Research Article

Mathias Muia* and Martial Longla

A point on discrete versus continuous state-space Markov chains

Abstract: This article investigates the effects of discrete marginal distributions on copula-based Markov chains. We establish results on mixing properties and parameter estimation for a copula-based Markov chain model with Bernoulli(p) marginal distributions, emphasizing some distinctions between continuous and discrete state-space Markov chains. We derive parameter estimators using the maximum-likelihood estimation (MLE) method and explore alternative estimators of p that are asymptotically equivalent to the MLE. Furthermore, we provide the asymptotic distributions of these parameter estimators. A simulation study is conducted to evaluate the performance of the various estimators for p. Additionally, we employ the likelihood ratio test to assess independence within the sequence.

Keywords: asymptotic normality, dependent Bernoulli trials, Frechet family of copulas, mixing for Markov chains

MSC 2020: 60J25

1 Introduction

Markov chain models have been widely used in the literature to represent the relationship between consecutive observations in a sequence of Bernoulli trials. Johnson and Klotz [18] employed a Markov chain generalization of a binomial model to analyze the crystal structure of a two-metal alloy in their metallurgical studies. Crow [9] investigated the applications of two-state Markov chains in telecommunications in his study focusing on approximating confidence intervals. Brainerd and Chang [7] used two-state Markov chains to address problems in linguistic analysis. These examples illustrate that Markov chain models can provide a solid foundation for modeling a wide range of two-state sequential phenomena.

Moreover, in the study of sequences consisting of different alternatives, there has been notable interest in using group or run tests to evaluate the randomness of the sequences. These tests are based on the count of changes between outcomes in the sequence, aiming to assess whether the sequence displays randomness (denoted as the null hypothesis, H_0) or exhibits a type of dependence similar to simple Markov chains (alternative hypothesis, H_1). These counts, denoted by n_{ij} (where $i, j = 0, \ldots, s$ denotes the states taken by the sequence), are commonly referred to as *transition numbers* of the sequence. The transition numbers have been widely used in the literature to construct run tests and in the estimation of model parameters (see David [10], Goodman [13], Anderson and Goodman [1], and Billingsley [4] for details). As used in this work, transition numbers prove to be very useful in computing the maximum-likelihood estimators (MLE) of parameters.

---

* Corresponding author: Mathias Muia, Department of Mathematics and Statistics, University of South Alabama, Mobile, Al 36688, United States of America, e-mail: mnmuia@southalabama.edu

Martial Longla: Department of Mathematics, University of Mississippi, University, MS 38677, United States of America, e-mail: mlongla@olemiss.edu

1.1 Copulas and Markov chains

A 2-copula is a bivariate function $C : [0, 1]^2 \to [0, 1]$ such that $C(0, u) = C(u, 0) = 0$, $C(u, 1) = C(1, u) = u$, $\forall u \in [0, 1]$, and $C(u_1, v_1) + C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) \geq 0$, $\forall [u_1, u_2] \times [v_1, v_2] \subset [0, 1]^2$. For some $k \geq 2$, let $C_1, ..., C_k$ be a set of k copulas and $a_1, ..., a_k$ a set of real constants such that $a_1 + \cdots + a_k = 1$ and $0 \leq a_i \leq 1$ for each $i = 1, ..., k$; then, the sum given by $C(u, v) = a_1 C_1(u, v) + \cdots + a_k C_k(u, v)$ is a copula referred to as the convex combination of $C_1, ..., C_k$ [24].

The copula $C(F(x), G(y)) = H(x, y)$, of the random variables X and Y, is known to be unique if F and G are continuous; otherwise, C is uniquely determined on Range(F) × Range(G) [27]. In the recent past, copulas have gained popularity in the modeling of temporal dependence within Markov processes. This is because, when used to represent joint distributions of random variables, they enable the detection of possible correlations or links between variables [11]. A copula-based Markov chain is a stationary Markov chain generated by the copula of its consecutive states and an invariant distribution. Copula-based Markov chains enable the exploration of dependence and mixing properties inherent in Markov processes. This makes them an invaluable tool for investigations and allows for the study of scale-free measures of dependence while remaining invariant under monotonic transformations of the variables. Consequently, copula methods have drawn a lot of interest across diverse fields, including technology, finance, and the natural world.

This article focuses on a stationary Markov chain based on a copula from the Frechet family of copulas. The Fréchet family of copulas has the form

$$C(u, v) = aM(u, v) + (1 - a - b)\Pi(u, v) + bW(u, v), \quad a, b \geq 0, \quad 0 \leq a + b \leq 1, \tag{1.1}$$

where $\Pi(u, v) = uv$ is the independence copula and $M(u, v) = \min(u, v)$ and $W(u, v) = \max(u + v - 1, 0)$ are the Fréchet-Hoeffding upper and lower bounds, respectively. We study two cases: a continuous state-space Markov chain with uniform marginal distributions and a discrete state-space Markov chain with Bernoulli(p) marginal distributions. MLEs of model parameters are derived in both cases, and their asymptotic distributions are provided. The likelihood ratio test is used to test for randomness in the model.

Studies involving inferential statistics for discrete cases closely similar to the one considered here have been presented in the works of Billingsley [4], Billingsley [3], Goodman [13], Klotz [19], Klotz [20], Price [29], Lindqvist [21], and others. Goodman [13] considered a single sequence of alternatives consisting of a long chain of observations to derive some long-sequence group tests. Klotz [20] presented small and large sample distribution theory for the sufficient statistics of a model for Bernoulli trials with Markov dependence. He provided estimators for the model parameters and showed that the uniform most powerful unbiased test of independence is the run test. On the other hand, Price [29] and Lindqvist [21] focused on estimating the model parameters only. Price [29] used Monte Carlo techniques to investigate finite sample properties for the parameters of a dependent Bernoulli process. Lindqvist [21] considered the weaker assumption of non-Markovian dependence and showed that the MLE, $(\hat{p}, \hat{a})$ is a strongly consistent and asymptotically normally distributed estimator for (p, a). In the context of Klotz [20] and Lindqvist [21], as well as in this article, p represents the Bernoulli (frequency) parameter of a sequence of Bernoulli trials, while a is an additional dependence parameter. Lindqvist [21] further showed that $(\hat{p}, \hat{a})$ is asymptotically equivalent to $(\bar{p}, \bar{a})$, where $\bar{p}$ is the sample mean and $\bar{a}$ is the empirical correlation coefficient of X_{t-1} and X_t.

The rest of this article is organized as follows: in Section 2, we introduce our model and detail the transition probabilities of the discrete Markov chain model considered. Mixing properties of the Markov chain model are discussed in Section 2.1. Section 3 covers parameter estimation, where we derive MLEs and their asymptotic distributions. In Section 4, we provide a test for independence in the sequence. This article concludes with a simulation study in Section 5.

2 Model

Consider a stationary Markov chain model based on the copula

$$C(u, v) = aM(u, v) + (1 - a)W(u, v), \quad a \in (0, 1). \tag{2.1}$$

Copula (2.1) is from the Frechet/Mardia family of copulas given by (1.1) with $a + b = 1$. Note that a Mardia copula with a parameter $\theta \in [-1, 1]$ is a Frechet copula with $a + b = \theta^2$. For details on these copula families, see Nelsen [27], Joe [17], or Durante and Sempi [12].

Longla [23] showed that a stationary Markov chain with copula (2.1) and uniform marginals has n-steps joint cumulative distribution function (cdf) of the form

$$C_n(u, v) = 1/2[1 + (2a - 1)^n]M(u, v) + 1/2[1 - (2a - 1)^n]W(u, v). \tag{2.2}$$

If the marginals are Bernoulli(p), the joint distributions can be characterized by the transition probabilities of the Markov chain and the stationary distribution. Note that Sklar's theorem states that C is only identifiable on $\text{Ran}(F) \times \text{Ran}(G)$. Theorem 1 provides the transition matrices for the Markov chain.

Theorem 1. *Let $X = (X_t)_{t=0,\dots,n}$ be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginals. Then, X has transition matrices given by*

$$A = \begin{pmatrix} \dfrac{ap + 1 - 2p}{1 - p} & \dfrac{p(1 - a)}{1 - p} \\ 1 - a & a \end{pmatrix}, \quad p < 1/2, \tag{2.3}$$

and

$$A = \begin{pmatrix} a & 1 - a \\ \dfrac{(1 - p)(1 - a)}{p} & \dfrac{a(1 - p) + 2p - 1}{p} \end{pmatrix}, \quad p \geq 1/2. \tag{2.4}$$

Proof. Due to stationarity of the Markov chain, $P(X_t = 1) = 1 - P(X_t = 0) = p$. It suffices to determine $P(X_0 = 0, X_1 = 0)$.

$$P(X_0 = 0, X_1 = 0) = P(X_0 \leq 0, X_1 \leq 0) = C(1 - p, 1 - p)$$
$$= a(1 - p) + (1 - a)\max(1 - 2p, 0)$$
$$= \begin{cases} ap + 1 - 2p, & \text{if } p < 1/2 \\ a(1 - p), & \text{if } p \geq 1/2, \end{cases}$$

where the second equality follows from Sklar's theorem (Sklar [30]). The rest of the probabilities are computed using the hypothesis that the Markov chain is stationary with Bern(p) marginals. For $p < 1/2$, the joint distribution is given in Table 1.

Using the definition of conditional probability, $P(X_1 = j | X_0 = i) = P(X_1 = j, X_0 = i)/P(X_0 = i)$, in Table 1 we obtain transition matrix (2.3). A similar argument is used for $p \geq 1/2$ to obtain transition matrix (2.4). $\quad\square$

The statistical behavior of the distribution at $p = 1/2$ reflects a symmetric transition structure where the probability of transitioning between states is balanced by the parameter a. This symmetry simplifies the

Table 1: Distribution of (X_0, X_1) when $p < 1/2$

X_0	X_1		$P_{X_0}(x)$
	0	**1**	
0	$ap + 1 - 2p$	$p(1 - a)$	$1 - p$
1	$p(1 - a)$	ap	p
$P_{X_1}(y)$	$1 - p$	p	1

interpretation and analysis of the Markov chain, making it an interesting case for studying the dynamics governed by the mixture copula model.

