Statistical Applications of CLT for Dependent Data

Martial Longla*

Abstract

This paper presents a survey of central limit theorems for dependent data, emphasizing on cases when the dependence is not quantified. New central limit theorems are provided for some time series examples with better properties than known results (when mixing is assumed). The given results are used to estimate the mean and provide confidence intervals for the mean of several populations. Several statistical models are considered and tests are provided to show the importance of the results. Some of these theorems use the CLT of Kipnis and Varadhan for reversible Markov chains and other results use the Lindeberg's condition for arrays of independent data. The use of a smoothing kernel allows us to prove a theorem that provides confidence intervals for an ARFIMA model without explicit use of the fractional difference parameter or its estimate. Several setups are used to illustrate the use of the results. While developing these concepts, we use simulations to show that even when some assumptions of the theorems are violated, the estimators that are proposed still perform well on large samples. We provide some comparisons that can help applied statisticians and encourage the use of these methods. Several statistical models are considered.

Key Words: Keywords: Reversible Markov chains, long range dependence, central limit theorem, ARFIMA Models, Dependence, Testing hypotheses.

1. Introduction

1.1 Motivation and Problem Setup

We often need to have inference and do some hypothesis testing for the mean of a population. To address these questions, it is common to either assume a model for the data or use non-parametric methods. But the most difficult challenge comes when the data is not independent. Perhaps, one usually asks if anything is "really independent". Statistical applications lack tools for analyzing data with dependence. Most of the available results rely on various assumptions on the dependence structure of the sequence of observations, making their use very difficult because of the need to quantify properly the dependence to check the underlying conditions. Without a proper central limit theorem, it is very hard to do anything other than estimating the parameter of interest of any problem. We try here to propose an alternative to usual notions that are heavily based on assumptions of independence or some measures of dependence. In the work Longla and Peligrad (2018) is proposed a new central limit theorem. The proposed central limit theorem is applied here to various statistical models and the outputs are compared with known results. It is assumed that we have a stationary and ergodic sequence $(Y_i, i \in Z)$ with finite variance ($\text{var}(Y_0) = \sigma_Y^2 < \infty$). Let $\mu_Y = EY_0$. The sample mean $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}S_n^Y$ (by the Birkhoff ergodic theorem) satisfies $\lim_{n\to\infty} \bar{Y}_n = \mu_Y$. Without other information on the dependence structure of $(Y_i)_{i\in Z}$, obviously, such a sequence might not satisfy the central limit theorem. Therefore, it is impossible to use to use common methods for estimation

*University of Mississippi, University Ave, University, 38677 MS

or testing problems. The new central limit theorem was inspired by the Nadaraya-Watson estimators. This theorem doesn't require a dependence structure for the sequence of Y observations, but rather introduces a new independent sequence of random variables that allows to control for the unknown structure of the dependence in the Y observations. We propose a study of the performance of the given theorem in applications on large and relatively small sample data sets.

1.2 Structure of the paper

In the first section of this paper, we provide the purpose of the work. In section 2 we provide necessary definitions and notations. In section 3 we propose a review of central limit theorems and in section 4 we have applications to various statistical models. Section 5 covers some simulations and discussion. The appendix of SAS codes id proposed in section 6.

2. Defintions, notations and useful notions

For the purpose of this paper, as in Longla and Peligrad (2018), we shall say that a sequence $(Y_i)_{i\in Z}$ has long range dependence if $\text{var}(S_n^Y)/n \to \infty$ and short range if $\text{var}(S_n^Y)$ behaves linearly in n.

Given a sample $(X_i, Y_i)_{1\le i\le n}$ from a random vector (X, Y) on a probability space (Ω, K, P), the well-known Nadaraya-Watson estimator (see Nadaraya (1964) and Watson (1964), or pages 126-127 in Härdle (1991)) is defined by

$$\hat{m}_n(x) = \frac{1}{nh_n\hat{f}_n(x)}\sum_{i=1}^{n} Y_i K(\frac{1}{h_n}(X_i - x)),$$

where

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{1}{h_n}(X_i - x)).$$

Longla and Peligrad (2018) mention that under various smoothness assumptions on (X, Y) and various dependence assumptions on the process $(X_i, Y_i)_{i\in Z}$, the speed of convergence of $\hat{m}_n(x)$ to the conditional mean of Y was pointed out in numerous papers. The dependence structure considered in the literature is rather restrictive, of the weak dependence type, such as mixing conditions, function of mixing sequences or martingale-like conditions. They mention for instance results in Bradley (1983), Collomb (1984), Peligrad (1992), Yoshihara (1994), Bosq (1996), Bosq et.al. (1999), Long and Qian (2013), and Hong and Linton (2016) among many others.

2.1 Definitions and notations

We assume that $(\xi_n, n \in \mathbb{Z})$ is a stationary Markov chain defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a general state space $(S, \mathcal{A})$. The marginal distribution is denoted by $\pi(A) = \mathbb{P}(\xi_0 \in A)$. Assume there is a regular conditional distribution for ξ_1 given ξ_0 denoted by $Q(x, A) = \mathbb{P}(\xi_1 \in A|\, \xi_0 = x)$. Let Q also denote the Markov operator acting via $(Qf)(x) = \int_S f(s)Q(x, ds)$. Next, let $\mathbb{L}_0^2(\pi)$ be the set of measurable functions on S such that $\int f^2 d\pi < \infty$ and $\int f d\pi = 0$. For some function $f \in\mathbb{L}_0^2(\pi)$, let

$$X_i = f(\xi_i),\ S_n = \sum_{i=1}^{n} X_i,\ \sigma_n = (\mathbb{E}S_n^2)^{1/2}. \qquad (1)$$

Denote by $\mathcal{F}_k$ the σ–field generated by ξ_i with $i \leq k$.

For any integrable random variable X we denote $\mathbb{E}_k(X) = \mathbb{E}(X|\mathcal{F}_k)$. Under this notation, $\mathbb{E}_0(X_1) = (Qf)(\xi_0) = \mathbb{E}(X_1|\xi_0)$. We denote by $||X||_p$ the norm in $\mathbb{L}_p(\Omega, \mathcal{F}, \mathbb{P})$.

All throughout the chapter $\Rightarrow$ denotes weak convergence, $[x]$ is the integer part of x and $\to^{\mathbb{P}}$ denotes convergence in probability. The notation $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$; $a_n = o(b_n)$ means $a_n/b_n \to 0$ as $n \to \infty$. By conditional convergence in distribution, denoted by $Y_n|\mathcal{F}_0 \Rightarrow Y$, we understand that for any function g which is continuous and bounded

$$\mathbb{E}_0(g(Y_n)) \to^{\mathbb{P}} \mathbb{E}g(Y) \text{ as } n \to \infty.$$

In other words, let $\mathbb{P}^x$ be the probability associated with the Markov chain started from x and let $\mathbb{E}^x$ be the corresponding expectation. Then, for any $\varepsilon > 0$

$$\mathbb{P}\{x : |\mathbb{E}^x g(Y_n) - \mathbb{E}g(Y)| > \varepsilon\} \to 0.$$

We also use the notion of slowly varying function in the following sense.

Definition 1 A function h defined on positive integers is said to be slowly varying in the strong sense if there exist a continuous function $f : (0, \infty) \to (0, \infty)$ such that $h(n) = f(n), \forall n \in \mathbb{N}$, and $\forall t > 0, \lim_{x \to \infty} f(tx)/f(x) = 1$.

Definition 2 Let $(\Omega, \mathscr{K}, \mathbb{P})$ be a probability space. A measurable function $T : \Omega \to \Omega$ is said to be measure-preserving if for all $A \in \mathscr{K}$, $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$.

