Proving Newman's CLT via Stein's Method

Bruno Barbosa*, Wasamon Jantai*, Yevgeniy Kovchegov*, and Guochen Xu*

Abstract

In 1980 C. M. Newman has proven the Central Limit Theorem for a sequence of strictly stationary random variables $\{X_n\}_{n\in\mathbb{N}}$ that are linearly positive or negative quadrant dependent with convergent $\sum_{i=2}^{\infty} Cov(X_1, X_i)$. In this paper, we present a new proof of the Newman's CLT based on Stein's method under additional assumptions.

1 Introduction

In the past century, a lot of the effort had been dedicated to proving the Central Limit Theorem (CLT) under a variety of dependence restrictions. C. M. Newman [18, 19, 20] has proven the CLT for the sequences of strictly stationary linearly positive or linearly negative quadrant dependent random variables under finite second moment assumption. In this paper, we will revisit the Newman's CLT, and reprove it using the Stein's method under additional association and finite third moment assumptions. Note that the finite third moment assumption is usual for the Stein's method.

In 1966, Lehmann [14] introduced the notion of positive and negative quadrant dependencies.

Definition 1. *A pair of random variables, X and Y, is* ***positive quadrant dependent (PQD)*** *if*

$$P[X > x, Y > y] - P[X > x]P[Y > y] \geq 0, \ \forall x, y \in \mathbb{R}. \tag{1}$$

Consequently, a pair of random variables, X and Y, is said to be ***negative quadrant dependent (NQD)*** *if X and $-Y$ are positive quadrant dependent.*

There are stronger dependencies defined as follows.

Definition 2. *A sequence of random variables X_n is* ***linearly positive quadrant dependent (LPQD)*** *if for any pair of disjoint sets of indices A and B, and a positive sequence $\{\lambda_i\}$, the random variables $\Sigma_{i\in A}\lambda_i X_i$ and $\Sigma_{j\in B}\lambda_j X_j$ are positive quadrant dependent.*

Similarly, X_n is ***linearly negative quadrant dependent (LNQD)*** *if for any pair of disjoint sets of indices A and B, and a positive sequence $\{\lambda_i\}$, the random variables $\Sigma_{i\in A}\lambda_i X_i$ and $\Sigma_{j\in B}\lambda_j X_j$ are negative quadrant dependent.*

The above defined LPQD and LNQD are weaker types of dependencies than the positive and negative associations that were originally introduced by Esary *at al.* in [10].

*Department of Mathematics, Oregon State University, Corvallis, OR 97331

Definition 3. *A finite set of random variables* $\{X_1, X_2, \ldots, X_n\}$ *is said to be* ***positively associated*** *(or simply,* ***associated****) if for any pair* A *and* B *of subsets of* $\{1, 2, \ldots, n\}$ *the following holds*

$$Cov\big(f(X_i,\ i \in A), g(X_j,\ j \in B)\big) \geq 0$$

for all coordinate-wise increasing functions f *on* $\mathbb{R}^{|A|}$ *and* f *on* $\mathbb{R}^{|B|}$*, whenever the respective covariances exist.*

A finite set of random variables $\{X_1, X_2, \ldots, X_n\}$ *is said to be* **negatively associated** *if for any pair* A *and* B *of subsets of* $\{1, 2, \ldots, n\}$ *the following holds*

$$Cov\big(f(X_i,\ i \in A), g(X_j,\ j \in B)\big) \leq 0$$

for all coordinate-wise increasing functions f *on* $\mathbb{R}^{|A|}$ *and* f *on* $\mathbb{R}^{|B|}$*, whenever the respective covariances exist.*

An infinite sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ *is said to be* **positively associated** *or* **associated** *(respectively,* **negatively associated***) if every finite subfamily is positively associated (respectively, negatively associated).*

Next, we recall the definition of strict stationarity.

Definition 4. *A sequence* $\{X_n\}_{n\in\mathbb{N}}$ *is said to be* ***strictly stationary*** *if for all* $k, n \in \mathbb{N}$*, and all indices* $t_1 < t_2 < \cdots < t_n$ *in* $\mathbb{N}$*, the random vector* $(X_{t_1+k}, X_{t_2+k}, \ldots, X_{t_n+k})$ *has the same joint distribution as* $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$*.*

In 1980, Newman [18] established the following version of CLT for positively associated strictly stationary sequences.