Transition matrix (2.3) is similar to the one considered by Klotz [20]. In our case, we have a frequency parameter $p = P(X_t = 1)$ and an additional dependence parameter $a = P(X_t = 1|X_{t-1} = 1)$, when $p < 1/2$ or $a = P(X_t = 0|X_{t-1} = 0)$, when $p \geq 1/2$. In Klotz [20], the condition $\max(0, (2p - 1)/p) \leq a \leq 1$ was required so that the transition probabilities of the model remain bounded between 0 and 1. The model presented in this work does not require this condition because the copula conditions sufficiently define the transition probabilities based on whether $p < 1/2$ or $p \geq 1/2$. Specifically, the transition probabilities remain bounded between 0 and 1 for any value of $0 < a < 1$, and they satisfy $\sum_j p_{ij} = 1$ for $j = 0, 1$.

Proposition 2. *The nth transition matrices for* (2.3) *and* (2.4) *are, respectively,*

$$A^n = \begin{pmatrix} 1 - p + p\left(\dfrac{a - p}{1 - p}\right)^n & p - p\left(\dfrac{a - p}{1 - p}\right)^n \\ 1 - p - (1 - p)\left(\dfrac{a - p}{1 - p}\right)^n & p + (1 - p)\left(\dfrac{a - p}{1 - p}\right)^n \end{pmatrix}, \qquad p < 1/2, \tag{2.5}$$

and

$$A^n = \begin{pmatrix} 1 - p + p\left(\dfrac{a + p - 1}{p}\right)^n & p - p\left(\dfrac{a + p - 1}{p}\right)^n \\ 1 - p - (1 - p)\left(\dfrac{a + p - 1}{p}\right)^n & p + (1 - p)\left(\dfrac{a + p - 1}{p}\right)^n \end{pmatrix}, \qquad p \geq 1/2. \tag{2.6}$$

Proof. For $p < 1/2$, the eigenvalues of A are obtained by solving the characteristic equation:

$$\lambda^2 - \frac{a + 1 - 2p}{1 - p}\lambda + \frac{a - p}{1 - p} = 0.$$

Solving this equation yields $\lambda_1 = 1$ and $\lambda_2 = \frac{a - p}{1 - p}$. Then,

$$A^n = \begin{pmatrix} a_{00} + b_{00}\lambda_2^n & a_{01} + b_{01}\lambda_2^n \\ a_{10} + b_{10}\lambda_2^n & a_{11} + b_{11}\lambda_2^n \end{pmatrix}, \tag{2.7}$$

where $a_{ij}, b_{ij}, i, j = 0, 1$ are the coefficients to be determined, and λ_2 is the non-simple eigenvalue of A. Using (2.7) to write A^0 and A^1, we can solve for a_{ij} and $b_{ij}, i, j = 0, 1$, by forming four simultaneous equations, which are easy to solve. This yields the nth transition matrix (2.5). Similar steps are used to obtain the nth transition matrix for (2.4). □

2.1 Mixing

Most of the dependence and mixing coefficients are heavily influenced by the copula of the model of interest [22]. Longla [23] proposed a set of copula families that generate exponential ρ-mixing Markov chains. He showed that stationary Markov chains generated by copulas from the the Frechet (Mardia) families and uniform marginals are exponentially ϕ-mixing, therefore ρ-mixing and geometrically β-mixing if $a + b \neq 1$. However, when $a + b = 1$, mixing does not occur [23]. Furthermore, Longla [24] showed that copulas from these families fail to generate ψ-mixing continuous state-space strictly stationary Markov chains for any a and b, achieving at best "lower ψ-mixing." Longla [24] also noted the absence of ψ-mixing for copulas from Mardia and Frechet families for any a and b. Different results emerge when discrete marginal distributions are employed.

We define two mixing coefficients in this article: the ϕ-mixing coefficient, introduced by Ibragimov [15] and further studied by Cogburn [8], and the ψ-mixing coefficient, first introduced by Blum et al. [5] and refined

by Philip [28]. In the context of a probability space $(\Omega, \mathcal{F}, P)$ with sigma algebras $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$, the coefficients are defined in the literature as follows (see [6]):

$$\phi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0} \left| \frac{P(B \cap A)}{P(A)} - P(B) \right|,$$

$$\psi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0, P(B) > 0} \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right|.$$

For random variables n-steps apart such that $X_0 \in A$ and $X_n \in B$, we define the mixing coefficients

$$\phi(n) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0} \left| \frac{P^n(B \cap A)}{P(A)} - P(B) \right|, \tag{2.8}$$

$$\psi(n) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0, P(B) > 0} \left| \frac{P^n(A \cap B)}{P(A)P(B)} - 1 \right|, \tag{2.9}$$

where $P^n(A \cap B) = P(X_0 \in A, X_n \in B)$. A random process is said to be ψ-mixing if $\psi(n) \to 0$ as $n \to \infty$ and ϕ-mixing if $\phi(n) \to 0$ as $n \to \infty$. For the case involving discrete state-space Markov chains generated by the copula given by (2.1) and Bernoulli(p) marginal distribution, we establish the following theorem:

Theorem 3. *Let $X = (X_t)_{t=0,\dots,n}$ be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginal distributions. Then, X is exponentially ψ-mixing with $\psi(n) = \frac{1-p}{p} \left| \frac{a-p}{1-p} \right|^n$, for $p \geq 1/2$, and $\psi(n) = \frac{p}{1-p} \left| \frac{a+p-1}{p} \right|^n$, for $p < 1/2$. Hence, X is also exponentially ϕ-, ρ-, β-, and α-mixing.*

Proof. Consider the coefficient defined by equation (2.9) with $\Omega = \{0, 1\}$. For $p < 1/2$, we examine

$$\left| \frac{P^n(A \cap B)}{P(A)P(B)} - 1 \right|,$$

where A and B are chosen only from $\{0\}$ or $\{1\}$ since $P(\{0, 1\} \cap A) = P(A)$. The joint distribution of (X_0, X_n) when $p < 1/2$ is obtained from the nth transition matrix (2.5) and using $P(X_1 = j, X_0 = i) = P(X_1 = j | X_0 = i)P(X_0 = i)$, where $i, j = 0, 1$. This is given in Table 2.

After computing the four quantities,

$$\left| \frac{P(X_0 = i, X_n = j)}{P(X_0 = i)P(X_n = j)} - 1 \right|.$$

From Table 2, we end up with

$$\psi(n) = \sup \left\{ \frac{p}{1-p} \left| \frac{a-p}{1-p} \right|^n, \left| \frac{a-p}{1-p} \right|^n, \frac{1-p}{p} \left| \frac{a-p}{1-p} \right|^n \right\} = \frac{1-p}{p} \left| \frac{a-p}{1-p} \right|^n.$$

For $p \geq 1/2$, a similar argument leads to $\psi(n) = \frac{p}{1-p} \left| \frac{a+p-1}{p} \right|^n$. Therefore, for all $a, p \in (0, 1)$, $\psi(n) \to 0$ at an exponential rate as $n \to \infty$. The rest of the proof follows from the fact that ψ-mixing implies.   □

Table 2: Distribution of (X_0, X_n) when $p < 1/2$

X_0	X_n		$P_{X_0}(x)$
	0	1	
0	$(1-p)^2 + p(1-p)\left[\frac{a-p}{1-p}\right]^n$	$p(1-p) - p(1-p)\left[\frac{a-p}{1-p}\right]^n$	$1-p$
1	$p(1-p) - p(1-p)\left[\frac{a-p}{1-p}\right]^n$	$p^2 + p(1-p)\left[\frac{a-p}{1-p}\right]^n$	p
$P_{X_n}(y)$	$1-p$	p	1

3 Parameter estimation asymptotic normality

The characteristics of the model parameter estimators, including the shape of their asymptotic distributions, rely on the chosen marginal distribution.

3.1 Markov chain with uniform marginals

Assuming $(X_t)_{t=0,\dots,n}$ is a Markov chain generated by copula (2.1) with uniform$(0, 1)$ marginal distribution and utilizing Sklar's theorem, the joint cdf of (X_t, X_{t+1}) is represented as:

$$F(x_t, x_{t+1}) = C(F(x_t), F(x_{t+1})) = aM(x_t, x_{t+1}) + (1 - a)W(x_t, x_{t+1}). \tag{3.1}$$

The copulas M and W have support on the main diagonal and secondary diagonal of $[0, 1]^2$, respectively [27]. The joint distribution (3.1) implies $P(X_t = X_{t+1}) = a$ and $P(X_t = 1 - X_{t+1}) = 1 - a$ due to the properties of mixture distributions. The indicator random variable $I(X_t = X_{t+1})$, $t \in 0, 1, \dots, n - 1$, taking the value 1 when $X_t = X_{t+1}$ and 0 otherwise, follows a Bernoulli(a) distribution with probability of success a and probability of failure $1 - a$. The parameter a can be estimated by the sample mean:

$$Y_n = \frac{1}{n} \sum_{t=0}^{n-1} I(X_t = X_{t+1}), \tag{3.2}$$

as demonstrated in Proposition 4.

Proposition 4. *Let $(X_t)_{t=0,\dots,n}$ be a Markov chain generated by copula* (2.1) *and the uniform* $(0, 1)$ *distribution. We have the following*:

  i. *The random variables $I(X_t = X_{t+1})$, $t = 0, \dots, n - 1$ are independent and identically distributed (i.i.d) Bernoulli random variables.*
 ii. *The random variable* (3.2) *is an unbiased and consistent estimator of a. Moreover, Y_n satisfies the central limit theorem.*
iii. *The MLE of a is the method of moment estimator given by* (3.2).

