Statistics and Applications {ISSN 2454-7395 (online)}
Volume 21, No. 1, 2023 (New Series), pp 187–205

Asymptotic Results for Generalized Runs in Higher Order Markov Chains

Anuradha
Department of Statistics, Lady Shri Ram College for Women
University of Delhi, Lajpat Nagar-IV, New Delhi – 110024, India

Received: 06 February 2022; Revised: 29 May 2022; Accepted: 08 July 2022

Abstract

Consider an m^{th} order Markov chain $\{X_j : j \geq -m+1\}$ taking values in $\{\mathbf{0}, \mathbf{1}\}$. Fix $k \geq 1$ and $r \geq 0$. A r-look-back run of length k, is defined as a run of $\mathbf{1}$'s, provided that there are at least r trials in between the ending point of the current run and the ending point of the previous occurrence of the r-look-back run of length k. The r-look-back run of length k encompasses the non-overlapping counting, the overlapping counting as well as the l-overlapping counting for $0 \leq l \leq k-1$ (defined by Aki and Hirano (2000)). We show that the waiting time for the n^{th} occurrence of the r-look-back run of length k converges in distribution to an extended Poisson distribution under the assumption that the model exhibits a strong propensity towards success. This generalizes similar results on l-overlapping runs of length k obtained under the Markov dependent set-up. We obtain a central limit theorem for the number of r-look-back runs of length k till the n^{th} trial. Further, we show that the rate of convergence in the central limit theorem is at least a fractional power of n with a logarithmic correction factor. We support our findings on the rate of convergence with some simulation results.

Key words: Success runs; Waiting time; Markov chain; Extended Poisson distribution; Central limit theorem; Rate of convergence.

AMS Subject Classifications: 60C05, 60E05, 60F05

1. Introduction

Let $\{X_i : i \geq 1\}$ be a sequence of $\{\mathbf{0}, \mathbf{1}\}$-valued random variables. Here X_n stands for the outcome of an experiment at the n-th trial and $\mathbf{1}$ and $\mathbf{0}$ imply success and failure respectively of the experiment. A *run of length* k is an occurrence of k (≥ 1) consecutive $\mathbf{1}$'s. In the literature, there are several schemes of counting runs of length k; two of the most commonly used ones are (a) *Non-overlapping counting* and (b) *Overlapping counting*. In the non-overlapping counting scheme a trial can contribute to only one possible run, while in the overlapping counting scheme a trial can contribute towards the counting of more than one run. Another method of l-overlapping counting has been introduced by Aki and Hirano (2000) where they allow an overlap of at most l successes between two consecutive runs of length k where $0 \leq l < k$. It is easy to observe that when $l = 0$ and $l = k-1$, this definition is equivalent to the non-overlapping counting and the overlapping counting respectively. Han

Corresponding Author: Anuradha
Email: sarkar.anuradha@gmail.com

and Aki (2000) have extended this counting scheme to the case where l assumes negative values. For $l < 0$, there is at least $|l|$ trials difference between the two runs of length k.

In this paper, we consider a new scheme of counting runs of length k. We will refer to this new scheme of counting as *look-back counting*. Let $r \geq 0$ be a fixed number. In the r-look-back counting scheme, the starting points (hence the ending points) of the two consecutive runs of length k should be separated by at least r trials in between, *i.e.*, a new run of length k can be counted only after r trials have elapsed since the starting point of the last counted run. Suppose that we are at the trial i such that it is a starting point of a run of length k, *i.e.*, $X_i = X_{i+1} = \ldots = X_{i+k-1} = 1$. Now, suppose that i' is the trial where the last enumerated r-look-back run of length k started. In order to enumerate the run starting at the i^{th} trial as a r-look-back run, we must have $i - i' > r$. The definition of r-look-back run of length k, encompasses the above definitions of overlapping runs and non-overlapping runs, in the sense that when $r = 0$ and $r = k - 1$, the r-look-back run of length k matches exactly with the overlapping counting and the non-overlapping counting respectively. Moreover, the l-overlapping run of length k can be identified as a r-look-back run of length k with $l = k - r - 1$ for $0 \leq l \leq k - 1$. However, when l assumes negative values the definitions do not match. To illustrate this, we quote the example from Han and Aki (2000):

$$\mathbf{1111011000111111110000111}.$$

In this example, for $k = 3$ and $r = 3$, we see that there are four 3-look-back runs of length 3 starting at trials $1, 11, 15$ and 23, while there are only three (-1)-overlapping runs of length 3, starting at $1, 11$ and 15. This is because the number of remaining trials (0 here) after the last run of length k starting at trial 23, is less than $|l|$, in the l-overlapping counting of runs of length k (for $l < 0$), such a run cannot be counted (see Han and Aki (2000)). But, in the look-back counting scheme, we do not put such a restriction. This will be clear from the mathematical definition given in the next section.

Practical usage of this scheme of counting can be illustrated from the following examples. In many counters for detection of cosmic rays and α-particles, the counter records a hit (detection of a particle) whenever the frequency recorded lies in a particular region (depending on the particle under detection). We refer to the detection of the particle as a success, while the non-detection is regarded as a failure. However, if particles are detected for $k(\geq 1)$ successive time points, the counter loses its power and is locked; hence it cannot record anything in the next $r - k + 1$ $(r > k - 1)$ time points. The number of r-look-back runs of length k is exactly the number of times the instrument loses its power and is locked. Another example is seen in the congestion model of computer networks, where a network receives packets of information from other networks and sends information back to the originating network. Each of these processes consumes certain processing resource. If the network receives packets at k consecutive time points, all its resources are spent in processing the information received; as a result it can not receive any information for the next $r - k + 1$ $(r > k - 1)$ time points. In one of the models of computer networking, the packets are rejected for these time points and are required to be re-sent by the originating network at a later time point. This situation is called the congestion of the network. Here also, the number of r-look-back runs of length k is exactly the number of times congestion occurs. In a drug administration model, observations are taken every hour for the presence or absence (success or failure) of a particular symptom, say, fever exceeding a specified temperature.

If we observe the presence of the symptom for k consecutive points (hours), a drug has to be administered; however, as is the case with most drugs, once the drug is administered, we have to wait for r-hours for the next administration of the drug with $r < k$. But the process of the observation for the presence or absence of the symptom is continued as ever. In such a case, the number of administrations of the drug until time point n, is the number of r-look-back runs of length k up to time n. In the first two examples, we have $r > k-1$ while in the last example $0 \leq r \leq k-1$.

The theory of runs plays a vital role in diverse fields of statistics, such as, non-parametric inference, statistical quality control, reliability theory *etc.*. Runs and run-related statistics have engaged researchers since the time of De Moivre (see Feller (1968)). In recent years, this field has seen tremendous growth, with researchers contributing to the theory as well as their practical applications to various disciplines. Systematic study of the theory of distributions of non-overlapping runs was initiated by Feller (1968). Feller studied the distribution of the number of non-overlapping runs up to the *n-th* trial and obtained its asymptotic distribution using the renewal theory techniques where the underlying trials were i.i.d. Bernoulli random variables. Aki (1985), Hirano (1986), Philippou and Makri (1986) *etc.* studied various run-statistics based on the non-overlapping counting of runs. Ling (1988) obtained the distribution of the number of overlapping runs of length k for a sequence of n i.i.d. Bernoulli trials. Aki and Hirano (1988), Godbole (1990, 1992) also studied the properties of the distribution of the number of overlapping runs up to time n. Hirano *et al.* (1991) obtained the probability generating function of the number of overlapping runs and also obtained the asymptotic distribution. Several generalization of the underlying model has also been considered (see Aki and Hirano (1995), Fu and Koutras (1994), Koutras (1996), Uchida and Aki (1995), Uchida (1998) and references therein). The waiting time distributions for the occurrence of runs of various types has been studied extensively by several authors (see, for example, Koutras (1996), Aki, Balakrishnan and Mohanty (1996), Balasubramanian, Viveros and Balakrishnan (1993) and references therein). Uchida (1998) has also investigated the waiting time problems for patterns under m^{th} order Markov set up. Makri and Psillakis (2015) also studied l-overlapping runs of successes of length k and obtained recurrence relations for probability mass functions for the case of Bernoulli trials ordered in line as well as in circle. For a more detailed account on the theory of runs and its applications, we refer the reader to Balakrishnan and Koutras (2002) and Makri and Psillakis (2015).