Theorem 1 (**Newman's CLT,**[18, 21])**.** *Consider an LPQD strictly stationary sequence* $\{X_n\}_{n\in\mathbb{N}} \in L^2$ *such that the series* $\sum_{j=2}^{\infty} Cov(X_1, X_j)$ *converges. Then,*

$$\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} \big(X_j - E[X_j]\big) \xrightarrow{d} Z, \tag{2}$$

where

$$\sigma^2 = Var(X_1) + 2\sum_{j=2}^{\infty} Cov(X_1, X_j),$$

Z *denotes the standard normal random variable, and* $\xrightarrow{d}$ *in* (2) *refers to convergence in distribution.*

Moreover, Newman [18] proved a corresponding version of the above CLT in the general context of random fields, $(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{Z}^d)$. The above Theorem 1 corresponds to the case when $d = 1$. In 1983, Newman [20] also extended the CLT result to sequence of functions, not necessarily monotone, of positively associated random variables.

In 1972, Stein [25] introduced what is now known as Stein's method, which is a powerful tool for approximations and convergences of random variables. Specifically, for the standard normal distribution, the characterizing (Stein) operator $\mathcal{A}$ is defined as follows

$$\mathcal{A}f(x) = f'(x) - xf(x).$$

The operator $\mathcal{A}$ is used to characterize the standard normal random variable.

Lemma 1 ([25])**.** *A random variable W is standard normal if and only if $E[\mathcal{A}f(W)] = 0$ for all piecewise continuously differentiable f satisfying $E|f'(Z)| < \infty$ for $Z \sim N(0,1)$.*

The following lemma is a consequence of the classical Sturm–Liouville theory.

Lemma 2. *For any real-valued function g such that $E|g(Z)| < \infty$ for $Z \sim N(0,1)$, there is a function f solving the Stein equation*

$$f'(x) - xf(x) = g(x) - E[g(Z)]. \tag{3}$$

Moreover, if g is Lipschitz, then the solution f of the Stein equation (3) satisfies

$$\|f\|_\infty \le \|g'\|_\infty; \qquad \|f'\|_\infty \le \sqrt{\frac{2}{\pi}}\|g'\|_\infty; \qquad \|f''\|_\infty \le 2\|g'\|_\infty \tag{4}$$

where $\|f\|_\infty := \sup\{|f(x)| : x \in \mathbb{R}\}$.

Then for all function g belonging to a large class of test functions $\mathcal{D}$, there exists a function f belonging to a suitably large class $\mathcal{D}'$ such that $E[g(X) - g(Z)] = E[f'(X) - Xf(X)]$ by taking expectation in (3). It implies that

$$\sup_{h\in\mathcal{D}} E[h(X) - h(Z)] \le \sup_{f\in\mathcal{D}'} E[f'(X) - Xf(X)]. \tag{5}$$

Thus, if $\mathcal{D}$ consists of all 1-Lipschitz functions, then we will need the Wasserstein distance also known as the Kantorovich-Monge-Rubinstein metric defined by

$$d_W(X,Z) = \sup_{h\in\mathcal{D}} \big|E[h(X)] - E[h(Z)]\big|.$$

As the right hand side of (5) is easier to bound than the left hand side of (5), Stein's method can be used for finding the rates of convergence.

Theorem 2 (Theorem 3.1 in [24])**.** *Let W be a random variable and Z the standard normal distribution. Define $\mathcal{D}' = \{f : \|f\|_\infty \le 1, \|f'\|_\infty \le \sqrt{\frac{2}{\pi}}$ and $\|f''\|_\infty \le 2\}$. Then*

$$d_W(W,Z) \le \sup_{f\in\mathcal{D}'} E[f'(W) - Wf(W)]. \tag{6}$$

Therefore, in order to prove $S_n \xrightarrow{d} Z$ it suffices to show that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{D}'} E[f'(S_n) - S_n f(S_n)] = 0.$$

The Stein's method was expanded beyond Gaussian distributions. The so called Stein-Chen method is based on the work of Chen (a Ph.D. student Stein at the time) who adapted the Stein's method for approximating the Poisson distribution. See Chen[6] and Barbour et al. [4]. The Stein's method was later adapted for the Gaussian processes by Barbour [1], the binomial distribution by Ehm [9], the gamma distribution by Luk [16], the uniform distribution by Diaconis [7], the compound Poisson distribution by Barbour et al. [3], Barbour and

Utev [2] and Roos [23], the multinomial distribution by Loh [15], the geometric distribution by Peköz [22], and so forth.

