Journal of Mathematics *Tạp chí Toán học* **21**(1 & 2)(1993), 13–19

ON THE RATE OF CONVERGENCE IN THE CONDITIONAL CENTRAL LIMIT THEOREM FOR STATIONARY MIXING SEQUENCES

BUI KHOI DAM

Abstract. *Let $(X_n, n \in N)$ be a stationary, φ - mixing sequence of real valued random variables with $E(X_n) = 0$ and $E(X_n^2) = 1$. Put $S_n = \sum_{i=1}^{n} X_i$, $s_n^2 = ES_n^2$. In this note, the rate of covergence in the conditional central limit theorem for stationary mixing sequences is shown as follows:*

$$\Delta_n(B) = \sup_{t \in R} |P(s_n^{-1} S_n < t|B) - \Phi(t)| = 0(n^{-(1/2-\gamma)})$$

$$E|X_1|^{p+\varepsilon} < \infty \quad \textit{for some } p \geq 8,\ \varepsilon > 0 \quad \phi(n) \leq C.n^{-\theta}, \quad \theta > \frac{p + 4\delta p - 4}{2}$$

and

$$d_1(Y, \sigma(X_1, \ldots, X_n)) = 0(n^{-(1/2+\delta)}(\log n)^{-r}),\ r > 1$$

where Φ is the standard normal distribution and

$$\gamma = \gamma(p, \delta) = 1/p + \varepsilon + 1/(p + 4\delta p).$$

1. INTRODUCTION

Let $(X_n, n \in N)$ be a stationary, φ - mixing sequence of real valued random variables with $E(X_n) = 0$ and $E(X_n^2) = 1$. Put $S_n = \sum_{i=1}^{n} X_i$, $s_n^2 = ES_n^2$. Let

$B \in \sigma(X_n, n \in N)$ with $P(B) > 0$. The aim of this note is to investigate the convergence rate to the normal law of conditioned distributions. Namely, we show that

$$\Delta_n(B) = \sup_{t\in R} |P(s_n^{-1} S_n < t|B) - \Phi(t)| = 0(n^{-(1/2-\gamma)})$$

under the assumptions

$$E|X_1|^{p+\varepsilon} < \infty \quad \text{for some} \quad p \geq 8,\ \varepsilon > 0,\ \varphi(n) \leq C.n^{-\theta},\ \theta > 0$$

$$d_1(Y, \sigma(X_1, \ldots, X_n)) = 0(n^{-(1/2+\delta)}(\log n)^{-r}),\ r > 1$$

where Φ is the standard normal distribution

$$\gamma = \gamma(p, \delta) = \frac{1}{p} + \varepsilon + \frac{1}{p + 4\delta p}.$$

If (X_n) is a sequence of i.i.d. random variables with mean 0 and variance 1 then the classical theorem of Berry-Esseen gave the best estimation of the rate of convergence to the normal law (see Petrov [4]):

$$\Delta_n(\Omega) = \sup_{t\in R} |P(n^{-1/2} S_n < t) - \Phi(t)| = 0(n^{-1/2}).$$

If (X_n) is a stationary, φ - mixing sequence of random variables with $E(X_1) = 0$, $E(X_1^2) = 1$, $E(X_1^8) < \infty$, and

$$0 < \lim \frac{E(S_n^2)}{n} < +\infty$$

then Stein [6] proved that $\Delta_n(\Omega) = 0(n^{-1/2})$.

The conditional central limit theorem obtained firstly by Renyi [5] asserts that

$$\lim_n \Delta_n(B) = \lim_n \sup_{t\in R} |P(S_n.n^{-1/2} < t|B) - \Phi(t)| = 0$$

for all $B \in \sigma(X_n, n \in N)$ with $P(B) > 0$, where (X_n) is i.i.d. sequence of random variables. The conditional central limit theorem plays an important role in the theory of random summation, in random walk problems, in sequential estimation and in the field of Monte Carlo methods.

