International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n3.25
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Applications of the CLT for Positively Associated Random Process in Time Series Analysis

Zongmin Ma

School of Equipment Management & Support, Engineering University of PAP, Xi'an, 710086, Shaanxi, China

ABSTRACT

This paper explores the implications of the Central Limit Theorem (CLT) within the framework of positively associated stationary random fields, which are pivotal in mathematical statistics, reliability theory, percolation, and statistical physics. It delves into the challenges of extending the CLT's convergence rate to these complex fields, building upon the foundational work by Newman and subsequent contributions. The study presents a novel approach to constructing ARMA models that align with the CLT, offering a robust framework for time series analysis. The paper concludes with the significance of these findings for statistical modeling and forecasting in various disciplines.

KEYWORDS

CLT, Positively Associated random process, Time Series, ARMA model

1. INTRODUCTION

The Central Limit Theorem (CLT) is a cornerstone of probability theory, providing a powerful tool for analyzing the behavior of sums of random variables. While the CLT has traditionally been applied to independent and identically distributed (i.i.d.) variables, the real-world phenomena often exhibit dependencies that cannot be captured by the i.i.d. assumption. This paper addresses this gap by investigating the role of the CLT in the context of positively associated random fields, which are characterized by their non-negative covariances and are prevalent in various applications such as signal processing, image analysis, and network theory.

The paper is structured as follows: after this introduction, we delve into the theoretical underpinnings of positively associated random fields and their relevance to the CLT. We then present a detailed analysis of the convergence rates and conditions under which the CLT holds for these fields. Subsequently, we construct and discuss ARMA models that are consistent with the CLT, demonstrating their practical utility in time series analysis. Finally, we conclude with a summary of our findings and potential implications for future research and applications.

There are a number of stochastic models involving families of dependent random variables and within the framework of such models the main limit theorems of Probability Theory have been established. In this regard, we refer, e.g., to [1]. The Central Limit Theorem (CLT) for independent identically distributed (i.i.d.) random variables, with its upper and lower bounds established by Berry-Esseen, has been a cornerstone of probability theory. However, extending the rate of convergence of the CLT to associated random fields presents considerable challenges. Newman [2] first proved the CLT for associated random fields, formulating necessary and sufficient conditions for its application, although finite susceptibility does not always hold. Subsequent work by Wood [3] provided Berry-Esseen type estimates Δ_n for associated random fields with a rate of $O(n^{-1/5})$, and Birkel [4] achieved a rate of

$O(n^{-1/2}\log^2 n)$ whenever the associated process has a finite third moment, and the Cox-Grimmett coefficient (see [5])

$$u(n)=\sup_{k\in\mathbb{N}}\sum_{j:|j-k|\geq n}\mathrm{Cov}(X_j,X_k)=O(e^{-\lambda n}) \tag{1}$$

for some $\lambda>0$ as $n\to\infty$.

For random fields with finite susceptibility, the Stein-Tihomilov method has been used by Bulinski [3] to derive Berry-Esseen type estimates for positively and negatively associated random fields.

Let $X=\{X_j, j\in\mathbb{Z}^d\}$ be a PA or NA centered random field such that for some $s\in(2,3]$ and $\lambda>0$ such that $D_s=\sup_j \mathrm{E}|X_j|^s<\infty$ and the general Cox-grimmett coefficient

$$u_k=\sup_{r\in\mathbb{Z}^d}\sum_{j\in\mathbb{Z}^d,|j-r|\geq k}|\mathrm{Cov}(X_j,X_r)|=O(k^{-\lambda}),\ \lambda>0 \tag{2}$$

as $k\to\infty$ Then

$$\Delta_n\leq\frac{C|U_n|}{(\mathrm{var}\,S_n)^{(s-\gamma)/2}} \tag{3}$$

with any $\gamma>\gamma^*=(1/(s-1)+(\lambda+d)/(ds(s-2))^{-1}$ and $C>0$ is independent of n.

2. ORGANIZATION OF THE TEXT

2.1. CLT for Positively Associated Random Process

Definition 1. For $n\in\mathbb{N}$, let $\mathrm{M}(n)$ denote the class of real-valued bounded coordinate-wise nondecreasing Borel functions on $\mathbb{R}^n$, then a family $X=\{X_t,t\in T\}$ is associated (abbreviated to A), if for every finite set $I\subset T$ and any functions $f,g\in\mathrm{M}(|I|)$, one has

$$\mathrm{Cov}(f(X_I),g(X_I))\geq 0 \tag{4}$$

Definition 2. A family X is weakly associated, or positively associated PA if

$$\mathrm{Cov}(f(X_I),g(X_J))\geq 0 \tag{5}$$

for all disjoint finite sets $I,J\subset T$ and all functions $f\in\mathrm{M}(|I|),g\in\mathrm{M}(|J|)$ such that covariance exists.