Recently, Goldstein and Wiroonsri [11] established a non-asymptotic bound on the Wasserstein distance between the sum $\sum_{i=1}^{m} X_i$ and the standard normal random variable in case when $\{X_1, X_2, \ldots, X_m\}$ is a finite set of positively associated mean zero random variables satisfying $Var\left(\sum_{i=1}^{m} X_i\right) = 1$ and $|X_i| < B$ for some $B > 0$ and all $i \in \{1, 2, \ldots, m\}$. Later, Wiroonsri [26] extended the results of [11] by replacing the positive association requirement with the negative association, and obtained a non-asymptotic bound on the Wasserstein distance. Both papers, [11] and [26], employ the Stein's method for obtaining their main results, that we state in the theorem below.

Theorem 3 ([11], [26])**.** *Let $\{X_1, X_2, \ldots, X_m\}$ be a finite set of a positively associated (resp. negatively associated) mean zero obeying $|X_i| < B$ for some $B > 0$ for all i. Set $W = \sum_{i=1}^{m} X_i$ and additionally assume that the variance of W is 1. Then, in the positively associated case,*

$$d_W(W, Z) \leq 5B + \sqrt{\frac{8}{\pi}} \sum_{i \neq j} Cov(X_i, X_j)$$

and, in the negatively associated case,

$$d_W(W, Z) \leq 5B - 5.2 \sum_{i \neq j} Cov(X_i, X_j)$$

where Z is the standard normal random variable.

Note that the above theorem provides a Berry-Esséen type bound for the Wasserstein metric in the case of independent and identically distributed mean zero bounded random variables X_i.

In the original Newman's CLT (1980), the proof uses the characteristic function of the sum $S_n = X_1 + X_2 + \cdots + X_n$ In this paper, we will prove a version of Newman's CLT for a positively associated strictly stationary processes via Stein's method.

2 Preliminaries

The pointwise ergodic theorem is a generalization of the strong law of large numbers (SLLN). We state the ergodic theorem, preceded by a few necessary definitions.

Definition 5. *A measure-preserving transformation $(\Omega, \mathcal{F}, P, T)$ is defined by a the probability space $(\Omega, \mathcal{F}, P)$, and a measurable transformation $T : \Omega \to \Omega$, for which $P(E) = P(T^{-1}(E))$ for all $E \in \mathcal{F}$.*

An event E is invariant under a transformation T if $E = T^{-1}(E)$, where equality is defined if the symmetric difference has measure 0. The collection of all T-invariant sets forms a σ-algebra, denoted by $\mathcal{I}$.

Definition 6. *A measure preserving transformation $(\Omega, \mathcal{F}, P, T)$ is said to be ergodic if $\mathcal{I}$ is trivial, that is, if $P(A) = 0$ or 1 for all $A \in \mathcal{I}$.*

Theorem 4. *(Ergodic theorem) Let $(\Omega, \mathcal{F}, P, T)$ be a measure preserving system. For any random variable $X \in L^p(\Omega, \mathcal{F}, P)$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ T^k \to E[X|\mathcal{I}],$$

where the convergence is a.s. and in L^p and $\mathcal{I}$ is the set of all T-invariant sets. If the system is ergodic, then

$$\frac{1}{n} \sum_{k=0}^{n-1} X \circ T^k \to E[X]$$

a.s. and in L^p as $n \to \infty$.

The *a.s.* convergence in Theorem 4 is called the pointwise ergodic theorem and the L^2 convergence in Theorem 4 is known as the mean ergodic theorem. The mean ergodic theorem is established by John von Neumann [17] and the pointwise ergodic theorem is established by George Birkhoff [5].

Recall that for a strictly stationary sequence X_j,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Cov(X_1, X_j) = 0. \tag{7}$$

is the necessary and sufficient condition for

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = E[X_1] \quad \text{in } L^2. \tag{8}$$

See [27, 12].

It is known (see [21]) that (7) is also a necessary and sufficient condition for a strictly stationary sequence of associated random variables with finite second moment to be ergodic. This statement appeared implicitly in Lebowitz [13], and later formalized in Theorem 7 of [21]. By Thm. 4 the ergodicity is a stronger property than L^2 convergence in (8).

Theorem 5 (Theorem 7 in [21])**.** *Let $\{X_n\}$ be a strictly stationary sequence which is either associated or negatively associated and let T be the usual shift transformation. Then T is ergodic if and only if (7) is satisfied. Consequently (Thm. 4), if (7) is satisfied, then for any function $f(x)$ such that $f(X_1) \in L^1$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_j) = E[f(X_1)] \quad \text{a.s. and in } L^1. \tag{9}$$

Curiously Thm. 5 gives a partial answer to the question asked on pp. 20-21 in Yaglom [27].