For any measure-preserving transformation T and any random variable Y, $(X_n = YoT^{n-1}, n \in \mathbb{N})$ is a strictly stationary stochastic process. On the other hand, any strictly stationary stochastic process $X = (X_1, X_2, \cdots)$ can be determined by the left shift transformation on $\mathbb{R}^\infty$ defined by $T((x_1, x_2, \cdots)) = (x_2, x_3, \cdots)$.

Definition 3 Let T be a measure-preserving transformation an Y a random variable on $(\Omega, \mathscr{K}, \mathbb{P})$. The invariant σ-field of the stationary process defined by T and Y is $\mathscr{I} = \{A \in \mathscr{K} : T^{-1}(A) = A\}$

Definition 4 The partial sum S_n of random variables $X_1, \cdots, X_n$ is said to be attracted (respectively, partially attracted) to a distribution D as $n \to \infty$, if there are sequences of real numbers a_n and b_n with $a_n > 0$, $a_n \to \infty$, such that

$$\frac{S_n - b_n}{a_n} \Rightarrow D \quad \text{as} \quad n \to \infty \quad \text{(respectively, along a subsequence of those integers).}$$

Definition 5 A distribution function F is stable if its characteristic function is of the form

$$\phi(t) = e^{itd - c|t|^a(1 + ib\frac{t}{|t|}\omega(t,a))}, \text{where} \quad 0 < a \leq 2, -1 \leq b \leq 1, c \geq 0, d \in \mathbb{R} \quad \text{and}$$

$$\omega(t, a) = \begin{cases} \tan \frac{\pi a}{2} & \text{if} \quad a \neq 1 \\ -\frac{2}{\pi} \ln |t| & \text{if} \quad a = 1. \end{cases}$$

The number a is called the exponent of the stable distribution F. The normal distribution is a stable distribution with exponent $a = 2$, and the Cauchy distribution is a stable distribution with exponent $a = 1$. A stable distribution is an infinitely-divisible distribution.

Definition 6 A cumulative distribution F on the real line is said to be infinitely divisible if for every positive integer n, there exist n independent identically distributed random variables $X_{n1}, \cdots, X_{nn}$, whose sum S_n has the distribution F.

These notions are often used in proofs of central limit theorems. When dealing with partial sums of random variables, the following Lindeberg condition is often used as in Longla and Peligrad (2018).

Condition 1 The sequence of random variables $(X_n, n \in \mathbb{N})$ is said to satisfy the Lindeberg condition if for every $\varepsilon > 0$, $\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \mathbb{E}X_i^2 \mathbb{I}(|X_i| > \varepsilon\sigma_n) = 0$.

2.2 Martingales

The theory of martingales is undoubtedly one of the most important topics in probability theory. Its use in the proof of convergence theorems is due to the martingale decomposition of sums of random variables. A martingale is a mathematical model for a fair wager. It takes its name from "la grande martingale", the strategy for even-odds bets in which one doubles the bet after each loss, trying to recover his loss on the first win. To model this mathematically, keep track of the succession of fortune after each bet $(X_i, i \in \mathbb{N})$. Having fortune X_i at the time we place the i^{th} bet, we will know certain things, including our fortune at the time (information, σ-algebra of the past and present), but we will not know the result of the bet (future). Our winnings at the i^{th} bet are $(X_i - X_{i-1}, 1 \le i \le n)$. The martingale requirement is that the expected winning knowing the past is zero (see Williams (1991) for more on Martingales).

Definition 7 Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let X_k be $\mathscr{F}_k$-measurable, where $\mathscr{F}_k \subset \mathscr{F}_{k+1} \subset \mathscr{F}$. The sequence $(\mathscr{F}_k, k \in \mathbb{N})$ is called a filtration. Moreover, if X_k is integrable and $\mathbb{E}(X_{k+1}|\mathscr{F}_k) = X_k$, then $(X_k, \mathscr{F}_k)$ is called a martingale; and if X_k is integrable and $\mathbb{E}(X_{k+1}|\mathscr{F}_k) = 0$, then $(X_k, \mathscr{F}_k)$ is called a sequence of martingale differences.

2.3 Tightness conditions for random measures

An important step in the proof of functional central limit theorems is the use of tightness conditions. Tightness conditions ensure continuity of every limiting process. To prove that a sequence of S-valued random variables X_n with distributions μ_n converges weakly to a limiting random variable X with distribution μ, the first step is to show that $\{\mu_n, n \in \mathbb{N}\}$ is a relatively compact subset of $\mathscr{M}_1(S)$, equipped with the topology of weak convergence (called weak topology). This means that any sequence from the set $\{\mu_n, n \in \mathbb{N}\}$ has a weakly convergent subsequence. The second step is then to show that all such weak limits are equal to μ. For more on this topic, see Billingsley (1968) and Durrett (1996). Theorem 8.1 of Billingsley (1968) reads as follows.

Theorem 2 Let $(\mathbb{P}, \mathbb{P}_n, n \in \mathbb{N})$ be probability measures on $(\mathbb{C}[0,1], \mathscr{C})$. If $\{\mathbb{P}_n, n \in \mathbb{N}\}$ is tight and finite dimensional distributions of $\mathbb{P}_n$ converge weakly to those of $\mathbb{P}$, then $\mathbb{P}_n \Rightarrow \mathbb{P}$.

3. Central limit theorems

A result from Ibragimov (1962) states that if the sequence $(X_n, n \in \mathbb{N})$ is a strictly stationary strongly mixing sequence, then S_n can be attracted only by a stable law. Moreover, if this law has exponent α, then $a_n = n^{1/\alpha}h(n)$ with $h(n)$ a slowly varying function. Cogburn (1960) showed, that S_n can be partially attracted only to infinitely divisible laws.

3.1 Theorems related to mixing conditions

This section contains the definitions and a short list of results based on mixing coefficients. We provide results including a functional central limit theorem for an additive functional associated to a Metropolis-Hastings algorithm, with the variance of partial sums behaving asymptotically like $nh(n)$ (where h is a slowly varying function). We will use CLT as short for central limit theorem throughout the rest of this paper. We introduce the historical background linking mixing coefficients to the theory of CLT for functionals of stationary Markov chains as in Longla (2013). A special situation where analogs of the classical results are valid without additional assumptions is Theorem 1.1 from Kipnis and Varadhan (1986) that reads as follows.

Theorem 3 (A functional CLT for martingale differences) Let $(X_j, j \in \mathbb{Z})$ be a stationary ergodic process such that $E[X_{n+1}|\mathcal{F}_n] = 0$ a.e., where $\mathcal{F}_n$ is the σ-field generated by $(X_j, j \leq n)$. For such a martingale difference sequence $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_{[nt]})$ converges weakly to the Brownian motion with variance σ^2, provided $E[X^2] = \sigma^2 < \infty$.

Earlier, Ibragimov (1975) proved the following.

Theorem 4 Suppose the stationary process $(X_i, i \in \mathbb{N})$ satisfies ρ-mixing. Let $\mathbb{E}(X_0) = 0$, $\mathbb{E}(|X_0|^{2+\delta}) < \infty$ for some $\delta > 0$, and let $\sigma_n^2 \to \infty$. Then, the random variables S_n/σ_n are asymptotically normal and the functional CLT holds.

The problem in applications is that we don't have the value of σ_n. Therefore, it is difficult to use this result, even when this variance is known up to a parameter. A review of central limit theorems can be found in the survey by Jones (2004), where Corollary 2 states consequences of existing theorems for Harris ergodic Markov chains from which the following holds.

Corollary 1 Suppose $(X_i, i \in \mathbb{N})$ is a Harris ergodic Markov chain with stationary distribution π. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $\mathbb{E}_\pi f(X_0) = 0$. Assume one of the following conditions holds.

