ASCLT remains valid at least until 'halfway' from logarithmic to ordinary averaging. In view of Hardy's result the 'true' form of the ASCLT is given by relation (1.7) where the D_N is maximal, i.e. D_N is the weakest summation method. This maximal D_N is unknown. In the following section we will give substantial improvements of the result in Berkes and Csáki: we will show that the ASCLT (1.7) holds under a fairly general growth condition on (d_k) , similar to the condition in Kolmogorov's LIL. In particular, our results will imply that the sequence (1.8) obeys the ASCLT for all $0 \leq \alpha < 1$, and thus the critical summation procedure in a.s. central limit theory lies much closer, in some sense, to ordinary averaging. We will then determine fairly sharply the critical summation method **D**.

2 Results

Our first theorem is formulated in a more general setting. We will assume that X_n are independent, but we drop the assumption that they are identically distributed or have finite second moments.

Theorem 1 *Let X_1, X_2, \dots be independent random variables with partial sums S_n and assume that for some numerical sequences $a_n > 0$ and b_n we have*

$$\frac{S_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H \quad (2.1)$$

with some (possibly degenerate) distribution function H . Assume that

$$E \left| \frac{S_n - b_n}{a_n} \right|^p = O(1) \quad (n \rightarrow \infty), \quad (2.2)$$

and

$$a_k/a_l \leq C(k/l)^\beta \quad (1 \leq k \leq l) \quad (2.3)$$

for some positive constants p, β, C . Assume finally that $kd_k \gg 1$ and $d_k k^\alpha$ is eventually non-increasing for some $0 < \alpha < 1$ and that for some $\rho > 0$

$$d_k = O \left(\frac{D_k}{k(\log D_k)^\rho} \right). \quad (2.4)$$

Then if f is a bounded Lipschitz 1 function or an indicator function of a Borel set A with $\lambda(\partial A) = 0$, we have

$$\lim_{N \rightarrow \infty} D_N^{-1} \sum_{k=1}^N d_k f\left(\frac{S_k}{a_k} - b_k\right) = \int_{-\infty}^{\infty} f(x) dH(x) \quad \text{a.s.} \quad (2.5)$$

Conditions (2.1)-(2.3) are satisfied e.g. if X_k are i.i.d. with finite second moments or if X_k are i.i.d. random variables belonging to the domain of attraction of a stable distribution H . In this case (a_k) is regularly varying with exponent $1/\alpha$ for some $0 < \alpha \leq 2$ and the validity of relation (2.3) follows easily from the representation theorem for regularly varying functions. Conditions (2.1)-(2.2) are also satisfied if X_k are independent r.v.'s with $EX_k = 0$, $EX_k^2 < \infty$ satisfying the Lindeberg condition

$$\lim_{n \rightarrow \infty} s_n^{-2} \sum_{k=1}^n EX_k^2 I(|X_k| \geq \varepsilon s_n) = 0 \quad \text{for all } \varepsilon > 0,$$

where $s_n^2 = \sum_{k=1}^n EX_k^2$. In this case we can choose $a_k = s_k$, $b_k = 0$ and the additional assumption (2.3) is essential: as an example in Berkes and Dehling (1993) shows, without (2.3) the theorem fails even if $d_k = 1/k$.

Proposition 1 *Assume that relation (2.4) of Theorem 1 is satisfied for some sequence (D_N) . Then it is also satisfied for any other sequence $D_N^* = \psi(D_N)$ provided $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is differentiable, $\psi'(x) = O(\psi(x)/x)$ for $x \rightarrow \infty$ and $\log \psi'(x)$ is uniformly continuous on (A, ∞) for some $A > 0$. In particular, (2.4) is satisfied for $D_N = \exp((\log N)^\alpha)$, $0 \leq \alpha < 1$.*

We assumed here implicitly that $\psi(x) \rightarrow \infty$ since otherwise D_N^* defines of course no summation method. Typical examples for functions ψ permitted in the Proposition are $\psi(x) = x^\alpha$ or $\psi(x) = (\log x)^\alpha$ when $\alpha > 0$.

