A Review on asymptotic normality of sums of associated random variables

Gane Samb Lo\textsuperscript{1,2,*}, Harouna Sangare\textsuperscript{2,3} and Cheikhna Hamallah Ndiaye\textsuperscript{2}

\textsuperscript{1}LSTA, Université Paris VI, France
\textsuperscript{2}LERSTAD, Université Gaston Berger de Saint-Louis, Sénégal
\textsuperscript{3}DER MI, Université des Sciences, des Techniques et des Technologies de Bamako, Mali

Received January 30, 2016; Accepted March 15, 2016

Copyright © 2016, Afrika Statistika and Statistics and Probability African Society (SPAS). All rights reserved

Abstract. Association between random variables is a generalization of independence of these random variables. This concept is more and more commonly used in current trends in any research fields in Statistics. In this paper, we proceed to a simple, clear and rigorous introduction to it. We will present the fundamental asymptotic normality theorem on stationary and associated sequences of random variables. A coherent and modern frame is used. This review will be profitable to new researchers in the topic.

Résumé. Le concept de variables aléatoires associées est une généralisation de l’indépendance entre variables aléatoires. Il devient de plus en plus important dans les probabilités et les applications statistiques de tous les jours. Dans ce papier, nous introduisons à ce concept et présentons le théorème fondamental de la normalité asymptotique de sommes partielles d’une suite stationnaire de variables aléatoires associées. Nous utilisons un cadre moderne et cohérent qui sera profitable aux chercheurs débutants dans ce domaine.

Key words: Positive and Negative Dependence; Association; Central Limit Theorem.

AMS 2010 Mathematics Subject Classification: 60F05; 62G20; 62H20.

1. A brief reminder of association

We begin to introduce to the associated random variables concept which goes back to Lehmann (1966) in the bivariate case. The concept of association for random variables generalizes that of independence and seems to model a great variety of stochastic models.

*Corresponding author Gane Samb LO: ganesamblo@ganesamblo.net
Harouna Sangare : harounasangareusttb@gmail.com
Cheikhna Hamallah Ndiaye : ham111266@yahoo.fr
This property also arises in Physics, and is quoted under the name of FKG property (Fortuin et al., 1971), in percolation theory and even in Finance (see Jiazhu, 2002).

The definite definition is given by Esary et al. (1967) as follows.

**Definition 1.** The rv’s $X_1, ..., X_n$ are associated, or equivalently the finite sequence $X_1, ..., X_n$ is associated, if for any couple of real and coordinatewise nondecreasing functions $h$ and $g$ defined on $\mathbb{R}^n$, we have

$$\text{Cov}(h(X_1, ..., X_n), g(X_1, ..., X_n)) \geq 0$$

An infinite sequence of rv’s is associated whenever all its finite subsequences are associated.

We have a few number of interesting properties to be found in (Prakasa Rao, 2012):

**(P1)** A sequence of independent rv’s is associated.

**(P2)** Partial sums of associated rv’s are associated.

**(P3)** Order statistics of independent rv’s are associated.

**(P4)** Non-decreasing functions and non-increasing functions of associated variables are associated.

**(P5)** Let the sequence $Z_1, Z_2, ..., Z_n$ be associated and let $(a_i)_{1 \leq i \leq n}$ be positive numbers and $(b_i)_{1 \leq i \leq n}$ real numbers. Then the rv’s $a_i(Z_i - b_i)$ are associated.

As immediate other examples of associated sequences, we may cite Gaussian random vectors with nonnegatively correlated components (see Pitt, 1982) and homogenous Markov chains (Daley, 1968).

Demimartingales are set from associated centered variables exactly as martingales are derived from partial sums of centered independent random variables. We have

**Definition 2.** A sequence of rv’s $\{S_n, n \geq 1\} \subset L^1(\Omega, \mathcal{A}, \mathbb{P})$ is a demimartingale if for any $j \geq 1$, for any coordinatewise nondecreasing function $g$ defined on $\mathbb{R}^j$, we have

$$\mathbb{E}( (S_{j+1} - S_j) g(S_1, ..., S_j)) \geq 0.$$  \hspace{1cm} (1)