In [3], Landers and Rogge showed that

$$\Delta_n(B) = 0(n^{-1/2})$$

if

$$E|X_1|^p < +\infty \quad \text{for some} \quad p > 3, \; d(B, \sigma(X_1, \dots, X_n)) = \inf\{P(B\Delta A), A \in \sigma(X_1, \dots, X_n)\} = 0\Big(\frac{1}{n^{1/2}(\log n)^{3/2}}\Big).$$

Therefore, our result can be considered as a simultaneous generalization of the results of Renyi, Stein, Landers and Rogge.

2. RESULT

We shall present the result in a slightly generalized form by considering a bounded random variable Y instead of the set B.

Theorem. *Let $(X_n, n \in N)$ be a stationary, φ-mixing sequence of random variables such that*

(i) $E|X_1|^{p+\varepsilon} < \infty$ for some $p > 8$, $\varepsilon > 0$,

(ii) mixing coefficient satisfies $\varphi(n) \leq C.n^{-\theta}$, C being constant and $\theta > 0$.

(iii) $d_1(Y, \sigma(X_1, \dots, X_n)) = \inf\{\|Y - Z\|_1 : Z \text{ is } \sigma(X_1, \dots, X_n) \text{ measurable}\} = 0(n^{-(1/2+\delta)}(\log n)^{-r})$, $r > 1$.

Then

$$\sup_{t\in R} |E(I_{\left(\frac{\delta_n}{s_n} < t\right)} .Y) - \Phi(t)E(Y)| = 0(n^{-(1/2 - \varepsilon(p,\delta))})$$

where $\varepsilon(p, \delta) = \dfrac{1}{p} + \varepsilon + \dfrac{1}{p + 4\delta p}$.

Proof.

Without loss of generality, we can assume that $E(S_n^2) = n$, (see, e.g. Billingsley [1, page 172]). Denote $N_1 = \{2^i, i \in N\}$, $Y_k = E(Y|\mathcal{F}_k)$ where $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$, $k \in N_1$. Then $(Y_k, \mathcal{F}_k)$ is martingale satisfying $E|Y - Y_k| = 0(n^{-(1/2+\delta)})$. Choose

$$n_0 = \max\{k \in N_1,\ k \leq \frac{n}{2}\ , n_1 = \left[n^{\frac{1}{p+4\delta p-4}}\right] + 1$$

we have

$$\left|E\Big[I\Big(\frac{S_n}{\sqrt{n}} < t\Big).Y\Big] - \Phi(t)E(Y)\right| = \left|E\Big\{\Big[I\Big(\frac{S_n}{\sqrt{n}} < t\Big) - \Phi(t)\Big]Y\Big\}\right|$$

$$= \left|E\{[I(\frac{S_n}{\sqrt{n}} < t) - \Phi(t)]\,[(Y - Y_{n_0}) + \sum_{\substack{k\in N_1\\ n_0\geq k\geq 2n_1}} (Y_k - Y_{k/2}) + Y_{n_1}]\}\right|$$

$$\leq 2E|Y - Y_{n_0}| + \sum_{\substack{k\in N_1\\ 2n_1\leq k\leq n_0}} E\Big|E[I(\frac{S_n}{\sqrt{n}} < t) - \Phi(t)\big|\mathcal{F}_k](Y_k - Y_{k/2})\Big|$$

$$+ \Big|E\{Y_{n_1}E[I(\frac{S_n}{\sqrt{n}} < t) - \Phi(t)\big|\mathcal{F}_{n_1}]\}\Big| = I_1 + I_2 + I_3.$$

For estimating the term $E\big(I(\frac{S_n}{\sqrt{n}} < t) - \Phi(t)\big|\mathcal{F}_k\big)$ we write $S_{ij} = \sum_{\nu=i+1}^{j} X_\nu$, $i < j$.

$$P\Big(\frac{S_n}{\sqrt{n}} < t\big|\mathcal{F}_k\Big) = P\Big(\frac{S_{2k,n}}{\sqrt{n-2k}} < \frac{t\sqrt{n}}{\sqrt{n-2k}} - \frac{S_{2k}}{\sqrt{n-2k}}\big|\mathcal{F}_k\Big)$$

$$= P\Big(\big(\frac{S_{2k,n}}{\sqrt{n-2k}} \leq \frac{t\sqrt{n}}{\sqrt{n-2k}} - \frac{S_{2k}}{\sqrt{n-2k}}\big)\big(|S_{2k}| < (2k)^{1/2+\delta}\big)\big|\mathcal{F}_k\Big)$$

$$+ P\Big(\big(\frac{S_{2k,n}}{\sqrt{n-2k}} < \frac{t\sqrt{n}}{\sqrt{n-2k}} - \frac{S_{2k}}{\sqrt{n-2k}}\big)\big(|S_{2k}| \geq (2k)^{1/2+\delta}\big)\big|\mathcal{F}_k\Big).$$