As we have mentioned above, replacing the weights d_k by smaller ones (subject to regularity conditions) leads to a stronger averaging procedure and thus it preserves the validity of the ASCLT (1.7). The second statement of Proposition 1 illustrates the type of regularity condition required in this context. Proposition 1 can be deduced from a theorem of Hirst in summation theory (see Chandrasekharan and Minakshisundaram, 1952: 37-38), but can also be deduced directly from the proof of Theorem 1. In view of the possibility of replacing the d_k with smaller ones, the assumption that $kd_k \gg 1$ is no real restriction of generality in Theorem 1. While excluding some irregular weight sequences the assumption $d_k k^\alpha$ is non-increasing for some $0 < \alpha < 1$ does also not make any restriction on the order of magnitude of (d_k) . Indeed, as we have seen, the sequence $d_k = k^{-\alpha}$, $\alpha > -1$, is already too large to imply the ASCLT.

Until now, on the independent sequence (X_n) we assumed only that its partial sums S_n , properly centered and normalized, converge weakly. Under the assumption that X_n are i.i.d. with finite variance, the sequence $D_N = \exp((\log N)^\alpha)$ obtained in Proposition 1 can be made essentially larger, and in fact a nearly optimal summation method can be obtained. Let us note, as we observed above, that the sequence $D_N = N^\alpha$, $\alpha > 0$, is too large in relation (1.7), and thus it is no restriction on the order of magnitude of D_N

to assume that D_N is slowly varying. By the theory of regular variation, D_N can then be represented in the form

$$D_N = c_N \exp \left(\int_A^N \epsilon(u)/u \, du \right), \tag{2.6}$$

where $A > 0$, $c_N \rightarrow c \in (0, +\infty)$, $\epsilon(x) \rightarrow 0$ for $x \rightarrow \infty$. Let \mathcal{W} denote the set of sequences (D_N) that can be represented as in (2.6) such that ϵ is monotone decreasing and obeys additionally

$$\epsilon(x)/\epsilon(x^2) = O(1) \quad (x \rightarrow \infty).$$

Since $\epsilon(x) = (\log x)^{-2}$ is already leading to a bounded D_N , the last assumption on ϵ is satisfied in all cases of interest.

Theorem 2 *Let X_1, X_2, \dots be i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$ and let $S_n = X_1 + \dots + X_n$. Let $\mathbf{D} \in \mathcal{W}$. Assume that*

$$d_k = O \left(\frac{D_k}{k(\log \log k)^\alpha} \right) \quad \text{with } \alpha > 3, \tag{2.7}$$

and

(kd_k) is eventually non-decreasing.

Then for any bounded function f which is either Lipschitz 1 or an indicator function of a Borel set $A \subset \mathbb{R}$ with $\lambda(\partial A) = 0$, we have

$$\lim_{N \rightarrow \infty} D_N^{-1} \sum_{k=1}^N d_k f(k^{-1/2} S_k) = \int_{-\infty}^{\infty} f(x) \, d\Phi(x) \quad \text{a.s.}$$

Theorem 2 yields a strong law of large numbers for the sums $\sum_{k=1}^N d_k f(k^{-1/2} S_k)$. Actually, much more is valid: under conditions similar to Theorem 2, the weighted sums of $f(k^{-1/2} S_k)$ satisfy also the central limit theorem and the law of the iterated logarithm.

Theorem 3 *Assume that the conditions of Theorem 2 are satisfied with $\alpha > 3$ in (2.7) replaced by $\alpha > 1$. Then we have for every non-constant Lipschitz 1 function f*

$$\lambda_N^{-1/2} \sum_{k=1}^N d_k \left(f \left(\frac{S_k}{\sqrt{k}} \right) - Ef \left(\frac{S_k}{\sqrt{k}} \right) \right) \xrightarrow{\mathcal{L}} N(0, 1),$$

where

$$\lambda_N := \text{Var} \left(\sum_{k=1}^N d_k f \left(\frac{S_k}{\sqrt{k}} \right) \right). \tag{2.8}$$

Theorem 4 *Assume that the conditions of Theorem 2 are satisfied. Then we have for every non-constant Lipschitz 1 function f*

$$\limsup_{N \rightarrow \infty} (2\lambda_N \log \log \lambda_N)^{-1/2} \sum_{k=1}^N d_k \left(f \left(\frac{S_k}{\sqrt{k}} \right) - Ef \left(\frac{S_k}{\sqrt{k}} \right) \right) = 1 \quad \text{a.s.}$$

where λ_N is defined by (2.8).