Two particular cases should be highlighted. First any martingale is a demimartingale. Secondly, partial sums $S_0 = 0$, $S_n = X_1 + ... + X_n$, $n \geq 1$, of associated and centered random variables $X_1, X_2, ...$ form a demimartingale for, in this case, (1) becomes:

$$\mathbb{E}( (S_{j+1} - S_j) g(S_1, ..., S_j)) = \mathbb{E}(X_{j+1} g(S_1, ..., S_j)) = \text{Cov} \{X_{j+1}, g(S_1, ..., S_j)\},$$

since $\mathbb{E}X_{j+1} = 0$. Since $(x_1, ..., x_{j+1}) \mapsto x_{j+1}$ et $(x_1, ..., x_{j+1}) \mapsto g(x_1, ..., x_j)$ are coordinate-wise nondecreasing functions and since the $X_1, X_2, ...$ are associated, we get

$$\mathbb{E}( (S_{j+1} - S_j) g(S_1, ..., S_j)) = \text{Cov} \{X_{j+1} g(S_1, ..., S_j)\} \geq 0.$$
2. Key results for associated sequences

**Lemma 1.** Let \((X, Y)\) be a bivariate random vector such that \(\mathbb{E}(X^2) < \infty\) and \(\mathbb{E}(Y^2) < \infty\). If \((X_1, Y_1)\) and \((X_2, Y_2)\) are two independent copies of \((X, Y)\), then we have

\[ 2 \text{Cov}(X, Y) = \mathbb{E}(X_1 - X_2)(Y_1 - Y_2). \]

We also have

\[ \text{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x, y) dxdy, \]

where,

\[ H(x, y) = \mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y). \]

Before the proof of the lemma, we observe that

\[ H(x, y) = \mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y) = \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y). \]

Indeed we have

\[
\begin{align*}
\mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y) &= \mathbb{E}(\mathbb{I}_{(X>x)}\mathbb{I}_{(Y>y)}) - \mathbb{E}(\mathbb{I}_{(X>x)})\mathbb{E}(\mathbb{I}_{(Y>y)}) \\
&= \text{Cov}(\mathbb{I}_{(X>x)}\mathbb{I}_{(Y>y)}) \\
&= \text{Cov}(1 - \mathbb{I}_{(X>x)}, 1 - \mathbb{I}_{(Y>y)}) \\
&= \text{Cov}(\mathbb{I}_{(X \leq x)}, \mathbb{I}_{(Y \leq y)}) \\
&= \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y).
\end{align*}
\]

**Proof.** We have

\[
\begin{align*}
\mathbb{E}(X_1 - X_2)(Y_1 - Y_2) &= \mathbb{E}(X_1 Y_1) - \mathbb{E}(X_1)\mathbb{E}(Y_1) - \mathbb{E}(X_2 Y_2) + \mathbb{E}(X_2)\mathbb{E}(Y_1) \\
&= 2\mathbb{E}(X_1 Y_1) - 2\mathbb{E}(X_1)\mathbb{E}(Y_1) \\
&= 2\text{Cov}(X_1, Y_1).
\end{align*}
\]

Next, for \(a \in \mathbb{R}\), by Fubini’s Theorem for nonnegative random variables,

\[
\begin{align*}
\int_{a}^{\infty} \int_{a}^{\infty} \mathbb{P}(X > x, Y > y) dxdy &= \mathbb{E} \int_{a}^{\infty} \int_{a}^{\infty} (\mathbb{I}_{(X>x)}\mathbb{I}_{(Y>y)}) dxdy \\
&= \mathbb{E} \left( \int_{a}^{X} dx \int_{a}^{Y} dy \right) \\
&= \mathbb{E}[(X - a)(Y - a)].
\end{align*}
\]

We have

\[
\begin{align*}
2\text{Cov}(X_1, Y_1) &= \mathbb{E}(X_1 - X_2)(Y_1 - Y_2) \\
&= \mathbb{E}[(X_1 - a) - (X_2 - a)](Y_1 - a) - (Y_2 - a))] \\
&= \mathbb{E}(X_1 - a)(Y_1 - a) - \mathbb{E}(X_1 - a)(Y_2 - a) - \mathbb{E}(X_2 - a)(Y_1 - a) + \mathbb{E}(X_2 - a)(Y_2 - a) \\
&= \int_{a}^{\infty} \int_{a}^{\infty} \mathbb{P}(X_1 > x, Y_1 > y) dxdy - \int_{a}^{\infty} \int_{a}^{\infty} \mathbb{P}(X_1 > x, Y_2 > y) dxdy
\end{align*}
\]
Review on asymptotic normality of sums of associated random variables.

