Statistics & Probability Letters 9 (1990) 201–205
North-Holland

March 1990

A NOTE ON THE ALMOST SURE CENTRAL LIMIT THEOREM

Michael T. LACEY * and Walter PHILIPP **
Department of Statistics, University of North Carolina, Chapel Hill, NC, USA

Received January 1989
Revised April 1989

Abstract: We give a new proof of the almost sure central limit theorem. It is based on an almost sure invariance principle and extends to weakly dependent random variables.

Keywords: Central limit theorem, weakly dependent random variables, almost sure invariance principle, summation methods.

1. Introduction

The theorem in the title refers to a recent result in Brosamler (1988a) and Schatte (1988) (also see Brosamler, 1988b), which is as follows: Let S_k be the kth partial sum of i.i.d. real-valued random variables (r.v.'s) X_i with mean 0, variance 1 and finite $(2+\delta)$th moments ($\delta > 0$ in Brosamler, 1988a; $\delta = 1$ in Schatte, 1988). Let the r.v.'s be defined on a probability space $(\Omega, \mathscr{F}, P)$. Then there is a P-null set $N \subset \Omega$ such that for all $\omega \in N^c$,

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} 1_A\left(k^{-1/2} S_k(\omega)\right) \to (2\pi)^{-1/2} \int_A \mathrm{e}^{-u^2/2}\, \mathrm{d}u \tag{1}$$

for all Borel sets $A \subset \mathbb{R}$ with $\lambda(\partial A) = 0$. If $\delta(x)$ denotes the point mass at $x \in \mathbb{R}$, then (1) can be restated in this way. For all $\omega \in N^c$,

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} \delta\left(k^{-1/2} S_k(\omega)\right) \xrightarrow{\mathscr{L}} \mathrm{N}(0, 1). \tag{2}$$

The functional version of (2), proved in Brosamler (1988a), is as follows. Define the usual 'broken line process' on [0, 1],

$$s_n(t, \omega) = \begin{cases} n^{-1/2} S_k, & \text{if } t = k/n,\ k = 0, 1, \ldots, n, \\ \text{linear in between,} & \end{cases} \tag{3}$$

and denote by $\delta(x)$ the point mass at $x \in C[0, 1]$. Then

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} \delta\left(s_k(\cdot, \omega)\right) \xrightarrow{\mathscr{L}} \mathrm{W} \quad P\text{-a.s.} \tag{4}$$

Here W is standard Brownian motion on [0, 1].

In this note we prove the following results.

* Current address: Department of Mathematics, Indiana University, Bloomington, IN 47405, USA.
** Current address: Departments of Mathematics and Statistics, University of Illinois, Urbana, IL 61801, USA.

0167-7152/90/$3.50

Theorem 1. *Let* $\{X_j,\ j \geqslant 1\}$ *be any sequence of real-valued r.v.'s defined on* $(\Omega, \mathscr{F}, P)$. *Suppose there exists a sequence* $\{Y_j,\ j \geqslant 1\}$ *of i.i.d.* N(0, 1) *r.v.'s such that with probability* 1,

$$\sum_{k \leqslant n} X_k - \sum_{k \leqslant n} Y_k = \mathrm{o}(n^{1/2}). \tag{5}$$

Then (4) *holds.*

Theorem 2. *Let* $\{X_j,\ j \geqslant 1\}$ *be a sequence of i.i.d. r.v.'s with mean* 0 *and variance* 1. *Then* (4) *holds.*

The difference between the proof of Theorem 1 and that in Brosamler (1988a) is two-fold. First, we assume an almost sure invariance principle (5) (ASIP) and present a simple reduction of the theorem to the Gaussian case. Second, the latter case is done in a straightforward, probabilistic way, with no appeal to ergodic theory, as in Brosamler (1988a). The proof of Theorem 2 is a corollary to the proof of Theorem 1. Further remarks on the theorems and some extensions are at the end of this note.