1. $(X_i, i \in \mathbb{N})$ is geometrically ergodic and $\mathbb{E}|f|^{2+\delta}(X) < \infty$ for some $\delta > 0$;
2. $(X_i, i \in \mathbb{N})$ is polynomially ergodic of order m, $\mathbb{E}_\pi M < \infty$ and $\mathbb{E}|f|^{2+\delta}(X) < \infty$, where $m\delta > 2+\delta$ and M is a nonnegative function such that $||P^n(x, .) - \pi(.)|| \leq M(x)n^{-m}$; or
3. $(X_i, i \in \mathbb{N})$ is polynomially ergodic of order $m > 1$, $\mathbb{E}_\pi M < \infty$ and there exists $B < \infty$ such that $|f(X)| < B$ π-almost surely.

Then, for any initial distribution of X_0, as $n \to \infty$, the functional CLT holds.

For ρ-mixing sequences, the conditions are less restrictive. This is due to the fact provided on page 190 of Ibragimov and Rozanov (1978) that we combine with a result of Shao Qi-Man (1989) in the following.

Theorem 5 Let $(X_n, n \in \mathbb{N})$ be a second order centered stationary sequence. If $\sum_{n=1}^{\infty} \frac{\rho_n}{n} < \infty$, then X_n has a continuous spectral density $f(\lambda)$. If, in addition, $f(0) \neq 0$, then $\sigma_n^2 = 2\pi f(0) n(1 + o(1))$. Moreover, the functional CLT holds.

The proof of the CLT was provided by Ibragimov (1975) and the proof of the functional CLT was given by Shao Qi-Man (1989). Ibragimov (1975) has also proved the following. According to this result, for any stationary ergodic Markov chain generated by copulas from the Frechet family (see Loongla (2015)), the CLT holds. Same as for Theorem 4 the difficulty lies in finding the value of the variance of partial sums. In Theorem 5, the variance of partial sums is of order n but has a parameter $f(0)$ that is in general not easy to find. We will show below that using a different estimator for the mean of a sample solves this issue for some sequences. For ϕ-mixing sequences, using $W_n(t) = n^{-1/2}\sigma_n^{-1} S_{[nt]}$, Peligrad (1985) has shown the following.

Theorem 6 Let $(X_n, n \in \mathbb{N})$ be a centered stationary second order ϕ-mixing sequence with $\sigma_n^2 \to \infty$. Assume the Lindeberg Condition 1 is satisfied, then $W_n(t) \Rightarrow W(t)$. If, in addition, $\phi_1 < 1$, then $W_n(t) \Rightarrow W(t)$ implies Condition 1.

Considering reversible Markov chains that we use as examples, Theorem 6 is equivalent to the following.

Theorem 7 Let $(X_n, n \in \mathbb{N})$ be a centered stationary ergodic second order sequence with $\sigma_n^2 \to \infty$. If $\phi_1 < 1$, then the Lindeberg Condition 1 holds if and only if $W_n(t) \Rightarrow W(t)$.

Peligrad and Utev (2006) also present a results on CLT and functional CLT. See Peligrad (2006) for a recent on the invariance principles (functional CLT) for stationary processes.

3.2 Theorems related to reversible Markov chains

A Markov chain is reversible if the operator its transition probabilities induce on $\mathbb{L}_2$ is self-adjoint. In other words, the Markov chain is called reversible if $Q = Q^*$, where Q^* is the adjoint operator of Q. The condition of reversibility is equivalent to requiring that (ξ_0, ξ_1) and (ξ_1, ξ_0) have the same distribution. For reversible Markov chains, theorems require less assumptions than in general. One of the classic theorems in this case is due to Kipnis and Varadhan (1986), who have shown the following:

Theorem 8 For any reversible stationary Markov chain $(\xi_j, j \in \mathbb{Z})$ defined on a state $\mathcal{X}$ with distribution π, and for any mean zero function f such that $\int f^2(x)\pi(dx) < \infty$ and $n^{-1}E(f(\xi_1) + \cdots + f(\xi_n))^2 \to \sigma_f^2 < \infty$, the reversible Markov chain defined by (1) satisfies the functional CLT.

Remark 1 If f is a vector valued function, then essentially the same result is valid. The vector valued process $W_n(t)$ converges weakly to a multidimensional Brownian motion with the corresponding covariance matrix.

Many authors have looked at this case. Zhao et al. (2010) analyzed the case when $\sigma_n^2 = nh(n)$, with h a slowly varying function. They have shown by example that the conditional distribution of $S_n/\sqrt{Var(S_n)}$ need not converge to the standard normal distribution in this case; and have developed sufficient conditions for convergence to a (possibly non-standard) normal distribution. In addition, they provided an example of reversible Markov chain satisfying (??), for which the central limit theorem holds with a different normalization. Some results consider mixing properties as that of Roberts and Rosenthal (1997):

Theorem 9 Let $(\xi_i, i \in \mathbb{N})$ be a reversible geometrically ergodic Markov chain with stationary distribution π, $(X_i = f(\xi_i), i \in \mathbb{Z})$ with $\mathbb{E}_\pi f(X_0) = 0$ and $\mathbb{E}_\pi f^2(X_0) < \infty$. Then for any initial distribution of ξ_0, as $n \to \infty$, $n^{-1/2} S_n \Longrightarrow N(0, \sigma_f^2)$.

Note that Theorem 9 uses geometric ergodicity and reversibility. These two conditions imply exponential ρ-mixing (see Longla and Peligrad (2012)), which implies convergence of $Var(S_n)/n$ to σ_f^2. So, the assumptions of this theorem are stronger than those of Kipnis and Varadhan (1986). There is a considerable amount of papers that further extend and apply this result to many models. Kipnis and Landim (1999) applied it to interacting particle systems, Tierney (1994) considered applications to Markov Chain Monte Carlo. Wu (1999), Zhao and Woodroofe (2008) tackled the law of the iterated logarithm, Derriennic and Lin (2001) and Cuny and Peligrad (2012) considered the CLT started at a point for a stationary irreducible and aperiodic Markov chain with uniform marginal distribution. This type of Markov chain is interesting since it can easily be transformed into Markov chains with different marginal distributions. Longla, Peligrad and Peligrad (2012) pointed out a functional CLT under a normalization other than the variance of partial sums. Markov chains of this type are often studied in the literature from different points of view(see Doukhan et al (1994), Rio (2000 and 2009) , Merlevède and Peligrad (2013)). Recently, Longla (2017) has shown that for a reversible Markov chain of the form $(Y_i = X_i\beta + \varepsilon_i, 1 \leq i \leq n)$, where ε_i is a mean zero reversible and ergodic Markov chain that is independent of the random sample of $(X_i, 1 \leq i \leq n)$ the following results holds:

Theorem 10 The least squares estimator of the slope in the given linear regression model $\hat{\beta}$ satisfies the CLT in the form

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, \frac{\sigma_\varepsilon^2}{\sigma_x^2}), \quad \text{where} \quad \hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}. \tag{2}$$

In the proof of Theorem 10, it was actually shown that for any stationary reversible and ergodic Markov chain $(\varepsilon_i, 1 \leq i \leq n)$, and any independent random sample $(X_i, 1 \leq i \leq n)$ such that $E(X) = 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i \to N(0, \sigma_x^2 \sigma_\varepsilon^2). \tag{3}$$

If we consider $\varepsilon_i = Z_i - \mu_Z$ for any reversible square integrable Markov chain with mean μ_z, $(Z_i, 1 \leq i \leq n)$, the following holds.

Theorem 11 For any ergodic reversible square integrable Markov chain $(Z_i, 1 \leq i \leq n)$, and for any independent mean zero random sample $(X_i, 1 \leq i \leq n)$, as $n \to \infty$, using $\hat{\mu}_z = \frac{1}{n\bar{X}} \sum_{i=1}^{n} X_i Z_i$ as estimator of μ_z,

$$\sqrt{n}\bar{X}(\hat{\mu}_z - \mu_z) \to N(0, \sigma_x^2 \sigma_z^2). \tag{4}$$

Note that $\hat{\mu}_z$ is an unbiased estimator of μ_z. The following assertion is based on consistency of $\bar{X}$ for random samples.