\[
- \int_a^\infty \int_a^\infty \mathbb{P}(X_2 > x, Y_1 > y) dx dy + \int_a^\infty \int_a^\infty \mathbb{P}(X_2 > x, Y_2 > y) dx dy.
\]

By the independence of \{X_1, Y_1\} and \{X_2, Y_2\}, \mathbb{P}(X_1 > x, Y_2 > y) = \mathbb{P}(X_1 > x) \times \mathbb{P}(Y_1 > y) and \mathbb{P}(X_2 > x, Y_1 > y) = \mathbb{P}(X_1 > x) \times \mathbb{P}(Y_1 > y),

\[
2\text{Cov}(X, Y) = 2 \left( \int_a^\infty \int_a^\infty \left[ \mathbb{P}(X_1 > x, Y_1 > y) - \mathbb{P}(X_1 > x) \times \mathbb{P}(Y_1 > y) \right] dx dy \right).
\]

We get the final result by letting \( a \to -\infty \).

\[ \square \]

**Lemma 2.** Suppose that \( X, Y \) are two random variables with finite variance and, \( f \) and \( g \) are \( C^1 \) complex valued functions on \( \mathbb{R}^1 \) with bounded derivatives \( f' \) and \( g' \). Then

\[ |\text{Cov}(f(X), h(Y))| \leq ||f'||_{\infty} ||g'||_{\infty} \text{Cov}(X, Y) \]

**Proof.** By Lemma 1, we have

\[
2\text{Cov}(f(X), g(Y)) = \mathbb{E}(f(X_1) - f(X_2))(g(Y_1) - g(Y_2))
= \mathbb{E} \left( \int_{X_1}^{X_2} f'(x) dx \int_{Y_1}^{Y_2} g'(x) dx \right).
\]

But

\[
\int_{X_1}^{X_2} f'(x) dx = \int_{X_1}^{+\infty} f'(x) dx - \int_{X_2}^{+\infty} f'(x) dx
= \int_{\mathbb{R}} f'(x) \{1_{(X_1 \leq x)} - 1_{(X_2 \leq x)}\} dx.
\]

Applying this to \( \int_{Y_1}^{Y_2} g'(x) dx \) and combining all that, leads to

\[
2\text{Cov}(f(X), g(Y)) = \mathbb{E} \int_{\mathbb{R}^2} f'(x)g'(y) \{1_{(X_1 \leq x)} - 1_{(X_2 \leq x)}\} \{1_{(Y_1 \leq y)} - 1_{(Y_2 \leq y)}\} dx dy. \tag{3}
\]

It is easy to see that

\[
\mathbb{E} \{1_{(X_1 \leq x)} - 1_{(X_2 \leq x)}\} \{1_{(Y_1 \leq y)} - 1_{(Y_2 \leq y)}\}
= 2(\mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y))
\]

and by (2), this is equal to \( 2H(x, y) \). By applying Fubini’s theorem in (3), we get

\[
2\text{Cov}(f(X), g(Y)) = 2 \int_{\mathbb{R}^2} f'(x)g'(y)H(x, y) dx dy.
\]

This gives, since \( H(x, y) \geq 0 \) for associated rv’s,

\[
|\text{Cov}(f(X), g(Y))| \leq ||f'||_{\infty} ||g'||_{\infty} \int_{\mathbb{R}^2} H(x, y) dx dy.
\]

And we complete the proof by applying Lemma 1.

\[ \square \]

**Remark:** We used the proof of Yu (1993) here.
Theorem 1. Let $X_1, X_2, \ldots, X_n$ be associated, then we have for all $t = (t_1, \ldots, t_n) \in \mathbb{R}^k$,

$$\left| \psi_{(X_1, X_2, \ldots, X_n)}(t) - \prod_{i=1}^n \psi_{X_i}(t_i) \right| \leq \frac{1}{2} \sum_{1 \leq i < j \leq n} |t_i t_j| \text{Cov}(X_i, X_j).$$ \hspace{1cm} (4)

