Теор. ймовірност. та матем. статист.
Вип. 66, 2002, стор. 78–81

Teor. Ĭmovirnost. ta Matem. Statyst.
No. 66, 2002, pp. 78–81

ON RATE OF CONVERGENCE IN ALMOST SURE CENTRAL LIMIT THEOREM

УДК 519.21

A. I. MARTIKAINEN

Резюме. We construct a version of the almost sure CLT with a power rate of convergence to the limit law. This rate is only logarithmic in the well-known logarithmic version of the almost sure CLT.

1. Introduction

Throughout $X, X_1, \ldots, X_n, \ldots$ is a sequence of independent and identically distributed random variables with $\mathsf{E}\, X_n = 0$ and $\mathsf{E}\, X_n^2 = 1$. Put $S_n = X_1 + \cdots + X_n$. The central limit theorem claims that

$$\mathsf{P}\left(S_n < x\sqrt{n}\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

uniformly in $x \in \mathbf{R}$. The rate of convergence in this limit theorem is well studied. For example, the famous Berry–Esseen's inequality gives for $n \geq 1$ the estimate

$$\sup_n \left|\mathsf{P}\left(S_n < x\sqrt{n}\right) - \Phi(x)\right| \leq \frac{\mathsf{E}\,|X|^3}{\sqrt{n}}. \tag{1}$$

Put $I_j = I_j(x) = I(S_j < x\sqrt{j})$, where $I(\cdot)$ is the standard indicator function of event. Starting with general ideas of the law of large numbers and the relation $\mathsf{E}\, I_n \to \Phi(x)$, one could suppose that

$$\frac{1}{n} \sum_{j=1}^{n} I_j(x) \to \Phi(x) \quad \text{a.s.}$$

This hypothesis appears to be wrong, the variance of sum of underlying indicators increases too fast. In 1988 Brosambler [3] and Schatte [6] discovered independently that a replacement of the averaging procedure can improve situation. They proved that for $x \in \mathbf{R}$

$$\frac{1}{\log n} \sum_{j=1}^{n} \frac{1}{j} I_j(x) \to \Phi(x) \quad \text{a.s.}, \tag{2}$$

if $\mathsf{E}\,|X|^{2+\delta} < \infty$ for some $\delta > 0$ ($\delta = 1$ in Schatte's paper). Their result was strengthened later to $\delta = 0$ by Fisher and independently by Lacey and Philipp (see further references in Berkes [2]). It is well known that a sequence of distribution functions which converges weakly to a continuous distribution function converges to it uniformly. Hence, the relation (2) holds with probability 1 uniformly in rational $x \in \mathbf{R}$ and therefore in all $x \in \mathbf{R}$. At presence, there is a great variety of papers investigating the a.s. CLT (2). A.s. limit theorems with other limit distributions are also extensively studied. Many of these papers can be found in Berkes [2] and his careful survey includes the non-i.i.d. case,

---

2000 *AMS Mathematics Subject Classification.* Primary 60F15, Secondary 60K05.
Key words and phrases. Almost sure central limit theorem, rate of convergence.
Supported in part by MinEduRF Grant E00-1.0-82 and RFBR Grant 00-15-96019.

local theorems, weakly dependent random variables, multidimensional case, more general classes of functions instead of indicators, large deviations, more general summation methods, rate of convergence in (2), etc.

It is clear that the rate of convergence in (2) can not be less than the weight of $I_1(x)$ in the logarithmic sum in (2). It means that the rate is $O(\log n)$ at best. In fact, it is even worse [2]:

$$\limsup(\log n \log\log n)^{-1/2} \sum_{j=1}^{n} \left(\frac{1}{j} I_j(x) - \mathsf{E}\, I_j(x)\right) = \sigma \quad \text{a.s.},$$

where

$$\sigma^2 = 2\int_0^{\infty}\int_{-\infty}^{x}\int_{-\infty}^{x} (\phi(u,v,s) - \phi(u)\phi(v))\, du\, dv\, ds,$$

$\phi(u,v,s)$ is the bivariate normal density with mean 0 and covariance $\exp\{-|s|/2\}$.