The basic new element in Theorems 2-4 is condition (2.7), which is weaker than the coefficient condition (2.4) in Theorem 1 and is very similar to Kolmogorov's classical condition for the LIL. Condition (2.7) permits a faster growing sequence D_N (and thus it yields a stronger result) than the sequence $D_N = \exp((\log N)^\alpha)$, $0 < \alpha < 1$ in Proposition 1. In fact, (2.7) is satisfied if

$$D_N = \exp(\log N / (\log \log N)^\alpha) \quad (\alpha > 3). \quad (2.9)$$

As a matter of fact, condition (2.7) is essentially sharp: for D_N as in (2.9) the conclusion of Theorem 4 fails with $0 < \alpha < 1$. Thus the 'dividing line' sequence in a.s. central limit theory is (2.9) and the change of behaviour happens in the interval $1 < \alpha < 3$. The critical value of α remains open.

3 Proofs

In what follows, we will only sketch the proofs of our results; for the detailed proof we refer to Hörmann (2005a) and (2005b). We use two different approaches to get Theorem 1 and Theorems 2-4.

3.1 First approach

To simplify the notation we set

$$\xi_k = f\left(\frac{S_k}{a_k} - b_k\right) - Ef\left(\frac{S_k}{a_k} - b_k\right).$$

Without loss of generality we may assume $|f| \leq 1$ and letting $X'_k = X_k - (a_k b_k - a_{k-1} b_{k-1})$ ($k = 1, 2, \dots$), we can also assume $b_k = 0$, $k \geq 1$. From (2.1) it follows that the conclusion of Theorem 1 is equivalent to

$$P\left\{\left|\sum_{k=1}^N d_k \xi_k\right| > \varepsilon D_N \quad \text{i.o.}\right\} = 0 \quad \text{a.s. for all } \varepsilon > 0.$$

The next lemma provides information on the covariances of the random variables ξ_k .

Lemma 2 *Let X_1, X_2, \dots be independent random variables with partial sums S_n and assume that for some numerical sequences $a_n > 0$ and $b_n = 0$ relations (2.2) and (2.3) are satisfied. Then there exist constants $c, \alpha > 0$ such that for every bounded Lipschitz 1 function f we have*

$$|E\xi_k \xi_l| \leq c(k/l)^\alpha \quad (1 \leq k \leq l). \quad (3.1)$$

Using the last lemma, the Markov inequality and (2.4) we can show for any $\eta < \rho$

$$P\left\{\left|\sum_{k=1}^N d_k \xi_k\right| > \varepsilon D_N\right\} \ll \frac{1}{\varepsilon^2} (\log D_N)^{-\eta}. \quad (3.2)$$

(Here $a_n \ll b_n$ means $\limsup_n |a_n/b_n| < \infty$.) Choose a sequence N_1, N_2, \dots such that

$$\sum_{i \geq 1} (\log D_{N_i})^{-\eta} < \infty. \tag{3.3}$$

Then the Borel-Cantelli Lemma implies that (2.5) holds along the subsequence N_1, N_2, \dots . In order to prove Theorem 1 it remains to estimate the oscillation of $\sum_{k=1}^N d_k \xi_k$ between the N_i 's. It is clear that if the gaps between N_i and N_{i+1} are too large, the oscillation will get out of control. On the other hand, we are forced to choose a rapidly growing sequence (N_i) in order to get (3.3). At this point the reason for assuming the condition $\alpha < 1/2$ for the weights of Berkes and Csáki (1.8) becomes clear. Simple analysis shows that for these d_k we have

$$d_k = O\left(\frac{D_k}{k(\log D_k)^{(1-\alpha)/\alpha}}\right).$$

If $\alpha < 1/2$, then we can choose $\eta > 1$ in (3.2)-(3.3) and thus for some $\delta > 0$ we have $\eta(1 - \delta) > 1$. This assures that for a subexponential sequence $D_{N_i} \sim e^{i^{1-\delta}}$ the series in (3.3) converges. The fact that the sequence (D_{N_i}) is subexponential provides that the oscillation of $\sum_{k=1}^N d_k \xi_k$ between the N_i 's is sufficiently small. The crucial tool for an improvement of the ASCLT is contained in the following moment inequality, which is similar to the well known *Rosenthal inequality*. (Cf. Petrov, 1995: p. 59).