Proof: First, we prove this for $n = 2$. Use the Newman inequality in Lemma 2. Let $X$ and $Y$ be two associated random variables. For $(s, t) \in \mathbb{R}^2$, put $U = f(X) = e^{isX}$ and $V = g(Y) = e^{itY}$. We have

$$\text{Cov}(U, V) = E(e^{isX + itY}) - E(e^{isX})E(e^{itY}) = \psi_{(X,Y)}(s, t) - \psi_X(s)\psi_Y(t).$$

But Lemma 2 implies

$$|\text{Cov}(U, V)| = |\text{Cov}(f(X), g(Y))| \leq |st| \|f'\|_\infty \|g'\|_\infty \|\text{Cov}(X,Y)\| = |st| |\text{Cov}(X,Y)|.$$

And (4) is valid for $n = 2$. Now we proceed by induction and suppose that (4) is true up to $n$. Consider associated random variables $X_1, X_2, \ldots, X_{n+1}$ and let $t = (t_1, \ldots, t_{n+1}) \in \mathbb{R}^{n+1}$. If all the $t_i$ are nonnegative, we have $U = t_1 X_1 + \ldots + t_n X_n$ and $V = X_{n+1}$ are associated. We have

$$\psi_{(X_1, X_2, \ldots, X_{n+1})}(t) = \psi_{(U,Y)}(1, t_{n+1})$$

and

$$\psi_{(X_1, X_2, \ldots, X_n)}(1) = \psi_{(X_1, X_2, \ldots, X_n)}(t_1, \ldots, t_n).$$

By the induction hypothesis, we have

$$\left| \psi_{(X_1, X_2, \ldots, X_{n+1})}(t) - \psi_{(X_1, X_2, \ldots, X_n)}(t_1, \ldots, t_n)\psi_{X_{n+1}}(t_{n+1}) \right| \leq |t_{n+1}| |\text{Cov}(X_{n+1}, t_1 X_1 + \ldots + t_n X_n)| \leq \frac{1}{2} \sum_{j=1}^n |t_j t_{n+1}| |\text{Cov}(X_{n+1}, X_j)|.$$

Next

$$\left| \psi_{(X_1, X_2, \ldots, X_{n+1})}(t) - \prod_{i=1}^{n+1} \psi_{X_i}(t_i) \right| \leq \left| \psi_{(X_1, X_2, \ldots, X_{n+1})}(t) - \psi_{(X_1, X_2, \ldots, X_n)}(t_1, \ldots, t_n)\psi_{X_{n+1}}(t_{n+1}) \right|$$

$$+ \left| \psi_{(X_1, X_2, \ldots, X_n)}(t_1, \ldots, t_n)\psi_{X_{n+1}}(t_{n+1}) - \prod_{i=1}^{n+1} \psi_{X_i}(t_i) \right|.$$

The first term in the right side member is bounded as in (5). The second term is bounded, due to the induction hypothesis, by

$$|\psi_{X_{n+1}}(t_{n+1})| \left| \psi_{(X_1, X_2, \ldots, X_n)}(t_1, \ldots, t_n) - \prod_{i=1}^n \psi_{X_i}(t_i) \right|$$

Journal home page: www.jafristat.net
Now use that we have proved (4) for \( n \) and get the same conclusion. This means that (4) is true. It remains the case where exactly

\[ \text{Review on asymptotic normality of sums of associated random variables.} \]


\[ \sum_{i=1}^{p} \psi_{X_i} (t_i) \]

and get the same conclusion. This means that (4) is true. It remains the case where exactly

\[ \frac{1}{2} \sum_{1 \leq i \neq j \leq n} |t_i t_j| |\text{Cov}(X_i, X_j)|. \]

(6)

By putting (5) and (6) together, we get that (4) is valid. By re-arranging the \( t_i \), we observe that we have proved (4) for \( n = 3 \), if at least \( n \) of the \( t_i \) are nonpositive. Also, if at least \( n \) of them are nonpositive, we consider the sequence \( -X_1, ..., -X_{n+1} \) that is also associated and get the same conclusion. This means that (4) is true. It remains the case where exactly

\[ \psi_{X_1, X_2, ..., X_n+1}(t) = \psi_{U, -V}(1, -1), \]

we have by the induction hypothesis

\[ \left| \psi_{X_1, X_2, ..., X_n+1}(t) - \psi_U(1) \psi_{-V}(-1) \right| \leq \frac{1}{2} |\text{Cov}(U, -V)| \leq \frac{1}{2} \sum_{i=1}^{p} \sum_{j=p+1}^{n+1} |t_i t_j| |\text{Cov}(X_i, X_j)| \]