The a.s. CLT based on weighted sums $b_n^{-1}\sum_{j=1}^{n} a_j I_j(x)$ with $b_n = \sum_{j=1}^{n} a_j \nearrow \infty$ have been investigated by several authors (see again [2] for details). In fact, the a.s. CLT holds for a class of methods which are similar to the logarithmic one (for example, $a_j = (\log j)^s/j$, $j \geq 2$, with $s \geq -1$ are quite acceptable). A simple calculation shows for the last example that the best rate of convergence in the a.s. CLT appears when $s = 0$. A different version of the a.s. CLT was also suggested by Schatte and Brosamler. They proved that

$$\frac{1}{n}\sum_{j=1}^{n} I(S_{2^j} < x2^{j/2}) \to \Phi(x) \quad a.s.$$

for distributions with finite moment of order $2+\delta$. The rate of convergence in this relation is at most the weight of the first indicator. Taking into account the fact that the sum on the left side involves 2^n elements of $X_1, X_2, \ldots$, we conclude that this rate of convergence is again logarithmic.

The main purpose of this paper is to suggest a version of the a.s. CLT with a power rate of convergence. We also prove that the rate of convergence in (2) remains at most logarithmic even if we replace the logarithmic averages by arbitrary functions of the same indicators.

2. Results and proofs

We recall that X_n are independent and identically distributed random variables with zero mean and variance 1 throughout this paper.

Theorem 1. *Assume that $H_n\colon \{0,1\}^n \to \mathbf{R}$ is a sequence of functions, and with probability 1*

$$b_n(H_n(I_1(x), \ldots, I_n(x)) - \Phi(x)) \to 0 \tag{3}$$

uniformly in $x \in \mathbf{R}$ where b_n is a sequence of constants. Then $b_n = o(\log n \log\log n)$.

We shall prove a slightly stronger result: theorem 1 remains true if we replace (3) (uniformly in $x \in \mathbf{R}$) by

$$\sup \mathsf{P}(|b_n(H_n - \Phi(x))| > \varepsilon) \to 0 \quad \text{for all } \varepsilon > 0, \tag{4}$$

where the supremum is taken over $x \in \Delta_n = ((2\,\mathrm{lln})^{1/2}, 2(\mathrm{lln})^{1/2})$, $\mathrm{lln} = \log\log(n \vee 16)$, $H_n = H_n(I_1(x), \ldots, I_n(x))$.

Proof. Assume now that (4) holds. Set $J_n = I_1(x)\cdots I_n(x)$. We have

$$\sup_{x\in\Delta_n} \mathsf{P}(|b_n(H_n - \Phi(x))| > \varepsilon, J_n = 1) \to 0 \quad \text{for all } \varepsilon > 0,$$

and since $H_n = H_n(1, \ldots, 1)$ on the event $\{J_n = 1\}$,

$$\sup_{x\in\Delta_n} \mathsf{P}(J_n = 1) I(|b_n(H_n(1,\ldots,1) - \Phi(x))| > \varepsilon) \to 0 \quad \text{for all } \varepsilon > 0. \tag{5}$$

Now we need a version of the Darling–Erdös theorem which is due to Einmahl [4]:

$$\mathsf{P}\left(\alpha_n \max_{1\le j\le n} S_j/B_j - \beta_n < y\right) \to \exp\left(-e^{-y}2\sqrt{\pi}\right)$$

uniformly in $y \in \mathbf{R}$, where $\alpha_n = (2\,\mathrm{lln})^{1/2}$, $\beta_n = 2\,\mathrm{lln} + \frac{1}{2}\log \mathrm{lln}$,

$$B_j^2 = \sum_{n=1}^{j} \mathsf{E}\, X^2 I\left(|X| \le \sqrt{n}/(\mathrm{lln})^2\right).$$

Evidently $B_n \le \sqrt{n}$ and $B_n \sim \sqrt{n}$. Put $x = \beta_n/\alpha_n$. Then

$$\mathsf{P}\left(\max_{1\le j\le n} S_j/\sqrt{j} < x\right) \ge \mathsf{P}\left(\max_{1\le j\le n} S_j/B_j < x\right) \to e^{-2\sqrt{\pi}}$$

and therefore the indicators $I(\cdot)$ in (5) equal zero for all sufficiently large n and such x. Hence $b_n(H_n(1,\dots,1)-\Phi(x)) \to 0$ and the same is true with $x' = x+1/x$ instead of x. We have $b_n(\Phi(x)-\Phi(x')) \to 0$. Note that $1-\Phi(x) \sim e^{-1}(1-\Phi(x')) \sim 1/(2\sqrt{\pi}\log n \log\log n)$, and so that the conclusion obtains. $\square$

