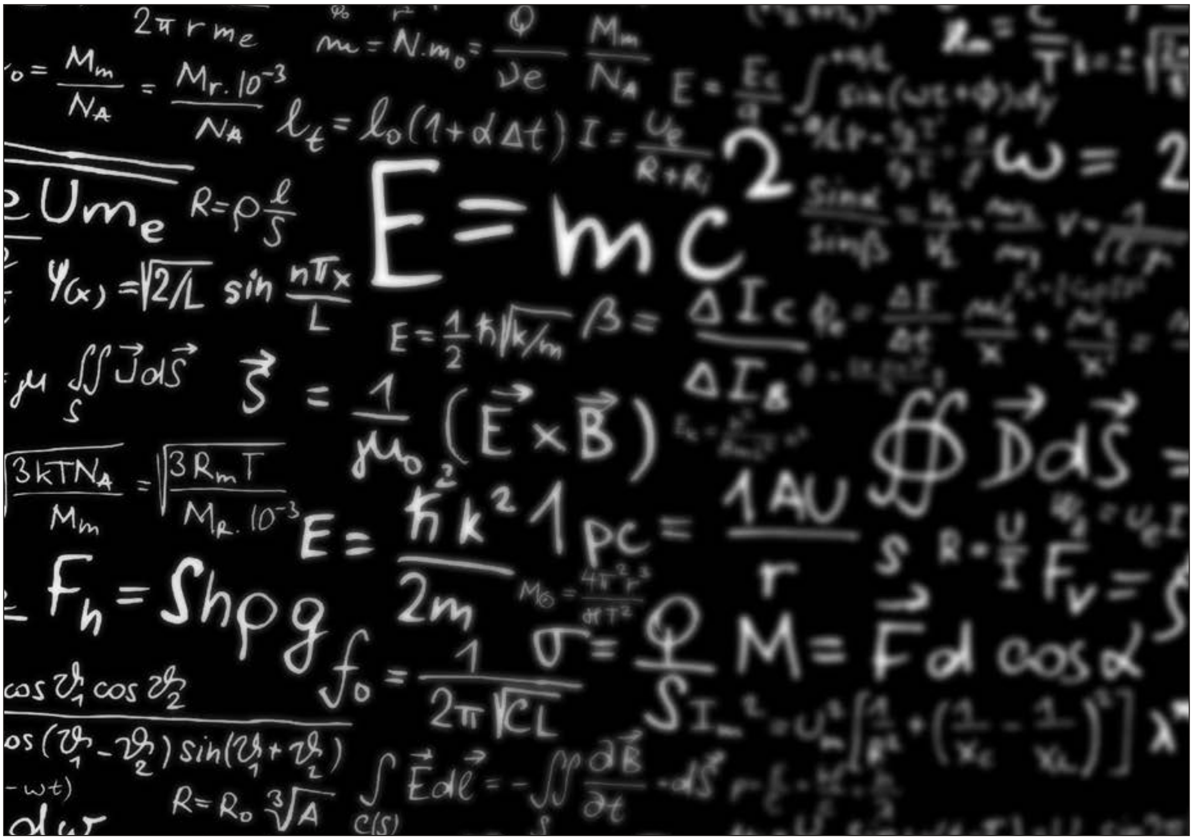

Asian Journal of Mathematics & Statistics

ISSN 1994-5418

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Mathematics and Statistics 4 (3): 168-173, 2011
ISSN 1994-5418 / DOI: 10.3923/ajms.2011.168.173

Central Limit Theorem for the Sum of a Random Number of Dependent Random Variables

Yilun Shang
Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

In this study, the central limit theorem for certain classes of dependent random variables is explored. The dependency structure, as defined in the class of random variables can be reflected in some physical phenomena. Sufficient conditions for the sum of a random number of dependent random variables tending to normality are provided.