Remark 2 It is also true that for the modified estimator $\hat{\mu}_z^* = \frac{1}{n\mu_x}\sum_{i=1}^n X_i Z_i$ when $\mu_x \neq 0$ and $var(\sum_1^n Z_n)/n \to \sigma_z^2$ as $n \to \infty$, under the conditions of Theorem 4 the central limit theorem holds in the form

$$\sqrt{n}(\hat{\mu}_z^* - \mu_z) \to N(0, \sigma_z^2). \tag{5}$$

These estimators $\hat{\mu}_z$ and $\hat{\mu}_z^*$ are unbiased estimators of μ_z. The asymptotic variance variance of $\hat{\mu}_z^*$ is of order σ_x^2 while that of $\hat{\mu}_z$ is a more complex. In Remark 2, we have a case to which applies the Kipnis and Varadhan central limit theorem. We will provide a comparison of the these estimators to $\bar{Z}$.

3.3 General theorems

In this section, we are concerned with central limit theorems for sequences for which the dependence structure is not defined in advance.

3.3.1 Mean estimation

Here, we are concerned with the problem of mean estimation in absence of modelled dependence. Assume that we have a population $(X_i, 1 \leq i \leq n)$ of dependent observations. We are interested in estimating the mean and build a confidence interval for the estimate or do some tests of hypotheses. We consider that K is a symmetric, bounded density function and

$$\sqrt{nh_n}(\bar{Y}_n - \mu_Y) \to^P 0, \tag{6}$$

$$nh_n^5 \to 0 \text{ and } nh_n \to \infty \text{ as } n \to \infty, \tag{7}$$

which is implied by

$$nh_n \text{var}(\bar{Y}_n) \to 0. \tag{8}$$

Note that we can always find a sequence $(h_n)_{n\geq 1}$ satisfying both conditions (7) and (8), provided $\text{var}(\bar{Y}_n) \to 0$. In this case, Longla and Peligrad (2018) have shown the following:

Theorem 12 Assume that $(Y_i, 1 \leq i \leq n)$ is a stationary and ergodic sequence with finite second moments and conditions (7) and condition (6) are satisfied. Also assume that K satisfies Condition A and that $(X_i, 1 \leq i \leq n)$ is an i.i.d. sequence of random variables, independent of Y, having a bounded density function $f(x)$, continuous and differentiable at the origin, with $f(0) \neq 0$. Then we have

$$\frac{\sqrt{nh_n}}{\sqrt{\overline{Y_n^2}}}(\hat{r}_n - \mu_Y) \Rightarrow N(0, \frac{1}{f(0)}\int K^2(x)dx).$$

where $\overline{Y_n^2} = \sum_{i=1}^n Y_i^2/n$ and $\hat{r}_n = \frac{1}{nhf(0)}\sum_{i=1}^n K(\frac{X_i}{h})Y_i$.

They have also shown that under the conditions $\mathrm{var}(\bar{Y}_n) = o(n^{-4/5})$ and $\mu_Y \neq 0$, the optimal bandwidth to be used in the confidence intervals is

$$h_o = [\frac{f(0)B\overline{Y_n^2}}{n(f"(0)A)^2(\bar{Y}_n)^2}]^{1/5}, \quad A = \int x^2K(x)dx \quad \text{and} \quad B = \int K^2(x)dx.$$

This theorem is in the spirit of Longla (2017) of works to provide a way to construct tools for analysis of data when the dependence structure is not known or easily quantifiable.

For applications, the following tables were provided in Longla and Peligrad (2018).

Kernel	Gaussian	Epanechnikov	Uniform	Quartic
$A\|B$	$1\|1/(2\sqrt{\pi})$	$1/5\|3/5$	$1/3\|1/2$	$1/7\|5/7$

Table 1: Values of A and B for various kernels

In Table 1 we have values of A and B for each of the provided kernels.

Distribution	Gaussian	$\chi^2(2)$	Cauchy
$f(0)\|\|f''(0)\|$	$1/\sqrt{2\pi}\|1/\sqrt{2\pi}$	$1/2\|1/8$	$1/\pi\|2/\pi$

Table 2: Values of $f(0)$ and $|f''(0)|$ for various distributions

Table 2 discloses the values of $f(0)$ and $|f''(0)|$ for the Gaussian, $\chi^2(2)$ and Cauchy distributions.

	Density of X		
Kernel	Gaussian	$\chi^2(2)$	Cauchy
Gaussian	$(\frac{\overline{y_n^2}}{n\sqrt{2}\bar{y}_n^2})^{1/5}$	$(\frac{16\overline{y_n^2}}{n\sqrt{\pi}\bar{y}_n^2})^{1/5}$	$(\frac{\sqrt{\pi}\overline{y_n^2}}{8n\bar{y}_n^2})^{1/5}$
Epanechnikov	$(\frac{15\sqrt{2\pi}\overline{y_n^2}}{n\bar{y}_n^2})^{1/5}$	$(\frac{480\overline{y_n^2}}{n\bar{y}_n^2})^{1/5}$	$(\frac{15\pi\overline{y_n^2}}{4n\bar{y}_n^2})^{1/5}$
Uniform	$(\frac{9\sqrt{\pi}\overline{y_n^2}}{n\sqrt{2}\bar{y}_n^2})^{1/5}$	$(\frac{144\overline{y_n^2}}{n\bar{y}_n^2})^{1/5}$	$(\frac{9\pi\overline{y_n^2}}{8n\bar{y}_n^2})^{1/5}$
Quartic	$(\frac{35\sqrt{2\pi}\overline{y_n^2}}{n\bar{y}_n^2})^{1/5}$	$(\frac{1120\overline{y_n^2}}{n\bar{y}_n^2})^{1/5}$	$(\frac{35\pi\overline{y_n^2}}{4n\bar{y}_n^2})^{1/5}$

Table 3: Optimal Bandwidths

With these tables, one can find the optimal strategy to minimize the variance by noticing that the asymptotic relative efficiency is independent of the the sample of observations of interest. Using the optimal bandwidths, if σ_i^2 is the asymptotic variance and A_i, B_i, $f_i(0)$ and $f_i''(0)$ are the parameters for the estimator i, then the relative efficiency e is

$$e = \frac{\sigma_1^2}{\sigma_2^2} = \frac{(B_1/B_2)^{4/5}(A_1/A_2)^{4/5}(f_1''(0)/f_2''(0))^{2/5}}{(f_1(0)/f_2(0))^{6/5}}.$$

Therefore, the following Table 4 provides the information to take into account for the choice of the density and Kernel in applications.

When not using the optimal bandwidths, the effect of bandwidth changes because the optimal bandwidths on depend of the kernel. We obtain Table 5.

It is clear that not using the optimal bandwidths prioritizes the Gaussian kernel and the use of $\chi^2(2)$ or standard normal distribution for X, while using the optimal

	Density of X		
Kernel	Gaussian	$\chi^2(2)$	Cauchy
Gaussian	47%	97%	51%
Epanechnikov	48%	100%	53%
Uniform	46%	95%	50%
Quartic	48%	100%	53%

Table 4: Asymptotic relative efficiencies (relative to: $\chi^2(2)$ for X and quartic kernel)

	Density of X		
Kernel	Gaussian	$\chi^2(2)$	Cauchy
Gaussian	80%	100%	66%
Epanechnikov	38%	47%	31%
Uniform	45%	56%	36%
Quartic	32%	39%	25%

Table 5: Asymptotic relative efficiencies (relative to: $\chi^2(2)$ for X and Gaussian kernel)

bandwidths prioritizes the use of $\chi^2(2)$ exclusively for X among checked distributions (see Table 4 and Table 5). These tables are exact and don't depend of the sample of observations to be analyzed.