In this paper, we assume that the underlying trials form a m^{th} order homogeneous Markov chain. We study the waiting time of the n^{th} occurrence of the r-look-back run of length k. We show that under the assumption of the model exhibiting a strong propensity towards success, *i.e.*, the probability of getting success converges to 1 in a certain sense, the waiting time for the n^{th} occurrence of r-look-back run of length k converges to a compound Poisson distribution as $n \to \infty$. This result generalises the results of Inoue and Aki (2003) where they considered that the underlying trials are from a homogeneous Markov chain ($m = 1$). Also, we show that the number of r-look-back runs of length k till the n^{th} trial, suitably normalised, converges to a normal distribution when the underlying process is an m^{th} order homogeneous Markov chain. Further, we obtain the (uniform) rate of convergence of the central limit theorem (Berry Essen type result). This result shows that the convergence rate is at least a fraction power of n with a logarithmic correction factor.

In the next section we give the formal definitions and the statement of results. The third section is devoted to showing the convergence of waiting times, while in the fourth section we obtain the rate of convergence results. In the final section, we demonstrate simulation results where the underlying trials are from a homogeneous Markov chain ($m = 1$) exhibiting the rate of convergence.

2. Definitions and statement of theorem

Let $X_{-m+1}, X_{-m+2}, \dots, X_0, X_1, \dots$ be a sequence of stationary m^{th} order $\{\mathbf{0},\mathbf{1}\}$-valued Markov chain. It is assumed that the values of $X_{-m+1}, X_{-m+2}, \dots, X_0$ are known, *i.e.*, we are given the initial condition $\{X_0 = x_0, X_{-1} = x_1, \dots, X_{-m+1} = x_{m-1}\}$, $x_i \in \{\mathbf{0},\mathbf{1}\}, i = -m+1, \dots, -1, 0$.

Define $N_l := \{0, 1, \dots, 2^l - 1\}$ for any $l \geq 0$. It is clear that $\{\mathbf{0},\mathbf{1}\}^l$ and N_l can be identified by the mapping $(x_0, x_1, \dots, x_{l-1}) \longrightarrow \sum_{j=0}^{l-1} 2^j x_j$. Thus, we will represent the initial condition by taking $x \in N_m = \{0, 1, \dots, 2^m - 1\}$ where $x = \sum_{j=0}^{m-1} 2^j x_j$.

We define, for any $n \geq 0$,

$$p_x := \mathbb{P}(X_{n+1} = 1 | X_n = x_0, X_{n-1} = x_1, \dots, X_{n-m+1} = x_{m-1}). \tag{1}$$

Consequently, $q_x := \mathbb{P}(X_{n+1} = 0 | X_n = x_0, X_{n-1} = x_1, \dots, X_{n-m+1} = x_{m-1}) = 1 - p_x$. We assume that $0 < p_x < 1$ for all $x \in N_m$. Define two functions $f_l, g_l : N_l \to N_l$ as

$$f_l(x) := 2x + 1 \pmod{2^l} \text{ and } g_l(x) := 2x \pmod{2^l}.$$

Note that, $f_m(x), g_m(x)$ can be interpreted as the two possible states which can be reached from the state x in a single step, provided we obtain a success, failure respectively in the next trial.

Let $R_i(k, r)$ be the indicator of the event that a r-look-back run of length k is completed at the i^{th} trial. In order that $R_i(k, r) = 1$, we must have $R_{i-1}(k, r) = R_{i-2}(k, r) = \dots = R_{i-r}(k, r) = 0$. Thus, the formal definition of $R_i(k, r)$ can be given inductively as follows:

Definition 1: Set $R_i(k, r) = 0$ for $i \leq k - 1$ and for any $i \geq k$, define

$$R_i(k, r) := \prod_{j=1}^{r} (1 - R_{i-j}(k, r)) \prod_{j=i-k+1}^{i} X_j. \tag{2}$$

When $r = 0$, the first product should be interpreted as 1. If $R_i(k, r) = 1$, we say that a *r-look-back run of length k* has been recorded at the i^{th} trial (*i.e.*, ending at trial i). Define

$$N_{n,k,r} := \sum_{i=1}^{n} R_i(k, r) = \sum_{i=k}^{n} R_i(k, r)$$

as the number of r-look-back runs of length k till the n^{th} trial. A sequence of stopping times is defined as follows: set $\tau_0(k, r) = 0$ and for $n \geq 1$,

$$\tau_n(k, r) := \inf\{i > \tau_{n-1}(k, r) : N_{i,k,r} = n\}. \tag{3}$$

Note, $\tau_n(k,r)$ is the waiting time for the n^{th} occurrence of r-look-back run of length k.

In the sequel, we say that a random variable ξ is of Poisson type with multiplicity p and parameter α, denoted by $\xi \sim \mathrm{Poi}(p,\alpha)$, if

$$\mathbb{P}(\xi = pt) = \frac{\exp(-\alpha)\alpha^t}{t!} \qquad \text{for } t = 0, 1, \ldots . \tag{4}$$

Note that, when $p = 1$ it is the usual Poisson distribution. Following Aki (1985), we say that a random variable ξ follows an extended Poisson distribution of order k with parameters $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, if its p. g. f. is given by,

$$\phi(z; \alpha_1, \alpha_2, \ldots, \alpha_k) = \exp\Big(-\sum_{j=1}^{k} \alpha_j + \sum_{j=1}^{k} \alpha_j z^j\Big).$$

It should be noted that if ξ follows an extended Poisson distribution of order k with parameters $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, then it can be represented as

$$\xi \stackrel{d}{=} \sum_{j=1}^{k} \xi_j$$

where $\{\xi_j : 1 \le j \le k\}$ are independent and $\xi_j \sim \mathrm{Poi}(j, \alpha_j)$.

The assumption, we impose on our model, is that the system has a strong tendency towards success. We formalize this by stating that, for $x \in N_m$, p_x (as a function of n) converges to 1 in such a way that

$$n(1 - p_x) \to \lambda_x \text{ as } n \to \infty \text{ where } \lambda_x > 0 \text{ is a positive constant.} \tag{5}$$

Our first theorem is:

Theorem 1: For any initial condition $x \in N_m$, if the condition (5) holds, we have

$$\begin{aligned} &(i)\ \ \tau_n(k,r) - (k-r-1) - n(r+1) \ \Rightarrow\ \sum_{i=0}^{r} \xi_i^{(1)} && \text{when } k \ge r+1 \\ &(ii)\ \ \tau_n(k,r) - k - (n-1)(r+1) \ \Rightarrow\ \sum_{i=0}^{k-1} \xi_i^{(2)} && \text{when } k < r+1 \end{aligned}$$

where $\left\{\xi_i^{(1)} : i = 0, 1, \ldots, r\right\}$ are independent random variables with $\xi_i^{(1)} \sim \mathrm{Poi}(k + i - r, \lambda_{2^m - 1})$ for $i = 0, 1, \ldots, r$ while $\left\{\xi_i^{(2)} : i = 0, 1, \ldots, k-1\right\}$ are independent random variables with $\xi_i^{(2)} \sim \mathrm{Poi}(i + 1, \lambda_{2^m-1})$ for $i = 0, 1, \ldots, k-1$.

From the above, when $0 \le r \le k-1$, the limiting distribution, $\sum_{i=0}^{r} \xi_i^{(1)}$, follows an extended Poisson distribution of order k with parameters $(\overbrace{0, 0, \ldots, 0}^{k-r-1}, \overbrace{\lambda_{2^m-1}, \lambda_{2^m-1}, \ldots, \lambda_{2^m-1}}^{r+1})$ and when $r \ge k$, $\sum_{i=0}^{k-1} \xi_i^{(2)}$ follows an extended Poisson distribution of order k with parameters $(\overbrace{\lambda_{2^m-1}, \lambda_{2^m-1}, \ldots, \lambda_{2^m-1}}^{k})$. Inoue and Aki (2003) have obtained a similar result for

l-overlapping counting under the Markov chain ($m = 1$) set-up. It should be noted that, Inoue and Aki (2003) have counted the run of length k as a l-overlapping run (for $l < 0$) even when the remaining number of trials after the run is completed, is less than $|l|$. Therefore, it matches with our counting scheme with $r = k - 1 + (-l)$ and hence their results can be deduced as a special case from our result. However, even if we follow the definition of Han and Aki (2000), a similar result can be established following our method.

Further, we establish a central limit theorem for $N_{n,k,r}$ and study the rate of convergence in the central limit theorem under the m^{th} order Markov chain set-up. Let $\sigma_n^2 = \mathrm{Var}(N_{n,k,r})$. We show that

Theorem 2: For any $r \geq 0$ and $k \geq 1$, we have

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\Big(N_{n,k,r} - \mathbb{E}(N_{n,k,r}) \leq t\sigma_n\Big) - \Phi(t)\right| = O(n^{-2/11}\log n)$$

where $O(f(n))$ is a function $g(n)$ such that $|g(n)/f(n)|$ remains bounded as $n \to \infty$ and $\Phi(\cdot)$ is the distribution function of the standard normal distribution.