Key words: Central limit theorem, dependent random variables, probability, characteristic function, random index

INTRODUCTION

The classical central limit theorem is considered as the heart of probability and statistics theory which has a number of applications (Lehmann, 1999; DasGupta, 2008). The assumption of independence for a sequence of observations $X_1,X_2,...$ is usually a technical convenience. Real data always exhibit some dependence and at least some correlations at small lags (Choi, 2004; Omekara, 2008; Dossou-Gbete *et al*., 2009; Osareh and Shadgar, 2009; Camminatiello and Lucadamo, 2010). One intensively studied class of dependence is the m-dependence case (Romano and Wolf, 2000; Christofides and Mavrikiou, 2003; Chaubey and Doosti, 2005; Francq and Zakoian, 2005) in which random variables are independent as long as they are m-step apart. More general measures of dependence are referred to as mixing conditions which are derived from the estimation of the difference between distribution functions of averages of dependent and independent random variables. Various mixing conditions have been proposed by Bradley (2005), Balan and Zamfirescu (2006), Kaminski (2007), Ould-Said and Tatachak (2009). The recent text book of DasGupta (2008) contains a survey of some related topics with an eye on statistics.

Recently, the exact value of difference between characteristic functions of sums of dependent and independent random variables is computed to derive central limit theorems for certain class of dependent random variables by Kaminski (2007). This class has some appealing physical interpretations and can be used to describe systems which are globally determined but locally random. His two main results are restated as follows.

Theorem 1: Let $\{X_i\}_{i\geq 1}$ be a sequence of identically distributed random variables such that $E|X_1|^{2+\varepsilon}<+\infty$ for some $\varepsilon>0$. Let $Var(X_1)=\sigma^2$ and ε_1 be a positive number such that $\varepsilon_1<\varepsilon/(2(1+\varepsilon))$. Denote by $S_n=\sum_{i=1}^{n}X_i$ the partial sum. Suppose that for sufficiently large k, the inequality:

$$\sup\left\{\left|P\left(\bigcap_{i=1}^{j}\{X_{v_i}\leq x_{v_i}\}\right)-\prod_{i=1}^{j}P(X_{v_i}\leq x_{v_i})\right|:(X_{v_1},\ldots,X_{v_j})\in\mathbb{R}^j\right\}\leq(1-k^{-\varepsilon_1})^{k-k^{\varepsilon_1}-j} \tag{1}$$

holds, where $v_1,...,v_j$ is any choice of indices such that $k^{\varepsilon_1}<v_1<...<v_j\le k$. Then:

$$\frac{S_n - ES_n}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1) \text{ as } n \to \infty \tag{2}$$

where "$\xrightarrow{D} N(0,1)$"denotes convergence in distribution to standard normal distribution.

Theorem 2: Let $\{X_i\}_{i\ge 1}$ be a sequence of identically distributed random variables such that $E|X_1|^{2+\varepsilon}<+\infty$ for some $\varepsilon>0$. Let Var $(X_1) = \sigma^2$ and ε_1 be a positive number such that $\varepsilon_1<\varepsilon/(2(1+\varepsilon))$ Suppose that for sufficiently large k:

$$\int_{\mathbb{R}^j}\left|P\left(\bigcap_{i=1}^{j}\{X_{v_i} \le x_{v_i}\}\right) - \prod_{i=1}^{j}P(X_{v_i} \le x_{v_i})\right|dx_{v_1}\cdots dx_{v_j} \le (1-k^{-\varepsilon_1})^{k-k^{\varepsilon_1}-j} \tag{3}$$

where $v_1,...,v_j$ is any choice of indices such that $k^{\varepsilon 1}<v_1<...v_j\le k$. Then:

$$\frac{S_n - ES_n}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1) \text{ as } n \to \infty \tag{4}$$

The aim of this study is to extend Theorem 1 and 2 in another direction, that is, consider the central limit theorem for partial sum of a random number of $\{X_1\}$. This question is critical not only in probability theory itself but in sequential analysis, random walk problems, Monte Carlo methods, etc. Central limit problems for the sum of a random number of random variables have been addressed in the study of Korolev (1992), Silvestrov (2006), Przystalski (2009) and Hung and Thanh (2010).