Theorem 1. Suppose $\{X_t,t\in\mathbb{N}\}$ is a centered stationary PA process with finite third moments,as $n\to\infty$, the rate of convergence for covariance function $\gamma_n=\mathrm{Cov}(X_t,X_{t+n})$ tends to $O(e^{-\lambda n})$, then

$$\frac{S_n-ES_n}{\sqrt{\sum_{i=1}^{n}\mathbf{E}X_i^2}}\Rightarrow N(0,1) \tag{6}$$

where $S_n:=\sum_{i=1}^{n}X_i$, and $\Rightarrow$ means convergence in distribution.

Proof: Given that $\{X_t,t\in\mathbb{N}\}$ is a centered stationary PA process, as $n\to\infty$, one obtains

$$u(n)=\sum_{|j-s|\geq n} Cov(X_j,X_s)=2\sum_{k=n+1}^{\infty} Cov(X_1,X_k)=2\sum_{k=n}^{\infty}\gamma_k \tag{7}$$

Taking into account $O(e^{-\lambda n})$, there exists a constant C such that as $n\to\infty$, one has $\gamma_n \leq Ce^{-\lambda n}$, leading to

$$u(n)=2\sum_{k=n}^{\infty}\gamma_k \leq 2C\sum_{k=n}^{\infty}e^{-\lambda k}=2C(e^{-\lambda n}+e^{-\lambda(n+1)}+e^{-\lambda(n+2)}+\cdots)=\frac{2Ce^{-\lambda n}}{1-e^{-\lambda}}=C^{*}e^{-\lambda n} \tag{8}$$

where $C^{*}=2C/(1-e^{-\lambda})$ is a constant, and which proves the desired result that as $n\to\infty$ $u(n)=O(e^{-\lambda n})$. Further more, as $n, n'<n$ tends to infinity,

$$\frac{\mathrm{var}S_n}{\sum_{i=1}^{n}EX_i^2}=\frac{\sum_{i=1}^{n}EX_i^2+2\sum_{k=2}^{n}\gamma_k}{\sum_{i=1}^{n}EX_i^2}=\frac{\sum_{i=1}^{n}EX_i^2+2\sum_{k=2}^{n'}\gamma_k+2\sum_{k=n'}^{n}\gamma_k}{\sum_{i=1}^{n}EX_i^2}=\frac{\frac{\sum_{i=1}^{n}EX_i^2}{n}+\frac{2(n'-2)\tilde{\gamma}}{n}+\frac{2\sum_{k=n'}^{n}\gamma_k}{n}}{\frac{\sum_{i=1}^{n}EX_i^2}{n}} \tag{9}$$

where n', n are also tends to 0, but $\lim_{n\to\infty} n/n'=\infty$, $\tilde{\gamma}=\max\gamma_k$, then one obtains

$$1\leq\frac{\mathrm{var}S_n}{\sum_{i=1}^{n}EX_i^2}\leq\frac{\frac{\sum_{i=1}^{n}EX_i^2}{n}+\frac{2(n'-2)\tilde{\gamma}}{n}+\frac{2\sum_{k=n'}^{n}\gamma_k}{n}}{\frac{\sum_{i=1}^{n}EX_i^2}{n}}\sim 1 \tag{10}$$

ensuring the application of the CLT, yielding the desired result.Therefore by means of CLT, one obtains the desired results.

Corollary: as a colloary of Theorem 1, it may generates the classical CLT for independent identical distributed random variables as the covariance function is 0, which satisfies the condition in Theorem 1. Which means $\frac{S_n-ES_n}{\sqrt{\mathrm{var}\,S_n}}\Rightarrow N(0,1)$.

2.2. Construction of ARMA Models satisfying CLT

ARMA models are widely used in economics for time series analysis, the typical ARMA(p,q) models are given below:

$$X_t=\sum_{i=1}^{p}\phi_i X_{t-i}+\varepsilon_t+\sum_{j=1}^{q}\theta_j\varepsilon_{t-j} \tag{11}$$

where $\{X_t, t\in\mathbb{N}\}$ is a time series, $\phi_i, i=1,\cdots,p$ are the coeffients, together with $X_{t-1},...,X_{t-p}$ called Auto Regression Parts, ε_t are i.i.d. white noise with $\theta_j, j=1,\cdots,q$ together consists of the Moving Average Parts.

When ARMA(p, q) models satisfy the unit root test and stationary conditions, and suppose that the ARMA(p, q) time series model $\{X_t, t\in\mathbb{N}\}$ satisfies $\mathrm{Cov}(X_1, X_{1+k})=O(e^{-\lambda k})$, one satisfies the CLT.