(7)

Now use

\[ \left| \psi_{X_1, X_2, ..., X_n+1}(t) - \prod_{i=1}^{n+1} \psi_{X_i} (t_i) \right| \leq \left| \psi_{X_1, X_2, ..., X_n+1}(t) - \psi_U(1) \psi_{-V}(-1) \right| \]

(8)

\[ \text{and get the same conclusion. This means that (4) is true. It remains the case where exactly} \]

\[ \text{a term already handled in (8). The second term is bounded as follows} \]

\[ \left| \psi_U(1) \psi_{-V}(-1) - \prod_{i=1}^{p} \psi_{X_i}(t_i) \psi_{-V}(-1)(t_i) \right| = |\psi_{-V}(-1)(t_i)| \times \left| \psi_U(1) - \prod_{i=1}^{p} \psi_{X_i}(t_i) \right| \]

(9)
where we used the induction hypothesis in the last formula. The last term is
\[ \left| \prod_{i=1}^{p} \psi_{\mathcal{N}}(t_i) - \prod_{i=1}^{n+1} \psi_{\mathcal{N}}(t_i) \right| \leq \left| \prod_{i=1}^{p} \psi_{\mathcal{N}}(t_i) \right| \times \left| \prod_{i=p+1}^{n+1} \psi_{\mathcal{N}}(t_i) \right| \leq \frac{1}{2} \sum_{p+1 \leq i \neq j \leq n+1} |t_i t_j| |\text{cov}(X_i, X_j)|, \tag{10} \]
where we used again the induction hypothesis. We complete the proof by putting (7), (8), (10) and (9) together, we arrive at the result (4). \( \square \)

3. Central limit theorem for a strictly stationary and associated sequence

In this section, we provide all the details of the sharpest result in this topic by Newman and Wright (1981). This came as a concluding paper for a series of papers by Newman.

We present here all the materials used in the proof of Newman and Wright in a detailed writing that makes it better understandable by a broad public.

First, we have this simple lemma.

**Lemma 3.** Let \( X \) and \( Y \) be finite variance random variables such that
\[ E(X, Y 1_{Y \leq 0}) \geq 0. \tag{11} \]
Then, we have
\[ E[(\max(X, X + Y))^2] \leq E(X + Y)^2. \tag{12} \]
If \( X \) and \( Y \) are associated and \( X \) is mean zero, then (11) holds and (12) is true.

**Proof.** We have
\[ \max(X, X + Y)^2 = \left\{ X 1_{Y \leq 0} + (X + Y) 1_{Y > 0} \right\}^2 \]
\[ = X^2 1_{Y \leq 0} + (X + Y)^2 1_{Y > 0} \]
\[ = X^2 1_{Y \leq 0} + X^2 1_{Y \leq 0} + Y^2 + 2XY 1_{Y > 0} \]
\[ = X^2 + Y^2 + 2XY - 2XY 1_{Y \leq 0} - Y^2 1_{Y \leq 0} \]
\[ = (X + Y)^2 - 2XY 1_{Y \leq 0} - Y^2 1_{Y \leq 0} \]
We get the desired result whenever
Now if $X$ and $Y$ are associated, we have

$$X_1(Y \leq 0) = (-X)(-Y)1(-Y \geq 0).$$

Since $(-X)$ and $(-Y)$ are associated too and $1(-Y \geq 0)$ is a nondecreasing function of $(-Y)$, and reminding that $X$ is mean zero, we get that

$$E(X_1(Y \leq 0)) = E((-X)(-Y)1(-Y \geq 0)) = Cov((-X), (-Y)1(-Y \geq 0)) \geq 0.$$ 

□

Theorem 2 (Maximal inequality of Newman and Wright). Let $X_1, X_2, \ldots, X_n$ be associated, mean zero, finite variance, random variables and $M_n = \max(S_1, S_2, \ldots, S_n)$ where $S_n = X_1 + X_2 + \cdots + X_n$, we have

$$E(M_n^2) \leq V(S_n). \quad (13)$$

Proof. Let us prove (13) by induction. It is obviously true for $n = 1$ and for $n = 2$ by Lemma 3. Let us suppose that it is true for $j, 2 \leq j < n$. By putting $L_j = X_2 + \cdots + X_j$, $j \geq 2$, we have

$$M_n = \max(X_1, X_1 + L_2, \ldots, X_1 + L_n) = X_1 + \max(0, L_2, \ldots, L_n).
$$