RESULTS

The main results are stated as follows.

Theorem 3: Let $\{X_i\}_{i\ge 1}$ be a sequence of identically distributed random variables such that $E|X_1|^{2+\varepsilon}<+\infty$ for some $\varepsilon>0$. Let Var $(X_1) = \sigma^2$ and ε_1 be a positive number such that $\varepsilon_1<\varepsilon/(2(1+\varepsilon))$. Denote by $S_n = \sum_{i=1} X_i$ the partial sum. Let $\{N_n\}_{n\ge 1}$ denote a sequence of positive integer-valued random variables such that

$$\frac{N_n}{\omega_n} \xrightarrow{P} \omega \text{ (in probability) as } n \to \infty \tag{5}$$

where, $\{\omega_n\}_{n\ge 1}$ is an arbitrary positive sequence tending to $+\infty$ and ω is a positive constant. Suppose that for sufficiently large k, the inequality:

$$\sup\left\{\left|P\left(\bigcap_{i=1}^{j}\{X_{v_i} \le x_{v_i}\}\right) - \prod_{i=1}^{j}P(X_{v_i} \le x_{v_i})\right| : (X_{v_1},\ldots,X_{v_j}) \in \mathbb{R}^j\right\} \le (1-k^{-\varepsilon_1})^{k-k^{\varepsilon_1}-j} \tag{6}$$

holds, where $v_1,...,v_j$ is any choice of indices such that $k^{\varepsilon 1}<v_1<...<v_j\le k$. If (A1) there exists some $k_0\ge 0$ and $c>0$ such that, for any $\lambda>0$ and k_0:

$$P\left(\max_{k_0<k_1\le k_2\le n}|S_{k_2}-S_{k_1}-(k_2-k_1)EX_1|\ge\lambda\right)\le\frac{c(n-k_0)Var(X_1)}{\lambda^2} \tag{7}$$

and (A2) Cov $(X_i,X_j)\ge 0$ for all i and j, then:

$$\frac{S_{N_n}-ES_{N_n}}{\sigma\sqrt{N_n}}\xrightarrow{D}N(0,1) \text{ as } n\to\infty \tag{8}$$

Theorem 4: Let $\{X_i\}_{i\ge1}$ be a sequence of identically distributed random variables such that $E|X_1|^{2+\varepsilon}<+\infty$ for some $\varepsilon>0$. Let Var $(X_1)=\sigma^2$ and ε_1 be a positive number such that $\varepsilon_1<\varepsilon/(2(1+\varepsilon))$. Let $\{N_n\}_{n\ge1}$ denote a sequence of positive integer-valued random variables such that:

$$\frac{N_n}{\omega_n}\xrightarrow{P}\omega \text{ (in probability) as } n\to\infty \tag{9}$$

where, $\{\omega_n\}_{n\ge1}$ is an arbitrary positive sequence tending to $+\infty$ and ω is a positive constant. Suppose that for sufficiently large k, the inequality:

$$\int_{\mathbb{R}^j}\left|P\left(\bigcap_{i=1}^{j}\{X_{v_i}\le x_{v_i}\}\right)-\prod_{i=1}^{j}P(X_{v_i}\le x_{v_i})\right|dx_{v_1}\cdots dx_{v_j}\le(1-k^{-\varepsilon_1})^{k-k^{\varepsilon_1}-j} \tag{10}$$

holds, where $v_1,\dots,v_j$ is any choice of indices such that $k^{\varepsilon_1}<v_1<\dots<v_j\le k$. If (A1) there exists some $k_0\ge0$ and $c>0$ such that, for any $\lambda>0$ and $n>k_0$:

$$P\left(\max_{k_0<k_1\le k_2\le n}|S_{k_2}-S_{k_1}-(k_2-k_1)EX_1|\ge\lambda\right)\le\frac{c(n-k_0)Var(X_1)}{\lambda^2} \tag{11}$$

and (A2) Cov $(X_i,X_j)\ge 0$ for all i and j, then:

$$\frac{S_{N_n}-ES_{N_n}}{\sigma\sqrt{N_n}}\xrightarrow{D}N(0,1) \text{ as } n\to\infty \tag{12}$$

