Filomat 26:4 (2012), 713–717
DOI 10.2298/FIL1204713S

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

A central limit theorem for randomly indexed m-dependent random variables

Yilun Shang[a]

[a] *Institute for Cyber Security, University of Texas at San Antonio, Texas 78249, USA*

Abstract. In this note, we prove a central limit theorem for the sum of a random number N_n of m-dependent random variables. The sequence N_n and the terms in the sum are not assumed to be independent. Moreover, the conditions of the theorem are not stringent in the sense that a simple moving average sequence serves as an example.

1. Introduction

The assumption of independence for a sequence of observations $X_1, X_2, \cdots$ is often a technical convenience. Real data frequently exhibit some dependence and at least some correlation at small lags or distances. A very important kind of dependence considering distance as a measure of dependence, is the m-dependence case. A sequence of random variables $\{X_i\}_{i\geq 1}$ is called m-dependent for a given fixed m if for any two subsets I and J of $\{1, 2, \cdots\}$ such that $\min(J) - \max(I) > m$, the families $(X_i)_{i\in I}$ and $(X_j)_{j\in J}$ are independent. Equivalently, the sequence is m-dependent if two sets of random variables $\{X_1, \cdots, X_i\}$ and $\{X_j, X_{j+1}, \cdots\}$ are independent whenever $j - i > m$.

The central limit theorem has been extended to the case of m-dependent random variables by Hoeffding and Robbins [20], Diananda [14], Orey [28] and Bergstrom [4]. Since then many other kinds of dependence in cental limit theorems have been investigated, among which one could mention the m-dependence with unbounded m [5, 31], finitely-dependence [11], multi-dimensional dependence [6, 12], classes of mixing conditions (e.g. $\alpha, \beta, \phi, \ell$-mixing) [9] and dependence inherit from stochastic processes such as stationary process [17, 19, 26], martingale [18, 32, 34, 35] and Markov process [10, 15, 22]. For more details, we refer the reader to the recent book [13] and references therein.

The classical central limit theorem for m-dependent random variables is the following. See [20] or [24] for a proof.

Theorem 1.1. *Let $\{X_i\}_{i\geq 1}$ be a stationary m-dependent sequence of random variables. Let $E(X_i) = \mu$, $0 < Var(X_i) = \sigma^2 < \infty$ and $S_n = \sum_{i=1}^n X_i$ be the partial sum. Then*

$$\frac{\sqrt{n}}{\tau}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{L} N(0, 1) \tag{1}$$

as $n \to \infty$, where $\tau^2 = \sigma^2 + 2\sum_{i=2}^{m+1} Cov(X_1, X_i)$ and "$\xrightarrow{L} N(0, 1)$" denotes convergence in distribution to standard normal distribution.

2010 *Mathematics Subject Classification*. Primary 60F05
Keywords. Central limit theorem, dependent random variables, stationary sequence
Received: 26 September 2011; Revised: 05 December 2011; Accepted: 12 December 2011
Communicated by Svetlana Janković
Email address: shylmath@hotmail.com (Yilun Shang)

In this note, we aim to generalize Theorem 1.1 in another direction, that is, consider the central limit theorem for partial sum of a random number of $\{X_i\}$. Studies on random central limit theorem have a long tradition and they are applicable in a wide range of problems including sequential analysis, random walk problems, and Monte Carlo methods. Central limit problems for the sum of a random number of independent random variables have been addressed in the pioneer work of Anscombe [3], Rényi [30], and Blum et. al. [8]. The main result in [8] has also been independently shown by Mogyoródi [27]. More recent studies can be found in e.g. [16, 21, 23, 29, 33], most of which, nevertheless, deal with independent cases.

The rest of the note is organized as follows. In Section 2, we present our cental limit theorem and in Section 3, we provide a proof. Finally, we mention briefly some future lines of research in Section 4.

2. The result

The random central limit theorem is established as follows.

Theorem 2.1. *Let $\{X_i\}_{i\geq 1}$ be a stationary m-dependent sequence of random variables. Let $E(X_i)=\mu$, $0<Var(X_i)=\sigma^2<\infty$ and $S_n=\sum_{i=1}^n X_i$ be the partial sum. Let $\{N_n\}_{n\geq 1}$ denote a sequence of positive integer-valued random variables such that*

$$\frac{N_n}{\omega_n}\xrightarrow{P}\omega \quad (\textit{in probability}) \tag{2}$$

as $n\to\infty$, where $\{\omega_n\}_{n\geq 1}$ is an arbitrary positive sequence tending to $+\infty$ and ω is a positive constant. If

(A1) there exists some $k_0\geq 0$ and $c>0$ such that, for any $\lambda>0$ and $n>k_0$, we have

$$P\left(\max_{k_0<k_1\leq k_2\leq n}|S_{k_2}-S_{k_1}-(k_2-k_1)\mu|\geq\lambda\right)\leq\frac{c\cdot Var(S_n-S_{k_0})}{\lambda^2} \tag{3}$$

and

(A2) $Cov(X_1,X_i)\geq 0$ for $i=2,\cdots,m+1$,

then

$$\frac{\sqrt{N_n}}{\tau}\left(\frac{S_{N_n}}{N_n}-\mu\right)\xrightarrow{L}N(0,1) \tag{4}$$

as $n\to\infty$, where $\tau^2=\sigma^2+2\sum_{i=2}^{m+1}Cov(X_1,X_i)$.