Proof.

  i. We prove independence in the sequence by induction. For $m = 2$ random variables, let us consider without loss of generality the random variables $I(X_1 = X_0)$ and $I(X_2 = X_1)$. Then,

$$P(X_1 = X_0, X_2 = X_1) \quad = \int_0^1 P(X_1 = X_0, X_2 = X_1 | X_0 = x) \, dP_{X_0}(x),$$

   (by the law of total probability, and since under the copula M, the joint distribution of (X_0, X_1) is concentrated on the main diagonal)

$$= \int_0^1 P(X_1 = x | X_0 = x) \cdot P(X_2 = x | X_0 = x, \quad X_1 = x) dx,$$

   (by the multiplication rule, and using that $X_0 \sim \text{Unif}(0, 1)$)

$$= \int_0^1 P(X_1 = x | X_0 = x) \cdot P(X_2 = x | X_1 = x) dx,$$

   (by the Markov property)

$$= \int_0^1 a \cdot a \, dx = a^2,$$

   since $P(X_{t+1} = x | X_t = x) = a$ for all $x \in [0, 1]$ and $t = 0, \dots, n - 1$. Therefore,

$$P(X_1 = X_0, X_2 = X_1) = P(X_1 = X_0)P(X_2 = X_1), \tag{3.3}$$

establishing independence for the base case.

ii. Induction step: Suppose independence holds for $m = k > 2$, i.e.,

$$P(X_1 = X_0, X_2 = X_1, ..., X_{k+1} = X_k) = \prod_{t=0}^{k} P(X_{t+1} = X_t) = a^{k+1}. \tag{3.4}$$

We now prove it for $m = k + 1$. Consider

$$P(X_1 = X_0, X_2 = X_1, ..., X_{k+1} = X_k, X_{k+2} = X_{k+1})$$
$$= P(X_{k+2} = X_{k+1}|X_{k+1} = X_k, ..., X_2 = X_1, X_1 = X_0)$$
$$\times P(X_1 = X_0, X_2 = X_1, ..., X_{k+1} = X_k) \text{ (by multiplication rule of probabilities)}$$
$$= P(X_{k+2} = X_{k+1}|X_{k+1} = X_k) \times \prod_{t=0}^{k} P(X_{t+1} = X_t) \quad \text{(by the Markov property and (3.4))}$$
$$= \prod_{i=0}^{k+1} P(X_{t+1} = X_t) = a^{k+2} \quad \text{(due to (3.3)).}$$

Thus, by induction, independence holds for all m. To show that the random variables $I(X_{t+1} = X_t)$ are Bernoulli(a), note that by the model assumption, for each t, we have $P(X_{t+1} = X_t) = a$. Therefore, the indicator variable

$$I(X_{t+1} = X_t) = \begin{cases} 1, & \text{if } X_{t+1} = X_t, \\ 0, & \text{otherwise}, \end{cases}$$

satisfies

$$P(I(X_{t+1} = X_t) = 1) = a, \quad P(I(X_{t+1} = X_t) = 0) = 1 - a,$$

which implies that $I(X_{t+1} = X_t) \sim \text{Bernoulli}(a)$.

iii. The proof follows from the properties of the estimator of the Bernoulli parameter for an i.i.d. Bernoulli(a) sample.

iv. Let $Y_t = I(X_t = X_{t+1})$,

$$L(X_t, X_{t+1}) = a^{Y_t}(1 - a)^{1-Y_t}. \tag{3.5}$$

The likelihood function is

$$L(a) = \prod_{t=0}^{n-1} L(X_t, X_{t+1}) = a^{\sum_{t=0}^{n-1} Y_t}(1 - a)^{n-\sum_{t=0}^{n-1} Y_t}. \tag{3.6}$$

It follows that $\hat{a} = \frac{1}{n}\sum_{t=0}^{n-1} Y_t = \frac{1}{n}\sum_{t=0}^{n-1} I(X_t = X_{t+1})$.   □

3.2 Stationary Markov chain with Bern(p) marginal distribution

The literature has extensively discussed the statistical inference for finite Markov chains. When transition probabilities of the Markov chain are expressed as $p_{ij}(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\theta_1, ..., \theta r)$, the log-likelihood is $\sum_{ij} n_{ij} \log p_{ij}(\boldsymbol{\theta})$ (see Billingsley [3,4]). If the r components of $\boldsymbol{\theta}$ are all real, then the maximum-likelihood equations are

$$\sum_{ij} \frac{n_{ij}}{p_{ij}(\boldsymbol{\theta})} \frac{\partial p_{ij}(\boldsymbol{\theta})}{\partial \theta_u} = 0, \quad u = 0, 1, ..., r. \tag{3.7}$$

Considering transition matrix (2.3) for instance, the likelihood function of the sample is

$$L(\boldsymbol{\theta}) = p^{x_0}(1 - p)^{1-x_0} \left(\frac{ap + 1 - 2p}{1 - p}\right)^{n_{00}} \left(\frac{p(1 - a)}{1 - p}\right)^{n_{01}} (1 - a)^{n_{10}} a^{n_{11}} \tag{3.8}$$

and the log-likelihood is given by

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\
&= x_0 \log(p) + (1 - x_0)\log(1 - p) + n_{00}\log(ap + 1 - 2p) \\
&\quad - n_{00}\log(1 - p) + n_{01}\log(p) + n_{01}\log(1 - a) - n_{01}\log(1 - p) \\
&\quad + n_{10}\log(1 - a) + n_{11}\log(a).
\end{aligned} \tag{3.9}
$$

Differentiating (3.9) with respect to a and p gives

$$
\begin{cases}
\dfrac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial a} = \dfrac{n_{00}p}{ap + 1 - 2p} - \dfrac{n_{01}}{1 - a} - \dfrac{n_{10}}{1 - a} + \dfrac{n_{11}}{a} = 0 \\[2ex]
\dfrac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial p} = \dfrac{x_0}{p} - \dfrac{1 - x_0}{1 - p} + \dfrac{n_{00}(a - 2)}{ap + 1 - 2p} + \dfrac{n_{00}}{1 - p} + \dfrac{n_{01}}{p} + \dfrac{n_{01}}{1 - p} = 0.
\end{cases} \tag{3.10}
$$

The equations in (3.10) can be rewritten in the form:

$$
\begin{cases}
npa^2 - [(2n - n_{00} + n_{11})p - n + n_{00}]a + n_{11}(2p - 1) = 0, \\
a[x_0 p - p^2 + n_{00}p + n_{01}p] + x_0 - 2x_0 p - p + 2p^2 - n_{00}p + n_{01} - 2n_{01}p = 0.
\end{cases} \tag{3.11}
$$

Solving for a in the second equation of (3.11) gives

$$
\hat{a} = \frac{p(\lambda_1 - 2p) - \lambda_2}{p(\lambda_3 - p)}, \tag{3.12}
$$

where $\lambda_1 = 2x_0 + 1 + n_{00} + 2n_{01}$, $\lambda_2 = x_0 + n_{01}$, and $\lambda_3 = x_0 + n_{00} + n_{01}$. Substituting (3.12) for a in the first equation of (3.11) and using $\lambda_4 = 2n - n_{00} + n_{11}$, and $\lambda_5 = n - n_{00}$, we obtain

$$
\begin{aligned}
&(4n - 2\lambda_4 + 2n_{11})p^4 + (2\lambda_3\lambda_4 - 4n\lambda_1 + \lambda_1\lambda_4 + 2\lambda_5 - n_{11} - 4n_{11}\lambda_3)p^3 \\
&\quad + (n\lambda_1^2 + 4n\lambda_2 - \lambda_1\lambda_3\lambda_4 - \lambda_2\lambda_4 - 2\lambda_3\lambda_5 - \lambda_1\lambda_5 + 2n_{11}\lambda_3 + 2n_{11}\lambda_3^2)p^2 \\
&\quad + (\lambda_1\lambda_3\lambda_5 - 2n\lambda_1\lambda_2 + \lambda_2\lambda_3\lambda_4 + \lambda_2\lambda_5 - n_{11}\lambda_3^2)p + (n\lambda_2^2 - \lambda_2\lambda_3\lambda_5) = 0.
\end{aligned} \tag{3.13}
$$

There is no closed-form solution for equation (3.13) but numerical solutions can be obtained.

The sufficient conditions for MLEs and their asymptotic properties are specified in Condition 5.1 of Billingsley [4], as presented in the following.

Condition 5. Let $X = (X_t)_{t=0,\dots,n}$ be a Markov chain with finite state space $S = \{0, \dots, s\}$. Suppose $P = (p_{ij}(\boldsymbol{\theta}))$ is the $s \times s$ stochastic matrix, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ ranges over an open subset Θ of $\mathbb{R}^r$. Suppose that the following conditions are satisfied:

i. The set E of (i, j) such that $p_{ij}(\boldsymbol{\theta}) > 0$ is independent of $\boldsymbol{\theta}$ and each $p_{ij}(\boldsymbol{\theta})$ has continuous partial derivatives of third order throughout Θ.

ii. The $d \times r$ matrix $D = (\frac{\partial p_{ij}(\boldsymbol{\theta})}{\partial \theta_k})$, $k = 1, \dots, r$, where d is the number of elements in E, has rank r throughout Θ.

iii. For each $\boldsymbol{\theta} \in \Theta$, there is only one ergodic set and no transient states.

Then, there exists $\hat{\boldsymbol{\theta}} \in \Theta$ such that, with probability tending to 1, $\hat{\boldsymbol{\theta}}$ is a solution of system (3.7) and converges in probability to the true value $\boldsymbol{\theta}$. Additionally, $\sqrt{(n+1)}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_r(\mathbf{0}, \Sigma^{-1})$, where Σ^{-1} is the inverse of matrix Σ with entries defined by

$$
\sigma_{uv}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\frac{\partial \log p_{x_1 x_2}(\boldsymbol{\theta})}{\partial \theta_u} \frac{\partial \log p_{x_1 x_2}(\boldsymbol{\theta})}{\partial \theta_v} \right], \quad u, v = 1, \dots, r. \tag{3.14}
$$

.   This condition can be used to verify the following theorem:

Theorem 6. *Let $X = (X_t)_{t=0,\dots,n}$ be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginal distribution. Then, there exists a consistent MLE, $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{p})$ of $\boldsymbol{\theta} = (a, p)$ such that $\sqrt{(n+1)}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_2(\mathbf{0}, \Sigma^{-1})$, where*

$$\Sigma^{-1} = \begin{cases} \begin{pmatrix} \dfrac{a(1-a)}{p} & a(1-p) \\[2ex] a(1-p) & \dfrac{p(1-p)(a+1-2p)}{1-a} \end{pmatrix}, & \text{when} \quad p < 1/2, \\[6ex] \begin{pmatrix} \dfrac{a(1-a)}{1-p} & -ap \\[2ex] -ap & \dfrac{p(1-p)(2p-1+a)}{1-a} \end{pmatrix}, & \text{when} \quad p > 1/2. \end{cases} \tag{3.15}$$

Note that $p = 1/2$ does not fall into either of these two cases because Condition (5) applies only to interior points of the range of (a, p).

Proof of Theorem (6). The proof of Theorem (6) follows by first verifying that the requirements outlined in Condition (5) are met for each of the cases, $p < 1/2$ and $p > 1/2$. The entries of the information matrices, Σ, are obtained by computing the expectations given in equation (3.14). Inverting Σ, gives (3.15). The complete proof can be found in Muia [26]. □

When $p = 1/2$, the stationary Markov chain based on copula (2.1) and Bernoulli(1/2) margins has a transition matrix of the form

$$P = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix}. \tag{3.16}$$

In this case, we are left with a to estimate. The likelihood function for the chain is given by

$$L(\mathbf{x}, a) = \frac{1}{2} a^{n_{00}+n_{11}} (1-a)^{n-(n_{00}+n_{11})},$$

and the log-likelihood is expressed as

$$\mathcal{L}(\mathbf{x}; a) = (n_{00} + n_{11}) \log(a) + (n - (n_{00} + n_{11})) \log(1-a). \tag{3.17}$$

The following proposition provides the MLE of a and its asymptotic properties in this case.