Since $n^{-2/11}\log n \to 0$ as $n \to \infty$, we obtain the standard central limit theorem from Theorem 2. Further, this result gives the uniform rate at which the normalised variable $N_{n,k,r}$ converges to normality. Since the number of l-overlapping runs of length k up to the n^{th} trial is at most one less than $N_{n,k,r}$, where $r = k - l - 1$, exactly same results will hold for the number of l-overlapping runs of length k up to the n^{th} trial.

3. Convergence of waiting time

In this section, we prove Theorem 1. We require the following lemmas on weak convergence of discrete random variables. The first lemma is an easy consequence of the Portmanteau Theorem (Billingsley (1968) p.p. 11) and the fact that all the random variables involved are discrete in nature; hence we omit its proof.

Lemma 1: If $\{\xi_n : n \geq 1\}$ and ξ are random variables taking values in $\mathbb{N} = \{0, 1, \dots\}$ such that for each $t \geq 0$

$$\liminf_{n\to\infty} \mathbb{P}(\xi_n = t) \geq \mathbb{P}(\xi = t),$$

then $\xi_n \Rightarrow \xi$.

Lemma 2: Suppose that $\{\xi_n : n \geq 0\}$ and $\{\xi_i^{(1)} : 1 \leq i \leq p_1\}$ and $\{\xi_i^{(2)} : 1 \leq i \leq p_2\}$ are random variables taking values in $\mathbb{N}$ and $\{\xi_1^{(1)}, \xi_2^{(1)}, \dots, \xi_{p_1}^{(1)}, \xi_1^{(2)}, \xi_2^{(2)}, \dots, \xi_{p_2}^{(2)}\}$ are independent. Suppose that, for each $n \geq 1$ and $t \geq 0$, $\Big\{A_n^t(u_1, u_2, \dots, u_{p_1}, v_1, v_2, \dots, v_{p_2}) : u_i \geq 0 \text{ for } 1 \leq i \leq p_1, v_i \geq 0 \text{ for } 1 \leq i \leq p_2 \text{ and } \sum_{i=1}^{p_1} u_i = t\Big\}$ is a collection of disjoint events, such that

$$\liminf_{n\to\infty} \mathbb{P}\Big(\xi_n = t, A_n^t(u_1, u_2, \dots, u_{p_1}, v_1, v_2, \dots, v_{p_2})\Big) \geq \prod_{i=1}^{p_1} \mathbb{P}(\xi_i^{(1)} = u_i) \prod_{i=1}^{p_2} \mathbb{P}(\xi_i^{(2)} = v_i).$$

Then

$$\xi_n \Rightarrow \sum_{i=1}^{p_1} \xi_i^{(1)}.$$

Note that, we require $p_1 \geq 1$ but $p_2 \geq 0$. In one of our applications, we will take $p_2 = 0$.

Proof: Clearly, for any fixed $t \in \mathbb{N}$,

$$\mathbb{P}\Big(\sum_{i=1}^{p_1} \xi_i^{(1)} = t\Big) = \sum_{\substack{u_1,\ldots,u_{p_1} \in \mathbb{N} \\ v_1,\ldots,v_{p_2} \in \mathbb{N} \\ \sum_{i=1}^{p_1} u_i = t}} \prod_{i=1}^{p_1} \mathbb{P}(\xi_i^{(1)} = u_i) \prod_{i=1}^{p_2} \mathbb{P}(\xi_i^{(2)} = v_i).$$

Fix any $\epsilon > 0$ and choose J so large that

$$\sum_{\substack{u_1,\ldots,u_{p_1} \in \mathbb{N} \\ v_1,\ldots,v_{p_2} \in \mathbb{N} \\ \sum_{i=1}^{p_1} u_i = t}} \prod_{i=1}^{p_1} \mathbb{P}(\xi_i^{(1)} = u_i) \prod_{i=1}^{p_2} \mathbb{P}(\xi_i^{(2)} = v_i) - \sum_{\substack{0 \leq u_1,\ldots,u_{p_1} \leq J \\ 0 \leq v_1,\ldots,v_{p_2} \leq J \\ \sum_{i=1}^{p_1} u_i = t}} \prod_{i=1}^{p_1} \mathbb{P}(\xi_i^{(1)} = u_i) \prod_{i=1}^{p_2} \mathbb{P}(\xi_i^{(2)} = v_i) < \epsilon.$$

Thus, we have,

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}(\xi_n = t) \\
&\quad\geq \liminf_{n\to\infty} \sum_{\substack{u_1,\ldots,u_{p_1} \in \mathbb{N} \\ v_1,\ldots,v_{p_2} \in \mathbb{N} \\ \sum_{i=1}^{p_1} u_i = t}} \mathbb{P}(\xi_n = t, A_n^t(u_1, u_2, \ldots, u_{p_1}, v_1, v_2, \ldots, v_{p_2})) \\
&\quad\geq \liminf_{n\to\infty} \sum_{\substack{0 \leq u_1,\ldots,u_{p_1} \leq J \\ 0 \leq v_1,\ldots,v_{p_2} \leq J \\ \sum_{i=1}^{p_1} u_i = t}} \mathbb{P}(\xi_n = t, A_n^t(u_1, u_2, \ldots, u_{p_1}, v_1, v_2, \ldots, v_{p_2})) \\
&\quad\geq \sum_{\substack{0 \leq u_1,\ldots,u_{p_1} \leq J \\ 0 \leq v_1,\ldots,v_{p_2} \leq J \\ \sum_{i=1}^{p_1} u_i = t}} \liminf_{n\to\infty} \mathbb{P}(\xi_n = t, A_n^t(u_1, u_2, \ldots, u_{p_1}, v_1, v_2, \ldots, v_{p_2})) \\
&\quad\geq \sum_{\substack{0 \leq u_1,\ldots,u_{p_1} \leq J \\ 0 \leq v_1,\ldots,v_{p_2} \leq J \\ \sum_{i=1}^{p_1} u_i = t}} \prod_{i=1}^{p_1} \mathbb{P}(\xi_i^{(1)} = u_i) \prod_{i=1}^{p_2} \mathbb{P}(\xi_i^{(2)} = v_i) \\
&\quad\geq \mathbb{P}\Big(\sum_{i=1}^{p_1} \xi_i^{(1)} = t\Big) - \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, by Lemma 1, the result follows. □

In the next lemma, we derive a lower bound of a particular event, defined below. This lower bound will be used in proving Theorem 1.

Definition 2: For any $K \geq m$, define $A_\gamma^\alpha(K)$ as the collection of all strings consisting of **0**'s and **1**'s, of length γ and having exactly α **0**'s, such that

(a) the number of **1**'s before the first occurrence of **0** is at least K,

(b) the number of **1**'s after the last occurrence of **0** is at least K,

(c) the number of **1**'s between any two occurrences of **0**'s is at least K.

For the initial condition $x \in N_m$, the probability of observing any given string $s \in A^{\alpha}_{\gamma}(K)$ is given by:

$$\delta_m(\gamma,\alpha,x) := h_m(x)\Big(h_m(2^m-2)\Big)^{\alpha}\Big(1-p_{2^m-1}\Big)^{\alpha}\Big(p_{2^m-1}\Big)^{\gamma-m-\alpha(m+1)} \tag{6}$$

where $h_m(x) := \prod_{j=0}^{m-1} p_{f^j_m(x)}$ where $f^0_m(x) := x$ and $f^{j+1}_m(x) := f_m(f^j_m(x))$ for $x \in N_m$.

For a string $s \in A^{\alpha}_{\gamma}(K)$, let β_0 be the number of **1**'s before the first occurrence of **0**, β_α be the number of **1**'s after the last occurrence of **0** and β_i be the number of **1**'s between the i^{th} and $(i+1)^{th}$ occurrences of **0** for $i = 1,2,\ldots,\alpha-1$.

- Case $r \le k-1$: Define

  $$\beta'_i := (\beta_i - (k-r-1)) \mod (r+1) \text{ for } i = 0,1,\ldots,\alpha.$$

  Clearly $0 \le \beta'_i \le r$. Set $S^{(1)}_j(s) := \#\Big\{i : \beta'_i = j\Big\}$ for $j = 0,1,\ldots,r$, then $\sum_{j=0}^{r} S^{(1)}_j(s) = \alpha$. Define the event

  $$B_\gamma(K,u_0,u_1,\ldots,u_r) := \Big\{s \in A^{\alpha}_{\gamma}(K) : S^{(1)}_j(s) = u_j \text{ for } j = 0,1,\ldots,r\Big\}.$$

  So, we must have $\alpha = \sum_{j=0}^{r} u_j$.