There is a large body of literature dealing with limiting theory of random sums, where it is often assumed that the random sequence N_n is independent of the terms in the sum; see e.g. [25, 36] for reference. The strength of Theorem 2.1 is that this independence is not assumed. Moreover, if we assume N_n to be a stopping time, the above result follows easily.

We further give some remarks here. Firstly, note that the assumption (A1) is for sufficiently large index of sequence X_i, i.e., $\{X_i\}_{i>k_0}$. Moreover, (A1) is closely related to the famous Anscombe condition [1, 3]. Secondly, if $\{X_i\}_{i\geq 1}$ is independent, then (A1) automatically holds for $k_0=0$ and $c=1$ by using the Kolmogorov inequality (see e.g. [7]). Therefore, the assumption (A1) may be regarded as a "relaxed" Kolmogrov inequality. Thirdly, the assumption (A2) says that each pair X_i,X_j of $\{X_n\}_{n\geq 1}$ are positively correlated since the sequence is stationary. Also note that (A2) is the only requirement pertinent to covariances in Theorem 2.1 so that our result can be used to describe systems which have strong correlation in short distance (less than $2m$). In view of the independent case [8], it seems likely that the assertion of Theorem 2.1 still holds when ω is a positive random variable.

In order to verify that our theorem is not vainly true, we present an example of the sequence $\{X_i\}_{i\geq 1}$ of random variables which satisfy the conditions of Theorem 2.1. Actually, a simple moving average process serves our purpose. Suppose that the sequence $\{Z_i\}_{i\geq 1}$ consists of i.i.d. random variables with a finite variance σ^2, expectation μ and let $X_i = (Z_i + Z_{i+1})/2$. Then, $\{X_i\}_{i\geq 1}$ is a stationary 1-dependent sequence with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2/2$. Obviously, $Cov(X_1, X_2) = \sigma^2/4$ and

$$S_n = \sum_{i=1}^{n} X_i = \frac{Z_1 + Z_{n+1}}{2} + \sum_{i=2}^{n} Z_i. \tag{5}$$

Hence, the assumption (A2) is satisfied. Furthermore, it is easy to see that (A1) is satisfied with $k_0 = 0$ and $c = 2$ by exploiting expression (5) and the Kolmogorov inequality.

3. Proof of Theorem 2.1

Without loss of generality, we can assume that X_i's are centered at 0, i.e., $\mu = 0$. Let $0 < \varepsilon < 1/2$. From (2) it follows that there exists some n_0, such that for any $n \geq n_0$,

$$P(|N_n - \omega\omega_n| > \varepsilon\omega\omega_n) \leq \varepsilon. \tag{6}$$

For any $x \in \mathbb{R}$, we have

$$P\left(\frac{S_{N_n}}{\tau\sqrt{N_n}} < x\right) = \sum_{n=1}^{\infty} P\left(\frac{S_n}{\tau\sqrt{n}} < x, N_n = n\right). \tag{7}$$

By (6) and (7), we have for $n \geq n_0$,

$$\left| P\left(\frac{S_{N_n}}{\tau\sqrt{N_n}} < x\right) - \sum_{|n-\omega\omega_n|\leq\varepsilon\omega\omega_n} P\left(\frac{S_n}{\tau\sqrt{n}} < x, N_n = n\right)\right| \leq \varepsilon. \tag{8}$$

Let $n_1 = [\omega(1-\varepsilon)\omega_n]$ and $n_2 = [\omega(1+\varepsilon)\omega_n]$. Since ω_n tends to infinity, we have $n_1 \geq k_0$ for large enough n. Note that $S_{n_1} + \sum_{n_1<k\leq n} X_k = S_n$. Then we have for $|n - \omega\omega_n| \leq \varepsilon\omega\omega_n$,

$$P\left(\frac{S_n}{\tau\sqrt{n}} < x, N_n = n\right) \leq P(S_{n_1} < \sqrt{n_2}\tau x + Y, N_n = n), \tag{9}$$

where

$$Y = \max_{n_1<n\leq n_2} \left| \sum_{n_1<k\leq n} X_k \right|. \tag{10}$$

Likewise, we get

$$P\left(\frac{S_n}{\tau\sqrt{n}} < x, N_n = n\right) \geq P(S_{n_1} < \sqrt{n_1}\tau x - Y, N_n = n), \tag{11}$$

Involving the assumption (A1) and (10), we obtain

$$P(Y \geq \varepsilon^{\frac{1}{3}}\sqrt{n_1}) \leq \frac{c(n_2 - n_1)\tau^2}{\varepsilon^{\frac{2}{3}} n_1} \leq 4c\tau^2\varepsilon^{\frac{1}{3}}, \tag{12}$$

the right-hand side of which is less than 1 when ε is small enough.