Proposition 7. *Let* $(X_t)_{t=0,\dots,n}$ *be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginal distribution. When* $p = 1/2$, *the parameter* a *has MLE* $\hat{a} = \frac{n_{00}+n_{11}}{n}$. *Moreover,* $\frac{\sqrt{(n+1)}(\hat{a}-a)}{\sqrt{a(1-a)}} \xrightarrow{d} N(0,1)$.

Proof. The proof of Proposition (7) follows from verifying the requirements of Condition (5). The variance of a is computed using formula (3.14) with $r = 1$. □

Klotz showed that the MLE, $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{p})$, is asymptotically equivalent to $(a^*(\bar{p}), \bar{p})$, where $a^*(\bar{p})$ is the MLE of a when p is replaced by $\bar{p} = 1/(n+1)\sum_{i=0}^{n} X_i$. The proof relies on the fact that $\hat{p}$ and $\bar{p}$ have the same asymptotic variance. Lindqvist [21] considers a general case of non-Markovian dependence and shows that the MLE, $\hat{p}$ is asymptotically equivalent to $\bar{p}$ given that the process is ergodic.

In this work, different asymptotically equivalent estimators for p are discussed. Since the Markov chain under consideration is stationary, it follows that $P(X_t) = p \ \forall \ t$. The sample proportion, $\bar{p} = 1/(n+1)\sum_{i=0}^{n} X_i$, is an unbiased estimator for p (see Bedrick and Aragon [2], Klotz [20] or Lindqvist [21]).

Proposition 8. *Let* $X = (X_t)_{t=0,\dots,n}$ *be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginal distribution. The asymptotic variance of the sample proportion,* $\bar{p}$, *is the same as that of* $\hat{p}$.

Proof. For the Markov chain generated when $p < 1/2$

$$\text{var}(\bar{p}) = \text{var}\left(\frac{1}{n+1}\sum_{t=0}^{n}X_t\right)$$

$$= \frac{1}{(n+1)^2}\left[(n+1)\text{var}(X_0) + 2\sum_{t=0}^{n-1}\sum_{k=1}^{n-t}\text{cov}(X_t, X_{t+k})\right]$$

$$= \frac{1}{(n+1)^2}\left[(n+1)p(1-p) + 2p(1-p)\sum_{t=0}^{n-1}\sum_{k=1}^{n-t}\left(\frac{a-p}{1-p}\right)^k\right].$$

Applying the properties of a geometric sum twice and simplifying, we obtain

$$\text{var}(\bar{p}) = \frac{p(1-p)}{n+1}\left[1 + 2\left(\frac{a-p}{1-a}\right)\left(\frac{n}{n+1} + \frac{1}{n+1}\left(\frac{a-p}{1-a}\right)\left(\left(\frac{a-p}{1-p}\right)^n - 1\right)\right)\right].$$

Therefore, for large n,

$$\text{var}(\bar{p}) \approx \frac{p(1-p)}{n+1}\left[1 + \frac{2a-2p}{1-a}\right] + O\left(\frac{1}{n}\right) = \frac{p(1-p)}{n+1}\left[\frac{a+1-2p}{1-a}\right] + O\left(\frac{1}{n}\right). \tag{3.18}$$

Similarly, for $p \geq 1/2$,

$$\text{var}(\bar{p}) \approx \frac{p(1-p)}{n+1}\left[1 + \frac{2a+2p-2}{1-a}\right] + O\left(\frac{1}{n}\right) = \frac{p(1-p)}{n+1}\left[\frac{2p-1+a}{1-a}\right] + O\left(\frac{1}{n}\right) \tag{3.19}$$

From (3.18) and (3.19), we note that the asymptotic variance of the sample mean is the same as that of $\hat{p}$ (refer to Theorem (6) for the asymptotic variance of $\hat{p}$). □

The following result provides the asymptotic properties of the sample mean for the stationary Markov chain based on copula (2.1).

Theorem 9. *Let $X = (X_t)_{t=0,\ldots,n}$ be a stationary Markov chain generated by copula* (2.1) *and Bernoulli(p) marginal distribution. Let $\bar{p}$ be the sample mean. It follows that*

$$\sqrt{(n+1)}(\bar{p} - p) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = p(1-p)(1+a-2p)/(1-a)$, $p < 1/2$ and $\sigma^2 = p(1-p)(a+2p-1)/(1-a)$, $p \geq 1/2$.

Proof. The proof of Theorem (9) follows from Theorem (18.5.2) of Ibragimov and Linnik [16] and the fact that the Markov chain is uniformly mixing. Using steps similar to those used in the proof of Theorem (3), the ϕ-mixing coefficient is found to be

$$\phi(n) = \begin{cases} (1-p)\left|\dfrac{a-p}{1-p}\right|^n, & p < 1/2, \\[4mm] p\left|\dfrac{a+p-1}{p}\right|^n, & p \geq 1/2. \end{cases}$$

Let $p < 1/2$. To check Condition (18.5.8) of Theorem (18.5.2) of Ibragimov and Linnik [16], we see that

$$\sum_{n=0}^{\infty}\{\phi(n)\}^{1/2} = \sum_{n=0}^{\infty}\sqrt{(1-p)}\left(\left|\frac{a-p}{1-p}\right|^{1/2}\right)^n \tag{3.20}$$

converges since $|(a-p)/(1-p)|^{1/2} < 1$. The proof for $p \geq 1/2$ is identical. The variance of $\bar{X}_n$ is given by the formulae (3.18) and (3.19). Hence, Condition (18.5.9) of Theorem (18.5.2) of Ibragimov and Linnik [16] is satisfied. □

4 Hypothesis testing

4.1 Test of independence

Different tests have been used in the literature to assess the independence of observations in a sequence. These include the χ^2 goodness-of-fit tests and the likelihood ratio test (LRT), as discussed by Anderson and Goodman [1], Goodman [14], Billingsley (1961a) [4], and others.

Having $p_{ij} = p_j$ for all j is sufficient for independence in a finite state-space Markov chain. The following proposition outlines, conditions for independence in the Markov chain based on copula (2.1) and Bernoulli(p) marginal distributions.

Proposition 10. *Let $(X_t)_{t=0,\ldots,n}$ be a stationary Markov chain with copula (2.1) and Bernoulli(p) marginal distribution. Then, $X_0,\ldots,X_n$ are independent when $a = p < 1/2$ or $a = 1 - p \leq 1/2$.*

Proof. For independent observations in a sequence, $p_{ij} = p_j$ for all states i, j, where p_j is the probability of outcome j. When $p < 1/2$, conditions for independence are given by $(ap + 1 - 2p)/(1 - p) = 1 - a$ and $p(1 - a)/(1 - p) = a$, leading to $a = p$ in both cases. For $p \geq 1/2$, independence holds if $a = (1 - p)(1 - a)/p$ and $1 - a = (a(1 - p) + 2p - 1)/p$, simplifying to $a = 1 - p$ in each case. □

The likelihood ratio approach for testing the hypothesis of independence is obtained as follows: for $p < 1/2$, the likelihood function is expressed as shown in (3.8). The hypotheses under consideration for the test are denoted as $H_0 : \theta \in \omega$ versus $H_1 : \theta \in \Theta \cap \omega^c$, where $\Theta = \{\theta = (a, p), 0 < a < 1, 0 < p < 1/2\}$ and $\omega = \{\theta = (p, p), 0 < p < 1/2\} \subset \Theta$. Let $\hat{\theta}$ be the MLE under H_a and $\hat{\theta}_0$ the MLE under H_0. The LR test statistic is then given by

$$\Lambda = \frac{\max_{\theta \in \omega} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}. \tag{4.1}$$

The likelihood under H_0 is expressed as

$$L(\omega) = p^{x_0 + n_{01} + n_{11}}(1 - p)^{1 - x_0 + n_{00} + n_{10}}. \tag{4.2}$$

The function (4.2) represents the likelihood of an i.i.d Bern(p) sequence. Note that $x_0 + n_{01} + n_{11} = X =$ number of ones. The MLE under H_0 is

$$\hat{p} = \frac{x_0 + n_{01} + n_{11}}{n + 1} = \frac{X}{n + 1} \quad \text{and} \quad \hat{\theta}_0 = (\hat{p}, \hat{p}).$$

The likelihood function evaluated at the MLE under H_0 simplifies to

$$L(\hat{\omega}) = \left(\frac{X}{n + 1}\right)^X \left(1 - \frac{X}{n + 1}\right)^{n + 1 - X}.$$

The LRT statistic is therefore

$$\Lambda_1 = \frac{\left(\frac{X}{n+1}\right)^X \left(1 - \frac{X}{n+1}\right)^{n+1-X}}{\hat{p}^{x_0}(1 - \hat{p})^{1-x_0}\left(\frac{\hat{a}\hat{p} + 1 - 2\hat{p}}{1 - \hat{p}}\right)^{n_{00}}\left(\frac{\hat{p}(1 - \hat{a})}{1 - \hat{p}}\right)^{n_{01}}(1 - \hat{a})^{n_{10}}\hat{a}^{n_{11}}}. \tag{4.3}$$

For a given significance level α, reject H_0 in favor of H_1 if $\Lambda_1 < c$, where c is such that $\alpha = P_{H_0}[\Lambda_1 \leq c]$.

A result due to Wilks [31] shows that under suitable regularity conditions, if Θ is a r-dimensional space and ω is a k-dimensional space, then $-2\log\Lambda_1$ has an asymptotically χ^2 distribution with $r - k$ degrees of freedom. The fulfillment of these desired conditions has been confirmed for our model by demonstrating that the

transition probabilities satisfy the requirements outlined in Condition (5). The asymptotic distribution of $-2\log\Lambda_1$ is χ^2 with 1 degree of freedom. The decision rule is to reject H_0 in favor of H_1 if $-2\log\Lambda_1 \geq \chi_a^2(1)$.