Also

$$\max(X_1, X_1 + \max(L_2, \ldots, L_n)) = X_1 + \max(0, \max(L_2, \ldots, L_n))$$

We obviously have

$$\max(0, \max(L_2, \ldots, L_n)) = \max(0, L_2, \ldots, L_n).$$

Then

$$EM_n^2 = E\max(X_1, X_1 + \max(L_2, \ldots, L_n))^2$$

Since $X_1$ and $\max(L_2, \ldots, L_n)$ are associated and $X_1$ is mean zero, use Lemma 3 to get

$$EM_n^2 = E\max(X_1, X_1 + \max(L_2, \ldots, L_n))^2 \leq EX_1^2 + E\max(L_2, \ldots, L_n)^2.$$ 

And then, apply (13) on $E\max(L_2, \ldots, L_n)^2$ for $(n - 1)$ mean zero associated rv’s to have

$$E\max(L_2, \ldots, L_n)^2 \leq EX_2^2 + \ldots + EX_n^2.$$ 

We conclude that

$$EM_n^2 \leq EX_1^2 + EX_2^2 + \ldots + EX_n^2.$$ 

□

Lemma 4. Let $X_1, X_2, \cdots, X_n$ be a second-order stationary sequence with $\sigma^2 = V(X_1) + 2 \sum_{j=2}^{\infty} |Cov(X_1, X_j)| < \infty$, then

$$V\left(\frac{S_n}{\sqrt{n}}\right) \to \sigma^2 = V(X_1) + 2 \sum_{j=2}^{\infty} Cov(X_1, X_j).$$
Proof. We have

\[ \alpha_n = V \left( \frac{S_n}{\sqrt{n}} \right) = \frac{1}{n} \left\{ \sum_{j=1}^{n} V(X_i) + \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \right\}. \]

By stationarity, we have

\[ V \left( \frac{S_n}{\sqrt{n}} \right) = V(X_1) + \frac{2}{n} \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) = V(X_1) + \frac{2}{n} \sum_{j=2}^{n} (n - j + 1) \text{Cov}(X_1, X_j). \]

Let \( \epsilon > 0 \). Since \( \sum_{j=2}^{\infty} \text{Cov}(X_1, X_j) < +\infty \), there exists \( K > 0 \) such that for any \( k \geq K \),

\[ \sum_{j \geq k+1} \text{Cov}(X_1, X_j) < \epsilon. \]

We fix that \( k \geq K \) and write,

\[ \alpha_n = V(X_1) + 2 \left( \sum_{j=2}^{k} \left( 1 - \frac{j-1}{n} \right) \text{Cov}(X_1, X_j) + \sum_{j=k+1}^{n} \left( 1 - \frac{j-1}{n} \right) \text{Cov}(X_1, X_j) \right) \]

and observe that

\[ \left| \alpha_n - V(X_1) - 2 \sum_{j=2}^{k} \left( 1 - \frac{j-1}{n} \right) \text{Cov}(X_1, X_j) \right| \leq 2\epsilon. \]

Thus, we get

\[ \liminf V(X_1) + 2 \sum_{j=2}^{k} \left( 1 - \frac{j-1}{n} \right) \text{Cov}(X_1, X_j) - 2\epsilon \leq \liminf \alpha_n \]

\[ \leq \limsup \alpha_n \leq \limsup V(X_1) + 2 \sum_{j=2}^{k} \left( 1 - \frac{j-1}{n} \right) \text{Cov}(X_1, X_j) + 2\epsilon. \]

Therefore, for any \( k \geq K \),

\[ V(X_1) + 2 \sum_{j=2}^{k} \text{Cov}(X_1, X_j) - 2\epsilon \leq \liminf \alpha_n \leq \limsup \alpha_n \]

\[ \leq V(X_1) + 2 \sum_{j=2}^{k} \text{Cov}(X_1, X_j) + 2\epsilon. \]