Lemma 3 *Let X_1, X_2, \dots be independent random variables with partial sums S_n . Assume that (2.2) and (2.3) are satisfied for some sequences $a_n > 0$ and $b_n = 0$. Further let f be a bounded Lipschitz 1 function. Then for every $p \in \mathbb{N}$ we have*

$$E \left| \sum_{k=1}^N d_k \left(f\left(\frac{S_k}{a_k}\right) - E f\left(\frac{S_k}{a_k}\right) \right) \right|^p \leq C_p \left(\sum_{1 \leq k \leq l \leq N} d_k d_l \left(\frac{k}{l}\right)^\alpha \right)^{p/2},$$

where $C_p = C(p, X_1, \alpha)$ (α is the constant in (3.1)) does not depend on N .

With Lemma 3 and the Markov inequality we can improve (3.2). We set $\eta = \rho/2$ and get

$$P \left\{ \left| \sum_{k=1}^N d_k \xi_k \right| > \varepsilon D_N \right\} \leq \frac{c(m)}{\varepsilon^{2m}} (\log D_N)^{-m\eta} \quad (m \geq 1).$$

Choosing N_j such that $D_{N_j} \sim \exp(\sqrt{j})$ and $m > 2/\rho$ gives in conjunction with the Borel-Cantelli lemma

$$\lim_{j \rightarrow \infty} \frac{1}{D_{N_j}} \sum_{k=1}^{N_j} d_k \xi_k = 0 \quad \text{a.s.}$$

A simple calculation shows that if $N_j \leq N < N_{j+1}$

$$\frac{1}{D_N} \left| \sum_{k=1}^N d_k \xi_k \right| \leq \frac{1}{D_{N_j}} \left| \sum_{k=1}^{N_j} d_k \xi_k \right| + \frac{D_{N_{j+1}}}{D_{N_j}} - 1 \quad \text{a.s.}$$

Since $D_{N_{j+1}}/D_{N_j} \rightarrow 1$ the convergence of the subsequence implies that the whole sequence converges a.s.

□

3.2 Second approach

The approach to Theorems 2-4 is based on a blocking technique. The proof and the corresponding lemmas are very technical and therefore we will just illustrate the main idea, which is simple.

Lemma 4 *Let X_1, X_2, \dots be i.i.d. r.v.'s with $EX_1 = 0$ and $EX_1^2 = 1$ and let $S_n = X_1 + \dots + X_n$. Further let $d_k = L(k)/k$, where L is slowly varying at infinity and let f be a bounded Lipschitz 1 function. Then*

$$\text{Var} \left(\sum_{k=1}^N d_k f \left(\frac{S_k}{\sqrt{k}} \right) \right) \gg \sum_{k=1}^N k d_k^2. \quad (3.4)$$

Lemma 4 shows that Theorem 2 is a consequence of Theorem 4, since (2.7) and (3.4) imply that

$$D_N / (\lambda_N \log \log \lambda_N)^{1/2} \rightarrow \infty.$$

The idea of the proof of Theorems 3 and 4 is this: We partition \mathbb{N} into disjoint blocks: $\mathbb{N} = A_1 \cup B_1 \cup A_2 \cup B_2 \dots$, where

$$B_j = \{2^{p_j} + 1, \dots, 2^{q_j}\} \quad \text{and} \quad A_j = \{2^{p'_j} + 1, \dots, 2^{q'_j}\}.$$

The exponents p_j, q_j, p'_j and q'_j are defined such that the number of elements in B_j and in the A_j is increasing very fast and that $|B_j|$ is much larger than $|A_j|$. Next we define the random variables

$$Z_j := \sum_{k \in B_j} d_k \xi_k \quad \text{and} \quad R_j := \sum_{k \in A_j} d_k \xi_k \quad (j \geq 1).$$

Then clearly

$$\sum_{j=1}^n (Z_j + R_j) = \sum_{k=1}^{2^{qn}} d_k \left(f \left(\frac{S_k}{\sqrt{k}} \right) - Ef \left(\frac{S_k}{\sqrt{k}} \right) \right).$$