Remark 3 For statistical applications, the condition $\mu_Y \neq 0$ is not an issue because we can always add an arbitrary constant to each of the data points without changing any of the conditions of the theorem. It is indeed just a condition to avoid the case when the MSE cannot be minimized over finite non-zero values of h.

4. Applications to statistical models

Recalling that for the normal kernel and the standard normal distribution for X the estimator of the mean of the sequence $(Y_i, 1 \leq i \leq n)$ is $\hat{r}_n = \frac{1}{nh}\sum_{i=1}^{n} Y_i exp(-\frac{1}{2}(\frac{X_i}{h})^2)$, we will define below test statistics for testing for the difference of means for some statistical models and build some confidence intervals. If the $\chi^2(2)$ distribution is used for X with the standard normal kernel, then the estimator is

$$\hat{r}_n = \frac{\sqrt{2}}{nh\sqrt{\pi}}\sum_{i=1}^{n} Y_i exp(-\frac{1}{2}(\frac{X_i}{h})^2).$$

Finally, using the quartic kernel and $\chi^2(2)$ for X, the estimator is

$$\hat{r}_n = \frac{15}{8nh}\sum_{i=1}^{n} Y_i(1-(\frac{X_i}{h})^2)^2 I(|X_i| < h).$$

4.1 Confidence interval for the mean

For a sample of observations $\{Y_i.1 \leq i \leq n\}$ with unquantified dependence, based on Theorem 12, a $(1-\alpha)100\%$ confidence interval for the mean μ_Y is

$$\left\{ \hat{r}_n - \left(\frac{\overline{Y_n^2} \int K^2(x)dx}{nhf(0)} \right)^{1/2} \times z_{\alpha/2},\ \hat{r}_n + \left(\frac{\overline{Y_n^2} \int K^2(x)dx}{nhf(0)} \right)^{1/2} \times z_{\alpha/2} \right\},$$

where $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$, and Z is a standard normal variable. Using the the standard normal distribution for X we obtain

$$\left\{ \hat{r}_n - \left(\frac{\overline{Y_n^2}}{nh\sqrt{2}} \right)^{1/2} \times z_{\alpha/2},\ \hat{r}_n + \left(\frac{\overline{Y_n^2}}{nh\sqrt{2}} \right)^{1/2} \times z_{\alpha/2} \right\}, \tag{9}$$

and using the Gaussian kernel and $\chi^2(2)$ as distribution for X, we obtain

$$\left\{ \hat{r}_n - \left(\frac{\overline{Y_n^2}}{nh\sqrt{\pi}} \right)^{1/2} \times z_{\alpha/2},\ \hat{r}_n + \left(\frac{\overline{Y_n^2}}{nh\sqrt{\pi}} \right)^{1/2} \times z_{\alpha/2} \right\}. \tag{10}$$

Using the quartic kernel and $\chi^2(2)$ for X, we obtain

$$\left\{ \hat{r}_n - \left(\frac{10\overline{Y_n^2}}{7nh} \right)^{1/2} \times z_{\alpha/2},\ \hat{r}_n + \left(\frac{10\overline{Y_n^2}}{7nh} \right)^{1/2} \times z_{\alpha/2} \right\}. \tag{11}$$

Finally, using the Epanechnikov kernel and $\chi^2(2)$ for X, we obtain

$$\left\{ \hat{r}_n - \left(\frac{6\overline{Y_n^2}}{5nh} \right)^{1/2} \times z_{\alpha/2},\ \hat{r}_n + \left(\frac{6\overline{Y_n^2}}{5nh} \right)^{1/2} \times z_{\alpha/2} \right\}. \tag{12}$$

4.2 Confidence interval for the mean of a reversible Markov chain

If the sample $(Y_i, 1 \le i \le n)$ is from a reversible Markov chain and we generate an independent random sample of $(X_i, 1 \le i \le n)$ from a mean zero square integrable distribution, the the following is a $(1-\alpha)100\%$ confidence interval for the mean of Y.

$$\left\{ \hat{\mu}_n - \frac{\sigma_x S_z}{\sqrt{n}|\bar{X}|} \times z_{\alpha/2}, \hat{\mu}_n + \frac{\sigma_x S_z}{\sqrt{n}|\bar{X}|} \times z_{\alpha/2} \right\}. \tag{13}$$

The drawback of this estimator is that the variance tends to be large, but it very quickly converges to the true value of the mean.

4.3 Testing for the mean of a single variable or difference of means.

For a test on the difference of means of a paired stationary sequence $((Z_i, W_i), i = 1 \cdots n)$, one can use $(Y_i = Z_i - W_i, i = 1 \cdots n)$, d - the difference of means, we use the test for a single variable on the difference of the two observations as in the standard case for independence.

$$Z = \frac{\hat{r}_n - d}{\sqrt{\frac{1}{\sqrt{2}h_n n^2} \sum_{i=1}^n Y_i^2}} - \text{is approximately standard normal.}$$

Reject the null hypothesis in the appropriate critical region. Notice that this test puts no assumptions on the distribution of Y, except for the existence of the variance.

4.4 Testing for difference of means for equal sample sizes

For a test on the difference of means of two independent stationary sequences $(M_i, i = 1 \cdots n)$ and $(W_i, i = 1 \cdots n)$, one can use $(Y_i = M_i - W_i, i = 1 \cdots n)$. If d is the difference of means, then

$$Z = \frac{\hat{r}_n - d}{\sqrt{\frac{1}{\sqrt{2} h_n n^2} \sum_{i=1}^{n} (W_i^2 + M_i^2)}} - \text{is approximately standard normal.}$$

Reject the null hypothesis in the appropriate critical region. Notice that in this case the same observations are used to smooth W and Z. We will compare below this test statistic to the one used in case of unequal sample siezes.

4.5 Testing for the mean for unequal sample sizes

For a test on the difference of means of two independent stationary sequences $(M_i, i = 1 \cdots n)$ and $(W_i, i = 1 \cdots m)$ d - the difference of means and the normal kernel and standard normal for X.

$$\hat{r}_{M,W} = \frac{1}{nh_n} \sum_{i=1}^{n} M_i \exp[-\frac{1}{2}(\frac{1}{h_n} X_i)^2] - \frac{1}{mh_m} \sum_{i=1}^{m} W_i \exp[-\frac{1}{2}(\frac{1}{h_m} X_{*i})^2],$$

$$Z = \frac{\hat{r}_{M,W} - d}{\sqrt{\frac{1}{\sqrt{2} h_n n^2} \sum_{i=1}^{n} M_i^2 + \frac{1}{\sqrt{2} h_m m^2} \sum_{i=1}^{m} W_i^2}}.$$

Reject the null hypothesis in the appropriate critical region. Notice that in this case the different observations are used to smooth W and M.

4.6 Testing for difference of more than two means for equal sample size

For a test on the difference of three independent treatment levels (observations within each of the treatments form a stationary ergodic sequence $(M_{ji}, i = 1 \cdots n, j = 1, 2, 3)$. We define

$$Z(j, n) = \frac{\hat{r}_n}{\sqrt{\frac{1}{\sqrt{2} h_n n^2} \sum_{i=1}^{n} M_{ji}^2}}.$$

$Z = \frac{1}{3}\Big([Z(1, n) - Z(2, n)]^2 + [Z(1, n) - Z(3, n)]^2 + [Z(2, n) - Z(3, n)]^2\Big)$ is approximately $\chi^2(2)$.

Reject the null hypothesis for large values.

5. Simulations and comparisons

Consider estimation of the mean of a sample of observations from $Y_i = Z_i + \mu$, where Z_i is an $ARFIMA(0, d, 0)$ (Auto-Regressive Fractionally Integrated Moving Average). In the literature, it is shown that this process is ergodic with variance of partial sums $var(\bar{Y}_n) = cn^{2d-1}$ for $0 \le d < .5$ (see Longla and Peligrad (2018) and Hosking (1981)). We will present here the results of simulations of this model and the conclusions.