It follows from the above equality that

$$\Big|P\Big(\frac{S_n}{\sqrt{n}} < t\big|\mathcal{F}_k\Big) - \Phi(t)\Big| \leq 2P\big(|S_{2k}| > (2k)^{1/2+\delta}\big|\mathcal{F}_k\big)$$

$$+ \Big|P\Big(\frac{S_{2k,n}}{\sqrt{n-2k}} \leq \frac{t\sqrt{n}}{\sqrt{n-2k}} + \frac{(2k)^{1/2+\delta}}{\sqrt{n-2k}}\big|\mathcal{F}_k\Big) - \Phi(t)\Big|$$

$$+ \Big|P\Big(\frac{S_{2k,n}}{\sqrt{n-2k}} \leq \frac{t\sqrt{n}}{\sqrt{n-2k}} - \frac{(2k)^{1/2+\delta}}{\sqrt{n-2k}}\big|\mathcal{F}_k\Big) - \Phi(t)\Big|$$

$$\leq 2P\big(|S_{2k}| > (2k)^{1/2+\delta}\big|\mathcal{F}_k\big) + 2\varphi(k) + 2\sup_{u\in R}\Big|P\Big(\frac{S_{2k,n}}{\sqrt{n-2k}} < u\Big)$$

$$- \Phi(u)\Big| + \frac{2(2k)^{1/2+\delta}}{\sqrt{2\pi}\sqrt{n-2k}} + \frac{1}{\sqrt{2\pi e}}\frac{\sqrt{n}}{\sqrt{n-2k}}.$$

Hence, we can deduce that

$$I_2 \le \sum_{n_1 \le k \in N_1} E\big[|Y_k - Y_{k/2}| P(|S_{2k}| > (2k)^{1/2+\delta})\big]$$

$$+ \sum_{n_1 \le k \in N_1} \varphi(k) E|Z_k| + \frac{2}{\sqrt{2\pi}} \sum_{n_1 \le k \in N_1} \frac{(2k)^{1/2+\delta} E|Y_k - Y_{k/2}|}{\sqrt{n-2k}}$$

$$+ \frac{2}{\sqrt{2\pi e}} \sum_{n_1 \le k \in N_1} \frac{2kE|Y_k - Y_{k/2}|}{\sqrt{n-2k}} + \frac{C}{\sqrt{n}} \sum_{n_1 \le k \in N_1} E|Y_k - Y_{k/2}|$$

$$\le C_{11} \sum_{\substack{k \in N_1 \\ k \ge n_1}} P(|S_{2k}| > (2k)^{1/2+\delta})$$

$$+ 2C_{12} \sum_{\substack{k \in N_1 \\ k \ge n_1}} \frac{1}{k^{1/2+\delta} (\log k)^r} \frac{1}{k^\theta}$$

$$+ \frac{C_{13}}{\sqrt{n}} \sum_{\substack{k \in N_1 \\ k \ge n_1}} \frac{1}{(\log k)^r} + \frac{C_{14}}{\sqrt{n}} \sum_{\substack{k \in N_1 \\ k \ge n_1}} \frac{k}{\sqrt{n}} \frac{1}{k^{1/2+\delta}} \frac{1}{(\log k)^r} + \frac{C_{15}}{\sqrt{n}}$$

$$= C_{11} \sum_{\substack{k \in N_1 \\ k \ge n_1}} P(|S_{2k}| > (2k)^{1/2+\delta})$$

$$+ C_{12} \sum_{\substack{k \in N_1 \\ k \ge n_1}} \frac{1}{k^{\theta+1/2+2\delta}} \frac{1}{(\log k)^r}$$

$$+ \frac{C_{21}}{\sqrt{n}}$$

where C denote constants depending only on p, δ.