We finish the proof by letting \( k \to \infty \) and next by letting \( \epsilon \to 0 \). \( \Box \)
Theorem 3. Let \( X_1, X_2, \cdots, X_m \) be a strictly stationary, mean zero, associated random variables such that

\[
\sigma^2 = V(X_1) + 2 \sum_{j=2}^{+\infty} Cov(X_1, X_j) < \infty,
\]

then

\[
\frac{S_n}{\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \rightarrow N(0, \sigma^2) \text{ as } n \rightarrow \infty.
\]

Proof. Let us fix \( \ell > 1 \) an integer and let us set \( m = \lfloor \frac{n}{\ell} \rfloor \), that is \( m\ell \leq n \leq m\ell + \ell \). Let us define \( \Psi_n(r) = E(e^{irS_n/\sqrt{n}}), r \in \mathbb{R} \). First, we have for \( r \in \mathbb{R} \),

\[
|\Psi_n(r) - \Psi_m(r)| = |E(e^{irS_n/\sqrt{n}}) - E(e^{irS_m/\sqrt{m\ell}})|
\]

\[
= \left| E \left[ e^{irS_m/\sqrt{m\ell}} \left( e^{ir(S_n/\sqrt{n})} - (S_m/\sqrt{m\ell}) - 1 \right) \right] \right| 
\]

\[
\leq E \left| e^{ir \left( \frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}} \right) - 1} \right|.
\] (14)

But for any \( x \in \mathbb{R} \),

\[
|e^{ix} - 1| = |(\cos x - 1) + i \sin x| = |2 \sin \frac{x}{2}| \leq |x|.
\]

Thus the second member of (14) is, by the Cauchy-Schwarz’s inequality, bounded by

\[
|r|E \left| \frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}} \right| \leq |r|V \left( \frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}} \right) \frac{1}{2}.
\]

Let us compute the quantity between brackets for fixed \( \ell \) and \( n \rightarrow \infty \) \((m \rightarrow \infty)\), we get

\[
\frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}} = \frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{n}} + \frac{S_m}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}}
\]

\[
= \frac{S_n - S_m}{\sqrt{n}} - \frac{\sqrt{n} - \sqrt{m\ell}}{\sqrt{nm\ell}} S_m
\]

and

\[
\delta_{m,\ell} = V \left( \frac{S_n}{\sqrt{n}} - \frac{S_m}{\sqrt{m\ell}} \right) = V \left( \frac{S_n - S_m}{\sqrt{n}} + \left( \frac{\sqrt{n} - \sqrt{m\ell}}{\sqrt{n}} \right)^2 \right) V \left( \frac{S_m}{\sqrt{m\ell}} \right)
\]

\[
-2 \frac{\sqrt{n} - \sqrt{m\ell}}{\sqrt{nm\ell}} Cov(S_n - S_m, S_m).
\]

\( Cov(S_n - S_m, S_m) \geq 0 \) by association. Thus

\[
\delta_{m,\ell} \leq V \left( \frac{S_n - S_m}{\sqrt{n}} \right) + \left( \frac{\sqrt{n} - \sqrt{m\ell}}{\sqrt{n}} \right)^2 V \left( \frac{S_m}{\sqrt{m\ell}} \right).
\]
Since $0 \leq n - m\ell \leq \ell$, and $\text{Cov}(X_1, X_j) \geq 0$ by association, we have

$$V(S_{n-m\ell}) = \sum_{i=1}^{n-m\ell} V(X_i) + \sum_{1 \leq i \neq j \leq n-m\ell} \text{Cov}(X_i, X_j)$$

$$\leq \sum_{i=1}^{\ell} V(X_i) + \sum_{1 \leq i \neq j \leq \ell} \text{Cov}(X_i, X_j) = A(\ell).$$

Further, $m\ell \leq n \leq (m+1)\ell$ implies

$$0 \leq \frac{\sqrt{n} - \sqrt{m\ell}}{\sqrt{n}} \leq \left(1 - \sqrt{\frac{m\ell}{n}}\right) \to 0 \text{ as } n \to +\infty.$$

Then when $m \to \infty \ (n \to \infty)$

$$V\left(\frac{S_{m\ell}}{\sqrt{m\ell}}\right) \to V(X_1) + 2 \sum_{j=2}^{\infty} \text{Cov}(X_1, X_j) < \infty$$

and

$$\delta_{m,\ell} \leq \frac{A(\ell)}{n} \left(1 - \sqrt{\frac{m\ell}{n}}\right)^2 V\left(\frac{S_{m\ell}}{\sqrt{m\ell}}\right) \to 0.$$