5.1 Using $\hat{\mu}_n$ to estimate the mean of shifted $ARFIMA(0, d, 0)$

On the example of the ARFIMA(0,d,0) given above, the conditions of the CLT for this estimator are violated, but it still estimates almost to perfection the mean of population Y under standard normal samples for the test population X. The standard error of this estimator behaves like an inverse normal random variable because $\sqrt{n}\bar{X} \to N(0, 1)$.

Table 6: Applications to $ARFIMA(0, d, 0)$

(a) d=.3 and $-.3 + Bernoulli(.3)$

Size	500	100	50
$\hat{\mu}_n$	-3.92	-3.79	-4.23
$Std(\hat{\mu}_n)$	11.09	2.81	1.89

(b) d=.3 and $Normal(0, 1)$

Size	500	100	50
$\hat{\mu}_n$	-99.38	- 98.8	-99.56
$Std(\hat{\mu}_n)$	23.66	12.49	4.16

(c) d=.7 and -2+$\chi^2(2)$

Size	500	100	50
$\hat{\mu}_n$	49.8	50.96	50.53
$Std(\hat{\mu}_n)$	52.67	12.0	13.8

(d) d=.9 and $Normal(0, 1)$

Size	500	100	50
$\hat{\mu}_n$	48.93	50.42	48.69
$Std(\hat{\mu}_n)$	11.16	6.24	6.35

5.2 Using $\hat{\mu}_n$ and $\hat{r}_n$ to estimate μ for some reversible Markkov chain

Longla (2015) proposed conditions for mixing properties of mixtures of copulas that generate reversible Markov chains. A class of copulas for such Markov chains was the Frechet family of copulas $C(x, y) = aW(x, y) + (1 - a)M(x, y)$, for $0 \le a \le 1$. This copula is the joint distribution of a bivariate random variable (U, V) with uniform marginals on $(0, 1)$. It generates reversible Markov chains with any initial distribution (see Longla and Peligrad (2012), Longla (2013) or Longla (2015) for more). Technically, any sample from any Markov chain generated by this copula will be a string made of two values X_0 and $1 - X_0$ with changes depending on the value of a. The number a is typically the probability to obtain $1 - x$ after obtaining x for the previous sample point. So, the larger a, the more flips we will have in the sample. Our aim here is to apply the results of the study to the population mean and compare the performances of various estimators.

Table 7 (a)-(d) indicate some sample results from the Markov chains with distributions given in the headings and having transition probabilities defined by the Frechet copula with parameter a. To generate observations from these stationary Markov chains, if F is the cumulative distribution of Y, we generate a Markov chain $(U_i, 1 \le i \le n)$ with uniform distribution as marginals ad Frechet copula for transitions, then set $Y_i = F^{-1}(U_i)$ for $i = 1, \cdots, n$. The estimators $\hat{r}_n$ and $\hat{mu}_n$ are applied to the same data set with the same sample $(X_i, 1 \le i \le n)$ from the standard normal distribution.

5.3 Estimation of μ for Markov Chain with Clayton copula

Copulas are bivariate distributions that are used to capture the strength of the dependence between random variables. When using a copula to model the dependence for a bivariate random variable with uniform marginals, the conditional distribution for transitions is the derivative with respect to the first variable of the copula (see Nelsen (2006) or Longla and Peligrad (2012)). Thus, to obtain the data we generate

Table 7: Applications to the Frechet family of copulas

(a) $a = .3$ and $Y = N(50, 1)$

Size	500	100	50
$\hat{\mu}_n$	50.42	49.75	49.87
$Std(\hat{\mu}_n)$	7.79	1.99	3.74
$\hat{r}_n$	45.09	53.04	45.03
$Std(\hat{r}_n)$	3.62	6.90	9.09
$\bar{Y}$	50	50.01	49.95
$Std(\bar{Y})$	.17	.25	.61

(b) $a = .5$ and $Y = Normal(6, 9)$

Size	500	100	50
$\hat{\mu}_n$	5.44	6.47	5.77
$Std(\hat{\mu}_n)$	14.54	15.92	3.79
$\hat{r}_n$	5.44	5.94	5.37
$Std(\hat{r}_n)$	.44	.83	1.11
$\bar{Y}$	5.97	5.91	6.06
$Std(\bar{Y})$	.80	1.42	.73

(c) $a = .7$ and $Y = N(200, 4)$

Size	500	100	50
$\hat{\mu}_n$	200.45	199.99	198.11
$Std(\hat{\mu}_n)$	52.37	.31	11.38
$\hat{r}_n$	201.12	166.53	233.94
$Std(\hat{r}_n)$	14.50	27.59	36.35
$\bar{Y}$	200.09	199.99	199.68
$Std(\bar{Y})$	1.51	.01	2.68

(d) $a = .7$ and $Y = Normal(-10, 4)$

Size	500	100	50
$\hat{\mu}_n$	-10.13	-9.87	-10.14
$Std(\hat{\mu}_n)$	2.31	1.32	2.80
$\hat{r}_n$	-9.77	-9.42	-8.68
$Std(\hat{r}_n)$	.72	1.38	1.82
$\bar{Y}$	-10.01	-10.01	-10.01
$Std(\bar{Y})$	.06	.28	.18

a Markov chain with uniform marginals and Clayton copula for transition probabilities $(Z_i, 1 \le i \le n)$. This is done using the Clayton copula and its derivative

$$C(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \quad C_u(u,v) = u^{-\alpha-1}(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha-1}.$$

Knowing the previous value u_0 of the Markov chain, the following is obtained by generating a value from the distribution $C_u(u_0, v)$ (see Nelsen (2006) for more). An independent observation v_i is generated from the uniform distribution. Then $Z_i = (u_0^{-\alpha}(v_i^{-\alpha/(\alpha+1)} - 1) + 1)^{-1/\alpha}$. We then set $(Y_i = F^{-1}(Z_i), 1 \le i \le n)$, where F is the cumulative distribution of the invariant distribution of the generated Markov chain. The estimators use the same sample of X values from the standard normal distribution.

Sample size mean	n=100	$\mu = -5$	n=100	$\mu = 50$
Estimate - Standard error	EST	STD	EST	STD
$\hat{\mu}_n$	-5.29	8.50	53.20	16.82
$\hat{r}_n$, Gaussian kernel, h_0	-5.28	.71	51.26	6.94
$\hat{r}_n$, Gaussian kernel, $h = n^{-.19}$	-5.14	.68	49.74	6.55
$\hat{r}_n$, Gaussian kernel, $h = n^{-.21}$	-5.27	.72	50.97	6.86
$\hat{r}_n$, Epanechnikov kernel, h_0	-4.16	.59	39.97	5.72
$\hat{r}_n$, Epanechnikov kernel, $h = n^{-.19}$	-4.97	.84	46.25	8.04
$\hat{r}_n$, Epanechnikov kernel, $h = n^{-.21}$	-5.18	.88	46.22	8.42

Table 8: Estimates and Standard deviations for $\alpha = 3$, Y follows t with $df = 2$.

Entries of Table 8 are estimates of the mean and their standard deviations for each of the estimators applied to the same data set. The fourth row for example means that $\hat{r}_n$ is computed using the optimal bandwidths h_o and the Gaussian kernel. The data set is a reversible Markov chain generated by the Clayton copula having shifted t-distribution with two degrees of freedom as invariant marginal distribution with mean μ. A SAS function is used to find $Y = F^{-1}(Z)$ for each of the Z observations generated via the Clayton copula.

