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 \( X_1, X_2, \ldots \) is usually a technical convenience. Real data always exhibit some dependence and at least some correlations at small lags (Choi, 2004; Omekara, 2008; Dossou-Ogbete 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 Mavrikic, 2003; Chaubey and Doosti, 2005; Franq 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-Saïd and Tatachak (2009). The recent textbook 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 \( \{X_i\}_{i=1}^{\infty} \) be a sequence of identically distributed random variables such that \( \mathbb{E}[|X_i|^{2\epsilon}] < \infty \) for some \( \epsilon > 0 \). Let \( \text{Var}(X_i) = \sigma^2 \) and \( \epsilon_i \) be a positive number such that \( \epsilon_i \leq \epsilon(2(1+i)) \). Denote by \( S_k = \sum_{i=1}^{k} X_i \) the partial sum. Suppose that for sufficiently large \( k \), the inequality:

\[
\sup \left\{ \mathbb{P}\left( \bigcap_{i=1}^{n} \{ X_i \leq x_i \} \right) - \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i) : (X_1, \ldots, X_n) \in \mathbb{R}^n \right\} \leq (1-k^{-\epsilon})^{2^{-n}} \quad (1)
\]
holds, where $v_1, ..., v_j$ is any choice of indices such that $k^{\varepsilon} < v_1 < ... < v_j \leq k$. Then:

$$\frac{S_k - ES_k}{\sigma \sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

(2)

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

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

$$\int_{\mathbb{R}^j} \left| P\left( \bigcap_{i=1}^{j} \{X_{v_i} \leq x_{v_i}\} \right) - \prod_{i=1}^{j} P(X_{v_i} \leq x_{v_i}) dx_{v_1} ... dx_{v_j} \right| \leq (1 - k^{-\varepsilon_i})^{j-1}$$

(3)

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

$$\frac{S_k - ES_k}{\sigma \sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

(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 $\{X_i\}$. 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 (2005), Przystalski (2009) and Hung and Thanh (2010).

**RESULTS**

The main results are stated as follows.

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

$$\frac{N_n}{\omega_n} \xrightarrow{p} 0 \text{ (in probability) as } n \to \infty$$

(5)

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

$$\sup_{(X_{v_1},...,X_{v_j})} \left| P\left( \bigcap_{i=1}^{j} \{X_{v_i} \leq x_{v_i}\} \right) - \prod_{i=1}^{j} P(X_{v_i} \leq x_{v_i}) \right| \leq (1 - k^{-\varepsilon_i})^{j-1}$$

(6)

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

169
\[
P\left( \max_{k_0 < k_1 \leq \ldots \leq k_n} |S_{k_2} - S_{k_1} - (k_2 - k_1)EX_j| \geq \lambda \right) \leq \frac{c(n-k_0)\text{Var}(X_j)}{\lambda^2} \tag{7}
\]

and (A2) \( \text{Cov}(X_i, X_j) \geq 0 \) for all \( i \) and \( j \), then:

\[
\frac{S_{k_n} - ES_{k_n}}{\sigma \sqrt{N_n}} \xrightarrow{d} N(0,1) \quad \text{as } n \rightarrow \infty \tag{8}
\]

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

\[
\frac{N_n}{\omega_n} \xrightarrow{p} \omega \quad \text{(in probability) as } n \rightarrow \infty \tag{9}
\]

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

\[
\int_{\mathbb{R}^\eta} P\left( \bigcap_{i=1}^{\eta} \{X_{\tau_i} \leq x_{\tau_i} \} \right) \prod_{i=1}^{\eta} P(X_{\tau_i} \leq x_{\tau_i}) \, dx_{\tau_\eta} \cdots dx_{\tau_1} \leq (1 - k^{-c})^{\sum_{i=1}^{\eta} x_{\tau_i}} \tag{10}
\]

holds, where \( \tau_1, \ldots, \tau_\eta \) is any choice of indices such that \( k^{\varepsilon_1} \leq \varepsilon_1 < \ldots < \varepsilon_k \leq k \). If (A1) there exists some \( k_0 > 0 \) and \( c > 0 \) such that, for any \( \lambda > 0 \) and \( n > k_0 \):

\[
P\left( \max_{k_0 < k_1 \leq \ldots \leq k_n} |S_{k_2} - S_{k_1} - (k_2 - k_1)EX_j| \geq \lambda \right) \leq \frac{c(n-k_0)\text{Var}(X_j)}{\lambda^2} \tag{11}
\]

and (A2) \( \text{Cov}(X_i, X_j) \geq 0 \) for all \( i \) and \( j \), then:

\[
\frac{S_{k_n} - ES_{k_n}}{\sigma \sqrt{N_n}} \xrightarrow{d} N(0,1) \quad \text{as } n \rightarrow \infty \tag{12}
\]