Remark 1. *The following result shows that for any stationary sequence, we can associate in a measure preserving system.*

Theorem 6. *If $\{X_n\}$ is strictly stationary then there exists a measure preserving transformation T such that $X_n(x) = X_1(T^{n-1}(x))$. Indeed, T is the shift transformation.*

Finally, we will need the following auxiliary results. First, Lemma 3.3 in [11] establishes the sufficient conditions under which a pair of random variables is PQD.

Lemma 3 (Lemma 3.3 in [11]). *A pair* $(X, Y) = (\phi(\xi), \varphi(\xi))$ *is PQD whenever* $\xi = (X_1, \ldots, X_n)$ *is positively associated and* $\phi(x_1, \ldots, x_n)$ *and* $\varphi(x_1, \ldots, x_n)$ *are coordinate wise increasing functions of* ξ.

Then, Lemma 3 in [18] presents an important upper bound for the covariance.

Lemma 4 (Lemma 3 in [18]). *Suppose* X *and* Y *are either PQD or NQD random variables with finite variances, and* f, g *are* C^1 *complex valued functions over* $\mathbb{R}$*, with bounded derivatives* f', g'*. Then*

$$|Cov(f(X), g(Y))| \leq \|f'\|_\infty \|g'\|_\infty \, |Cov(X, Y)| \, , \tag{10}$$

where $\|\cdot\|_\infty$ *denotes the sup norm on* $\mathbb{R}$.

3 Results

In this section, we present a new proof of the Newman's CLT which is based on applying Stein's Method. It is the main result of this paper, and appears in Theorem 7.

3.1 Main Result

The results from this section are derived for a sequence of random variables $\{X_k\}_{k\in\mathbb{N}}$ satisfying the following hypotheses.

Assumption 1. *Sequence* $\{X_k\}_{k\in\mathbb{N}}$ *satisfies the following conditions.*

(a) $\{X_k\}_{k\in\mathbb{N}}$ *is a strictly stationary sequence of positively associated random variables with finite third moment.*

(b) $\sum_{j=2}^{\infty} Cov(X_1, X_j) < \infty$.

Next, we formulate the main result of this paper. Its proof invokes a series of lemmas that will be presented in Subsection 3.2 that follows.

Theorem 7. *Suppose the sequence* $\{X_k\}_{k\in\mathbb{N}}$ *satisfies Hypotheses 1. Then*

$$\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} \left(X_j - E[X_j]\right) \xrightarrow{d} Z,$$

where

$$\sigma^2 = Var(X_1) + 2\sum_{j=2}^{\infty} Cov(X_1, X_j),$$

and Z *denotes the standard normal random variable.*

Proof. Without loss of generality, assume $E[X_k] = 0$. We will use the following notations. Denote $W = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i$. Next, we let $\mathcal{I}_n = \{1, 2, \ldots, n\}$, and for a given $i \in \mathbb{Z}$ and $K \in \mathbb{Z}_+$, define

$$\mathcal{J}_{i,K} = \{i - K, \ldots, i - 1, i, i + 1, \ldots, i + K\}.$$

Finally, for $i \in \mathcal{I}_n$, we define

$$Y_{i,K} = \sum_{j \in \mathcal{I}_n \cap \mathcal{J}_{i,K}} X_j \quad \text{and} \quad W_{i,K} = W - \frac{Y_{i,K}}{\sigma\sqrt{n}}.$$

The term $E[X_i Y_{i,K}]$ will appear many times in our calculations. For $K \in \mathbb{Z}_+$ and $i \in \{K + 1, \ldots, n - K\}$, we set

$$\gamma_K = E[X_i Y_{i,K}] = Var(X_1) + 2 \sum_{j=2}^{K+1} Cov(X_1, X_j). \tag{11}$$

Because of Theorem 2, it is sufficient to establish that

$$\lim_{n\to\infty} \sup_{f \in \mathcal{D}'} \left| E\left(f'(W) - Wf(W)\right) \right| = 0.$$