Here are some remarks for Theorem 3 and similar comments may apply to Theorem 4. Firstly, note that the assumption (A1) is for sufficiently large index of sequence X_i, i.e., $\{X_i\}_{i>k0}$. Secondly, if $\{X_i\}_{i\ge1}$ is independent, then (A1) automatically holds for $k_0=0$ and c=1 by using the Kolmogorov inequality Billingsley (1995). Therefore, the assumption (A1) may be regarded as a relaxed kolmogorov inequality. Thirdly, the assumption (A2) says that each pair X_i, X_j of $\{X_n\}_{n\ge1}$ are positively correlated. In view of the independent case studied in previous work, it seems likely that the assertion of Theorem 3 still holds when ω is a positive random variable.

Proof: In the sequel, the proof of Theorem 3 is provided and that of Theorem 4 is left for the interested reader.

Without loss of generality, assume that X_i are centered at 0, i.e., $EX_1 = 0$. Let $0<\eta<1/2$. From Eq. 5 it follows that there exists some n_0, for any $n\geq n_0$:

$$P(|N_n - \omega\omega_n| > \varepsilon\omega\omega_n) \leq \eta \tag{13}$$

For any $x\in\mathbb{R}^j$:

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{N_n}} < x\right) = \sum_{n=1}^{\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} < x, N_n = n\right) \tag{14}$$

By Eq. 13 and 14, for $n\geq n_0$:

$$\left|P\left(\frac{S_{N_n}}{\sigma\sqrt{N_n}} < x\right) - \sum_{|n-\omega\omega_n|\leq\eta\omega\omega_n} P\left(\frac{S_n}{\sigma\sqrt{n}} < x, N_n = n\right)\right| \leq \eta \tag{15}$$

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

$$P\left(\frac{S_n}{\sigma\sqrt{n}} < x, N_n = n\right) \leq P(S_{n_1} < \sqrt{n_2}\sigma x + Y, N_n = n) \tag{16}$$

where:

$$Y = \max_{n_1<n\leq n_2}\left|\sum_{n_1<k\leq n} X_k\right| \tag{17}$$

Likewise:

$$P\left(\frac{S_n}{\sigma\sqrt{n}} < x, N_n = n\right) \geq P(S_{n_1} < \sqrt{n_1}\sigma x - Y, N_n = n) \tag{18}$$

Involving the assumption (A1) and Eq. 17,

$$P(Y \geq \eta^{1/3}\sqrt{n_1}) \leq \frac{c(n_2 - n_1)\sigma^2}{\eta^{2/3} n_1} \leq 4c\sigma^2\eta^{1/3} \tag{19}$$

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

Denote by E the event that $Y < \eta^{1/3}\sqrt{n_1}$. By virtue of Eq. 15, 16 and 19:

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{N_n}} < x\right) \le P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < \sqrt{\frac{n_2}{n_1}}x + \frac{\eta^{1/3}}{\sigma}, E\right) + 4c\sigma^2\eta^{1/3} + \eta$$

$$\le P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < \sqrt{\frac{1+2\eta}{1-2\eta}}x + \frac{\eta^{1/3}}{\sigma}\right) + (4c\sigma^2 + 1)\eta^{1/3} \tag{20}$$

Similarly, from Eq. 15, 18 and 19 it follows that:

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{N_n}} < x\right) \ge P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma}, E\right) - \eta \tag{21}$$

Using (19), (21) and the assumption (A2):

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{N_n}} < x\right) \ge P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma}\right)P(E) - \eta$$

$$\ge (1 - 4c\sigma^2\eta^{1/3}) \cdot P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma}\right) - \eta \tag{22}$$

where, the first inequality is due to an application of the FKG inequality (Alon and Spencer, 2008).
Now by Theorem 1:

$$\lim_{n_1\to\infty} P\left(\frac{S_{n_1}}{\sigma\sqrt{n_1}} < x\right) = \Phi(x) \tag{23}$$

where, Φ (x) is the standard normal distribution function. Thus, the proof of Theorem 3 is completed by combining Eq. 20, 22 and 23.