In a similar manner, for $p > 1/2$, define $\Theta = \{\theta = (a, p) : 0 < a < 1, 1/2 < p < 1\}$ and $\omega = \{\theta = (1 - p, p) : 1/2 < p < 1\}$. The distribution does not change under H_0, but changes under H_a. The LRT statistic becomes

$$\Lambda_2 = \frac{\left(\frac{X}{n+1}\right)^X\left(1 - \frac{X}{n+1}\right)^{n+1-X}}{\hat{p}^{x_0}(1 - \hat{p})^{1-x_0}\left(\frac{(1 - \hat{p})(1 - \hat{a})}{\hat{p}}\right)^{n_{10}}\left(\frac{\hat{a}(1 - \hat{p}) + 2\hat{p} - 1}{\hat{p}}\right)^{n_{11}}(1 - \hat{a})^{n_{01}}\hat{a}^{n_{00}}}. \tag{4.4}$$

The limiting distribution of $-2\log\Lambda_2$ approaches a chi-squared distribution with one degree of freedom as $n \to \infty$. The level α decision rule is to reject H_0 in favor of H_1 if $-2\log\Lambda_2 \geq \chi_a^2(1)$.

Table 3: Average length of 400, 95% confidence intervals, and the coverage probabilities (CPs) for the true values of a and p when $p < 1/2$

n	Initial parameters	CIML for "a"	CIML for "p"	CP for "a"	CP for "p"
499	$a = 0.1, p = 0.1$	0.1586	0.0523	0.8925	0.9425
	$a = 0.1, p = 0.3$	0.0950	0.0598	0.9300	0.9500
	$a = 0.1, p = 0.4$	0.0821	0.0494	0.9475	0.9500
	$a = 0.2, p = 0.1$	0.2172	0.0584	0.9050	0.9425
	$a = 0.2, p = 0.3$	0.1273	0.0695	0.9475	0.9575
	$a = 0.2, p = 0.4$	0.1100	0.0605	0.9475	0.9525
	$a = 0.7, p = 0.1$	0.2671	0.1153	0.9675	0.9050
	$a = 0.7, p = 0.3$	0.1479	0.1530	0.9650	0.9625
	$a = 0.7, p = 0.4$	0.1275	0.1490	0.9500	0.9475
	$a = 0.9, p = 0.1$	0.2237	0.2000	0.9500	0.8350
	$a = 0.9, p = 0.3$	0.1014	0.2830	0.9675	0.9175
	$a = 0.9, p = 0.4$	0.0869	0.2797	0.9650	0.8825
999	$a = 0.1, p = 0.1$	0.1146	0.0370	0.9125	0.9325
	$a = 0.1, p = 0.3$	0.0674	0.0423	0.9525	0.9500
	$a = 0.1, p = 0.4$	0.0585	0.0351	0.9275	0.9350
	$a = 0.2, p = 0.1$	0.1554	0.0414	0.9300	0.9200
	$a = 0.2, p = 0.3$	0.0902	0.0491	0.9475	0.9350
	$a = 0.2, p = 0.4$	0.0782	0.0429	0.9500	0.9400
	$a = 0.7, p = 0.1$	0.1854	0.0818	0.9600	0.9225
	$a = 0.7, p = 0.3$	0.1042	0.1084	0.9550	0.9550
	$a = 0.7, p = 0.4$	0.0899	0.1050	0.9350	0.9575
	$a = 0.9, p = 0.1$	0.1373	0.1446	0.9625	0.8825
	$a = 0.9, p = 0.3$	0.0693	0.2035	0.9500	0.9275
	$a = 0.9, p = 0.4$	0.0597	0.1997	0.9550	0.9600
4,999	$a = 0.1, p = 0.1$	0.0524	0.0166	0.9450	0.9600
	$a = 0.1, p = 0.3$	0.0303	0.0189	0.9575	0.9525
	$a = 0.1, p = 0.4$	0.0263	0.0157	0.9475	0.9500
	$a = 0.2, p = 0.1$	0.0700	0.0186	0.9600	0.9525
	$a = 0.2, p = 0.3$	0.0404	0.0220	0.9500	0.9475
	$a = 0.2, p = 0.4$	0.0350	0.0192	0.9600	0.9325
	$a = 0.7, p = 0.1$	0.0809	0.0370	0.9600	0.9325
	$a = 0.7, p = 0.3$	0.0465	0.0486	0.9600	0.9650
	$a = 0.7, p = 0.4$	0.0402	0.0471	0.9675	0.9775
	$a = 0.9, p = 0.1$	0.0541	0.0678	0.9300	0.9425
	$a = 0.9, p = 0.3$	0.0306	0.0913	0.9425	0.9525
	$a = 0.9, p = 0.4$	0.0264	0.0899	0.9550	0.9525

5 Simulation study

5.1 MLEs of a and p

Tables 3 and 4 present the MLEs of a and p for various sample sizes and selected initial value pairs. Few pairs of a and p were selected for this study to avoid very lengthy tables. We used different combinations of true values for a and p and constructed 95% confidence intervals accordingly. Utilizing the asymptotic normality described in Theorem 6, we derived confidence intervals for a and p based on the standard normal distribution. Furthermore, we generated 400 replications for each sequence of length n, which were used to calculate the CPs of the true values by the confidence intervals, as well as the mean lengths of the confidence intervals (CIML). Note that we did not include the numerical values of the estimators because 400 numerical values were computed for each estimator in each case, and there is no criterion to choose only one representative from the 400. Relatively narrow confidence intervals reflect a higher level of precision in estimation, while wider confidence intervals indicate greater uncertainty in estimation.

Table 4: Average length of 400, 95% confidence intervals, and the CPs for the true values of a and p when $p > 1/2$

n	a	p	CIML for "a"	CIML for "p"	CP for "a"	CP for "p"
499	0.1	0.6	0.0821	0.0494	0.9475	0.9500
	0.1	0.7	0.0950	0.0598	0.9300	0.9500
	0.1	0.9	0.1593	0.0524	0.8925	0.9425
	0.3	0.6	0.1267	0.0726	0.9500	0.9575
	0.3	0.7	0.1460	0.0803	0.9275	0.9550
	0.3	0.9	0.2514	0.0658	0.9325	0.9550
	0.7	0.6	0.1275	0.1484	0.9475	0.9475
	0.7	0.7	0.1479	0.1530	0.9650	0.9625
	0.7	0.9	0.2671	0.1153	0.9675	0.905
	0.9	0.6	0.0868	0.2797	0.9600	0.9625
	0.9	0.7	0.1014	0.2830	0.9675	0.9175
	0.9	0.9	0.2237	0.2000	0.9500	0.8350
999	0.1	0.6	0.0585	0.0351	0.9275	0.9350
	0.1	0.7	0.0674	0.0423	0.9525	0.9500
	0.1	0.9	0.1146	0.0370	0.9125	0.9325
	0.3	0.6	0.0896	0.0512	0.9500	0.9450
	0.3	0.7	0.1034	0.0567	0.9450	0.9325
	0.3	0.9	0.1787	0.0464	0.9475	0.9350
	0.7	0.6	0.0899	0.1050	0.9350	0.9575
	0.7	0.7	0.1042	0.1084	0.9550	0.9550
	0.7	0.9	0.1854	0.0818	0.9600	0.9225
	0.9	0.6	0.0596	0.1997	0.9550	0.9550
	0.9	0.7	0.0693	0.2035	0.9500	0.9275
	0.9	0.9	0.1373	0.1446	0.9625	0.8825
4,999	0.1	0.6	0.0263	0.0157	0.9475	0.9500
	0.1	0.7	0.0303	0.0189	0.9575	0.9525
	0.1	0.9	0.0524	0.0166	0.9450	0.9600
	0.3	0.6	0.0401	0.0229	0.9300	0.9450
	0.3	0.7	0.0463	0.0254	0.9350	0.9425
	0.3	0.9	0.0803	0.0208	0.9525	0.9325
	0.7	0.6	0.0402	0.0471	0.9675	0.9775
	0.7	0.7	0.0465	0.0486	0.9600	0.9650
	0.7	0.9	0.0809	0.0370	0.9600	0.9325
	0.9	0.6	0.0264	0.0899	0.9550	0.9525
	0.9	0.7	0.0306	0.0913	0.9425	0.9525
	0.9	0.9	0.0541	0.0678	0.9300	0.9325

In Table 3, the confidence intervals for a decrease as p increases, reflecting reduced uncertainty in estimating a with higher values of $p < 0.5$. This reduction is attributed to the decreasing variance of $\hat{a}$ as $\hat{p} < 1/2$ increases. The accuracy in estimation improves with sample size. Moreover, as the sample size grows, the CP for a becomes more precise, indicating improved accuracy in capturing the true value of a by the confidence intervals.

Table 4 shows that confidence intervals lengthen as p increases due to the increasing variance of $\hat{a}$ for $\hat{p} > 1/2$. The CPs for both a and p vary with different combinations of initial values and sample sizes, underscoring the significance of careful parameter selection in statistical inference. Nevertheless, as sample size increases, both CPs tend to converge towards the desired nominal value of 95%.

Furthermore, the results presented in Tables 3 and 4 demonstrate a noteworthy symmetry in the CIMLs when comparing parameter pairs equidistant from $p = 0.5$ for fixed a, i.e., for each fixed value of a, and for p values equidistant from $p = 0.5$, the CIMLs for a and p are the same. Figure 1 shows this symmetry for $n = 9,999$.

The observed behavior of the CIML for a and p can be explained by considering the variance-covariance matrix Σ^{-1} from Theorem 6, which governs the precision of the MLEs $\hat{a}$ and $\hat{p}$. When p is closer to 0.5, the off-diagonal elements (representing the covariance between a and p) and the denominators in the variance terms become more influential. As p increases from 0.1 to 0.4 (for $p < 0.5$), the CIML for a decreases, indicating reduced uncertainty in estimating a due to lower variance. For $p > 0.5$, as p moves further away from 0.5, the CIML for a increases again because of increased variance terms and covariance between a and p. The symmetry in the CIML for a around $p = 0.5$ (e.g., similar CIML for $p = 0.1$ and $p = 0.9$) suggests that confidence intervals widen as p deviates from 0.5 in either direction, reflecting greater difficulty in estimating a at extreme values of p.

For p, the mean length of the confidence intervals is relatively stable across different values, with small fluctuations likely due to the interaction between a and p in the variance terms of Σ^{-1}. For small a (e.g., $a = 0.1$), the CIML for p is slightly lower when p is near 0.1 and 0.9 compared to intermediate values, indicating higher certainty in extreme probability scenarios. Given the large sample size ($n = 9,999$), the confidence intervals are generally narrow, as the variance of the MLEs decreases with increasing sample size, leading to more precise estimates.