Lemma 6 from Subsection 3.2 implies

$$\begin{aligned} \Big| E[f'(W)] - E[Wf(W)] \Big| &\le \left| \frac{1}{n} \sum_{i=1}^{n} \left(\frac{E[X_i Y_{i,K}]}{\sigma^2} - 1 \right) E\left[f'(W_{i,K})\right] \right| \\ &+ \left| \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} E\left[X_i f(W_{i,K})\right] \right| + \left| \frac{1}{\sigma^2 n} \sum_{i=1}^{n} Cov\left(X_i Y_{i,K},\ f'(W_{i,K})\right) \right| + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{12}$$

Now, we will show that each term on the right hand side of (12) converges to zero. Observe that $E[X_i Y_{i,K}] = \gamma_K \to \sigma^2$ as $K \to \infty$ and f' is bounded. Thus, the first term on the right hand side of (12) goes to zero because

$$\limsup_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^{n} \left(\frac{E[X_i Y_{i,K}]}{\sigma^2} - 1 \right) E\left[f'(W_{i,K})\right] \right| \le \left(1 - \frac{\gamma_K}{\sigma^2}\right) \|f'\|_\infty \longrightarrow 0 \ \text{ as } K \to \infty.$$

In order for us to bound the second term on the right hand side in (12), we note that the variables X_i and $W_{i,K}$ are coordinate wise increasing functions of $(X_1, \ldots, X_n)$. By Lemma 3, the pair $(X_i, W_{i,K})$ is PQD, and therefore, Lemma 4 implies

$$\left| \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} E\left[X_i f(W_{i,K})\right] \right| = \left| \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} Cov\left(X_i, f(W_{i,K})\right) \right| \le \frac{\|f'\|_\infty}{\sigma\sqrt{n}} \sum_{i=1}^{n} Cov\left(X_i, W_{i,K}\right). \tag{13}$$

Next, the bound in (13) can be rewritten as

$$\begin{aligned}\frac{\|f'\|_\infty}{\sigma\sqrt{n}}\sum_{i=1}^n Cov\left(X_i, W_{i,K}\right) &= \frac{\|f'\|_\infty}{\sigma\sqrt{n}}\sum_{i=1}^n E[X_iW] - \frac{\|f'\|_\infty}{\sigma^2 n}\sum_{i=1}^n \gamma_k + \mathcal{O}\left(\frac{K}{n}\right)\\ &= \|f'\|_\infty E\left[\sum_{i=1}^n \frac{X_iW}{\sigma\sqrt{n}}\right] - \frac{\|f'\|_\infty}{\sigma^2 n}\sum_{i=1}^n \gamma_k + \mathcal{O}\left(\frac{K}{n}\right)\\ &= \|f'\|_\infty\left(E[W^2] - \frac{\gamma_k}{\sigma^2}\right) + \mathcal{O}\left(\frac{K}{n}\right)\\ &\longrightarrow \|f'\|_\infty\left(1-\frac{\gamma_K}{\sigma^2}\right) \text{ as } n\to\infty, \end{aligned} \tag{14}$$

where (14) follows from Corollary 1, and finally,

$$1-\frac{\gamma_K}{\sigma^2}\longrightarrow 0 \text{ as } K\to\infty.$$

The last term of the right hand side in (12) can be bounded as follows:

$$\begin{aligned}\left|\sum_{i=1}^n Cov\left(\frac{X_iY_{i,K}}{\sigma^2 n}, f'(W_{i,K})\right)\right| &\le \left|Cov\left(\sum_{i=1}^n \frac{X_iY_{i,K}}{\sigma^2 n}, f'(W)\right)\right|\\ &\quad + \left|\sum_{i=1}^n Cov\left(\frac{X_iY_{i,K}}{\sigma^2 n}, f'(W_{i,K}) - f'(W)\right)\right|\\ &\le 2\|f'\|_\infty\left(E\left[\left|\sum_{i=1}^n \frac{X_iY_{i,K}}{\sigma^2 n} - \frac{\gamma_K}{\sigma^2}\right|\right]\right)\\ &\quad + \frac{1}{\sigma^2 n}\sum_{i=1}^n \left|Cov\left(X_iY_{i,K}, f'(W_{i,K}) - f'(W)\right)\right|, \end{aligned} \tag{15}$$

where the last inequality follows from the Hölder inequality. Next, we observe that in (15),

$$\begin{aligned}&\frac{1}{\sigma^2 n}\sum_{i=1}^n \left|Cov\left(X_iY_{i,K}, f'(W_{i,K}) - f'(W)\right)\right|\\ &\qquad \le \frac{\|f''\|_\infty}{\sigma^3 n\sqrt{n}}\sum_{i=1}^n E\left[|X_i|\, Y_{i,K}^2\right] \longrightarrow 0 \quad \text{as } N\to\infty. \end{aligned} \tag{16}$$