CONCLUSION

The central limit theorem for dependent random variables is one of the most active areas of research over the past decades. In this study, the central limit theorems for the sum of a random number of certain classes of dependent random variables are treated. The dependency structure may be reflected in some physical phenomena. Other issues such as the convergence rates and other dependent structures are possible future research.

REFERENCES

Alon, N. and J.H. Spencer, 2008. The Probabilistic Method. 3rd Edn., Wiley-Interscience, New Jersey.

Balan, R. and I.M. Zamfirescu, 2006. Strong approximation for mixing sequences with infinite variance. Elect. Comm. Probab., 11: 11-23.

Billingsley, P., 1995. Probability and Measure. Wiley-Interscience, New York.

Bradley, R.C., 2005. Basic properties of strong mixing conditions: A survey and some open questions. Probab. Surveys, 2: 107-144.

Camminatiello, I. and A. Lucademo, 2010. Estimating multinomial logit model with multicollinear data. Asian J. Math. Stat., 3: 93-101.

Chaubey, Y.P. and H. Doosti, 2005. Wavelet based estimation of the derivatives of a density for m-dependent random variables. J. Iran. Stat. Soc., 4: 97-105.

Choi, J.R., 2004. Unitary transformation of the time-dependent hamilton system for the linear, the V-shape and the triangular well potentials into the quadratic hamiltonian system. J. Applied Sci., 4: 636-643.

Christofides, T.C. and P.M. Mavrikiou, 2003. Central limit theorem for dependent multidimensionally indexed random variables. Stat. Probab. Lett., 63: 67-78.

DasGupta, A., 2008. Asymptotic Theory of Statistics and Probability. Springer, New York.

Dossou-Gbete, S., B. Some and D. Barro, 2009. Modeling the dependence of parametric bivariate extreme value copulas. Asian J. Mathematics Stat., 2: 41-54.

Francq, C. and J.M. Zakoian, 2005. A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. Econometric Theory, 21: 1165-1171.

Hung, T.L. and T.T. Thanh, 2010. Some results on asymptotic behaviors of random sums of independent identically distributed random variables. Commun. Korean Math. Soc., 25: 119-128.

Kaminski, M., 2007. Central limit theorem for certain classes of dependent random variables. Theory Probab. Appl., 51: 335-342.

Korolev, V.Y., 1992. The asymptotics of randomly indexed random sequences: Independent indices. J. Math. Sci., 59: 926-938.

Lehmann, E.L., 1999. Elements of Large Sample Theory. Springer, New York.

Omekara, C.O., 2008. Detecting non-linearity using squares of time series data. Asian J. Math. Statist., 1: 43-49.

Osareh, A. and B. Shadgar, 2009. Classification and diagnostic prediction of cancers using gene microarray data analysis. J. Applied Sci., 9: 459-468.

Ould-Said, E. and A. Tatachak, 2009. Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left truncation. Commun. Stat. Theory Methods, 38: 1154-1169.

Przystalski, M., 2009. Asymptotics for products of a random number of partial sums. Bull. Polish Acad. Sci. Math., 57: 163-167.

Romano, J.P. and M. Wolf, 2000. A more general central limit theorem for m-dependent random variables with unbounded m. Stat. Probab. Lett., 47: 115-124.

Silvestrov, D., 2006. Limit theorems for randomly stopped stochastic processes. J. Math. Sci., 138: 5467-5471.