- Case $r \ge k$: Define

  $$\beta'_0 := \beta_0 \mod (r+1).$$

  Having specified $\beta'_0, \beta'_1, \ldots, \beta'_i$, define

  $$\beta'_{i+1} := \begin{cases} \beta_{i+1} \mod (r+1) & \text{if } \beta'_i < k \\ \Big(\beta_{i+1} - (r-\beta'_i)\Big) \mod (r+1) & \text{if } \beta'_i \ge k. \end{cases}$$

  Set $S^{(2)}_j(s) := \#\Big\{i : \beta'_i = j\Big\}$ for $j = 0,1,\ldots,r$. Here also, $\sum_{i=0}^{r} S^{(2)}_j(s) = \alpha$. Define the event

  $$\begin{aligned} C_\gamma(K,u_0,\ldots,u_{k-1},v_0,\ldots,v_{r-k}) := \Big\{s \in A^{\alpha}_{\gamma}(K) : S^{(2)}_j(s) = u_j &\text{ for } j = 0,1,\ldots,k-1 \\ \text{and } S^{(2)}_j(s) = v_{j-k} &\text{ for } j = k,k+1,\ldots,r\Big\}. \end{aligned}$$

  Here, we must have $\alpha = \sum_{j=0}^{k-1} u_j + \sum_{j=0}^{r-k} v_j$.

The following lemma gives a lower bound of the probability of the events defined above. When $r \le k-1$, we choose $K = K(r) = (k-r-1) + t_0(r+1)$ where $t_0 \ge 1$ is so large that $K \ge m$. When $r \ge k$, we choose $K = K(r) = t_0(r+1)$ so that $t_0(r+1) \ge m$.

Lemma 3: (a) For $r \le k-1$ and given non-negative integers $u_0, u_1, \ldots, u_r$ and n such that $n \ge t_0(1+\sum_{i=0}^r u_i)$, define $\gamma(n) = (k-r-1) + n(r+1) + \sum_{i=0}^r u_i(k+i-r)$. Then for any $x \in N_m$, we have

$$\mathbb{P}_x\Big(B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)\Big) \ge \delta_m\Big(\gamma(n), \sum_{i=0}^r u_i, x\Big) \frac{\Big(n - t_0(1+\sum_{i=0}^r u_i) + \sum_{i=0}^r u_i\Big)!}{\prod_{i=0}^r u_i! \Big(n - t_0(1+\sum_{i=0}^r u_i)\Big)!}. \tag{7}$$

(b) For $r > k-1$ and given non-negative integers $u_0, u_1, \ldots, u_{k-1}, v_0, v_1, \ldots, v_{r-k}$ and n such that $n \ge t_0 + t_0\sum_{i=0}^{k-1} u_i + (1+t_0)\sum_{i=0}^{r-k} v_i + 1$, define $\gamma(n) = k + (n-1)(r+1) + \sum_{i=0}^{k-1} u_i(i+1)$. Then for any $x \in N_m$, we have

$$\begin{aligned}\mathbb{P}_x\Big(C_{\gamma(n)}(K, u_0 \ldots, u_{k-1}, v_0, \ldots, v_{r-k})\Big) &\ge \delta_m\Big(\gamma(n), \sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i, x\Big) \\ &\times \frac{\Big(n-1-(t_0 + t_0\sum_{i=0}^{k-1} u_i + (1+t_0)\sum_{i=0}^{r-k} v_i) + \sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i\Big)!}{\prod_{i=0}^{k-1} u_i! \prod_{i=0}^{r-k} v_i! \Big(n-1-(t_0 + t_0\sum_{i=0}^{k-1} u_i + (t_0+1)\sum_{i=0}^{r-k} v_i)\Big)!}.\end{aligned} \tag{8}$$

Proof: As we have noted, the probability of any string is independent of the positions of the $\mathbf{0}$'s and is given by (6). So, by multiplying this with the number of strings in $B_{\gamma(n)}(K, u_0, \ldots, u_r)$ and $C_{\gamma(n)}(K, u_0, \ldots, u_{k-1}, v_0, \ldots, v_{k-r})$ respectively, we get the probability of the respective events. Now we describe a method for obtaining a lower bound of the number of strings in the respective events.

(a) We define $r+2$ objects as follows: $O_0 = \mathbf{0}\underbrace{\mathbf{11\cdots1}}_{K}, O_1 = \mathbf{10}\underbrace{\mathbf{11\cdots1}}_{K}, \ldots, O_r = \underbrace{\mathbf{1\cdots1}}_{r}\mathbf{0}\underbrace{\mathbf{11\cdots1}}_{K}$ and $O_{r+1} = \underbrace{\mathbf{1\cdots1}}_{r+1}$.

First, we put K $\mathbf{1}$'s in the beginning of the string. Next we distribute u_i objects of type O_i for $i = 0, 1, \ldots, r$ and $(n - t_0(1+\sum_{i=0}^r u_i))$ objects of type O_{r+1} in any way we like. It is evident, from the construction of the objects, that any arrangement given above will result in a string in $B_{\gamma(n)}(K, u_0, \ldots, u_r)$. Thus the number of arrangements of the above objects will provide a lower bound of the number of strings in $B_{\gamma(n)}(K, u_0, \ldots, u_r)$. The number of arrangements is given by $(n - t_0(1+\sum_{i=0}^r u_i) + \sum_{i=0}^r u_i)!/\Big(\prod_{i=0}^r u_i!(n - t_0(1+\sum_{i=0}^r u_i))!\Big)$. This proves part (a).

(b) Here we define the objects as follows: $O_0 = \mathbf{0}\underbrace{\mathbf{11\cdots1}}_{K}$, $O_1 = \mathbf{10}\underbrace{\mathbf{11\cdots1}}_{K}$, $\ldots, O_{k-1} = \underbrace{\mathbf{1\cdots1}}_{k-1}\mathbf{0}\underbrace{\mathbf{11\cdots1}}_{K}, O_k = \underbrace{\mathbf{1\cdots1}}_{k}\mathbf{0}\underbrace{\mathbf{1\cdots1}}_{r-k}\underbrace{\mathbf{11\cdots1}}_{K}, O_{k+1} = \underbrace{\mathbf{1\cdots1}}_{k}\mathbf{10}\underbrace{\mathbf{1\cdots1}}_{r-1-k}\underbrace{\mathbf{11\cdots1}}_{K}, \ldots, O_r = \underbrace{\mathbf{1\cdots1}}_{k}\underbrace{\mathbf{1\cdots1}}_{r-k}\mathbf{0}\underbrace{\mathbf{11\cdots1}}_{K}$ and $O_{r+1} = \underbrace{\mathbf{1\cdots1}}_{r+1}$.

First, we put K $\mathbf{1}$'s in the beginning of the string. Now, we distribute u_i objects of type O_i for $i = 0, 1, \ldots, k-1$ and v_i objects of type O_{k+i} for $i = 0, 1, \ldots, r-k$ and $(n - 1 - (t_0 + t_0\sum_{i=0}^{k-1} u_i + (1+t_0)\sum_{i=0}^{r-k} v_i))$ objects of type O_{r+1}. Again, it is evident from the construction of the objects that any arrangement of the objects will result in a string in $C_{\gamma(n)}(K, u_0, \ldots, u_{k-1}, v_0, \ldots, v_{r-k})$. Thus a lower bound of the number of strings in $C_{\gamma(n)}(K, u_0, \ldots, u_{k-1}, v_0, \ldots, v_{r-k})$ is obtained by counting the number of such possible arrangements, which is given by $(n-1-(t_0+t_0\sum_{i=0}^{k-1}u_i+(1+t_0)\sum_{i=0}^{r-k}v_i)+\sum_{i=0}^{k-1}u_i+\sum_{i=0}^{r-k}v_i)!/(\prod_{i=0}^{k-1}u_i!\prod_{i=0}^{r-k}v_i!(n-1-(t_0+t_0\sum_{i=0}^{k-1}u_i+(t_0+1)\sum_{i=0}^{r-k}v_i))!)$. □

Now we are in a position to prove the Theorem 1.