2. Proofs

(a) Denote by $\mathrm{BL} = \mathrm{BL}(C[0,1],\ \|\cdot\|_{\mathrm{BL}})$ the class of functions $f: C[0,1] \to \mathbb{R}$ with $\|f\|_{\mathrm{BL}} := \|f\|_L + \|f\|_\infty < \infty$. Here

$$\|f\|_L := \sup\{|f(x) - f(y)|/\|x-y\|_\infty : x,\ y \in C[0,1],\ x \neq y\}.$$

We first prove that (4) is equivalent to the following statement: For each $f \in \mathrm{BL}$,

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} f((s_k(\cdot, \omega)) \to Ef(W(\cdot)) \quad P\text{-a.s.} \tag{6}$$

Although the left-hand side of (4) does not define a probability measure, it is so close to one that the theory of weak convergence still applies. But (6) does not yet imply (4) since, in general, the exceptional P-null set may depend on f. However, it is clear from the proof of (Dudley, 1989, Theorem 11.3.3, b $\Rightarrow$ c) that if we have (6) for a certain countable $\|\cdot\|_\infty$-dense subset of BL, which is not difficult to write down explicitly, then we have only a countable union N of P-null sets to contend with, and, outside N, the convergence in (6) is uniform in $f \in \mathrm{BL}$. Thus (4) and (6) are equivalent.

(b) Next, by (3) and (5), there is a P-null set, call it also N, such that for all $\omega \in N^c$,

$$|f(s_k(\cdot, \omega)) - f(t_k(\cdot, \omega))| \leqslant \|f\|_L \|s_k(\cdot, \omega) - t_k(\cdot, \omega)\|_\infty$$
$$\leqslant \|f\|_L k^{-1/2} \max_{j \leqslant k} \left| \sum_{i \leqslant j} (X_i - Y_i) \right| = \mathrm{o}(1).$$

Here we denoted by t_n the broken line process, defined in (3) with X_j replaced by Y_j. Thus (4) is equivalent to

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} f(t_k(\cdot, \omega)) \to Ef(W(\cdot)) \quad P\text{-a.s.} \tag{7}$$

for each $f \in \mathrm{BL}$. Hence we have reduced the proof to the case of Gaussian r.v.'s.

(c) For the proof of (7) we note that $t_k \xrightarrow{\mathscr{L}} W$ by Donsker's theorem applied to $\{Y_j,\ j \geqslant 1\}$, i.e. $Ef(t_k) \to Ef(W)$. Hence it is enough to prove that

$$(\log n)^{-1} \sum_{k \leqslant n} k^{-1} \xi_k \to 0 \quad P\text{-a.s.} \tag{8}$$

where

$$\xi_k := f(t_k(\cdot)) - Ef(t_k(\cdot)). \tag{9}$$

We need the following estimate.

Lemma. *There is a constant $C > 0$ depending only on $\| f \|_{BL}$ such that for all $j < k$,*

$$| E\xi_j \xi_k | \leqslant C(j/k)^{1/2}. \tag{10}$$

We postpone the proof of the lemma and finish the proof of (8). Let

$$Z_l := \sum_{4^{l-1} \leqslant j < 4^l} j^{-1} \xi_j, \quad l = 1, 2, \ldots. \tag{11}$$

Then, for the proof (8), we obviously only need to show

$$\frac{1}{n} \sum_{l \leqslant n} Z_l \to 0 \quad P\text{-a.s.} \tag{12}$$

But this follows from a strong law of large numbers for bounded quasi-orthogonal r.v.'s, since by (10), $EZ_l Z_m \ll 2^{l-m}$ for all $l < m$. (See e.g. Chung, 1974, Theorem 5.1.2 and its proof. Of course, the Gaal–Koksma strong law of large numbers also can be applied, see e.g. Philipp and Stout, 1975, p. 134.)

It remains to prove the lemma. Write $T_k := \sum_{j \leqslant k} Y_j$. We fix $j < k$ and define for $0 \leqslant t \leqslant 1$,

$$r(t) = \begin{cases} 0, & 0 \leqslant t \leqslant j/k, \\ t_k(t) - k^{-1/2} T_j, & j/k \leqslant t \leqslant 1. \end{cases}$$

Then

$$k^{1/2} E \| t_k(\cdot) - r(\cdot) \|_\infty \leqslant E | T_j | + E \max_{i \leqslant j} | T_i | \ll j^{1/2}.$$

Thus

$$E | f(t_k(\cdot)) - f(r(\cdot)) | \ll (j/k)^{1/2}. \tag{13}$$

The lemma follows now easily from the fact that r only depends on $Y_{j+1}, \ldots, Y_k$ and thus is independent of t_j.

(d) For the proof of Theorem 2 we show (6) directly. Since by Donsker's theorem $Ef(s_k) \to Ef(W)$ it is still enough to prove (8) but with t_k replaced by s_k in the definition (9). Thus to finish the proof we only need to show (10) in its modified form. The argument still works when we replace T_j by S_j.