Let $V_i = X_iY_{i,K}$ for $i\in\{K+1,\dots,n-K\}$. Since $\{X_n\}_{n=1}^\infty$ is stationary, then $\{V_i\}_{i=1}^\infty$ is stationary for all $i\in\{K+1,\dots,n-K\}$. Letting $n>K$, we have

$$\frac{1}{n}\sum_{i=1}^n X_iY_{i,K} - \gamma_K = \frac{1}{n}\sum_{i=K+1}^{n-K} V_i - \gamma_K + \sum_{\mathcal{I}_n\setminus\{K+1,\dots,n-K\}} \frac{X_iY_{i,K}}{n}.$$

Consequently, using the fact that $|a+b|^p \leq 2^p(|a|^p + |b|^p)$ for $p > 0$, we obtain

$$E\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i Y_{i,K} - \gamma_K\right|\right] \leq 2\sqrt{2}\, E\left[\left|\frac{1}{n}\sum_{i=K+1}^{n-K} V_i - \gamma_K\right|\right] + 2\sqrt{2}\, E\left[\left|\sum_{\mathcal{I}_n \setminus \{K+1,\dots,n-K\}} \frac{X_i Y_{i,K}}{n}\right|\right] = 2\sqrt{2}\, E\left[\left|\frac{1}{n-2K}\sum_{i=K+1}^{n-K} V_i - \gamma_K\right|\right] + \mathcal{O}\left(\frac{1}{n\sqrt{n}}\right).$$

By Lemma 8, we have $\lim_{n\to\infty} E\left[\left|\frac{1}{n-2K}\sum_{i=K+1}^{n-K} V_i - \gamma_K\right|\right] = 0$, which implies

$$\lim_{n\to\infty} E\left[\left|\frac{1}{n}\sum_{i=1}^{n} V_i - \gamma_K\right|\right] = 0.$$

Therefore, the right hand side in (15) goes to zero as $n \to \infty$, which completes the proof. $\square$

3.2 Auxiliary Results

Lemma 5. *In the set-up of the proof of Thm. 7 we have*

$$E\left[f'\left(W_{i,K}\right)\right] = E\left[f'\left(W\right)\right] + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

and

$$E\left[X_i f\left(W_{i,K} + \frac{Y_{i,K}}{\sigma\sqrt{n}}\right)\right] = E\left[X_i f(W_{i,K})\right] + \frac{1}{\sigma\sqrt{n}} E\left[X_i Y_{i,K} f'(W_{i,K})\right] + \frac{1}{\sqrt{n}}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Using Lagrange's form of the remainder term, there exists c between $W_{i,K}$ and W such that

$$f\left(W_{i,K} + \frac{Y_{i,K}}{\sigma\sqrt{n}}\right) = f\left(W_{i,K}\right) + f'\left(W_{i,K}\right)\frac{Y_{i,K}}{\sigma\sqrt{n}} + f''(c)\frac{Y_{i,K}^2}{2\sigma^2 n}.$$

Hence, taking the expectations, we obtain

$$E\left[f'\left(W\right)\right] = E\left[f'\left(W_{i,K} + \frac{Y_{i,K}}{\sigma\sqrt{n}}\right)\right] = E\left[f'(W_{i,K})\right] + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

Next, since f'' is bounded and X_i has finite third moment, we have

$$E\left[X_i f\left(W_{i,K} + \frac{Y_{i,K}}{\sigma\sqrt{n}}\right)\right] = E\left[X_i f\left(W_{i,K}\right)\right] + \frac{1}{\sigma\sqrt{n}} E\left[X_i Y_{i,K} f'(W_{i,K})\right] + \frac{1}{\sqrt{n}}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

$\square$

Lemma 6. *In the set-up of the proof of Thm. 7 we have*

$$E[f'(W)] - E[Wf(W)] = -\frac{1}{\sigma^2 n}\sum_{i=1}^{n} Cov\left(X_i Y_{i,K},\, f'(W_{i,K})\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E[X_i Y_{i,K}]}{\sigma^2} - 1\right) E\left[f'(W_{i,K})\right] - \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} E\left[X_i f(W_{i,K})\right] + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{17}$$

Proof. By Lemma 5,

$$E[f'(W)] - E[Wf(W)] = \frac{1}{n}\sum_{i=1}^{n} E\left[f'(W_{i,K})\right] - \frac{1}{\sigma^2 n}\sum_{i=1}^{n} E\left[X_i Y_{i,K} f(W_{i,K})\right] - \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} E\left[X_i f(W_{i,K})\right] + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{18}$$

Equation (17) is obtained by rearranging the terms in the right hand side of (18). □

The following is a result from [21], which implies Corollary 1, used in the proof of the main theorem.