Proof: (a) Fix any $t \geq 0$. We consider the collection of events $\Big\{B_{\gamma(n)}(K, u_0,\ u_1, \ldots, u_r) : u_i \geq 0, t = \sum_{i=0}^{r} u_i(k+i-r)\Big\}$. Clearly, these events are disjoint. By Lemma 2, it is enough to show that for any $x \in N_m$,

$$\liminf_{n\to\infty} \mathbb{P}_x\Big(\tau_n(k,r)-(k-r-1)-n(r+1)=t, B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)\Big) \geq \prod_{i=0}^{r} \frac{\exp(-\lambda_{2^m-1})(\lambda_{2^m-1})^{u_i}}{u_i!}.$$

It is clear that if $\omega \in B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)$, $\tau_n(k,r) = (k-r-1)+n(r+1)+\sum_{i=0}^{r}u_i(k+i-r) = (k-r-1)+n(r+1)+t$ and $\gamma(n) = (k-r-1)+n(r+1)+t$. Thus, $\Big\{\tau_n(k,r)-(k-r-1)-n(r+1)=t, B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)\Big\} = B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)$. So, by part (a) of Lemma 3, we have

$$\begin{aligned}
&\mathbb{P}_x\Big(\tau_n(k,r)-(k-r-1)-n(r+1)=t, B_{\gamma(n)}(K, u_0, u_1, \ldots, u_r)\Big)\\
&\geq \delta_m\Big(\gamma(n), \sum_{i=0}^{r} u_i, x\Big)\frac{\Big(n-t_0(1+\sum_{i=0}^{r}u_i)+\sum_{i=0}^{r}u_i\Big)!}{\prod_{i=0}^{r}u_i!\Big(n-t_0(1+\sum_{i=0}^{r}u_i)\Big)!}\\
&= \Big(1-p_{2^m-1}\Big)^{(\sum_{i=0}^{r}u_i)}\Big(p_{2^m-1}\Big)^{n(r+1)}\Big(1+o(1)\Big)\times\frac{n^{(\sum_{i=0}^{r}u_i)}}{\prod_{i=0}^{r}u_i!}\Big(1+o(1)\Big)\\
&\to \prod_{i=0}^{r}\frac{\exp(-\lambda_{2^m-1})(\lambda_{2^m-1})^{u_i}}{u_i!} \qquad \text{as } n\to\infty.
\end{aligned}$$

This establishes part (a).

For part (b), fix any $t \geq 0$. Consider the collection of events $\Big\{C_{\gamma(n)}(K, u_0, u_1, \ldots, u_{k-1}, v_0, v_1, \ldots, v_{r-k}) : u_i, v_i \geq 0, t = \sum_{i=0}^{k-1}u_i(i+1)\Big\}$. Again these events are disjoint. Further, we have $\Big\{\tau_n(k,r)-k-(n-1)(r+1)=t, C_{\gamma(n)}(K, u_0, u_1, \ldots, u_{k-1}, v_0, v_1, \ldots, v_{r-k})\Big\} =$

$C_{\gamma(n)}(K, u_0, u_1, \ldots, u_{k-1}, v_0, v_1, \ldots, v_{r-k})$. Therefore, by part (b) of Lemma 3, we have

$$
\begin{aligned}
&\mathbb{P}_x\Big(\tau_n(k,r) - k - (n-1)(r+1) = t, C_{\gamma(n)}(K, u_0, \ldots, u_{k-1}, v_0, \ldots, v_{r-k})\Big) \\
&\geq \frac{\Big(n - 1 - (t_0 + t_0\sum_{i=0}^{k-1} u_i + (1+t_0)\sum_{i=0}^{r-k} v_i) + \sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i\Big)!}{\prod_{i=0}^{k-1} u_i! \prod_{i=0}^{r-k} v_i! \Big(n - 1 - (t_0 + t_0\sum_{i=0}^{k-1} u_i + (t_0+1)\sum_{i=0}^{r-k} v_i)\Big)!} \\
&\qquad \times \delta_m\Big(\gamma(n), \sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i, x\Big) \\
&= \Big(1 - p_{2^m-1}\Big)^{(\sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i)} \Big(p_{2^m-1}\Big)^{n(r+1)} \Big(1 + o(1)\Big) \\
&\qquad \times \frac{n^{(\sum_{i=0}^{k-1} u_i + \sum_{i=0}^{r-k} v_i)}}{\prod_{i=0}^{k-1} u_i! \prod_{i=0}^{r-k} v_i!} \Big(1 + o(1)\Big) \\
&\to \prod_{i=0}^{k-1} \frac{\exp(-\lambda_{2^m-1})(\lambda_{2^m-1})^{u_i}}{u_i!} \prod_{i=0}^{r-k} \frac{\exp(-\lambda_{2^m-1})(\lambda_{2^m-1})^{v_i}}{v_i!} \qquad\qquad \text{as } n \to \infty.
\end{aligned}
$$

This, by Lemma 2, completes the proof of the Theorem. □

Remark 1: Since the limiting distribution is independent of the initial condition, we can assume any distribution on the initial conditions. Suppose that μ is the probability distribution on $\{0,1\}^m$. As we have already discussed, μ can be identified as a probability measure on N_m by the mapping $(x_0, x_1, \ldots, x_{m-1}) \to x = \sum_{j=0}^{m-1} 2^j x_j$ where each $x_i \in \{\mathbf{0}, \mathbf{1}\}$. Let $\mathbb{P}_\mu$ be the probability measure governing the Markov chain with initial distribution μ. From theorem 1, we can easily conclude that, under $\mathbb{P}_\mu$

$$
\begin{aligned}
&(a) \qquad \tau_n(k,r) - (k - r - 1) - n(r+1) \;\Rightarrow\; \sum_{i=0}^{r} \xi_i^{(1)} \text{ when } r \leq k-1 \\
&(b) \qquad \tau_n(k,r) - k - (n-1)(r+1) \;\Rightarrow\; \sum_{i=0}^{k-1} \xi_i^{(2)} \text{ when } r > k-1
\end{aligned}
$$

by first conditioning on $x \in N_m$ and then summing over all possible values of $x \in N_m$. The random variables, $\xi_i^{(1)}$ and $\xi_i^{(2)}$ are as defined in the Theorem 1.

4. Central limit theorem

In this section, we prove the central limit theorem for $N_{n,k,r}$ and obtain the uniform rate of convergence for the central limit theorem. The result can be generalized for a wider class of processes; however we concentrate only on the m^{th} order Markov chain set-up described in this paper.

We define two new sequences of random variables: the first one, Y_n, captures the sequence of $\mathbf{1}$'s observed in last k trials (going back from trial n) and the second one, Z_n, keeps track whether the end point of the last r-look-back run of length k is within r trials (going back) from the trial n. Both these random variables assume values in finite sets. Further, the random vector (Y_n, Z_n) jointly form a homogeneous Markov chain taking values in a finite set (for sufficiently large n). Next, we translate the description of r-look-back runs of length k, from the original random variables $\{X_n\}$ to the new set of random vectors $\{(Y_n, Z_n)\}$. Further, the newly defined Markov chain will be a irreducible chain; hence will satisfy the properties of ϕ-mixing sequence. This allows us to apply the central limit theorem and the rate of convergence results for the ϕ-mixing sequence to this case to yield Theorem 2.

Define $s = \max(k, m)$. Set $X_{-m} = X_{-m-1} = \cdots = X_{-s+1} = 0$ provided $s > m$. Define a sequence of random variables Y_n as follows:

$$Y_n = \sum_{j=0}^{s-1} 2^j X_{n-j}$$

for $n \geq 1$. Since $X_i \in \{0, 1\}$ for all i, Y_n assumes values in the set N_s. It is clear that Y_n captures the last s observations $\{X_n, X_{n-1}, \dots, X_{n-s+1}\}$. Indeed, from the binary expansion of Y_n, one can easily retrieve the values of X_n's.

Since the sequence of random variables X_n is stationary and form a m^{th} order Markov chain, we have that the random variables $\{Y_i : i \geq 0\}$ form a homogeneous Markov chain with initial distribution δ_x, where δ_x is the Dirac measure at $x \in N_m$, and transition probabilities given by

$$\mathbb{P}(Y_{n+1} = y_1 | Y_n = y_0) = \begin{cases} p_{\theta_m(y_0)} & \text{if } y_1 = f_s(y_0) \\ 1 - p_{\theta_m(y_0)} & \text{if } y_1 = g_s(y_0) \\ 0 & \text{otherwise} \end{cases}$$

where $\theta_m : N_s \to N_m$ is given by $\theta_m(x) = x \pmod{2^m}$.

Let $N'_r = \{0, 1, 2, \dots, 2^{r-1}\}$. Now, we define another sequence of random variables $\{Z_n\}$ taking values in the set N'_r. Set $Z_n = 0$ for $n < k$. For $n \geq k$, we define

$$Z_n = \begin{cases} 2Z_{n-1} \pmod{2^r} & \text{if } Z_{n-1} > 0 \\ 1\{Y_n \pmod{2^k} = 2^k - 1\} & \text{otherwise} \end{cases}$$

where $1\{Y_n \pmod{2^k} = 2^k - 1\}$ is the indicator variable for the event $\{Y_n \pmod{2^k} = 2^k - 1\}$.