Denote by E the event that $Y < \varepsilon^{1/3}\sqrt{n_1}$. By virtue of (8), (9) and (12), we get

$$\begin{aligned} P\left(\frac{S_{N_n}}{\tau\sqrt{N_n}} < x\right) &\leq P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < \sqrt{\frac{n_2}{n_1}}x + \frac{\varepsilon^{\frac{1}{3}}}{\tau}, E\right) + 4c\sigma^2\varepsilon^{\frac{1}{3}} + \varepsilon \\ &\leq P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < \sqrt{\frac{1+2\varepsilon}{1-2\varepsilon}}x + \frac{\varepsilon^{\frac{1}{3}}}{\tau}\right) + (4c\sigma^2 + 1)\varepsilon^{\frac{1}{3}}. \end{aligned} \tag{13}$$

Similarly, from (8), (11) and (12) it follows that

$$P\left(\frac{S_{N_n}}{\tau\sqrt{N_n}} < x\right) \geq P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < x - \frac{\varepsilon^{\frac{1}{3}}}{\tau}, E\right) - \varepsilon. \tag{14}$$

Using (14), (12) and the assumption (A2), we may derive

$$\begin{aligned} P\left(\frac{S_{N_n}}{\tau\sqrt{N_n}} < x\right) &\geq P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < x - \frac{\varepsilon^{\frac{1}{3}}}{\tau}\right) \cdot P(E) - \varepsilon \\ &\geq \left(1 - 4c\sigma^2\varepsilon^{\frac{1}{3}}\right) \cdot P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < x - \frac{\varepsilon^{\frac{1}{3}}}{\tau}\right) - \varepsilon, \end{aligned} \tag{15}$$

where the first inequality is due to an application of the FKG inequality (see e.g. [2]). In general, if Ω is a finite distributive lattice, and μ is a positive measure satisfying the lattice condition, then any two monotonically increasing functions f and g on Ω have the positive correlation inequality:

$$\langle fg \rangle \geq \langle f \rangle \langle g \rangle, \tag{16}$$

where $\langle f \rangle$ is the expected value with respect to μ.

Now by Theorem 1.1 we obtain

$$\lim_{n_1\to\infty} P\left(\frac{S_{n_1}}{\tau\sqrt{n_1}} < x\right) = \Phi(x), \tag{17}$$

where $\Phi(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-u^2/2}\mathrm{d}u$ is the standard normal distribution function. Combining (13), (15) and (17), we then conclude the proof of Theorem 2.1.

4. Concluding remarks

As remarked in Section 2, it is likely that Theorem 2.1 holds even when the limit in (2) is a positive random variable, and not necessarily a constant. From the practical point of view, this would be much more useful, as many random sums that appear in practical applications have such N_n. Take, for instance, geometric random sums, where N_n is geometric with parameter $1/n$ and N_n/n converges to standard exponential distribution, and not to a constant. Thus, it would be desirable to extend our result to the case of random ω in (2).

A more general question could be: what happens if one only assumes that N_n tends to infinity?

In addition, it would be useful to evaluate the limits of random sums normalized by constants rather than random quantities, as this leads to practical way of approximating the random sums. So, it would be interesting to modify our main result to include the limiting distribution of

$$a_n \sum_{i=1}^{N_n} (X_i - b_n) \tag{18}$$

for some suitable sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$.

Acknowledgements

The author wishes to thank the referees for their useful and detailed comments which improved the presentation.