Lemma 7 ([21])**.** *Consider a stationary sequence $\{X_n\}_{n\in\mathbb{N}}$ of positively associated random variables. Then, letting $\sigma_n = Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)$, we have $\sigma^2 = \lim_{n\to\infty} \sigma_n$.*

Corollary 1. *In the set-up of the proof of Thm. 7 we have $\lim_{n\to\infty} E[W^2] = 1$.*

Proof. Since $\{X_i\}_{i\in\mathbb{N}}$ is stationary and $E(X_i) = 0$ for all $i \in \mathbb{N}$, then $E[X_i^2] = Var(X_1)$ and $E[X_i X_{i+k}] = Cov(X_1, X_k)$, for $k \geq 1$. Thus,

$$E[W^2] = E\left[\frac{1}{\sigma^2 n}\left(\sum_{i=1}^{n} X_i\right)^2\right] = \frac{\sigma_n}{\sigma^2}.$$

Then, $\lim_{n\to\infty} E(W^2) = 1$ follows from Lemma 7. □

Lemma 8. *In the set-up of the proof of Thm. 7, for a given integer $K \geq 0$, we have*

$$\lim_{n\to\infty} E\left[\left|\frac{1}{n-2K}\sum_{i=1+K}^{n-K} V_i - \gamma_K\right|\right] = 0 \tag{19}$$

where $V_i = X_i Y_{i,K}$ and $\gamma_K = E[V_i]$ for $i \in \{K+1, \ldots, n-K\}$.

Proof. Theorem 5 implies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 = E[X_1^2] \quad a.s. \tag{20}$$

Since $\{X_n + X_{n+K}\}$ is a stationary sequence and $\sum_{j=1}^{\infty} Cov(X_1, X_j) < \infty$, we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Cov(X_1 + X_{1+K}, X_i + X_{i+K}) = 0,$$

and once again, by Theorem 5,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_i + X_{i+K})^2 = E[(X_1 + X_{1+K})^2] \quad a.s. \tag{21}$$

Since $X_i X_{i+K} = \frac{1}{2}\big((X_i + X_{i+K})^2 - X_i^2 - X_{i+K}^2\big)$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i X_{i+K} = E[X_1 X_{1+K}] \quad a.s. \tag{22}$$

Next, since $\{V_i\}_{i\in\{K+1,\ldots,n-K\}}$ is a stationary sequence, we have

$$\sum_{i=1+K}^{n-K} V_i = \sum_{j=-K}^{K} \sum_{i=1+K}^{n+K} X_i X_{i+j}.$$

By formula (22),

$$\lim_{n\to\infty} \frac{1}{n-2K} \sum_{i=1+K}^{n-K} V_i = \sum_{j=-K}^{K} E[X_{1+K} X_{1+j+K}] = E[V_{K+1}] = \gamma_K \quad a.s. \tag{23}$$

Observe that, by the Pointwise Ergodic Theorem (Thm. 4),

$$\lim_{n\to\infty} \frac{1}{n-2K} \sum_{i=1+K}^{n-K} V_i = E[V_{k+1}|\mathcal{I}] \quad a.s.$$

where $\mathcal{I}$ is the class of invariant events. Since $E\left[|V_{K+1}|\right] < \infty$, Theorem 4 implies

$$\lim_{n\to\infty} \frac{1}{n-2K} \sum_{i=1+K}^{n-K} V_i = E[V_{k+1}|\mathcal{I}] \quad \text{ in } L_1. \tag{24}$$

By the equivalence of the limits in (23) and (24), one obtains $E[V_{k+1}|\mathcal{I}] = \gamma_K$. Hence, (24) yields

$$\lim_{n\to\infty} \frac{1}{n-2K} \sum_{i=1+K}^{n-K} V_i = \gamma_K \text{ in } L_1.$$

□

Acknowledgements

The authors would like to thank Ed Waymire, Enrique Thomann, and Radu Dascaliuc for helpful comments. This research was supported by FAPESP award 2018/07826-5 and by NSF award DMS-1412557.