Model 1: Next we considers AR(1) models to illustrate the application of the CLT. For a random process $\{X_t, t \in \mathbb{N}\}$, AR(1) models can be described as follows:

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{12}$$

where ρ is a coefficient, ε_t is a white noise with mean 0 and variance σ_ε^2. Assume $\{X_0\}$ has an arbitrary distribution with mean μ and variance σ^2. Firstly one should determinate the ρ, ε_t in order to hold the conditions that the random process is stationary, PA and satisfies $\mathrm{Cov}(X_t, X_{t+k}) = O(e^{-\lambda k})$ for some $\lambda > 0$ as $k \to \infty$ that

$$\mathrm{Cov}(X_t, X_{t+k}) = \rho \mathrm{Cov}(X_t, X_{t+k-1}) = \cdots = \rho^k \mathrm{Cov}(X_t, X_t) = \rho^k \sigma^2 \tag{13}$$

is only dependent of k, and is easy to hold the $\mathrm{Cov}(X_t, X_{t+k}) = O(e^{-\lambda k})$ whenever $\rho = C^{1/k} e^{-\lambda}$, for convenience, one may set $C = 1$.

To ensure stationarity and PA, we normalize the process by $Y_t = X_t - \mu$, resulting in:

$$X_t = \mu(1 - e^{-\lambda}) + e^{-\lambda} X_{t-1} + \varepsilon_t \tag{14}$$

Without loss of generality, the foregoing models already satisfies the CLT conditions, we further assume that $\{X_t, t \in \mathbb{N}\}$ holds the homogeneity of variance, then the variance of ε_t is determined by:

$$\mathrm{var}(X_t) = \mathrm{var}(\mu(1 - e^{-\lambda}) + e^{-\lambda} X_{t-1} + \varepsilon_t) = e^{-2\lambda} \mathrm{var}(X_{t-1}) + \mathrm{var}(\varepsilon_t) \tag{15}$$

which means

$$\sigma^2 = e^{-2\lambda}\sigma^2 + \sigma_\varepsilon{}^2 \Rightarrow \sigma_\varepsilon{}^2 = (1 - e^{-2\lambda})\sigma^2 \tag{16}$$

Now, the foregoing models is stationary PA with constant expectation μ, which leads to

$$X_t = \mu(1 - e^{-\lambda}) + e^{-\lambda} X_{t-1} + \varepsilon_t, \tag{17}$$

where $\varepsilon \sim N(0, (1 - e^{-2\lambda})\sigma^2)$ and the initial random variable X_0 comes from an arbitrary distribution with mean 0 and variance σ^2, $\lambda > 0$, $\{\varepsilon_t, t \in \mathbb{N}\}$ are white noise satisfies the CLT conditions.

Model 2: Suppose the random process $\{X_t, t \in \mathbb{N}\}$ has the same condition like in Model 1 unless $\mathrm{Cov}(X_t, X_{t+k}) = O(k^{-\lambda})$ for some $\lambda > 0$ as $k \to \infty$, it is obviously that $\mathrm{Cov}(X_t, X_{t+k}) = \rho^k \sigma^2$ is only dependent of k. As we know for any $\alpha > 0$ and $a > 1$, $\lim_{n \to \infty} \frac{n^\alpha}{a^n} = 0$, for any constant $\rho < 1$, one obtain $\gamma(k)$ decreases rapidly than $O(k^{-\lambda})$. Similar to the same assumption of the homogeneity of variance, one obtains

$$\mathrm{var}(X_t) = \mathrm{var}(\mu(1 - \rho) + \rho X_{t-1} + \varepsilon_t) = \rho^2 \mathrm{var}(X_{t-1}) + \mathrm{var}(\varepsilon_t) \tag{18}$$

which means

$$\sigma^2 = \rho^2 \sigma^2 + \sigma_\varepsilon{}^2 \Rightarrow \sigma_\varepsilon{}^2 = (1 - \rho^2)\sigma^2 \tag{19}$$

Thus,

$$X_t = \mu(1 - \rho) + \rho X_{t-1} + \varepsilon_t, \tag{20}$$

$\varepsilon_t \sim N(0, (1 - \rho^2)\sigma^2)$

3. CONCLUSION

The research presented in this paper has successfully bridged the gap between the classical CLT and its application to positively associated random fields. By constructing ARMA models that adhere to the CLT, we have provided a comprehensive framework for time series analysis, which can be utilized across various fields. The findings underscore the importance of considering the dependencies within random fields when applying the CLT, thereby enhancing the accuracy and reliability of statistical inferences. Future work could extend this framework to other types of random fields and further refine the convergence rates for practical applications.

REFERENCES

[1] Bulinski, A.V., et al. (2007) Limit theorems for associated random fields and related systems, World Scientific, volume 10.

[2] Newmann, C.M. (1980) Normal fluctuations and the FKG inequalities. Communications in Mathematical Physics, 74(2):119–128.

[3] Wood, T.E. (1983) A Berry-Esseen theorem for associated random variables. The Annals of Probability, 1042–1047.

[4] Birkel, T. (1988). On the convergence rate in the central limit theorem for associated processes. The Annals of Probability, 1685–1698.

[5] Cox, J.T., Grimmett, G. (1984) Central limit theorems for associated random variables and the percolation model.The annals of probability, 514–528.