References

[1] D. J. Aldous, Weak convergence of randomly indexed sequences of random variables, Math. Proc. Cambridge Philos. Soc. 83 (1978) 117–126.
[2] N. Alon, J. H. Spencer, The Probabilistic Method, Wiley-Interscience, 2008.
[3] F. J. Anscombe, Large-sample theory of sequential estimation, Proc. Cambridge Philos. Soc. 48 (1952) 600–607.
[4] H. Bergstrom, A comparison method for distribution functions of sums of independent and dependent random variables, Theor. Probab. Appl. 15 (1970) 430–457.
[5] K. N. Berk, A central limit theorem for m-dependent random variables with unbounded m, Ann. Probab. 1 (1973) 352–354.
[6] M. Beśka, H. Klopotowski, L. Slomiński, Limit theorems for random sums of dependent d-dimensional random vectors, Z. Wahrsch. verw. Gebiete 61 (1982) 43–57.
[7] P. Billingsley, Probability and Measure, Wiley-Interscience, 1995.
[8] J. R. Blum, D. L. Hanson, J. I. Rosenblatt, On the central limit thoerem for the sum of a random number of independent random variables, Probability Theory and Related Fields 1 (1963) 389–393.
[9] R. C. Bradley, Introduction to Strong Mixing Conditions, Vol. 1–3, Kendrick Press, Heber City, 2007.
[10] R. C. Bradley, A note on two measures of dependence, Stat. Prob. Lett. 81 (2011) 1823–1826.
[11] L. H. Y. Chen, Two central limit problems for dependent random variables, Z. Wahrsh. Verw. Geb. 43 (1978) 223–243.
[12] T. C. Christofides, P. M. Mavrikiou, Central limit theorem for dependent multidimensionally indexed random variables, Stat. Prob. Lett. 63 (2003) 67–78.
[13] A. DasGupta, Asymptotic Theory of Statistics and Probability, Springer, New York, 2008.
[14] P. H. Diananda, The central limit theorem for m-dependent variables, Proc. Cambridge Philos. Soc. 51 (1955) 92–95.
[15] R. L. Dobrushin, Central limit theorem for nonstationary Markov chains. I., Theory Probab. Appl. 1 (1956) 65–80.
[16] I. Fakhre-Zakeri, J. Farshidi, A central limit theorem with random indices for stationary linear processes, Stat. Prob. Lett. 17 (1993) 91–95.
[17] M. I. Gordin, A remark on the martingale method for proving the central limit theorem for stationary sequences, J. Math. Sci. 133 (2006) 1277–1281.
[18] P. Hall, C. C. Heyde, Rates of convergence in the martingale central limit theorem, Ann. Probab. 9 (1981) 395–404.
[19] C. C. Heyde, On the central limit theorem for stationary processes, Z. Wahrsch. verw. Gebiete 30 (1974) 315–320.
[20] W. Hoeffding, H. Robbins, The central limit theorem for dependent random variables, Duke Math. J. 15 (1948) 773–780.
[21] T. L. Hung, T. T. Thanh, Some results on asymptotic behaviors of random sums of independent identically distributed random variables, Commun. Korean Math. Soc. 25 (2010) 119–128.
[22] G. L. Jones, On the Markov chain central limit theorem, Probab. Surveys 1 (2004) 299–320.
[23] V. Y. Korolev, The asymptotics of randomly indexed random sequences: Independent indices, J. Math. Sci. 59 (1992) 926–938.
[24] E. L. Lehmann, Elements of Large Sample Theory, Springer, New York, 1999.
[25] Z. Y. Liu, C. R. Lu, C. L. Lu, On the central limit theorem for random sums of stationary sequences, Acta Math. Sinica 23 (1980) 566–571.
[26] F. Merlevède, M. Peligrad, S. Utev, Recent advances in invariance principles for stationary sequences, Probab. Survey 3 (2006) 1–36.
[27] J. Mogyoródi, A central limit theorem for the sum of a random number of independent random variables, Publ. Math. Hungar. Acad. Sci., Ser. A 7 (1962) 409–424.
[28] S. A. Orey, Central limit theorems for m-dependent random variables, Duke Math. J. 25 (1958) 543–546.
[29] M. Przystalski, Asymptotics for products of a random number of partial sums, Bull. Polish Acad. Sci. Math. 57 (2009) 163–167.
[30] A. Rényi, On the asymptotic distribution of the sum of a random number of independent random variables, Acta. Math. Acad. Sci. Hung. 8 (1957) 193–199.
[31] J. P. Romanoa, M. Wolf, A more general central limit theorem for m-dependent random variables with unbounded m, Stat. Prob. Lett. 47 (2000) 115–124.
[32] Y. Shang, On the degree sequence of random geometric digraphs, Appl. Math. Sci. 4 (2010) 2001–2012.
[33] Y. Shang, Central limit theorem for the sum of a random number of dependent random variables, Asian J. Math. Stat. 4 (2011) 168–173.
[34] Y. Shang, A martingale central limit theorem with random indices, Azerb. J. Math. 1 (2011) 109–114.
[35] Y. Shang, On the central limit theorems for forward and backward martingales, Proc. World Acad. Sci. Eng. Tech. 76 (2011) 860–863.
[36] I. Sugiman, A random CLT for dependent random variables, J. Multivariate Anal. 20 (1986) 321–326.