Overall, the behavior of the CIML for a and p reflects the interplay between the values of a and p, their interaction in the variance-covariance structure, and the large sample size ensuring narrow confidence intervals. The symmetry around $p = 0.5$ and the increasing interval lengths as p approaches the extremes illustrate the underlying statistical properties of the MLEs and their variances.

This symmetry reflects the underlying properties of the mixture copula model. The behavior suggests that the model treats the transition probabilities in a symmetric manner around $p = 0.5$, ensuring that the statistical characteristics such as interval lengths and CPs remain stable for p values that are symmetrically opposite
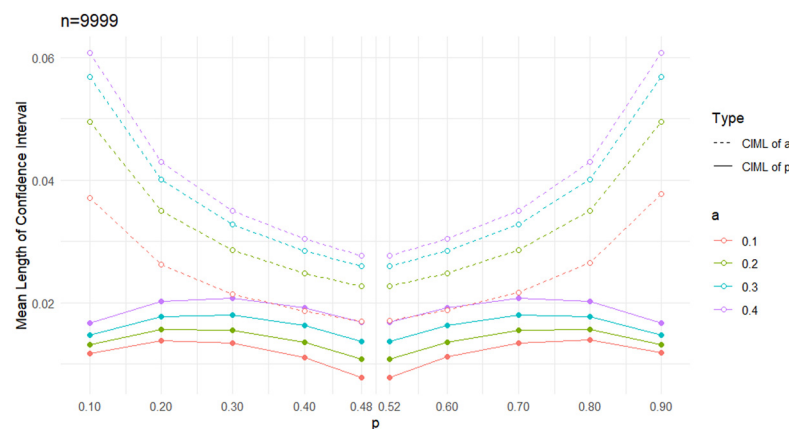

Figure 1: Symmetry in mean lengths of confidence intervals for "p" about $p = 0.5$.

about 0.5. This property enhances the robustness of the parameter estimation process within the specified range of p.

5.2 Test for independence

5.2.1 Likelihood ratio test for independence

Tables 5 and 7 present the results of the likelihood ratio test for independence in the Markov chain generated by copula model (2.1) and Bernoulli(p) marginal distribution. These tables provide separate results for $p < 1/2$ and $p > 1/2$, respectively.

The Markov chain is generated using true values of a and p, and subsequently, a chi-squared independence test is performed based on the test statistics $-2\log\Lambda_1$ and $-2\log\Lambda_2$, where Λ_1 and Λ_2 are given by formulae (4.3) and (4.4), respectively. A significance level of $\alpha = 0.05$ is chosen for the test, implying that

Table 5: Likelihood ratio test results

a	p	$-2 \log \Lambda_1$	Decision on H_0
0.1	0.1	1.9648	Fail to reject
0.1	0.2	75.3318	Reject
0.1	0.3	465.3507	Reject
0.1	0.4	1302.2148	Reject
0.2	0.1	62.5624	Reject
0.2	0.2	0.5426	Fail to reject
0.2	0.3	91.6928	Reject
0.2	0.4	537.1743	Reject
0.3	0.1	178.0430	Reject
0.3	0.2	85.5790	Reject
0.3	0.3	−0.6683	Fail to reject
0.3	0.4	143.3914	Reject
0.4	0.1	315.6163	Reject
0.4	0.2	282.6401	Reject
0.4	0.3	93.5413	Reject
0.4	0.4	−0.7403	Fail to reject
0.5	0.1	598.6498	Reject
0.5	0.2	678.1999	Reject
0.5	0.3	443.8051	Reject
0.5	0.4	175.6443	Reject
0.6	0.1	986.4705	Reject
0.6	0.2	1060.7808	Reject
0.6	0.3	908.5222	Reject
0.6	0.4	520.2437	Reject
0.7	0.1	1243.0918	Reject
0.7	0.2	1596.5124	Reject
0.7	0.3	1581.8311	Reject
0.7	0.4	1243.7897	Reject
0.8	0.1	1548.5890	Reject
0.8	0.2	2291.3934	Reject
0.8	0.3	2551.6879	Reject
0.8	0.4	2361.9013	Reject
0.9	0.1	2466.6171	Reject
0.9	0.2	3305.7161	Reject
0.9	0.3	3631.3346	Reject
0.9	0.4	3835.4538	Reject

the null hypothesis of independence will be rejected if $-2\log\Lambda_i \geq 3.842$, $i = 1, 2$. Conversely, if $-2\log\Lambda_i < 3.842$, the null hypothesis of independence is not rejected.

The chain is simulated for a total of $n + 1$ steps, where n is set to 9,999. The iterative process for Table 5 has true values of a ranging from 0.1 to 0.9 with a step of 0.1 and $p < 1/2$ ranging from 0.1 to 0.4 with a step of 0.1. On the other hand, Table 7 has true values of a ranging from 0.1 to 0.9 with a step of 0.1 and $p > 1/2$ ranging from 0.6 to 0.9 with a step of 0.1.

The test indicates that when $a = p < 1/2$ or $a = 1 - p \leq 1/2$, the Markov chain model generated by copula (2.1) and the Bernoulli(p) marginal distribution transforms into independent Bernoulli trials.

5.2.2 Kolmogorov-Smirnov (KS) distance

We employed the KS distance to compare the copula of the sample with the independence copula. Specifically, we used it to test whether the empirical joint distribution of (X_t, X_{t+1}) deviates significantly from the product of their marginal distributions. The KS distance is given by

Table 6: KS distance values

a	p	KS distance	Decision on H_0
0.1	0.1	0.0027	Fail to reject
0.1	0.2	0.0181	Reject
0.1	0.3	0.0603	Reject
0.1	0.4	0.1188	Reject
0.2	0.1	0.0111	Reject
0.2	0.2	0.0009	Fail to reject
0.2	0.3	0.0297	Reject
0.2	0.4	0.0775	Reject
0.3	0.1	0.0203	Reject
0.3	0.2	0.0211	Reject
0.3	0.3	0.0013	Fail to reject
0.3	0.4	0.0433	Reject
0.4	0.1	0.0284	Reject
0.4	0.2	0.0403	Reject
0.4	0.3	0.0286	Reject
0.4	0.4	0.0020	Fail to reject
0.5	0.1	0.0379	Reject
0.5	0.2	0.0611	Reject
0.5	0.3	0.0617	Reject
0.5	0.4	0.0405	Reject
0.6	0.1	0.0507	Reject
0.6	0.2	0.0786	Reject
0.6	0.3	0.0887	Reject
0.6	0.4	0.0781	Reject
0.7	0.1	0.0606	Reject
0.7	0.2	0.0992	Reject
0.7	0.3	0.1177	Reject
0.7	0.4	0.1200	Reject
0.8	0.1	0.0701	Reject
0.8	0.2	0.1189	Reject
0.8	0.3	0.1510	Reject
0.8	0.4	0.1623	Reject
0.9	0.1	0.0916	Reject
0.9	0.2	0.1445	Reject
0.9	0.3	0.1718	Reject
0.9	0.4	0.1969	Reject

$$KS = \sup_{x} |\hat{F}(X_t, X_{t+1}) - \hat{F}(X_t) \cdot \hat{F}(X_{t+1})|,$$

where $\hat{F}(X_t, X_{t+1})$ is the empirical joint distribution function of consecutive states, and $\hat{F}(X_t)$ and $\hat{F}(X_{t+1})$ are the empirical marginal distribution functions of X_t and X_{t+1}, respectively.

A large KS distance suggests a significant departure from independence, indicating the presence of dependence in the Markov chain. The results of this test are summarized in Tables 6 and 8.

For this test, we defined the null hypothesis H_0: the consecutive states in the Markov chain are independent, against the alternative hypothesis H_1: the consecutive states exhibit some form of dependence. We conducted the test at a significance level of $\alpha = 0.05$ for a sequence of length $n = 9{,}999$. The test shows that when $a = p < \frac{1}{2}$ or $a = 1 - p \leq \frac{1}{2}$, the Markov chain model generated by copula (2.1) with Bernoulli(p) marginals reduces to a sequence of independent Bernoulli trials.

Table 7: Likelihood ratio test results

a	p	$-2 \log \Lambda_2$	Decision on H_0
0.1	0.6	279.8288	Reject
0.1	0.7	107.8140	Reject
0.1	0.8	16.0485	Reject
0.1	0.9	0.3192	Fail to reject
0.2	0.6	111.3087	Reject
0.2	0.7	17.5196	Reject
0.2	0.8	0.4450	Fail to reject
0.2	0.9	12.1072	Reject
0.3	0.6	38.3697	Reject
0.3	0.7	−1.3511	Fail to reject
0.3	0.8	15.9764	Reject
0.3	0.9	26.5985	Reject
0.4	0.6	0.0341	Fail to reject
0.4	0.7	17.6239	Reject
0.4	0.8	50.2535	Reject
0.4	0.9	62.1340	Reject
0.5	0.6	16.9398	Reject
0.5	0.7	97.3027	Reject
0.5	0.8	127.9626	Reject
0.5	0.9	165.1340	Reject
0.6	0.6	118.6838	Reject
0.6	0.7	195.6122	Reject
0.6	0.8	209.1521	Reject
0.6	0.9	222.3218	Reject
0.7	0.6	279.7745	Reject
0.7	0.7	328.9172	Reject
0.7	0.8	315.8419	Reject
0.7	0.9	263.6996	Reject
0.8	0.6	522.4635	Reject
0.8	0.7	557.5287	Reject
0.8	0.8	533.2392	Reject
0.8	0.9	284.6627	Reject
0.9	0.6	768.7074	Reject
0.9	0.7	688.8740	Reject
0.9	0.8	545.8894	Reject
0.9	0.9	411.5155	Reject

5.3 Comparison of different estimators for the mean

This section is dedicated to the performance of the different estimators of the Bernoulli parameter. The MLEs are obtained numerically, and their confidence intervals are constructed. The sample mean satisfies the central limit theorem (9). The $(1 - \alpha)100\%$ confidence intervals for p are therefore

$$
\begin{cases}
\bar{p} \pm z_{\alpha/2}\left[\dfrac{\bar{p}(1 - \bar{p})}{n + 1}\left(\dfrac{1 + a - 2\bar{p}}{1 - a}\right)\right]^{1/2}, & \text{when} \quad \bar{p} < 1/2, \\[4mm]
\bar{p} \pm z_{\alpha/2}\left[\dfrac{\bar{p}(1 - \bar{p})}{n + 1}\left(\dfrac{a + 2\bar{p} - 1}{1 - a}\right)\right]^{1/2}, & \text{when} \quad \bar{p} \geq 1/2.
\end{cases}
$$