It can be seen from Table 8 that $\hat{r}_n$ defined via the Epanechnikov kernel performs poorly on estimation compared to other scenarii no matter what bandwidths are used, but provided a smaller standard error comparable to that of the Gaussian. Table 8 also shows that moving away from the optimal bandwidths increases accuracy and estimation when the Epanechnikov kernel is used. Overall, $\hat{\mu}_n$ clearly estimates better than all these estimators as the sample size increases.

Sample size mean	n=1000	$\mu = 500$	n=1000	$\mu = 500$
Estimate - Standard error	EST	STD	EST	STD
$\hat{\mu}_n$	499.01	191.65	505.26	88.63
$\hat{r}_n$, Gaussian kernel, h_0	500.14	27.47	513.07	27.47
$\hat{r}_n$, Gaussian kernel, $h = n^{-.19}$	493.82	25.63	509.33	25.63
$\hat{r}_n$, Gaussian kernel, $h = n^{-.21}$	500.13	27.47	513.06	27.47
$\hat{r}_n$, Epanechnikov kernel, h_0	398.41	22.64	409.65	22.64
$\hat{r}_n$, Epanechnikov kernel, $h = n^{-.19}$	418.82	31.43	420.64	31.43
$\hat{r}_n$, Epanechnikov kernel, $h = n^{-.21}$	422.24	33.68	422.44	33.68

Table 9: Estimates and Standard deviations for $\alpha = 3$, Y follows t with $df = 2$.

Entries of Table 9 indicate that the variance of $\hat{r}_n$ stabilizes for large samples and depends only on the kernel and the distribution of X. This is due to the fact that the sequence of observations is ergodic (This implies that the average of squares observations converges to the mean of the population). It is clear that the estimators performs very well even in cases when some of the assumptions are violated. The optimal bandwidths are shown to be not necessarily best for estimation.

6. Appendix

6.1 SAS code for $\hat{\mu}$ and $ARFIMA(0, d, 0)$ of Table 6

```
%MACRO ARIMA(n);
** Delete possible present sets;
proc datasets lib= work memtype=data nolist; delete BOSS; delete XYSet; delete
Xset; delete Last; quit;
/**Creating The random sample of Innovations**/
Data AR;
%do t=1 %to &n+1;
e=rand('normal',0,1); output; %end; run;
/**Creating The ARFIMA Sample of Y for estimation**/
proc iml;    use AR; read all var e;    call fdif(Y, e, d);
create ARMA var {Ye}; /** create data set **/
append; /** write vectors into the data set**/
close ARMA; /** close the data set **/
Data ARMA; Set Arma;    If Y='.' then delete; Y=Y+50; run;
proc univariate data=ARMA noprint;
var Y ; output out=BOSS MEAN=M1 stddev= S1; run;
DATA XSet; /**Creating The random sample of X **/
%do t=1 %to &n;
X=rand('Normal', 0,1); output;%end; run;
```

```
/**Creating data set of X and Y**/
proc iml;   use XSet; read all var{X}; use ARMA; read all var {Y};
use Boss; read all var{S1   M1};
create XYSet var {X   Y   S1   M1}; /** create data set **/
append; /** write vectors into the data set**/
close XYSet; /** close the data set **/
DATA XYhSet; Set XYSet; Z=X*Y; run;
proc univariate data=XYhSet noprint;
var X Z;   output out=FINAL MEAN= M4 M3; run;
Data Last; set Final; set Boss;
Sigma=S1/abs(M4); muhat=M3/M4; run;
proc print data=laST;run;
%MEND;
%ARIMA(n); run; ***Running the macro.*
```

This macro can be used to generate samples from any distribution of X and any $ARMA(0, d, 0)$. In the DATA set ARMA, $Y = Y + 50$ adds the value of μ to be estimated. This value can be changed to see the variation for various values of the mean. Using the code, the variance of X is assumed to be 1. Thus, if a a distribution is chosen, the values need to be rescaled to have variance 1. The innovations of the ARFIMA are given as e, the fractional difference parameter is d. The sample size is n, argument of the Macro shall be typed in to run the Macro after inputting the parameters of interest.

6.2 SAS code for Table 7

```
%MACRO FRECHET(a, n); ** Delete possible present sets;
proc datasets lib= work memtype=data nolist; delete BOSS; delete Begin; quit;
/**Creating The random from Markov chain with Frechet transition copula**/
Data Begin;
do i=1 to &n;
Y1=0; output; end; run;
proc iml; Use Begin; read all var{Y1}; Y1[1]= rand('uniform');
do i=2 to &n;
U=rand('uniform'); j=i-1; v1=Y1[j];
if U< &a then do; Y1[i]=v1;end;
If u >&a then do; Y1[i]=1-v1; end; end;
create Frechet var{Y1}; /** create data set **/
append; /** write vectors into the data set**/
close Frechet; /** close the data set **/
Data SampleY; Set Frechet;
Y=Probit(Y1)+50; YY=Y*Y; run;
proc univariate data=SampleY noprint;
var Y YY; output out=BOSS MEAN=M1 M2 stddev= S1 S2; run;
DATA H; Set Boss;
%do i=1 %to &n;
h=(M2/(&n*(M1**2)*(2**.5)))**.2; output; %end; run;
DATA XSet; /**Creating The random sample of X **/
%do t=1 %to &n;
X=rand('Normal', 0,1); output; %end; run;
/**Creating data set of X and Y**/
```

```
proc iml;
use XSet; read all var{X}; use SampleY;
read all var{Y}; use H; read all var{h  S1  M1};
create XYSet var{X  Y  h  S1  M1}; /** create data set **/
append; /** write vectors into the data set**/
close XYSet; /** close the data set **/
DATA XYhSet; Set XYSet;
Z1=Y*exp(-(X/h)**2/2)/h; Z=X*Y; Y2=Y**2; run;
proc univariate data=XYhSet noprint;
var X Z Z1 Y2 h; output out=FINAL MEAN= M4 M3 Rhat MY2 h; run;
Data Last; set Final; set H; set Boss;
Sigmamuhat=S1/abs(M4); muhat=M3/M4; SIGMARhat= (MY2/(&n*h*(2)**.5))**.5;
run;
%MEND;
%FRECHET(.3,500); run;
```