Now, for all $n \geq s$, the joint distribution of (Y_n, Z_n) is Markovian since Y_n is Markovian and independent of $\{Z_i : i \leq n-1\}$ and the value of Z_n depends only on the values of Z_{n-1} and Y_n. The transition probabilities are easy to compute: for $y_0, y_1 \in N_s, z_0, z_1 \in N'_r$, we have

$$\begin{aligned} &\mathbb{P}\Big((Y_{n+1}, Z_{n+1}) = (y_1, z_1) \mid (Y_n, Z_n) = (y_0, z_0)\Big) \\ &= \mathbb{P}\Big(Z_{n+1} = z_1 | Y_{n+1} = y_1, (Y_n, Z_n) = (y_0, z_0)\Big)\mathbb{P}\Big(Y_{n+1} = y_1 | (Y_n, Z_n) = (y_0, z_0)\Big) \\ &= \mathbb{P}\Big(Z_{n+1} = z_1 | Y_{n+1} = y_1, Z_n = z_0\Big)\mathbb{P}\Big(Y_{n+1} = y_1 | Y_n = y_0\Big). \end{aligned}$$

Note that $\mathbb{P}\Big(Z_{n+1} = z_1 | Y_{n+1} = y_1, Z_n = z_0\Big)$ is a deterministic function taking values in $\{0, 1\}$. We also assume that $p_x > 0$ for all $x \in N_m$. Therefore we can conclude that the sequence of random variables $\{(Y_n, Z_n) : n \geq s\}$ is a homogeneous Markov chain with transition probabilities specified by the above formula. However, not all states of $N_s \times N'_r$

are feasible for the Markov chain $\{(Y_n, Z_n) : n \geq s\}$. For example, the state $(0, 1)$ can never be reached. Therefore, we need to restrict our attention to a smaller set.

Let $\mathcal{S}$ be the collection of all the feasible states, *i.e.*, all states of $N_s \times N'_r$ which can be reached by this Markov chain. Formally, we define $\mathcal{S} := \{(y, z) : (0, 0) \rightsquigarrow (y, z)\}$ where $(y_1, z_1) \rightsquigarrow (y_2, z_2)$ implies that (y_2, z_2) can be reached from (y_1, z_1) (in the usual Markov chain sense). Since $1 > p_x > 0$ for all $x \in N_m$, it is easy to see that if $(0, 0) \rightsquigarrow (y, z)$, then $(y, z) \rightsquigarrow (0, 0)$. Therefore, if we restrict our attention to the set $\mathcal{S}$, we get an irreducible Markov chain. More formally, define the following stopping time:

$$\tau_S := \inf\{n \geq 1 : (Y_n, Z_n) \in \mathcal{S}\}. \tag{9}$$

Now, the process after the stopping time, *i.e.*, $\{(Y_n, Z_n) : n \geq \tau_S + 1\}$, using the strong Markov property, is a homogeneous, irreducible Markov chain with state space $\mathcal{S}$ with transition probabilities as specified. The initial distribution of this chain is given by the distribution of the random variable (Y_{τ_S}, Z_{τ_S}) starting from $x \in N_m$ *i.e.*, the measure $\mu^{(x)}$ on $\mathcal{S} \subseteq N_s \times N'_r$ where

$$\mu^{(x)}(\{y, z\}) = \mathbb{P}((Y_{\tau_S} = y, Z_{\tau_S} = z)|Y_0 = x) \text{ for } (y, z) \in \mathcal{S}.$$

Further, observe that $\tau_S \leq s$ almost surely. Indeed, suppose that (y, z) is any possible value of (Y_s, Z_s) which has been obtained through the observations $X_1 = x_1, X_2 = x_2, \ldots, X_s = x_s$. Now, for any $n \geq \tau_S$ with $Y_n = 0$ and $Z_n = 0$, the probability of the event $\{X_{n+1} = x_1, X_{n+2} = x_2, \ldots, X_{n+s} = x_s\}$ is positive. Clearly, in that case, $Y_{n+s} = y$ and $Z_{n+s} = z$. Therefore, $(y, z) \in \mathcal{S}$ which implies that $(Y_s, Z_s) \in \mathcal{S}$.

Note that $R_i(k, r) = 1$ if and only if $Z_i = 1$ for any $i \geq 1$. Therefore, we may define $R_i(k, r) = 1\{Z_i = 1\}$ for $i \geq 1$. Now, we claim that $\{R_i(k, r) : i \geq \tau_S + 1\}$ is a ϕ-mixing sequence. Since $\{(Y_i, Z_i) : i \geq \tau_S + 1\}$ is an irreducible homogeneous Markov chain with finite state space, it is a ϕ-mixing sequence with mixing coefficients given by $\phi_n = C\rho^n$ for $n \geq 1$ where $C > 0$ and $0 < \rho < 1$ are constants. Since, $R_i(k, r)$ is function of Z_i only, the same mixing coefficients will satisfy the mixing condition for the sequence $\{R_i(k, r) : i \geq \tau_S + 1\}$.

However, note here that $\{R_i(k, r) : i \geq \tau_S + 1\}$ need not be a stationary sequence of random variables. Babu, Ghosh and Singh (1978) have studied the convergence rates of central limit theorem for non-stationary ϕ-mixing sequences. We state theorem 1 of Babu, Ghosh and Singh (1978) here for the sake of completeness.

Theorem 3: (Babu, Ghosh and Singh) For a ϕ-mixing sequence $\{X_n\}$, let $S_n = \sum_{i=1}^n X_i$, $\sigma_n^2 = \mathrm{Var}(S_n)$ and $F_n(t) = \mathbb{P}(S_n \leq t\sigma_n)$. Suppose that

$$\mathbb{E}(X_n) = 0 \quad \text{ for all } n \geq 1, \tag{10}$$

$$\sum_{n=1}^{\infty} \phi_n^{1/2} < \infty, \tag{11}$$

$$\inf_{n \geq 1} n^{-1/2}\sigma_n > 0, \tag{12}$$

and for some $c > 0$ and $M > 1$

$$\mathbb{E}(|X_n|^{2+c}) < M, \quad \text{ for all } n \geq 1. \tag{13}$$

Then,

$$\sup_{t\in\mathbb{R}} \mid F_n(t) - \Phi(t) \mid = O(n^{-\gamma(c)} \log n)$$

where $\gamma(c) = 2c^\star/(6+5c^\star)$ and $c^\star = \min(1, c)$.

We will use the above result to prove the Theorem 2. Let us define, for $i \geq \tau_S + 1$,

$$R'_i(k,r) = R_i(k,r) - \mathbb{E}(R_i(k,r))$$

as the centred sequence of random variables. It is clear that the conditions (10), (11) and (13) of above result holds with $c = 1$. Since $c = 1$, we have that $\gamma(c) = 2/11$. To prove (12) we proceed as follows:

Define a sequence of stopping times in the following way: $\zeta_0 := \inf\{n > \tau_S : (Y_n, Z_n) = 0\}$ and for $i \geq 1$, set $\zeta_i := \inf\{n > \zeta_{i-1} : (Y_n, Z_n) = 0\}$. Further define a sequence of random variables, $U_0 := \sum_{j=\tau_S+1}^{\zeta_0} R'_j(k,r)$ and for $i \geq 1$, $U_i := \sum_{j=\zeta_{i-1}+1}^{\zeta_i} R'_j(k,r)$. In the following lemma, we prove independence of the collection of random variables $\{U_i : i \geq 0\}$.

Lemma 4: The collection of random variables $\{U_t : t \geq 0\}$ are independent. Further, $\{U_t : t \geq 1\}$ are identically distributed.

We will proceed to prove theorem 2 assuming the result of Lemma 4 and prove this lemma in the end. Define, $N(n) := \inf\{t : \zeta_t > n\}$. Next we need the following result

Lemma 5:

$$\frac{\mathrm{Var}(\sum_{j=\tau_S+1}^{n} R'_j(k,r))}{n} \to C_1 \text{ as } n \to \infty$$

where $C_1 > 0$ is a constant.