3. Further remarks

(a) ASIP's are known for broad classes of sequences $\{X_j, j \geqslant 1\}$, such as mixing, martingale differences, lacunary trigonometric r.v.'s, etc. (see Philipp, 1986, for a survey). But Theorem 1 does not include Theorem 2 since $EX_1^2 = 1$ yields the error term $o((n \log\log n)^{1/2})$ in (5). To obtain the error term $o(n^{1/2})$ one needs to assume $EX_1^2(\log\log | X_1 |)^{1/2} < \infty$ (see Einmahl, 1987).

(b) The proof of Theorem 2 can be adapted so as to cover sequences of r.v.'s which fall between Theorems 1 and 2. For instance, stationary ϕ-mixing sequences can be treated in this way. The case $E | X_1 |^{2+\delta} < \infty$, $\delta > 0$ and $\phi(n) \downarrow 0$ is more delicate as $\operatorname{Var} S_n = nL(n)$ where L is a slowly varying function.

(c) The r.v.'s in (5) can also be Banach space valued, or even random elements (not necessarily measurable) in a non-separable Banach space. ASIPs are known (Dudley and Philipp, 1983; Philipp, 1982) and the proof above requires only easy modifications. The same can be said for $\mathbb{R}^d$-valued r.v.'s in the domain of normal attraction to a stable law of index α, $0<\alpha<2$ (see e.g. Philipp, 1986, p. 246). We only sketch the proof of the lemma above in the case of the (ordinary) CLT, that is the version corresponding to (1), and the X_ν, $\nu=1,2,\ldots,$ are i.i.d. symmetric stable r.v.'s of index α. If $1<\alpha<2$, the expectation $E|X_1|$ exists and so the lemma follows as above. If $0<\alpha\leqslant 1$ we argue as follows. Let $f\in \mathrm{BL}(\mathbb{R})$ with $\|f\|_{\mathrm{BL}}\leqslant 1$. Set for $j<k$,

$$s_j := j^{-1/\alpha}\sum_{\nu\leqslant j} X_\nu, \qquad r := k^{-1/\alpha}\sum_{j<\nu\leqslant k} X_\nu.$$

Then

$$E|f(s_k)-f(r)| \leqslant E\{|s_k-r|\wedge 1\} = E\left|k^{-1/\alpha}\sum_{\nu\leqslant j} X_\nu\right|\wedge 1$$

$$= (j/k)^{1/\alpha} E\left|X_1\wedge (k/j)^{1/\alpha}\right| \ll (j/k)\left(1+|\log(j/k)|\right).$$

(d) The use of the logarithmic mean looks perhaps peculiar at first glance, but it appears that it is essentially the only summation method that works. The Cesaro mean certainly does not, since

$$(1/n)\#\{k\leqslant n;\ S_k>0\} \overset{\mathscr{L}}{\to} \operatorname{arcsin}.$$

A similar argument shows that the weights $k^{-\alpha}$, $0<\alpha<1$, don't work either. Indeed, choose $\varepsilon>0$ so small that $(1-\alpha)^{-1}(1-\varepsilon^{1-\alpha})>\frac{3}{4}$. Then, by the arcsin law, for all sufficiently large n,

$$\#\{k\leqslant n: S_k>0\}\geqslant (1-\varepsilon)n$$

on a set C with $P(C)>\varepsilon^{1/2}/(2\pi)>0$. Thus on this set C,

$$(1-\alpha)^{-1}n^{\alpha-1}\sum_{k\leqslant n} k^{-\alpha}1_{\{S_k>0\}} \geqslant (1-\alpha)^{-1}n^{\alpha-1}\sum_{\varepsilon n\leqslant k\leqslant n} k^{-\alpha}$$

$$\sim (1-\alpha)^{-1}(1-\varepsilon^{1-\alpha}) > \tfrac{1}{2} = (2\pi)^{-1/2}\int_0^\infty e^{-u^2/2}\,du.$$

Hence the properly modified version of (1) cannot hold for the set $A=(0,\infty)$. Moreover, the covariance estimate (10) is sharp, and the present method will not allow the use of any summation method which is not close to the logarithmic mean.