6.3 SAS code for Table 8 and Table 9

```
%MACRO Clayton(alpha, nu, n);
** Delete possible present sets; proc datasets lib= work memtype=data nolist; delete
BOSS; delete Begin; quit;
/**Creating The random from Markov chain with Clayton transition copula**/
Data Begin;
do i=1 to &n; Y1=0; output; end; run;
proc iml; Use Begin; read all varY1; Y1[1]= rand('uniform');
do i=2 to &n; U=rand('uniform'); j=i-1; v1=Y1[j];
Y1[i]=(v1**(-&alpha)*(u**(-&alpha/(&alpha+1))-1)+1)**(-1/&alpha);end;
create Clayton var{Y1}; /** create data set **/
append; /** write vectors into the data set**/
close Clayton; /** close the data set **/
Data SampleY; Set Clayton; Y=TINV(Y1, &nu)+500; YY=Y*Y; run;
proc univariate data=SampleY noprint;
var Y YY; output out=BOSS MEAN=M1 M2 stddev= S1 S2; run;
DATA H; Set Boss; pi=constant('pi');
%do i=1 %to &n; h=(M2/(&n*(M1**2)*(2**.5)))**.2;
h1=(15*((2*pi)**.5)*M2/(&n*(M1**2)))**.2; h2=(&n**(-.19)); h3=(&n**(-.21));
output; %end; run; DATA XSet; /**Creating The random sample of X **/
%do t=1 %to &n; X=rand('Normal', 0,1); output; %end; run;
/**Creating data set of X and Y**/
proc iml; use XSet; read all var{X}; use SampleY; read all var{Y}; use H; read all
var{h h1 h2 h3 S1 M1 pi};
create XYSet var{X Y h h1 h2 h3 S1 M1 pi}; /** create data set **/
append; /** write vectors into the data set**/
close XYSet; /** close the data set **/
DATA XYhSet; Set XYSet;
Z1=Y*exp(-(X/h)**2/2)/h; Z=X*Y; Y2=Y**2;
Z2=Y*exp(-(X/h2)**2/2)/h2; Z3=3*((2*pi)**.5)*Y*(1-(X/h1)**2)*(-h1<X<h1)/(5*h1);
Z4=3*((2*pi)**.5)*Y*(1-(X/h2)**2)*(-h2<X<h2)/(5*h2); Z5=Y*exp(-(X/h3)**2/2)/h3;
Z6=3*((2*pi)**.5)*Y*(1-(X/h3)**2)*(-h3<X<h3)/(5*h3); run;
proc univariate data=XYhSet noprint;
```

```
var X Z Z1 Z2 Z3 Z4 Y2 h h1 h2 Z5 Z6;
output out=FINAL MEAN= MX MZ Rhat1 Rhat2 Rhat3 Rhat4 MY2 h h1 h2
RHAT5 RHAT6; run;
Data Last(Drop=MY2 M2 S2 h h1 h2 MX MZ pi ); set Final; set H; set Boss;
pi=constant('pi'); Sigmamuhat=S1/abs(MX); muhat=MZ/MX;
SIGMARhat1= (MY2/(&n*h*(2**.5)))**.5; SIGMARhat2= (MY2/(&n*h2*(2**.5)))**.5;
SIGMARhat3= (3*((2*pi)**.5)*MY2/(&n*5*h1*(2**.5)))**.5;
SIGMARhat4= (3*((2*pi)**.5)*MY2/(&n*5*h2*2**.5))**.5;
SIGMARhat5= (MY2/(&n*h3*(2**.5)))**.5;
SIGMARhat6= (3*((2*pi)**.5)*MY2/(&n*5*h3*(2**.5)))**.5; run;
%MEND;
%Clayton(3, 2, 1000); run;
proc print data=Last;run;
```

Acknowledgments

The author has been supported by the CLA of the University of Mississippi Summer Grant for this work. The author thanks Dr. Isidore Seraphin Ngongo for useful discussions on the topic.

REFERENCES

Billingsley P. (1968), Convergence of probability measures. John Wiley, New York.

Billingsley, P. (1995), Probability and measure. 3-rd edition, John Willey & Sons.

Bradley, R. (1983), "Asymptotic normality of some kernel-type estimators of probability density". Statist. Probab. Lett. 1, 295-300.

Bosq, D. (1996), Nonparametric statistics for stochastic processes: estimation and prediction. New York: Springer-Verlag.

Bosq, D., Merlevède, F. and Peligrad, M. (1999), "Asymptotic normality for density kernel estimators in discrete and continuous time." Journal of Multivariate Anal. 68, 79-95.

Cogburn R. (1960), " Asymptotic properties of stationary sequences." Univ. Calif. publ. Statis. 3, 99–146.

Collomb, G. (1984), " Propriétés de convergence presque complète du prédicteur à noyau." Z. Wahr. Verwandte Gebiete 66, 441–460.

Cuny C. and Peligrad M. (2012), " Central limit theorem started at a point for additive functional of reversible Markov Chains." Journal of Theoretical Probability, 25, 171–188.

Derriennic Y. and Lin M. (2001). " The central limit theorem for Markov chains with normal transition operators started at a point." Probability Theory and Related Fields, 119, 508–528.

Doukhan P., Massart P. and Rio E. (1994), " The functional central limit theorem for strongly mixing processes." Annales de l'I.H.P., section B, tome 30, 1, 63–82.

Durrett R. (1996), Stochastic Calculus: a Practical Introduction. CRC Press.

Härdle, W. (1991). Smoothing Techniques With Implementation in S. Springer-Veralg.

Hong, S. Y. and Linton, O. (2016), "Asymptotic properties of a Nadaraya-Watson type estimator for regression functions of infinite order". arXiv:1604.06380

Hosking, J. R. M. (1981), "Fractional differencing." Biometrika 68, 165–176.

Ibragimov I.A. (1962), " Some limit theorems for stationary processes." Theory Probab. Appl., 7(4), 349вЪ"382;

Ibragimov I.A. (1975), " A note on the central limit theorem for dependent random variables." Theory of probability and its applications, volume XX, 1.

Ibragimov I.A. and Rozanov Y.A. (1978), " Gaussian random processes." Theory of probability and its application, volume VII, 4, 349–382.

Jones G. L. (2004), " On the Markov chain central limit theorem", Probability Surveys, volume 1, 299–320.

Kipnis C. and Landim C. (1999), Scaling Limits of Interacting Particle Systems. Springer, New York.

Kipnis C. and Varadhan S. R. S.(1986), " Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions." Comm. Math. Phys. 104, 1–19.

Long, H. and Qian, L. (2013), "Nadaraya-Watson estimator for stochastic processes driven by stable Lévy motions", Electronic Journal of Statistics 7, 1387–1418.

Longla M. and Peligrad M. (2012), "Some aspects of modeling dependence in copula-based Markov chains". Journal of Multivariate Analysis 111, 234-240.

Longla M. , Peligrad M. and Peligrad C. (2012). " On the functional CLT for reversible Markov Chains with nonlinear growth of the variance", Journal of Applied Probability. 49 1091-1105.

Longla M. (2013), "Remarks on the speed of convergence of mixing coefficients and applications", Statistics & Probability Letters 83 (10), 2439-2445.

Longla M. (2015), "On mixtures of copulas and mixing coefficients", Journal of Multivariate Analysis 139, 259-265.

Longla M. (2017), " Remarks on limit theorems for reversible Markov processes and their applications", Journal of Statistical Planning and Inference 187, 28-43

Longla M., Peligrad M. (2018), "New robust confidence intervals for the mean under dependence", arXiv preprint arXiv:1801.00175.

Merlevède F., and Peligrad M. (2013), " Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples", Ann. Probab volume 41, 2, 914–960.

Nadaraya, E.A. (1964), "On estimating regression". Theory of Probability & Its Applications, 9, 141-142.

Nelsen, R.B. (2006), An introduction to copulas. Springer.

Peligrad M. and Utev S. (2006), " A new maximal inequality and invariance principle for stationary sequences", Annals of Probability, 33, 798–815.

Rio E. (2009), " Moment inequalities for sums of dependent random variables under projective conditions." Journal of Theoretical Probability 22, 146–163.

Rio E. (2000). "Théorie asymptotique des processus aléatoires faiblement dépendants", Mathématiques et Applications. 31, Springer, Berlin.

Roberts G.O., Rosenthal J.S. (1997). " Geometric ergodicity and hybrid Markov chains", Electronic Communications in Probability 2, 13–25.

Shao Q. (1989), " On the invariance principle for stationary ρ–mixing sequences of random variables." Chinese Ann. Math. 10B, 427–433.

Tierney L. (1994). " Markov chains for exploring posterior distribution (with discussion)", Annals of Statistics, 22, 1701–1762.

Watson, G.S. (1964), "Smooth regression analysis", Sankhya Series A, 26, 359-372.

Williams D. (1991), Probability with Martingales. Cambridge mathematical textbooks.

Wu L. (1999), "Forward-backward martingale decomposition and compactness results for additive functionals of stationary ergodic Markov processes." Ann. Inst. H. Poincaré Probab. Stat. 35, 121–141.

Yoshihara, K. (1994), Weakly Dependent Stochastic Sequences and their Applications: Volume IV: Curve Estimation Based on Weakly Dependent Data. Sanseido, Tokyo, Japan.

Zhao O. and Woodroofe M. (2008). " Law of the iterated logarithm for stationary processes", Annals of Probability. 36, 127–142.

Zhao, O., Woodroofe, M. and Volný, D. (2010), "A central limit theorem for reversible processes with nonlinear growth of variance", J. Appl. Prob. 47, 1195-1202.