DISTRIBUTION THEORY OF RUNS AND RUN-RELATED STATISTICS IN SEQUENCES OF DEPENDENT TRIALS

FEMİN YALÇIN

JULY 2013

DISTRIBUTION THEORY OF RUNS AND RUN-RELATED STATISTICS IN SEQUENCES OF DEPENDENT TRIALS

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL

AND APPLIED SCIENCES OF

IZMIR UNIVERSITY OF ECONOMICS

BY

FEMİN YALÇIN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

JULY 2013

Ph.D. DISSERTATION EXAMINATION RESULT FORM

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Murat Aşkar
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. İsmihan Bayramoğlu
Head of Department

We have read the dissertation entitled **"Distribution Theory of Runs and Run-related Statistics in Sequences of Dependent Trials"** completed by **FEMİN YALÇIN** under supervision of **Prof. Dr. İsmihan Bayramoğlu** and **Prof. Dr. Serkan Eryılmaz** and we certify that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Dr. Serkan Eryılmaz
Co-Supervisor

Prof. Dr. İsmihan Bayramoğlu
Supervisor

Examining Committee Members

Date: July 12, 2013

Prof. Dr. İsmihan Bayramoğlu
Dept. of Mathematics, İUE

Assoc. Prof. Dr. Alexei Stepanov
Dept. of Mathematics, İUE

Assoc. Prof. Dr. Burcu H. Üçer
Dept. of Statistics, DEU

Asst. Prof. Dr. Uğur Madran
Dept. of Mathematics, İUE

Asst. Prof. Dr. Halil Tanıl
Dept. of Statistics, EU

ABSTRACT

DISTRIBUTION THEORY OF RUNS AND RUN-RELATED STATISTICS IN SEQUENCES OF DEPENDENT TRIALS

FEMİN YALÇIN
Ph.D. in Applied Mathematics and Statistics
Graduate School of Natural and Applied Sciences
Supervisor: Prof. Dr. İsmihan Bayramoğlu
Co-Supervisor: Prof. Dr. Serkan Eryılmaz
July 2013

Runs and run-related statistics have been successfully used in various applied fields including system reliability, statistical process control, and hypothesis testing. Various methods have been developed in the literature to investigate both exact and asymptotic distributions of runs not only for independent trials but also for a sequence of dependent trials.

In this thesis, two different types of dependence have been considered one of which is the Markovian type of dependence. According to the other dependence model, the outcome of the present trial depends on the total number of successful trials so far. Distributions of some run statistics under these dependence types have been obtained and applications of the theoretical results have been established with illustrative examples.

Keywords: Runs, Bernoulli trials, Exact distributions, Markov chain, Previous-sum dependent model.

ÖZ

BAĞIMLI DENEME DİZİLERİNDE TEKRARLAR VE TEKRARLARLA İLGİLİ İSTATİSTİKLERİN DAĞILIM TEORİSİ

FEMİN YALÇIN
Uygulamalı Matematik ve İstatistik, Doktora
Fen Bilimleri Enstitüsü
Tez Danışmanı: Prof. Dr. İsmihan Bayramoğlu
İkinci Tez Danışmanı: Prof. Dr. Serkan Eryılmaz
Temmuz 2013

Tekrarlar ve tekrarlarla ilgili istatistikler, sistem güvenilirliği, istatistiksel süreç kontrolü ve hipotez testi gibi çeşitli uygulamalı alanlarda kullanılmaktadır. Literatürde hem bağımsız hem de bağımlı deneme dizileri üzerinde tanımlanan tekrarların dağılımlarını elde etmek için çeşitli yöntemler geliştirilmiştir.

Bu tezde, biri Markov bağımlılık olmak üzere iki farklı bağımlılık türü ele alınmıştır. Diğer bağımlılık modeline göre bir denemenin başarılı ya da başarısız olması, kendisinden önceki denemelerdeki toplam başarılı deneme sayısına bağlıdır. Bazı tekrar istatistiklerinin dağılımları elde edilmiş ve teorik sonuçların uygulamaları açıklayıcı örneklerle pekiştirilmiştir.

Anahtar Kelimeler: Tekrarlar, Bernoulli denemeleri, Dağılımlar, Markov zinciri, Geçmiş toplama dayalı bağımlılık modeli.