Theorem 1 illustrates the slow rate of convergence in any variant of the a.s. CLT based on indicators $I_1,\dots,I_n$. To construct a variant with a power rate of convergence, we need first to replace this system of indicators by a different one.

Put $l_n = [\sqrt{n\log n}]$ and $k_n = [\sqrt{n/\log n}]$, where $[x]$ denotes the integer part of x. Further, set $I_{n,j} = I_{n,j}(x) = I(S_{jk_n} - S_{(j-1)k_n} < x\sqrt{k_n})$ for $1 \le j \le l_n$. Here $S_0 = 0$. For each $n \ge 1$, such defined indicators are independent.

Theorem 2. *Assume that* $\mathsf{E}\,|X|^3 < \infty$. *Then for any sequence of constants* x_n

$$\limsup n^{1/4}(\log n)^{-1/4}\left|\frac{1}{l_n}\sum_{j=1}^{l_n} I_{n,j}(x_n) - \Phi(x_n)\right| < \infty \quad a.s.$$

Comparing this result with (1) one finds here a power but worse rate of convergence. The deviation of random variable $l_n^{-1}\sum_{j=1}^{l_n} I_{n,j}(x)$ from $\Phi(x)$ appears here as a sum of two quantities. One of them is the deviation of this random variable from its mean, i.e., from $\mathsf{P}(S_{k_n} < x\sqrt{k_n})$. It can be estimated with the help of the Borel–Cantelli lemma and some results of the theory of large deviations. To get better rate on this step, we need faster increasing l_n. On the other hand, we need faster increasing k_n to make $\mathsf{P}(S_{k_n} < x\sqrt{k_n})$ closer to $\Phi(x)$ (with the help of (1)). Note that $l_nk_n \le n$ so far as we involve only first n elements of sequence $\{X_n\}$ in our construction. Our choice of k_n and l_n seems to be optimal for the proof below.

Proof. We start with recalling of Hoeffding's inequality which can be found for example in Petrov [5, p. 78]: if $Y_1,\dots,Y_n$ are independent random variables, $0 \le Y_j \le 1$, then for any $y > 0$,

$$\mathsf{P}\left(\sum_{j=1}^{n}(Y_j - \mathsf{E}\,Y_j) > ny\right) \le \exp\left(-2ny^2\right).$$

Applying it to indicators $I_{n,1},\dots,I_{n,l_n}$, we conclude that

$$\mathsf{P}\left(|T_n(x_n) - \mathsf{E}\,T_n(x_n)| > n^{-1/4}(\log n)^{1/4}\right) \le 2\exp(-2\log n) = n^{-2},$$

where $T_n(x) = l_n^{-1}\sum_{j=1}^{l_n} I_{n,j}(x)$. Hence the probabilities on the left form a convergent series. By the Borel–Cantelli lemma with probability 1 the inequalities

$$|T_n(x) - \mathsf{E}\,T_n(x)| \le n^{-1/4}(\log n)^{1/4}$$

hold for all $n > n_\omega$. Together with the estimate

$$|\mathsf{E}\,T_n(x) - \Phi(x)| \le \mathsf{E}\,|X|^3 n^{-1/4}(\log n)^{1/4}$$

which follows immediately from (1), this concludes the proof. $\square$

Our final result estimates the best possible convergence rate in variants of the a.s. CLT based on indicators $I_{n,j}$.