The other estimator for the mean (p) considered here is the robust estimator proposed by Longla and Peligrad [25]. It is based on a sample of dependent observations and requires some mild conditions to be satisfied on the variance of partial sums. The procedure is as follows: given the sample data $(X_t)_{0 \leq t \leq n}$ from a stationary sequence $(X_t)_{t \in \mathbb{Z}}$, we generate another random sample $(Y_t)_{0 \leq t \leq n}$ independently of $(X_t)_{0 \leq t \leq n}$ and

Table 8: KS distance values

a	p	KS distance	Decision on H_0
0.1	0.6	0.1188	Reject
0.1	0.7	0.0603	Reject
0.1	0.8	0.0181	Reject
0.1	0.9	0.0027	Fail to reject
0.2	0.6	0.0775	Reject
0.2	0.7	0.0297	Reject
0.2	0.8	0.0009	Fail to reject
0.2	0.9	0.0111	Reject
0.3	0.6	0.0433	Reject
0.3	0.7	0.0013	Fail to reject
0.3	0.8	0.0211	Reject
0.3	0.9	0.0203	Reject
0.4	0.6	0.0020	Fail to reject
0.4	0.7	0.0286	Reject
0.4	0.8	0.0403	Reject
0.4	0.9	0.0284	Reject
0.5	0.6	0.0365	Reject
0.5	0.7	0.0574	Reject
0.5	0.8	0.0559	Reject
0.5	0.9	0.0409	Reject
0.6	0.6	0.0781	Reject
0.6	0.7	0.0887	Reject
0.6	0.8	0.0786	Reject
0.6	0.9	0.0507	Reject
0.7	0.6	0.1200	Reject
0.7	0.7	0.1177	Reject
0.7	0.8	0.0992	Reject
0.7	0.9	0.0606	Reject
0.8	0.6	0.1623	Reject
0.8	0.7	0.1510	Reject
0.8	0.8	0.1189	Reject
0.8	0.9	0.0701	Reject
0.9	0.6	0.1969	Reject
0.9	0.7	0.1718	Reject
0.9	0.8	0.1445	Reject
0.9	0.9	0.0916	Reject

following a standard normal distribution. The Gaussian kernel and the optimal bandwidth h_n are used. The optimal bandwidth, as proposed in Longla and Peligrad [25], for our case is

$$h_n = \left(\frac{1}{(n+1)\sqrt{2}}\right)^{1/5}. \tag{5.1}$$

To use this estimator, the following conditions must be satisfied (i.) $(X_t)_{t \in \mathbb{Z}}$ is an ergodic sequence, (ii.) $(X_t)_{t \in \mathbb{Z}}$ has finite second moments, and (iii.) $(n+1)h_n\text{var}(\bar{p}) \to 0$ as $n \to \infty$. These conditions are easy to verify: (i.) the Markov chain X is ψ-mixing and hence ergodic. (ii.) The second condition follows from that $\text{Var}(X_0) = p(1-p) < \infty$ and $E[X_0] = p$. (iii.) Using the variance of $\bar{X}$ given by (3.18), we see that

$$(n+1)\left(\frac{1}{(n+1)\sqrt{2}}\right)^{1/5}\left[\frac{p(1-p)}{(n+1)}\left(\frac{1+a-2p}{1-a}\right) + O(1/n)\right] \to 0,$$

as $n \to \infty$. The proposed robust estimator for the mean (p) of X is

$$\tilde{p} = \frac{1}{(n+1)h_n}\sum_{t=0}^{n}X_t \exp\left[-\frac{1}{2}\left(\frac{Y_t}{h_n}\right)^2\right].$$

Table 9: Comparison of the performance of $\hat{p}$, $\bar{p}$, and $\tilde{p}$ for $p < 0.5$

n	Estimator	p_0			
		0.1	**0.2**	**0.3**	**0.4**
99	$\hat{p}$	CIML: 0.1862	CIML: 0.2258	CIML: 0.2355	CIML: 0.2245
		CP: 0.8625	CP: 0.9100	CP: 0.9275	CP: 0.9400
	$\bar{p}$	CIML: 0.1896	CIML: 0.2325	CIML: 0.2410	CIML: 0.2272
		CP: 0.9625	CP: 0.9475	CP: 0.9575	CP: 0.9625
	$\tilde{p}$	CIML: 0.1656	CIML: 0.2399	CIML: 0.2946	CIML: 0.3416
		CP: 0.7875	CP: 0.9000	CP: 0.9325	CP: 0.9500
499	$\hat{p}$	CIML: 0.0828	CIML: 0.1030	CIML: 0.1074	CIML:0.1012
		CP: 0.9050	CP: 0.9150	CP: 0.9450	CP: 0.9300
	$\bar{p}$	CIML: 0.0848	CIML: 0.1040	CIML: 0.1078	CIML:0.1016
		CP: 0.9450	CP: 0.9550	CP: 0.9375	CP: 0.9275
	$\tilde{p}$	CIML: 0.0889	CIML: 0.1264	CIML: 0.1550	CIML: 0.1797
		CP: 0.8825	CP: 0.9325	CP: 0.9250	CP: 0.9675
999	$\hat{p}$	CIML: 0.0591	CIML: 0.0732	CIML: 0.0761	CIML: 0.0717
		CP: 0.9375	CP: 0.9400	CP: 0.9375	CP: 0.9475
	$\bar{p}$	CIML: 0.0600	CIML: 0.0735	CIML: 0.0762	CIML: 0.0719
		CP: 0.9375	CP: 0.9700	CP: 0.9575	CP: 0.9550
	$\tilde{p}$	CIML: 0.0678	CIML: 0.0962	CIML: 0.1179	CIML: 0.1362
		CP: 0.8825	CP: 0.9200	CP: 0.9300	CP: 0.9550
4,999	$\hat{p}$	CIML: 0.0267	CIML: 0.0328	CIML: 0.0341	CIML: 0.0321
		CP: 0.9425	CP: 0.9575	CP: 0.9475	CP: 0.9350
	$\bar{p}$	CIML: 0.0268	CIML: 0.0329	CIML: 0.0341	CIML: 0.0321
		CP: 0.9625	CP: 0.9600	CP: 0.9400	CP: 0.9525
	$\tilde{p}$	CIML:0.0356	CIML: 0.0505	CIML: 0.0619	CIML: 0.0716
		CP: 0.9000	CP: 0.9200	CP: 0.9425	CP: 0.9450
9,999	$\hat{p}$	CIML: 0.0189	CIML: 0.0232	CIML: 0.0241	CIML: 0.0227
		CP: 0.9525	CP: 0.9550	CP: 0.9500	CP: 0.9300
	$\bar{p}$	CIML: 0.0190	CIML: 0.0232	CIML: 0.0241	CIML: 0.0227
		CP: 0.9500	CP: 0.9570	CP: 0.9625	CP: 0.9450
	$\tilde{p}$	CIML: 0.0270	CIML: 0.0383	CIML: 0.0470	CIML: 0.0542
		CP: 0.9025	CP: 0.9000	CP: 0.9500	CP: 0.9500

A fixed value of $a = 0.5$ was used in all the cases, and 95% confidence intervals were constructed. *Est* denotes the estimator and p_0 denotes the true value of p.

Table 10: Comparison of the performance of $\hat{p}$, $\bar{p}$ and $\tilde{p}$ for $p \geq 0.5$

n	Estimator	p_0			
		0.6	0.7	0.8	0.9
99	$\hat{p}$	CIML: 0.2242	CIML: 0.2381	CIML: 0.2286	CIML: 0.1912
		CP: 0.9575	CP: 0.9275	CP: 0.9250	CP: 0.8900
	$\bar{p}$	CIML: 0.2272	CIML: 0.2410	CIML: 0.2326	CIML: 0.1896
		CP: 0.9450	CP: 0.9575	CP: 0.9450	CP: 0.9500
	$\tilde{p}$	CIML: 0.4163	CIML: 0.4529	CIML: 0.4841	CIML: 0.5124
		CP: 0.9450	CP: 0.9525	CP: 0.9575	CP: 0.9700
499	$\hat{p}$	CIML: 0.1013	CIML: 0.1071	CIML: 0.1032	CIML: 0.0832
		CP: 0.9200	CP: 0.9425	CP: 0.9550	CP: 0.9075
	$\bar{p}$	CIML: 0.1016	CIML: 0.1077	CIML: 0.1040	CIML: 0.0848
		CP: 0.9575	CP: 0.9475	CP: 0.9475	CP: 0.9400
	$\tilde{p}$	CIML: 0.2200	CIML: 0.2381	CIML: 0.2543	CIML: 0.2695
		CP: 0.9450	CP: 0.9650	CP: 0.9500	CP: 0.9775
999	$\hat{p}$	CIML: 0.0716	CIML: 0.0759	CIML: 0.0731	CIML: 0.0595
		CP: 0.9275	CP: 0.9175	CP: 0.9400	CP: 0.9225
	$\bar{p}$	CIML: 0.0719	CIML: 0.0762	CIML: 0.0735	CIML: 0.0600
		CP: 0.9475	CP: 0.9350	CP: 0.9625	CP: 0.9175
	$\tilde{p}$	CIML: 0.1664	CIML: 0.1803	CIML: 0.1926	CIML: 0.2042
		CP: 0.9525	CP: 0.9725	CP: 0.9550	CP: 0.9375
4,999	$\hat{p}$	CIML: 0.0321	CIML: 0.0341	CIML: 0.0329	CIML: 0.0268
		CP: 0.9400	CP: 0.9425	CP: 0.9500	CP: 0.9400
	$\bar{p}$	CIML: 0.0321	CIML: 0.0341	CIML: 0.0329	CIML: 0.0268
		CP: 0.9550	CP: 0.9275	CP: 0.9525	CP: 0.9625
	$\tilde{p}$	CIML: 0.0876	CIML: 0.0946	CIML: 0.1012	CIML: 0.1073
		CP: 0.9475	CP: 0.9525	CP: 0.9500	CP: 0.9525
9,999	$\hat{p}$	CIML: 0.0227	CIML: 0.0241	CIML: 0.0233	CIML: 0.0190
		CP: 0.9500	CP: 0.9500	CP: 0.9475	CP: 0.9425
	$\bar{p}$	CIML: 0.0227	CIML: 0.0241	CIML: 0.0233	CIML: 0.0190
		CP: 0.9500	CP: 0.9450	CP: 0.9450	CP: 0.9500
	$\tilde{p}$	CIML: 0.0664	CIML: 0.0717	CIML: 0.0767	CIML: 0.0813
		CP: 0.9700	CP: 0.9725	CP: 0.9700	CP: 0.9650

A fixed value $a = 0.5$ was used in all the cases, and 95% confidence intervals were constructed. *Est* denotes the estimator, and p_0 denotes the true value of p.