ACKNOWLEDGEMENT

In the beginning of the present thesis I would like to express my deepest gratitude to my PhD supervisors, Prof. Dr. İsmihan Bayramoğlu and Prof. Dr. Serkan Eryılmaz, for their excellent support and supervision during the entire phase of this doctoral program. They dedicated their precious time to guide me during these past four years.

I want to thank Assoc. Prof. Dr. Gözde Yazgı Tütüncü for her endless encouragement and support. She has been a strong and supportive mentor to me throughout my both undergraduate and graduate school career.

I am very grateful to Assoc. Prof. Dr. Alexei Stepanov and Asst. Prof. Dr. Güvenç Arslan for many helpful discussions and comments.

I have to thank the other members of my thesis monitoring committee, Asst. Prof. Dr. Uğur Madran and Asst. Prof. Dr. Halil Tanıl for their helpful advices and suggestions in general.

I also extend my thanks to my current and former colleagues Demet Ersoy Özdek, Ceki Franko, Cihangir Kan, Cemal Murat Özkut, and Hande Tunçel for providing support and friendship that I needed.

I am deeply indebted to my parents Nilhan and Mehmet Akif Yalçın and to my grandparents Kadriye and Gökhan Toker for supporting and motivating me during all stages of my life.

Special thanks to my fiance, best friend and soul mate Yusuf Ziya Güleç. I couldn't be able to finish this journey without his love and encouragement. Finding him and my lovely cat Sushi and putting them at the center of my life are the best outcomes from these four years.

Finally, I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting me with its national scholarship programme.

To my parents and my fiance

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

Studies based on runs in a sequence consisting of two-state (Bernoulli) trials were started in the areas of hypothesis testing ([67] and [89]) and statistical quality control ([68] and [90]) in 1940's. During the late 1980's and early 1990's, the concept of runs became very popular again in many fields including reliability theory (see [16] and the review articles [15] and [24]), start-up demonstration testing ([44]), DNA sequence matching ([43]), and radar astronomy ([79]).

Various definitions of a success run of length k in a sequence consisting of successes and failures have been proposed in the literature (see [36]). These definitions differ from each other through the counting scheme used. Which counting scheme should be used depends on the problem. The four best-known counting schemes are:

(i) Counting the number of success runs of length *exactly* k (Mood [67]);

(ii) Counting the number of *non-overlapping* success runs of length k (Feller [33]);

(iii) Counting the number of *overlapping* success runs of length k (Ling [56]);

(iv) Counting the number of success runs of length *at least* k (Goldstein [43]).

In this thesis, we define the term *success run* to be an uninterrupted sequence of successes ("1"s) bordered at each end by failures ("0"s) or by the beginning or by the end of the complete sequence. This definition is a reduced (to Bernoulli trials) version of run definition that belongs to Balakrishnan and Koutras [11] and it is based on the first counting scheme stated above. The term *failure run* can be defined analogously. For example, in a sequence of twelve trials 110110001111 we have a success run of length two followed by a failure run of length one, then a success run of length two again, a failure run of length three and finally a success run of length four.

Some well-known and often used run-statistics defined in a sequence of n Bernoulli trials are:

(i) S_n, total number of successes;

(ii) $N_{n,k}$, number of non-overlapping consecutive k successes;

(iii) $M_{n,k}$, number of overlapping consecutive k successes;

(iv) $E_{n,k}$, number of success runs of length exactly k;

(v) $G_{n,k}$, number of success runs of length greater than or equal to k;

(vi) $T_k^{(r)}$, waiting time for the r-th occurrence of consecutive k successes;

(vii) L_n, length of the longest success run.

In order to make these definitions clear consider the sequence of $n = 12$ trials 110110001111. Then, for $k = 2$, $S_{12} = 8$, $N_{12,2} = 4$, $M_{12,2} = 5$, $E_{12,2} = 2$, $G_{12,2} = 3$, $T_2^{(1)} = 2$, $T_2^{(2)} = 5$, $T_2^{(3)} = 10$, and $L_{12} = 4$. The following relationships between these run statistics always hold (see [37]):