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., \( \{X_i\}_{i=1} \). Secondly, if \( \{X_i\}_{i=1} \) is independent, then (A1) automatically holds for \( k_0 = 0 \) and \( c = 1 \) by using the Kolmogorov inequality. Thirdly, the assumption (A2) says that each pair \( X_i, X_j \) of \( \{X_i\}_{i=1} \) 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 \( \omega \) 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., \( \mu = 0 \). 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 \( x \in \mathbb{R}^n \):

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

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

\[
\left| P\left( \frac{S_n}{\sqrt{N_n}} < x \right) - \sum_{|n|=\omega, |n| \geq n_0} P\left( \frac{S_n}{\sqrt{n}} < x, N_n = n \right) \right| \leq \eta \tag{15}
\]

Let \( n_1 = [\omega(1-\eta)\omega_n] \) and \( n_2 = [\omega(1+\eta)\omega_n] \). Since \( \omega_n \) tends to infinity, then \( n_1 \geq k_0 \) for large enough \( n \). Note that \( S_n + \sum_{n_1 < k \leq n_2} X_k = S_k \). Therefore, for \( n_1 \omega \omega_n \leq \eta_0 \omega_n \):

\[
P\left( \frac{S_n}{\sqrt{n}} < x, N_n = n \right) \leq P(S_n < \sqrt{n_1 \omega x} + Y, N_n = n) \tag{16}
\]

where:

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

Likewise:

\[
P\left( \frac{S_n}{\sqrt{n}} < x, N_n = n \right) \geq P(S_n < \sqrt{n_1 \omega x} - Y, N_n = n) \tag{18}
\]

Involving the assumption (A1) and Eq. 17,

\[
P(Y \geq \eta^{\nu_3} \sqrt{n_1}) \leq \frac{c(n_2 - n_1)\sigma^2}{\eta^{\nu_3} n_1} \leq 4\sigma^2 \eta^{\nu_3} \tag{19}
\]

the right-hand side of which is less than 1 when \( \eta \) is small enough.

Denote by \( E \) the event that \( Y < \eta^{\nu_3} \sqrt{n_1} \). By virtue of Eq. 15, 16 and 19:
\[
P\left( \frac{S_{n_1}}{\sigma \sqrt{N_n}} < x \right) \leq P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < \sqrt{\frac{1}{n_1} x + \frac{\eta^{1/3}}{\sigma}} \right) + 4\sigma^2 \eta^{1/3} + \eta \\
\leq P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < \sqrt{\frac{1 + 2\eta}{1 - 2\eta} x + \frac{\eta^{1/3}}{\sigma}} \right) + (4\sigma^2 + 1)\eta^{1/3}
\]

Similarly, from Eq. 15, 18 and 19 it follows that:
\[
P\left( \frac{S_{n_1}}{\sigma \sqrt{N_n}} < x \right) \geq P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma} \right) - \eta
\]

Using (19), (21) and the assumption (A2):
\[
P\left( \frac{S_{n_1}}{\sigma \sqrt{N_n}} < x \right) \geq P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma} \right) P(E) - \eta
\]
\[
\geq (1 - 4\sigma^2 \eta^{1/3}) \cdot P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < x - \frac{\eta^{1/3}}{\sigma} \right) - \eta
\]

where, the first inequality is due to an application of the FKG inequality (Alon and Spencer, 2008).

1. By Theorem 1:
\[
\lim_{n_1 \to \infty} P\left( \frac{S_{n_1}}{\sigma \sqrt{n_1}} < x \right) = \Phi(x)
\]

where, \( \Phi(\cdot) \) 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