A Note on the almost everywhere central limit theorem

Michael T. Lacey and Walter Philipp
Department of Statistics, UNC, Chapel Hill

1. Introduction. The theorem in the title refers to a recent result in [2] and [11], (also see [3]), which is as follows: Let $S_k$ be the $k$-th 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 [2] and $\delta = 1$ in [11]). Let the r.v.s be defined on a probability space $(\Omega, \mathcal{F}, P)$. Then there is a $P$-null set $N \subset \Omega$ such that for all $\omega \in N^c$,

\begin{equation}
(\log n)^{-1} \sum_{k \leq n} k^{-1} 1_{A}[k^{-\delta} S_k(\omega)] \to (2\pi)^{-\frac{1}{2}} \int_A e^{-\frac{1}{2}u^2} du
\end{equation}

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 $\omega \in N^c$

\begin{equation}
(\log n)^{-1} \sum_{k \leq n} k^{-1} \delta(k^{-\delta} S_k(\omega)) \overset{\mathbb{P}}{\rightarrow} N(0,1).
\end{equation}

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

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

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

\begin{equation}
(\log n)^{-1} \sum_{k \leq n} k^{-1} \delta(s_k(\cdot, \omega)) \overset{\mathbb{P}}{\rightarrow} W \quad P\text{-a.s.}
\end{equation}
Here $W$ is standard Brownian motion on $[0,1]$.

In this note we prove the following results.

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

\begin{equation}
\sum_{k \leq n} X_k - \sum_{k \leq n} Y_k = o(n^{1/2}).
\end{equation}

Then (4) holds.

**Theorem 2.** Let $\{X_j, j \geq 1\}$ 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 [2] 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 [2]. 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 $BL = BL(C[0,1], \|\cdot\|_\infty)$ the class of functions $f : C[0,1] \to \mathbb{R}$ with $\|f\|_{BL} = \|f\|_L + \|f\|_\infty < \infty$. 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 $f \in BL$

\begin{equation}
(\log n)^{-1} \sum_{k \leq n} k^{-1} f((s_k(\cdot, \omega))) \to Ef(W(\cdot)) \quad P\text{-}a.s.
\end{equation}
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 [5, Theorem 8.3, II \Rightarrow III] that if we have (6) for a certain countable 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 \( f \in \text{BL} \). Thus (4) and (6) are equivalent.

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

\[
|f(s_k(\cdot, \omega)) - f(t_k(\cdot, \omega))| \leq \|f\|_{BL} \|s_k(\cdot, \omega) - t_k(\cdot, \omega)\|_{\infty}
\]

\[
\leq \|f\|_{BL} \sum_{j=1}^{\infty} \max_{j \leq k} \max_{i \leq j} |Y_i - Y_j| = o(1).
\]

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

\[
(7) \quad (\log n)^{-1} \sum_{k \in \mathbb{N}} k^{-1} f(t_k(\cdot)) \Rightarrow E f(W(\cdot)) \quad P\text{-a.s.}
\]

for each \( f \in \text{BL} \). Hence we have reduced the proof to the case of Gaussian r.v.s.

For the proof of (7) we note that \( t_k \overset{d}{\Rightarrow} W \) by Donsker's theorem applied to \( \{Y_j, j \geq 1\} \), i.e. \( E f(t_k) \Rightarrow E f(W) \). Hence it is enough to prove that

\[
(8) \quad (\log n)^{-1} \sum_{k \in \mathbb{N}} k^{-1} \xi_k \to 0 \quad P\text{-a.s.}
\]

where

\[
\xi_k := f(t_k(\cdot)) - E f(t_k(\cdot)).
\]

We need the following estimate.
e. 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

\[ \frac{1}{n} \#\{k \leq n; S_k > 0\} \leq \arcsin. \]

A similar argument shows that the weights \( k^{-\alpha}, 0 < \alpha < 1 \) don't work either. 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.

f. 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).

REFERENCES


Michael T. Lacey
Walter Philipp
Departments of Mathematics
and Statistics
University of Illinois
Urbana, IL 61801