Proof: Let us define $N'_{n,k,r} = \sum_{j=\tau_S+1}^{\zeta_{N(n)}} R'_j(k,r) = U_0 + \sum_{j=1}^{N(n)} U_j$. Thus, we have,

$$\begin{aligned}
\mathbb{E}(N'_{n,k,r}) &= \mathbb{E}(U_0) + \mathbb{E}(\sum_{j=1}^{N(n)} U_j) = \mathbb{E}(U_0) + \mathbb{E}(N(n))\mathbb{E}(U_1) \\
\mathrm{Var}(N'_{n,k,r}) &= \mathrm{Var}(U_0 + \sum_{j=1}^{N(n)} U_j) \\
&= \mathrm{Var}(U_0) + \mathbb{E}(N(n))\mathrm{Var}(U_1) + \mathrm{Var}(N(n))\mathbb{E}(U_1^2).
\end{aligned}$$

The stop times $\{\zeta_t : t \geq 0\}$ represent the visits of the Markov chain to the state $(0,0)$. Thus it is a renewal event. So, $N(n)$ represents the number of renewals till time $n - \tau_S$. Since $\tau_S \leq s$, we have

$$\frac{\mathbb{E}(N(n))}{n} \to C_2 \text{ and } \frac{\mathrm{Var}(N(n))}{n} \to C_3 \text{ as } n \to \infty$$

where $C_2, C_3 > 0$ (see Feller (1968)). Thus, we have that

$$\frac{\mathrm{Var}(N'_{n,k,r})}{n} \to C_4 \text{ for some constant } C_4 > 0. \tag{14}$$

Now, we have,

$$
\begin{aligned}
&\left|\mathrm{Var}(\sum_{j=\tau_S+1}^{n} R'_j(k,r)) - \mathrm{Var}(N'_{n,k,r})\right| \\
&\quad \leq \mathrm{Var}\left(\sum_{j=n+1}^{\zeta_{N(n)}} R'_j(k,r)\right) + 2\left|\mathrm{Cov}\left(\sum_{j=n+1}^{\zeta_{N(n)}} R'_j(k,r), N'_{n,k,r}\right)\right| \\
&\quad \leq 4\mathbb{E}(\zeta_1^2) + 2\left(\mathrm{Var}\left(\sum_{j=n+1}^{\zeta_{N(n)}} R'_j(k,r)\right)\mathrm{Var}\left(N'_{n,k,r}\right)\right)^{1/2} \\
&\quad \leq C_5 n^{1/2}
\end{aligned}
$$

for some constant $C_5 > 0$. This coupled with (14) proves the Lemma. $\square$

Now, we are in a position to prove the Theorem 2.

Proof: For $n > s$, we have that

$$
\begin{aligned}
&\frac{N_{n,k,r} - \mathbb{E}(N_{n,k,r})}{\sqrt{\mathrm{Var}(N_{n,k,r})}} \\
&= \frac{\sum_{i=k}^{\tau_S} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} + \frac{\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} - \frac{\sum_{i=n+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} \\
&= \frac{\sum_{i=k}^{\tau_S} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} + \frac{\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r))}} \times \frac{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r))}}{\sqrt{\mathrm{Var}(N_{n,k,r})}} - \frac{\sum_{i=n+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} \\
&= E_1 + E_2 + E_3 + \frac{\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r))}}
\end{aligned}
$$

where

$$
\begin{aligned}
E_1 &= \frac{\sum_{i=k}^{\tau_S} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} \\
E_2 &= -\frac{\sum_{i=n+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(N_{n,k,r})}} \\
E_3 &= \Big(\frac{\sigma'_n}{\sigma_n} - 1\Big)\frac{\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r)}{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r))}}
\end{aligned}
$$

with $\sigma_n^2 = \mathrm{Var}(N_{n,k,r})$ and $(\sigma'_n)^2 = \mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R'_i(k,r))$.

First, using Lemma 5, we show that

$$n^{-1/2}\sigma'_n \to C_6 \text{ as } n \to \infty$$

where $C_6 > 0$ is a constant. Indeed, we have

$$\begin{aligned}
&\left|(\sigma_n')^2 - \mathrm{Var}(N'_{n,k,r})\right| \\
&\leq \mathrm{Var}(\sum_{j=n+1}^{\tau_S+n} R_i'(k,r)) + 2\left|\mathrm{Cov}(\sum_{j=n+1}^{\tau_S+n} R_i'(k,r), N'_{n,k,r})\right| \\
&\leq C_7 n^{1/2}
\end{aligned}$$

using Lemma 5 and the fact that $\tau_S \leq s$. This implies the condition (12) of Babu, Ghosh and Singh (1978) is satisfied. Hence, using their result, we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sum_{i=\tau_S+1}^{\tau_S+n} R_i'(k,r) \leq t\sigma_n'\right) - \Phi(t)\right| = O(n^{-2/11}\log n). \tag{15}$$

To conclude the result of the Theorem 2, we need to show that for some $K > 0$,

$$\mathbb{P}(|E_1 + E_2 + E_3| > Kn^{-2/11}\log n) = O(n^{-2/11}\log n).$$

Note that, using a similar argument as above, we get

$$\frac{\sigma_n^2}{n} \to C_8 \text{ as } n \to \infty$$

where $C_8 > 0$. Thus, we have, for constants C_9 and C_{10}, $|E_1| \leq C_9 n^{-1/2}$ and $|E_2| \leq C_{10} n^{-1/2}$. Finally, again using similar arguments, $|\sigma_n'/\sigma_n - 1| \leq C_{11} n^{-1/2}$. Thus, for n sufficiently large, we have,

$$\begin{aligned}
&\mathbb{P}(|E_1 + E_2 + E_3| > Kn^{-2/11}\log n) \\
&\leq \mathbb{P}(|E_3| > K'n^{-2/11}\log n) \\
&\leq \mathbb{P}\left(\left|\frac{\sum_{i=\tau_S+1}^{\tau_S+n} R_i'(k,r)}{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R_i'(k,r))}}\right| > K''n^{1/2-2/11}\log n\right) \\
&\leq \Phi(K''n^{1/2-2/11}\log n) + 2\sup_{t\in\mathbb{R}}\left|\Phi(t) - \mathbb{P}\left(\frac{\sum_{i=\tau_S+1}^{\tau_S+n} R_i'(k,r)}{\sqrt{\mathrm{Var}(\sum_{i=\tau_S+1}^{\tau_S+n} R_i'(k,r))}} \leq t\right)\right| \\
&= O(n^{-2/11}\log n)
\end{aligned}$$

using (15) and property of the normal distribution function, where K', K'' are positive constants. This proves the theorem. □

Now, we prove the Lemma 4.

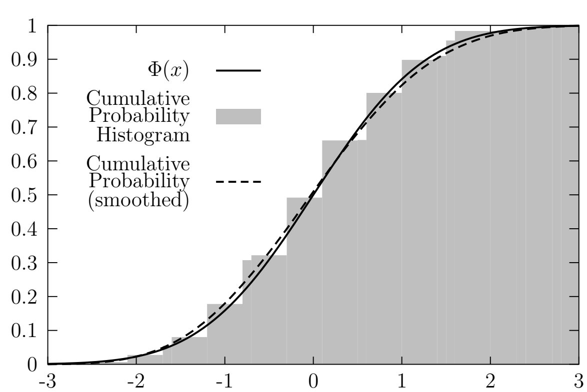
Proof: It is clear, from the definition of U_0, that U_0 is determined by the process $\{(Y_j, Z_j) : \tau_S + 1 \leq j \leq \zeta_0\}$, *i.e.*, U_0 is a $\mathcal{F}_{\zeta_0} = \sigma((Y_j, Z_j) : \tau_S + 1 \leq j \leq \zeta_0)$ measurable random variable. Further, for $i \geq 1$, the random variable U_i is determined by the process $\{(Y_j, Z_j) : j > \zeta_{i-1} + 1\}$. Therefore, the sequence of random variables $\{U_i : i \geq 1\}$ are measurable with respect to the sigma algebra generated by $\{(Y_j, Z_j) : j \geq \zeta_0 + 1\} = \mathcal{F}_{\zeta_0+}$. Now, the conditional distribution of the process $\{(Y_j, Z_j) : j \geq \zeta_0 + 1\}$, given the process up to time $\zeta_0(\mathcal{F}_{\zeta_0})$, using the strong Markov property, is same as that of $\{(Y_j, Z_j) : j \geq 0\}$ with the initial condition that $(Y_0, Z_0) = (0, 0)$. Therefore, it is independent of the process up to time ζ_0. Hence, it is independent of the random variables which are measurable with respect to the process $\{(Y_j, Z_j) : j \leq \zeta_0\}$. Thus, U_0 is independent of $\{U_i : i \geq 1\}$. Now this argument can be carried out inductively to prove the result. Further, the distribution $\{U_i : i \geq 1\}$ depends only on the initial condition $(Y_{\zeta_0}, Z_{\zeta_0}) = (0, 0)$ and transition matrix of the Markov chain. Since, for any $i \geq 1$, the sequence $\{U_j : j \geq i\}$ will have the same initial condition $((Y_{\zeta_{i-1}}, Z_{\zeta_{i-1}}) = (0, 0))$ and the same transition probabilities, we have that $\{U_i : i \geq 1\}$ are identically distributed. However, the initial condition of the sequence $\{U_i : i \geq 0\}$ is given by the distribution of (Y_{τ_S}, Z_{τ_S}). There is no reason to expect that, $(Y_{\tau_S}, Z_{\tau_S}) = (0, 0)$. Therefore, U_0 may have a different distribution. $\square$