$$E_{n,k} \leq G_{n,k} \leq N_{n,k} \leq M_{n,k},$$

$$E_{n,k} = G_{n,k} - G_{n,k+1},$$

$$L_n < k \iff N_{n,k} = 0,$$

$$T_k^{(1)} \leq n \iff L_n \geq k.$$

Distribution theory of runs has been studied whenever the elements of the sequence are independent and identically distributed (i.i.d.) (see, e.g. [46], [74], [41], and [42]), independent but nonidentically distributed (i.n.i.d) (see, e.g. [1], [17], [36], and [53]) , exchangeable ([78], [28], [61], [58], [62], and [63]), partially exchangeable ([26] and [31]), homogeneous Markov dependent (see, e.g. [76], [47], [2], [3], and [57]), nonhomogeneous Markov dependent ([19]), and dependent in a form which is other than the Markovian type dependence ([85] and [18]).

In 1983, with the paper of Schwager [79] runs were started to be studied intensively in a sequence of multi-state trials as well as two-state trials. In a sequence consisting of multi-state trials the distributions of number of runs have been studied for i.i.d., i.n.i.d., homogeneous and nonhomogeneous Markov dependent sequences by Fu [35], for exchangeable sequences by Eryilmaz [22] and for partially exchangeable sequences by Inoue, Aki and Hirano [50]. Since the main interest of this thesis is runs defined in a sequence of two-state trials, for more details on runs in a sequence of multi-state trials see the above references and the papers of Han and Aki [45], Antzoulakos [6], Vaggelatou [84], and Inoue and Aki [49].

More recently, Eryilmaz [21] extended the concept of runs to the continuous-valued sequences. See also the works of Eryilmaz and Fu [29], Eryilmaz and Stepanov [30], Fan, Wang and Ding [32], and Stepanov [82] for contributions to runs in continuous-valued sequences.

The rest of the thesis is organized as follows. In Chapter 2 we present the distributions of most common and useful run statistics in independent sequences that are obtained in the literature using different techniques. Main results are given in Chapter 3 and Chapter 4. In Chapter 3 the distributions of some run statistics in Markov dependent sequences are obtained and in Chapter 4 runs are studied under a different dependence model. Numerical examples are also given to illustrate further the theoretical results in both Chapter 3 and Chapter 4.

Throughout the thesis, for integers n and m and real number x, let $\binom{n}{m}$ and $\lfloor x \rfloor$ denote the binomial coefficients and the greatest integer less than or equal to

x, respectively. We also assume for convenience that if $a > b$, then $\sum_{i=a}^{b} = 0$ and $\prod_{i=a}^{b} = 1$.

Chapter 2

Runs and Run-related Statistics in Independent Sequences

Distribution theory of runs and run-related statistics under the assumption of independent Bernoulli trials have been very popular in the literature. The trials may be either independent and identically distributed (i.i.d.) with common success probability p and failure probability $q = 1-p$ or independent but nonidentically distributed with varying success probabilities p_i and failure probabilities $q_i = 1 - p_i$, $i = 1, 2, \ldots$.

The simplest run statistic in a sequence of n trials may be S_n, the total number of successes. For i.i.d. trials the distribution of S_n is called Binomial distribution (for $n \geq 2$) and Bernoulli distribution (for $n = 1$).

$N_{n,k}$, the number of non-overlapping consecutive k successes, was introduced by Feller [33]. For i.i.d. Bernoulli trials its distribution is called *(type I) binomial distribution of order k* and its exact distribution was obtained independently by Hirano [46] and Philippou and Makri [74] as

$$P\{N_{n,k} = x\} = \sum_{i=0}^{k-1} \sum_{x_1+2x_2+\cdots+kx_k=n-i-kx} \binom{x_1+\cdots+x_k+x}{x_1,\ldots,x_k,x} p^n \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}$$

for $x = 0, 1, \ldots, \lfloor \frac{n}{k} \rfloor$.

Godbole [41] derived the following alternative formula for the distribution of $N_{n,k}$ which is computationally more efficient.

$$P\{N_{n,k} = x\} = \sum_{\lfloor \frac{n-kx}{k} \rfloor \leq y \leq n-kx} q^y p^{n-y} \binom{y+x}{x} \sum_{0 \leq j \leq \lfloor \frac{n-kx-y}{k} \rfloor} (-1)^j \binom{y+1}{j} \times \binom{n-kx-jk}{y}$$

for $x = 0, 1, \ldots, \lfloor \frac{n}{k} \rfloor$.