(e) For any sequence $\{X_j\}$ the rates of convergence in these theorems cannot be $o(1/\log n)$ as can be seen for instance by setting $A=\mathbb{R}$ in (1). This is due to the fact that $\log n=\sum_{k\leqslant n}k^{-1}+\gamma+o(1)$, as $n\to\infty$, where γ is Euler's constant. This can be corrected by a better choice of norming constants.

However, no matter how we change the norming constants we cannot get a rate of convergence better than $O((\log n)^{-1/2})$ in probability, for any sequence $\{X_j,\ j\geqslant 1\}$ as in Theorem 2, as we shall now demonstrate. Let $f\in\mathrm{BL}(\mathbb{R})$ be a non-decreasing function. Define

$$\xi_k := f\left(k^{-1/2}S_k\right) - Ef\left(k^{-1/2}S_k\right),$$

and with these ξ_k's define Z_l by (11). By Resnick (1987, Lemma 5.32) the sequence $\{Z_l,\ l\geqslant 1\}$ is associated. Notice that (10) still holds. Hence by Cox and Grimmett (1984, Theorem 1.2), which is a CLT for not necessarily stationary associated sequences, we have

$$\sigma_n^{-1} \sum_{l \leqslant n} Z_l \xrightarrow{\mathscr{D}} N(0, 1) \tag{14}$$

where $c^{-1}n \leqslant \sigma_n^2 \leqslant cn$ for some finite constant c. (The lower bound follows from the ordinary CLT and the fact that $\{Z_l\}$ is associated, the upper bound from (10).) Transforming (14) into the form (1) proves our assertion about the error term.

(f) Comparing (12) and (14) we are led to a technically difficult project. Relation (12) (and thus (2)) is a uniform SLLN, (14) is a CLT for a single nondecreasing $f \in \text{BL}(\mathbb{R})$. Assuming appropriate entropy conditions for a class $\mathscr{F}$ of functions $f \in \text{BL}(C[0, 1], \|\cdot\|_{\text{BL}})$ obtain, as refinements of (12), the corresponding versions of a CLT and LIL. Now try to prove almost sure versions of this conjectured CLT. It appears that a whole hierarchy of CLT's looms in the background. It also appears that we are getting carried away at this point.

(g) After this paper was submitted for publication we learned that Alby Fisher ("A pathwise central limit theorem for random walks", preprint) has obtained Theorem 2 one year earlier, and independently of Brosamler and Schatte. He proves that for the sequence $\{X_j,\ j \geqslant 1\}$ of Theorem 2, relation (5) holds except on a subset of the positive integers having logarithmic density 0. Note that the hypotheses of our Theorem 1 can be weakened to this assumption, without change in the proof. Also, our proof is quite a bit simpler.

References

Billingsley, P. (1968), *Convergence of Probability Measures* (Wiley, New York).

Brosamler, G.A. (1988a), An almost everywhere central limit theorem, *Math. Proc. Cambridge Philos. Soc.* **104**, 561–574.

Brosamler, G.A. (1988b), Un théorème limite presque sur pour les mesures d'occupation du mouvement brownien sur une variété riemannienne compacte, *C.R. Acad. Sci. Paris Sér. I Math.* **307**, 919–922.

Chung, K.L. (1974), *A Course in Probability Theory* (Academic Press, New York, 2nd ed.).

Cox, J.T. and G. Grimmett (1984), Central limit theorems for associated random variables and the percolation model, *Ann. Probab.* **12**, 514–528.

Dudley, R.M. (1989), *Real Analysis and Probability* (Wadsworth, Belmont, CA).

Dudley, R.M. and W. Philipp (1983), Invariance principles for Banach space valued random elements and empirical processes, *Z. Wahrsch. Verw. Gebiete* **62**, 509–552.

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

Philipp, W. (1982), Invariance principles for sums of mixing random elements and the multivariate emprirical process, *Colloq. Math. Soc. János Bolyai, Vol. 36, Limit Theorems in Probability and Statistics* (Veszprém, Hungary) pp. 843–873.

Philipp, W. (1986), Invariance principles for independent and weakly dependent random variables, in: E. Eberlein and M.S. Taqqu, eds., *Dependence in Probability and Statistics* (Birkhauser, Boston, MA).

Philipp, W. and W. Stout (1975), Almost sure invariance principles for partial sums of weakly dependent random variables, *Mem. Amer. Math. Soc.* **161**.

Resnick, S.I. (1987), *Extreme Values, Regular Variation and Point Processes* (Springer, New York).

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