Theorem 3. *Assume that*

$$\mathsf{P}(X > n) = o\Big((\log n)^{(c+1)/2}/n^{2+c}\Big), \qquad \log n\,\mathsf{E}\,X^2 I(|X| > n) \to 0 \tag{6}$$

for some $c > 0$. For integer $1 \le k_n \le n$ put $l_n = [n/k_n]$, and for $1 \le j \le l_n$ define $I_{n,j}(x) = I(S_{jk_n} - S_{(j-1)k_n} < x\sqrt{k_n})$ as before. Let $H_n\colon \{0,1\}^n \to \mathbf{R}$ be a sequence of functions, and with probability 1

$$b_n(H_n(I_{n,1}(x), \dots I_{n,l_n}(x)) - \Phi(x)) \to 0 \tag{7}$$

uniformly in $x \in \mathbf{R}$, where b_n is a sequence of constants. If $l_n \to \infty$ and

$$l_n^2/\log l_n = O\left(k_n^c\right) \tag{8}$$

then $b_n = o(l_n)$.

As in case of Theorem 1 we shall prove a stronger result: theorem 3 remains true if we replace (7) (uniformly in $x \in \mathbf{R}$) by

$$\sup \mathsf{P}\left(|b_n(H_n(I_{n,1}(x), \dots, I_{n,l_n}(x)) - \Phi(x))| > \varepsilon\right) \to 0 \quad \text{for all } \varepsilon > 0, \tag{9}$$

where the supremum is taken over $x \in [(2\log l_n - \log\log l_n)^{1/2}, (\log l_n)^{1/2})$.

Proof. We shall use a result on moderate deviations due to Amosova [1]: if (6) holds then $\mathsf{P}(S_n > x\sqrt{n}) \sim 1 - \Phi(x)$ uniformly in x, $0 \le x \le (c\log n)^{1/2}$. Setting $x = (2\log l_n - \log\log l_n - A)^{1/2}$ and noting that $x < (c\log k_n)^{1/2}$ and (8) holds if the constant A is sufficiently large we conclude

$$\mathsf{P}\left(S_{k_n} \ge x\sqrt{k_n}\right) \sim 1 - \Phi(x) \sim e^A \left(2\sqrt{\pi} l_n\right)^{-1}.$$

Therefore

$$\mathsf{P}(I_{n,1}(x) = \dots = I_{n,l_n} = 1) = \mathsf{P}\left(S_{k_n} < x\sqrt{k_n}\right)^{l_n} \to \exp\left(-e^A(2\sqrt{\pi})^{-1}\right)$$

Using a similar argument as in the proof of Theorem 1, it follows $b_n(1 - \Phi(x)) \to 0$, and we obtain the assertion.

Going back to $l_n = [\sqrt{n\log n}]$ and $k_n = [\sqrt{n/\log n}]$ from Theorem 2 we see that (8) holds with $c > 2$ and in turn $b_n = o(\sqrt{n\log n})$ if the condition (6) is satisfied with $c > 2$. Note that the moment condition $\mathsf{E}\,|X|^t < \infty$ with some $t > 2 + c$ implies (6).

A careful reading proofs of Theorems 1 and 3 shows that we can assume (4) and (9) only for two increasing sequences $x = x_n$ in each theorem or even for *one* sequence in each theorem if we assume additionally $H_n(1, \dots, 1) = 1$ for $n \ge 1$.

Література

1. N. N. Amosova, *On the probabilities of moderate deviations for sums of independent random variables*, Probab. Theory and Its Applic. **24** (1979), 858–865.
2. I. Berkes, *Results and problems related to the pointwise central limit theorem*, Asymptotic results in probability and statistics (B. Scyszkowicz, ed.), A volume in honour of Miklos Csörgő, Elsevier, Amsterdam, 1998, pp. 59–96.
3. G. Brosamler, *An almost everywhere central limit theorem.*, Math. Proc. Cambridge Phil. Soc. **104** (1988), 561–574.
4. U. Einmahl, *The Darling-Erdös theorem for sums of i.i.d. random variables*, Probab. Th. Rel. Fields **82** (1989), 241–25.
5. V. V. Petrov, *Limit theorems of probability theory. Sequences of independent random variables*, Clarendon Press, Oxford, 1995.
6. P. Schatte, *On strong versions of the central limit theorem*, Math. Nachr. **137** (1988), 249–256.

St. Petersburg State University, Bibliotechnaja pl., 2, St. Peterhof, 198904 St. Petersburg, Russia

Електронна адреса: Alexander.Martikainen@pobox.spbu.ru

Надійшла 4/04/2002