5. Simulation results

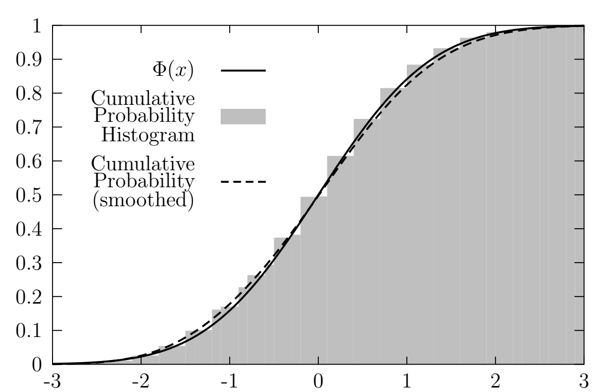
In this section, we provide some simulation results exhibiting the goodness of the approximation in the central limit theorem. These results have been obtained for a Markov chain, with the transition matrix P given by

$$P = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}.$$

Simulation has been performed for n number of trials where $n = 50, 100, 500$ and 1000. For $k = 4$ and $r = 2$, the values of $N_{n,4,2}$ have been computed. For each choice of n, the experiment is repeated 10000 times and then the mean and the variance of $N_{n,4,2}$ have been obtained. Normalizing these 10000 observations, cumulative probability histograms have been drawn for a grid of 0.1 using the computer package GNU PLOT. The smoothed version of the histogram have been plotted using bezier smoothing algorithm. From the following plots, it is indeed evident that the smoothed version of the cumulative probability histogram is a good approximation of the normal probability distribution function $(\Phi(x))$ even for value of n as small as 50.

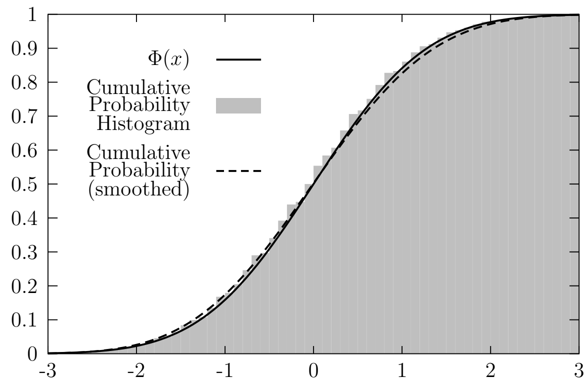


(a) $n = 50$

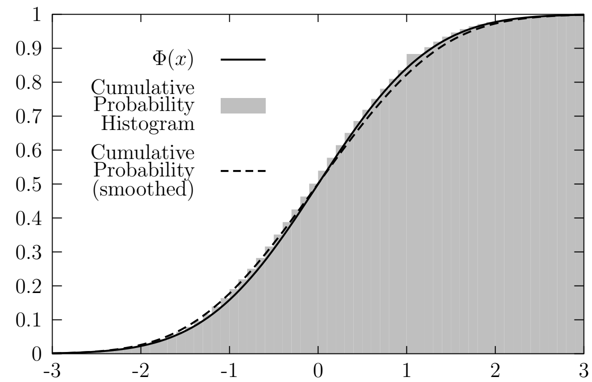


(b) $n = 100$

Further, we support the findings by illustrating the simulated values of the maximum difference between $\Phi(x)$ and $\mathbb{P}(N_{n,k,r} - \mathbb{E}(N_{n,k,r}) \leq \sigma_n x)$ for $-3.0 \leq x \leq 3.0$ and for various



(c) $n = 500$



(d) $n = 1000$

choices of (k, r) and n in Table 1. Here the underlying sequence of random variables constitute a Markov chain with transition probabilities $p_{01} = 0.4$ and $p_{11} = 0.8$. The table 1 shows a nice decay of the maximum difference as n grows.

Table 1

Sample size (n)	$k=4$ $r=2$	$k=5$ $r=7$	$k=5$ $r=0$	$k=6$ $r=5$	$k=6$ $r=8$	$k=7$ $r=3$	$k=8$ $r=2$
50	0.081278	0.164263	0.060424	0.132263	0.172504	0.127723	0.148725
75	0.064522	0.134955	0.056294	0.118998	0.140559	0.107202	0.105870
100	0.050815	0.117124	0.052313	0.092598	0.123940	0.091438	0.093857
125	0.045183	0.099382	0.041538	0.074459	0.107547	0.082336	0.085319
150	0.038893	0.083109	0.039858	0.069004	0.090610	0.060012	0.076902
200	0.036554	0.075693	0.028927	0.067591	0.088099	0.059561	0.062446
300	0.028691	0.061773	0.025912	0.047239	0.066568	0.047434	0.044028
400	0.020599	0.049510	0.019476	0.041114	0.060933	0.035864	0.035205
500	0.019384	0.047328	0.018392	0.039172	0.054056	0.024733	0.025782
1000	0.008500	0.032516	0.009471	0.022737	0.035285	0.022301	0.009905

References

Aki, S. (1985). Discrete distributions of order k on a binary sequence. *Annals of the Institute of Statistical Mathematics*, **37**, 205–224.

Aki, S., Balakrishnan, N. and Mohanty, S. G. (1996). Sooner and later waiting time problems for success and failure runs in higher order Markov dependent trials. *Annals of the Institute of Statistical Mathematics*, **48**, 773–787.

Aki, S. and Hirano, K. (1988). Some characteristics of the binomial distribution of order k and related distributions. *Statistical Theory and Data Analysis II, Proceedings of the 2^{nd} Pacific Area Statistical Conference,* North-Holland, Amsterdam, 211–222.

Aki, S. and Hirano, K. (1995). Joint distributions of numbers of success-runs and failures until the first k consecutive successes. *Annals of the Institute of Statistical Mathematics*, **47**, 225–235.

Aki, S. and Hirano, K. (2000). Number of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k. *Annals of the Institute of Statistical Mathematics*, **52**, 767–777.

Babu, G. J., Ghosh, M. and Singh. K. (1978) On rates of convergence to narmality for ϕ-mixing processes, *Sankhyã, Series A*, **40**, 278–293.

Balasubramanian, K., Viveros, R. and Balakrishnan, N. (1993). Sooner and later waiting time problems for Markovian Bernoulli trials. *Statistics and Probability Letters*, **18**, 153–161.

Billingsley P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

Feller W. (1968) *An introduction to Probability Theory and its Applications. Vol - I.* John Wiley, New York, 3rd ed.

Fu, J. C. and Koutras, M. V. (1994). Distribution Theory of runs: a Markov Chain approach. *Journal of the American Statistical Association*, **89**, 1050–1058.

Godbole, A. P. (1990). Specific formulæ for some success run distributions. *Statistics and Probability Letters*, **10**, 119–124.

Godbole, A. P. (1992). The exact and asymptotic distribution of overlapping success runs. *Communications in Statistics - Theory and Methods*, **21**, 953–967.

Han, S. and Aki, S. (2000). A unified approach to binomial-type distributions of order k. *Communications in Statistics - Theory and Methods*, **29**, 1929–1943.

Hirano, K. (1986). Some properties of the distributions of order k. in: A. N. Philippou and A. F. Horadam eds., *Fibonacci numbers and their applications*, (Reidel Dordrecht), 43–53.

Hirano, K. and Aki, S. (1993). On number of occurrences of success runs of specified length in a two-state Markov chain. *Statistica Sinica*, **3**, 313–320.

Hirano, K., Aki, S., Kashiwagi, N. and Kuboki, H. (1991). On Ling's binomial and negative binomial distributions of order k. *Statistics and Probability Letters*, **11**, 503–509.

Inoue, K. and Aki, S. (2003). Generalized binomial and negative binomial distributions of order k by the l-overlapping enumeration scheme. *Annals of the Institute of Statistical Mathematics*, **55**, 153–167.

Koutras, M. V. (1996). On a waiting time distribution in a sequence of Bernoulli Trials. *Annals of the Institute of Statistical Mathematics*, **48**, 789–806.

Ling, K. D. (1988). On binomial distributions of order k. *Statistics and Probability Letters*, **6**, 247–250.

Makri, F. S. and Psillakis, Z. M. (2015). On l-overlapping Runs of Ones of Length k in Sequences of Independent Binary Random Variables. *Communications in Statistics - Theory and Methods*, **44**, 3865–3884.

Philippou, A. N. and Makri, F. S. (1986). Successes, runs and longest runs. *Statistics and Probability Letters*, **4**, 101–105.

Uchida, M. (1998). On number of occurrences of success runs of specified length in a higher-order two-state Markov chain, *Annals of the Institute of Statistical Mathematics*, **50**, 587–601.

Uchida, M. and Aki, S. (1995). Sooner and later waiting time problems in a two-state Markov chain, *Annals of the Institute of Statistical Mathematics*, **47**, 415–433.