Fu and Koutras [36] proposed a new approach different from the combinatorial one to derive the distributions of run-related statistics. Their method is based on a finite Markov chain imbedding (*FMCI*) technique. This approach is an extension of an earlier study of Fu [34] and has also been used in the papers of Koutras and Alexandrou [53] and Lou [57], and in the book of Fu and Lou [37]. The results cover the case of i.n.i.d. trials in addition to the case of i.i.d. trials. Before giving the details of *FMCI* technique, it would be good to introduce the concepts of finite Markov chain and finite Markov chain imbeddable random variable.

Let $\Omega = \{1, 2, \ldots, m\}$ $(m < \infty)$ be a finite state space and $\{Y_t\}_{t \geq 0}$ a sequence of random variables defined on Ω. The sequence $\{Y_t\}_{t \geq 0}$ is called a *finite Markov chain* if, for any sequence $i_0, i_1, \ldots, i_t$, $t = 1, 2, \ldots$, we have

$$P\{Y_t = i_t | Y_{t-1} = i_{t-1}, \ldots, Y_0 = i_0\} = P\{Y_t = i_t | Y_{t-1} = i_{t-1}\},$$

that is, the future state depends only on the present state and it is independent of the past states. The conditional probabilities

$$P\{Y_t = j | Y_{t-1} = i\} = p_{ij}(t),$$

$i, j \in \Omega$, are called (one-step) transition probabilities at time t, the $m \times m$ matrices

$$M_t = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1m}(t) \\ p_{21}(t) & \cdots & \cdots & \cdots \\ \vdots & \ddots & \ddots & \cdots \\ p_{m1}(t) & \cdots & \cdots & p_{mm}(t) \end{bmatrix}_{m \times m},$$

$t = 1, 2, \ldots$, are called (one-step) transition probability matrices, and the probabilities at time 0, $P\{Y_0 = i\}$, $i = 1, \ldots, m$, are called the initial probabilities or the initial distribution of the Markov chain $\{Y_t\}_{t \geq 0}$.

The Markov chain $\{Y_t\}_{t \geq 0}$ is said to be homogeneous if the transition probabilities are constant in time, i.e., $P\{Y_t = j | Y_{t-1} = i\} = p_{ij}$ for any $i, j \in \Omega$, and $t = 1, 2, \ldots$. If the Markov chain is homogeneous, then its transition probability matrix is also independent of the time index t, that is,

$$M = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & \cdots & \cdots & \cdots \\ \vdots & \ddots & \ddots & \cdots \\ p_{m1} & \cdots & \cdots & p_{mm} \end{bmatrix}_{m \times m}.$$

Let $X_n(\Lambda)$ be the number of occurrences of a pattern Λ in a sequence of n Bernoulli trials, $\Gamma_n = \{0, 1, \ldots, n\}$ an index set, and $\Omega = \{a_1, a_2, \ldots, a_m\}$ a finite state space. The non-negative integer valued random variable $X_n(\Lambda)$ is said to be finite Markov chain imbeddable if:

(i) there exists a finite Markov chain $\{Y_t : t \in \Gamma_n\}$ defined on the finite state space Ω with initial probability vector ξ_0,

(ii) there exists a finite partition $\{C_x : x = 0, 1, \ldots, l_n\}$ on the state space Ω, and

(iii) for every $x = 0, 1, \ldots, l_n$, we have

$$P\{X_n(\Lambda) = x\} = P\{Y_n \in C_x | \xi_0\}.$$

Theorem 2.1 ([37]) *If $X_n(\Lambda)$ is finite Markov chain imbeddable, then*

$$P\{X_n(\Lambda) = x\} = \xi_0 \left(\prod_{t=1}^{n} M_t \right) U'(C_x),$$

where $U(C_x) = \sum_{r:a_r \in C_x} e_r$, e_r is a $1 \times m$ unit row vector corresponding to state a_r, ξ_0 is the initial probability vector, and $M_t, t = 1, \ldots, n$ are the transition probability matrices of the imbedded Markov chain.