A $(1 - \alpha)100\%$ confidence interval for p is

$$\tilde{p}\sqrt{1 + h_n^2} \pm z_{\alpha/2}\left[\frac{\bar{X}_n}{(n + 1)\sqrt{2}\,h_n}\right]^{1/2}.$$

Tables 9 and 10 present a comparison of the three estimators of p across varying true values of p. Samples of size $n + 1$ were utilized, and 95% confidence intervals were constructed. CPs were computed based on 400 replications of each of the chains of size $n + 1$. Subsequently, the means of the lengths of the 400 confidence intervals were calculated to determine the confidence interval mean length. The results in both tables demonstrate that the estimators are asymptotically equivalent and converge to the true value. Note that we did not include the numerical values of the estimators because 400 numerical values were computed for each estimator in each case, and there is no criterion to choose only one representative from the 400. Relatively narrow confidence intervals reflect a higher level of precision in estimation, while wider confidence intervals indicate greater uncertainty in estimation.

In both cases of $p < 1/2$ and $p > 1/2$, $\hat{p}$ yields the most precise estimates with the shortest mean confidence intervals. The sample mean closely follows $\hat{p}$ in precision but exhibits slightly longer mean confidence intervals. Although the robust estimator shows larger mean confidence intervals compared to $\hat{p}$ and $\bar{p}$, it still
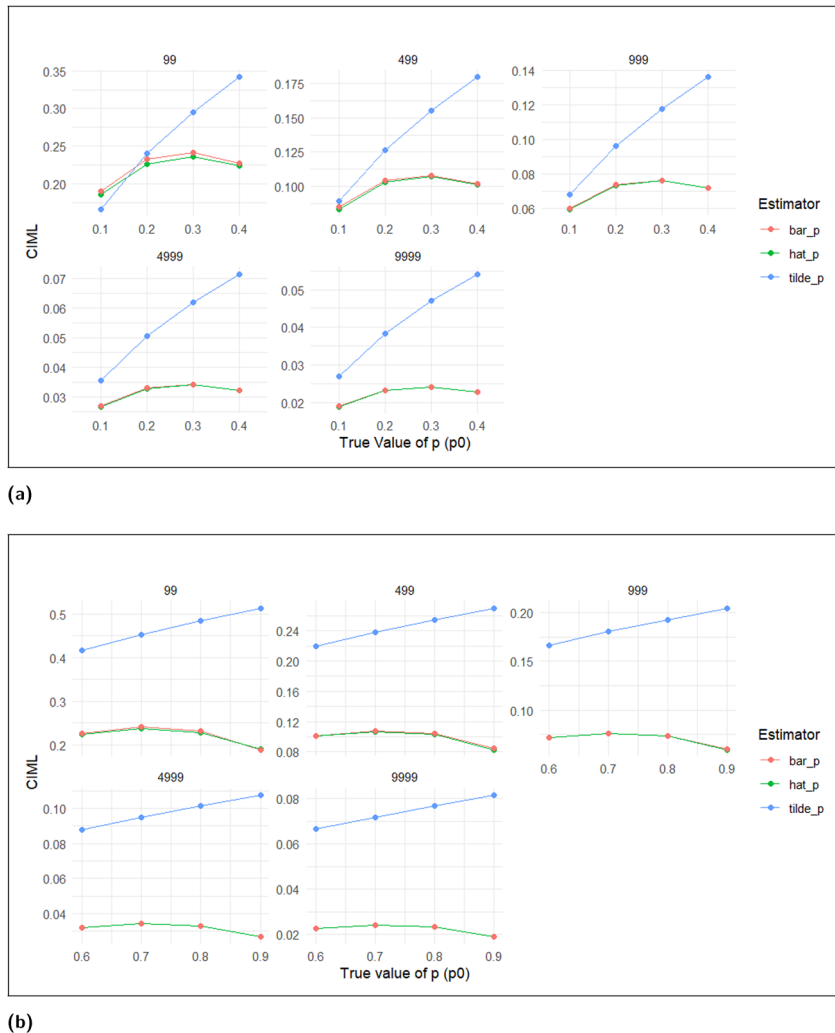
(a)

(b)

Figure 2: CIML for different values of p: (a) CIML for estimators when $p < \frac{1}{2}$ and (b) CIML for estimators when $p > \frac{1}{2}$.

provides reasonable estimates, especially with larger sample sizes. This comparison is presented on the plots in Figure 2(a) and (b).

Both the MLE and the robust estimator demonstrate sensitivity to sample size, resulting in initially low CPs. However, as the sample size increases, these probabilities approach the desired 95% level. The CPs for the sample mean ($\bar{p}$) consistently approach the desired nominal level, indicating reliable confidence intervals across different sample sizes.

The performance of the robust estimator is less optimal with smaller sample sizes but improves as the sample size increases. The robust estimator is designed to be competitive in data containing noise or outliers; however, the data considered here has neither.

While all three estimators provide asymptotically unbiased estimates for the Bernoulli parameter p, the sample mean ($\bar{p}$) generally strikes the best balance between precision and accuracy, making it the preferred choice for estimating p in this context.

6 Conclusion

The mixing properties of copula-based Markov chains are highly dependent on the chosen marginal distributions. Some copulas that do not generate mixing with continuous marginals may generate mixing with discrete marginals, so generalizations should be approached with caution. In this article, we have demonstrated that copulas from the Fréchet (Mardia) family generate ψ-mixing Markov chains for $a + b = 1$, which is not the case when the marginals are uniform$[0, 1]$. Understanding these mixing properties is crucial for deriving central limit theorems for sample means, which in turn facilitates the construction of confidence intervals for the sample means based on the standard normal distribution.

References

[1] Anderson, W., & Goodman, A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, *28*, 89–110.
[2] Bedrick, J., & Aragon, J. (1989). Approximate confidence intervals for the parameters of a stationary binary Markov chain. *Technometrics*, *31*, 437–448.
[3] Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, *32*, 12–40.
[4] Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Institute of Mathematical Statistics-University of Chicago Statistical Research Monographs, Chicago: University of Chicago Press.
[5] Blum, R., Hanson, L., & Koopmans, L. (1963). *On the strong law of large numbers for a class of stochastic processes*. Albuquerque, NM: Sandia Corporation.
[6] Bradley, R. (2007). *Introduction to strong mixing conditions* (vol. 1, 2), Kendrick Press.
[7] Brainerd, B., & Chang, M. (1982). Number of occurrences in two-state Markov chains, with applications in linguistics. *The Canadian Journal of Statistics*, *10*, 225–231.
[8] Cogburn, R. (1960). *Asymptotic properties of stationary sequences* (Vol. 3), Berkeley, CA: University of California Press.
[9] Crow, E. (1979). Approximate confidence intervals for a proportion from Markov dependent trials. *Communications in Statistics, Part B-Simulation and Computation*, *8*, 1–24.
[10] David, F. (1947). A power function for tests of randomness in a sequence of alternatives. *Biometrika*, *34*, 335–339.
[11] Darsow, F., Nguyen, B., & Olsen, E. (1992). Copulas and Markov processes. *Illinois Journal of Mathematics*, *36*(4), 600–642.
[12] Durante, F., & Sempi, C. (2015). *Principles of copula theory*. Boca Raton, FL: CRC Press.
[13] Goodman, A. (1958). Simplified run tests and likelihood ratio tests for Markoff chains. *Biometrika*, *45*(1/2), 181–197.
[14] Goodman, A. (1959). On some statistical tests for mth order Markov chains, *Ann. Math. Statist. 30*, 154–164.
[15] Ibragimov, A. (1959). Some limit theorems for stochastic processes stationary in the strict sense. *Dokl. Akad. Nauk SSSR*, *125*, 711–714.
[16] Ibragimov, A., & Linnik, V. (1975). *Independent and stationary sequences of random processes*, Wolters: Nordhoff Publications.
[17] Joe, H. (1997). *Multivariate models and multivariate dependence concepts.* Boca Raton, FL: CRC Press.
[18] Johnson, C., & Klotz, J. (1974). The atom probe and Markov chain statistics of clustering. *Technometrics*, *16*, 483–493.
[19] Klotz, J. (1972). Markov chain clustering of births by sex. *Proc. Sixth Berkeley Symp. Math. Statist. Prob* (vol. 4, pp. 173–185), Berkeley, CA: University of California Press.
[20] Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence. *Annals of Statistics*, *1*, 373–379.
[21] Lindqvist, B. (1976). A note on Bernoulli trials with dependence, *Scandinavian Journal of Statistics*, *5*, 205–208.
[22] Longla, M., & Peligrad, M. (2012). Some aspects of modeling dependence in copula based Markov chains. *Journal of Multivariate Analysis*, *111*, 234–240.
[23] Longla, M. (2014). On Dependence Structure of Copula-based Markov Chains. *ESAIM: Probability and Statistics*, *18*, 570–583, DOI: 10.1051/ps/2013052.

[24] Longla, M. (2015). On mixtures of copulas and mixing coefficients. *Journal of Multivariate Analysis*, *139*, 259–265.

[25] Longla, M., & Peligrad, M. (2021). New robust confidence intervals for mean under dependence, *Journal of Statistics Planning and Inference*, *211*, 90–106.

[26] Muia, M. (2024). *Dependence and Mixing for Perturbations of Copula-based Markov Chains*, Ph.D. thesis, ProQuest, Ann Arbor, MI: University of Mississippi.

[27] Nelsen, R. (2006). *Introduction to Copulas*. *Springer Series in Statistics* (2nd ed.), New York: Springer-Verlag.

[28] Philipp, W. (1969). The central limit problem for mixing sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *12*(2), 155–171.

[29] Price, B. (1976). A note on estimation in Bernoulli trials with dependence. *Communication in Statistics, Part A-Theory and Methods*, *5*, 661–671.

[30] Sklar, A. (1959). *Fonctions de repartition an dimensions et leurs marges* (vol. 8, pp. 119–231), Paris: Publ. inst. statist. University.

[31] Wilks, S. (1938). The large sample distribution of the likelihood ratio for testing com posite hypothese. *Ann. Math. Stat. 9*, 60–62.