For fixed $\ell$, $n \to \infty$, we get

$$|\Psi_n(r) - \Psi_{m\ell}(r)| \to 0.$$

Now, let us set $Y_j = (S_{j\ell} - S_{(j-1)\ell})/\sqrt{\ell}$, for a fixed $\ell$. By strict stationarity, the $Y_j$'s are associated and identically distributed. Let $\Psi_{\ell}$ be the common characteristic function of $Y_1, \cdots, Y_m$. Furthermore

$$\frac{S_{m\ell}}{\sqrt{m\ell}} = \frac{1}{\sqrt{m\ell}} \sum_{j=1}^{m} (S_{j\ell} - S_{(j-1)\ell}) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_j.$$

According to the Newman’s Theorem (see Theorem 1)

$$\left|\Psi_{m\ell}(r) - \left(\Psi_{\ell}\left(\frac{r}{\sqrt{m}}\right)\right)^m\right| \leq \frac{r^2}{2m} \sum_{1 \leq j \neq k \leq m} \text{Cov}(Y_j, Y_k),$$

and we know that

$$V\left(\sum_{j=1}^{m} Y_j\right) = \sum_{j=1}^{m} V(Y_j) + \sum_{1 \leq j \neq k \leq m} \text{Cov}(Y_j, Y_k).$$

Thus, by using the stationarity again, we get

$$\frac{1}{m} \sum_{1 \leq j \neq k \leq m} \text{Cov}(Y_j, Y_k) = \frac{1}{m} V\left(\sum_{j=1}^{m} Y_j\right) - \frac{1}{m} \sum_{j=1}^{m} V(Y_j)$$

Journal home page: www.jafristat.net
\[
= V \left( \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_j \right) - \frac{1}{m} \sum_{j=1}^{m} V(Y_j)
\]
\[
V \left( \frac{S_{m\ell}}{\sqrt{m\ell}} \right) - V \left( \frac{S_{\ell}}{\sqrt{\ell}} \right) = \sigma^2_{m\ell} - \sigma^2_{\ell},
\]
where for any \( p \geq 2, \)
\[
\sigma^2_p = \frac{1}{p} \sum_{i=1}^{p} V(Y_i) + \frac{1}{p} \sum_{1 \leq i \neq j \leq p} \text{Cov}(Y_i, Y_j)
\]
Now, when \( m \to \infty, \sigma^2_{m\ell} \to \sigma^2 \) and
\[
\left( \Psi_{\ell} \left( \frac{r}{\sqrt{m}} \right) \right)^m \to e^{-\sigma^2 r^2 / 2},
\]
where \( \sigma^2 \) is the common variance of \( Y_j \)'s,
\[
\sigma^2 = \sum_{i=1}^{\ell} V(X_i) + \frac{1}{\ell} \sum_{1 \leq i \neq j \leq m} \text{Cov}(X_i, X_j).
\]
Then it comes out that
\[
\lim_{m \to \infty} \left| \Psi_{m\ell}(r) - e^{-\sigma^2 r^2 / 2} \right| \leq \frac{r^2}{2} (\sigma^2 - \sigma^2).
\]
We complete the proof by letting \( \ell \to \infty \). Thus \( \sigma^2 - \sigma^2 \to 0 \) and we get
\[
\lim_{n \to \infty} \left| \Psi_n(r) - e^{-\sigma^2 r^2 / 2} \right| = 0.
\]

\[\square\]

**Remark.** We finish this exposition by these important facts. A number of CLT’s and invariance principles are available in the literature for strictly stationary sequences of associated random variables and not stationary ones. The most general CLT seems to be the one provided by Cox and Grimmett (1983) for arbitrary associated rv’s fulfilling a number of moment conditions. Dabrowski and co-authors (see Burton et al., 1986 and Dabrowski, 1958) considered weakly associated random variables to establish principle invariances in the lines of Newman and Wright (1981), as well as Berry-Esséntype results and functional LIL’s. But almost all these results use adaptations of the original method of Newman we have described here.

**References**


Journal home page: www.jafristat.net