References

[1] A. D. Barbour, *Stein's method for diffusion approximations*, Probab. Th. Rel. Fields. **84** (1990), no. 3, 297–322. **MR3765894**

[2] A. D. Barbour and S. Utev, *Compound Poisson approximation in total variation*, Stoch. Process. Their Appl. **82** (1999), no. 4, 89–125. **MR3765894**

[3] L. H. Y. Chen, A. D. Barbour, and W-L. Loh, *Compound Poisson approximation for nonnegative random variables via Stein's method*, Stoch. Process. Their Appl. **20** (1992), no. 4, 1843–1866. **MR3765894**

[4] L. Holst, A. D. Barbour, and S. Janson, *Poisson approximation*, Oxford University Press, Oxford, 1992. **MR1163825**

[5] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. **17** (1931), 656–660.

[6] L. H. Y. Chen, *Poisson approximation for dependent trials*, Ann. Probab. **3** (1975), no. 3, 534–545. **MR3765894**

[7] P. Diaconis, *An example for Stein's method*, 384, Stanford. Stat., 1991.

[8] R. Durrett, *Probability: theory and examples*, fourth ed., Cambridge Series in Statistical and Probabilistic Mathematics, vol. 31, Cambridge University Press, Cambridge, 2010. **MR2722836**

[9] W. Ehm, *Binomial approximation to the Poisson binomial distribution*, Stat. Probab. Lett. **11** (1991), no. 1, 7–16. **MR3765894**

[10] J. D. Esary, F. Proschan, and D. W. Walkup, *Association of random variables, with applications*, Ann. Math. Statist. **38** (1967), 1466–1474. **MR0217826**

[11] L. Goldstein and N. Wiroonsri, *Stein's method for positively associated random variables with applications to the Ising and voter models, bond percolation, and contact process*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 1, 385–421. **MR3765894**

[12] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, second ed., Academic Press, New York-London, (1975) **MR0356197**

[13] J. L. Lebowitz, *Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems*, Comm. Math. Phys. **28** (1972), 313–321. **MR0323271**

[14] E. L. Lehmann, *Some concepts of dependence*, Ann. Math. Statist. **37** (1966), 1137–1153. **MR0202228**

[15] W.-L. Loh, *Stein's method and multinomial approximation*, Ann. Appl. Probab. **2** (1992), no. 3, 536–554. **MR3765894**

[16] H. Luk, *Stein's method for the gamma distribution and related statistical applications*, Ph.D. thesis, University of Southern California, 1994.

[17] J. v. Neumann, *Proof of the quasi-ergodic hypothesis*, Proceedings of the National Academy of Sciences **18** (1932), no. 1, 70–82.

[18] C. M. Newman, *Normal fluctuations and the FKG inequalities*, Comm. Math. Phys. **74** (1980), no. 2, 119–128. **MR576267**

[19] C. M. Newman and A. L. Wright, *An invariance principle for certain dependent sequences*, Ann. Probab. **9** (1981), no. 4, 671–675. **MR624694**

[20] C. M. Newman, *A general central limit theorem for FKG systems*, Comm. Math. Phys. **91** (1983), no. 1, 75–80. **MR719811**

[21] C. M. Newman, *Asymptotic independence and limit theorems for positively and negatively dependent random variables*, Inequalities in statistics and probability (Lincoln, Neb., 1982), IMS Lecture Notes Monogr. Ser., vol. 5, Inst. Math. Statist., Hayward, CA, 1984, pp. 127–140. **MR789244**

[22] E. Peköz, *Stein's method for geometric approximation*, J. Appl. Probab. **33** (1996), no. 3, 707–713. **MR3765894**

[23] M. Roos, *Stein's method for compound Poisson approximation: the local approach*, Ann. Appl. Probab. **4** (1994), no. 4, 1177–1187. **MR3765894**

[24] N. Ross, *Fundamentals of Stein's method*, Probab. Surv. **8** (2011), 210–293. **MR2861132**

[25] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. **II**: Probability theory, Univ. California Press, Berkeley, Calif., (1972), 583–602. **MR0402873**

[26] N. Wiroonsri, *Stein's method for negatively associated random variables with applications to second-order stationary random fields*, J. Appl. Prob. **55** (2018), no. 1, 196–215. **MR3765894**

[27] A. M. Yaglom, *An introduction to the theory of stationary random functions*, (Revised English edition. Translated and edited by Richard A. Silverman) Prentice-Hall, Inc., Englewood Cliffs, N.J. (1962) **MR0184289**