Using result of Lai (see [2, Theorem 1, page 695]) and noticing that

$$n \in N_1, \quad n_1 \ge \big[n^{\frac{1}{p+4\delta p-4}}\big] + 1$$

we get

$$\sum_{\substack{k \in N_1 \\ k \ge n_1}} P(|S_{2k}| > (2k^{1/2+2\delta}) = \sum_{\substack{k \in N_1 \\ k \ge n_1}} \frac{(2k)^{p(1/2+2\delta)-2} P(|S_{2k}| > (2k)^{1/2+2\delta})}{(2k)^{p/2+2\delta p-2}} \le \frac{C_{21}}{\sqrt{n}}.$$

For the next term, we have

$$\sum_{\substack{k\in N_1\\ k\geq n_1}} \frac{1}{k^{\theta+1/2+2\delta}} \cdot \frac{1}{(\log k)^r} \leq \frac{C_{21}}{\sqrt{n}}$$

since

$$\theta > \frac{p+4\delta p-4}{2}.$$

Similarly, we can estimate I_3 as follows:

$$\begin{aligned} I_3 &\leq C_{31} P\left(|S_{2n_1}| > (C(n_1))^{1/2+2\delta}\right) + 2\varphi(n_1) \\ &+ 2\sup_{u\in R}\left|P\left(\frac{S_{2n_1,n}}{\sqrt{n-2n_1}} < u\right) - \Phi(u)\right| + \frac{C(n_1)}{\sqrt{2\pi}\sqrt{n-2n_1}} \\ &+ \frac{2}{\sqrt{2\pi e}}\frac{2n_1}{n-2n_1}. \end{aligned}$$

We choose $C(n_1) = n_1^{1/2} n^{\frac{1}{2(p+\varepsilon)}}$. Then we can estimate all terms of the right-hand side of the above inequality as follows: First, by Markov's inequality we have

$$P(|S_{2n_1}| > n_1^{1/2} n^{\frac{1}{2(p+\varepsilon)}}) \leq \frac{E|S_{2n_1}|^{p+\varepsilon}}{n_1^{\frac{p+\varepsilon}{2}} n^{1/2}} \leq \frac{C_{21}}{\sqrt{n}}.$$

Further, we have obviously

$$\begin{aligned} &\varphi(n_1) \leq n_1^{-\theta} \leq \frac{1}{n^{\frac{\theta}{p+4\delta p-4}}} \leq \frac{1}{n^{1/2}}, \\ &\frac{1}{\sqrt{2\pi}}\frac{n_1^{1/2} n^{\frac{1}{2(p+\varepsilon)}}}{\sqrt{n-2n_1}} = \frac{1}{n^{1/2(1-\frac{1}{p+\varepsilon}-\frac{1}{p+4\delta p-4})}}, \\ &\frac{1}{\sqrt{2\pi e}}\frac{2n_1}{n-2n_1} \leq \frac{2}{\sqrt{2\pi e}}\frac{1}{\sqrt{n}}. \end{aligned}$$

Therefore, we get finally

$$I_3 \leq \frac{C_4}{n^{1/2(1-\frac{1}{p+\varepsilon}-\frac{1}{p+4\delta p-4})}}.$$

To complete the proof, we must show that

$$I_1 = 2E|Y - Y_{n_0}| \leq \frac{C_5}{\sqrt{n}}.$$

But this is obvious by the property of martingale (Y_k).

REFERENCES

1. P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1968.
2. T. L. Lai, *Convergence rates and r-quick versions of the strong law for stationary mixing sequences*, Ann. Probability **5** (1977), 693-706.
3. D. Landers and L. Rogge, *Exact approximation orders in the conditional central limit theorem*, Z. Wahrsch. verv. Gebiete **66** (1984), 227-244.
4. V. V. Petrov, *Sums of independent random variables*, Springer, Berlin, 1975.
5. A. Renyi, *On mixing sequences of sets*, Acta. Math. Sci. Hungar. **9** (1958), 215-228.
6. C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proc. of the sixth Berkeley Symposium, vol. 2 (1972), 583-602.

Institute of Mathematics
P. O. Box 631 Bo Ho
10000 Hanoi, Vietnam

Received February 15, 1992