The idea is to approximate the random process $\{Z_j, j \geq 1\}$ by an *independent* process $\{Z_j^*, j \geq 1\}$. We define

$$Z_j^* = \sum_{k \in B_j} d_k \xi_k^j \quad (j \geq 1),$$

with

$$\xi_k^j = f \left(\frac{S_k - S_{l_j}}{\sqrt{k}} \right) - Ef \left(\frac{S_k - S_{l_j}}{\sqrt{k}} \right) \quad \text{for } k \in B_j,$$

where l_j is the largest integer contained in B_{j-1} . Then clearly by its definition the random variables Z_j^* are independent. Note that the faster the number of elements in the blocks A_j and B_j grow the smaller will be the error which we make by replacing ξ_k by ξ_k^j . However, too large blocks will cause a higher oscillation of the Z_j 's and R_j 's, which leads to a worse remainder term. Therefore an optimal choice of $|A_j|$ and $|B_j|$ is very important, in order to get optimal results.

Now we derive in a first step a CLT and an LIL for the sequences $\{Z_j^*, j \geq 1\}$ via Lyapunov's theorem and Kolmogorov's classical LIL (cf. Petrov, 1995: Theorem 4.9 and Theorem 7.1).

In the second step we estimate the difference between Z_j^* and Z_j . It turns out that it does not disturb the limiting behavior, i.e. the approximation is close enough. Further we show that the contribution of $\sum_{j=1}^n R_j$ to

$$\sum_{k=1}^{2^{q_n}} d_k \left(f \left(\frac{S_k}{\sqrt{k}} \right) - E f \left(\frac{S_k}{\sqrt{k}} \right) \right)$$

is negligible for our purposes. This proves the theorems along the subsequence $(2^{q_n})_{n \geq 1}$. Finally it remains to show that the fluctuation of the partial sums between 2^{q_n} and $2^{q_{n+1}}$ is small enough.

Acknowledgements

The author would like to thank Prof. István Berkes for his great support and two anonymous referees for carefully reading the paper and improving the first version.

References

- [1] Berkes, I. (1998): Results and problems related to the pointwise central limit theorem. In *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)* (pp. 59–96). Amsterdam: North-Holland.
- [2] Berkes, I. and Csáki, E. (2001): A universal result in almost sure central limit theory. *Stochastic Processes and their Applications*, **94**, 105-134.
- [3] Berkes, I. and Dehling, H. (1993): Some limit theorems in log density. *The Annals of Probability*, **21**, 1640-1670.
- [4] Brosamler, G. A. (1988): An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, **104**, 561-574.
- [5] Chandrasekharan, K. and Minakshisundaram, S. (1952): *Typical Means*. Oxford University Press.
- [6] Csörgő, M. and Révész, P. (1981): *Strong Approximations in Probability and Statistics*. Budapest: Academic Press.
- [7] Erdős, P. and Hunt, G. (1953): Changes of signs of sums of random variables. *Pacific J. Math.*, **3**, 673-687.
- [8] Hörmann, S. (2005a): Summation methods and the almost sure central limit theorem. *Preprint*.
- [9] Hörmann, S. (2005b): An extension of almost sure central limit theory. *Accepted for publication in Statist. Probab. Lett.*
- [10] Lacey, M. T. and Philipp, W. (1990): A note on the almost sure central limit theorem. *Statist. Probab. Lett.*, **9**, 201-205.
- [11] Lévy, P. (2003): *Théorie de l'addition des variables aléatoires* (12 ed.). Jacques Gabay.

- [12] Peligrad, M. and Révész, P. (1991): On the almost sure central limit theorem. In *Almost everywhere convergence, II (Evanston, IL, 1989)* (pp. 209–225). Boston, MA: Academic Press.
- [13] Petrov, V. (1995): *Limit Theorems of Probability Theory*. Oxford: Oxford Science Publications.
- [14] Schatte, P. (1988): On strong versions of the central limit theorem. *Math. Nachr.*, **137**, 249-256.
- [15] Weber, M. (2000): Un théorème central limite presque sûr à moments généralisés pour les rotations irrationnelles. *Manuscripta Math.*, **101**, 175-190.