Fu and Lou [37] defined a finite Markov chain $\{Y_t : t = 0, 1, \ldots, n\}$ on the state space $\Omega = \{(x, i) : x = 0, 1, \ldots, l_n \text{ and } i = 0, 1, \ldots, k-1\}$ by

$$Y_t = (N_{t,k}, E_t), \; t = 1, 2, \ldots, n, \tag{2.1}$$

where $l_n = \left\lfloor \frac{n}{k} \right\rfloor$, $N_{t,k}$ is the number of non-overlapping consecutive k successes in the first t trials, and $E_t = m \mod k$, where m, $0 \leq m < k$ represents the number of trailing successes that exist in the sequence after the first t trials. For example, in the sequence of $n = 10$ trials 0100111011, for $k = 2$, the realization of the Markov chain $\{Y_t : t = 1, 2, \ldots, 10\}$ is $\{Y_1 = (0,0), Y_2 = (0,1), Y_3 = (0,0),$ $Y_4 = (0,0)$, $Y_5 = (0,1)$, $Y_6 = (1,0)$, $Y_7 = (1,1)$, $Y_8 = (1,0)$, $Y_9 = (1,1)$, $Y_{10} = (2,0)\}$ and it is unique. Define

$$C_x = \{(x, i) : i = 0, 1, \ldots, k-1\}$$

for $x = 0, 1, \ldots, l_n$. The collection of subsets $\{C_x : x = 0, 1, \ldots, l_n\}$ forms a partition of the state space Ω.

From the definition of the imbedded Markov chain given in (2.1), the one-step transition probabilities in $M_t(N_{n,k})$ for i.n.i.d. trials are given by

$$p_{(x,i)(y,j)}(t) \;=\; P\{Y_t = (y, j) \,|Y_{t-1} = (x, i)\}$$

$$= \begin{cases} q_t & \text{if } y = x \text{ and } j = 0 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ & i = 0, 1, \ldots, k-1, \\ p_t & \text{if } y = x \text{ and } j = i + 1 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ & i = 0, 1, \ldots, k-2, \\ p_t & \text{if } y = x + 1 \text{ and } j = 0 \text{ for } i = k-1 \text{ and} \\ & x = 0, 1, \ldots, l_n - 1, \\ 1 & \text{if } y = x = l_n \text{ and } j = i = k-1, \\ 0 & \text{otherwise} \end{cases}$$

for $t = 1, 2, \ldots, n$. In general, $M_t\left(N_{n,k}\right)$ are matrices of the form

$$M_t\left(N_{n,k}\right) = \begin{bmatrix} A_t & B_t & 0 & \cdots & 0 \\ 0 & A_t & B_t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & A_t & B_t \\ 0 & 0 & \cdots & 0 & A_t^* \end{bmatrix}_{d \times d}$$

for $t = 1, 2, \ldots, n$, where

$$A_t = \begin{bmatrix} q_t & p_t & 0 & \cdots & 0 \\ q_t & 0 & p_t & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \cdots & \ddots & p_t \\ q_t & 0 & 0 & \cdots & 0 \end{bmatrix}_{k \times k}, \tag{2.2}$$

B_t is a $k \times k$ matrix having p_t at the entry $(k, 1)$ and zero elsewhere,

$$A_t^* = \begin{bmatrix} q_t & p_t & 0 & \cdots & 0 \\ q_t & 0 & p_t & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ q_t & 0 & \cdots & 0 & p_t \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{k \times k},$$

and $d = k\left(l_n + 1\right)$. Hence by Theorem 2.1,

$$P\left\{N_{n,k} = x\right\} = \xi_0 \left(\prod_{t=1}^{n} M_t\left(N_{n,k}\right)\right) U'\left(C_x\right)$$

for $x = 0, 1, \ldots, l_n$, where $\xi_0 = (1, 0, \ldots, 0)_{1\times d}$ and $U'\left(C_x\right)$ is the transpose of the vector $U\left(C_x\right) = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)$ with ones at the locations associated with the states in C_x.

$M_{n,k}$, the number of overlapping consecutive k successes, was introduced by Ling ([56]). For i.i.d. Bernoulli trials its distribution is called *type II binomial distribution of order* k and Ling obtained two formulae one of which is recursive. These are

$$P\left\{M_{n,k} = x\right\} = \begin{cases} p^n & \text{if } x = n-k+1, \\ 2p^{n-1}q & \text{if } x = n-k\,(>0), \\ \sum_{j=1}^{x+k} p^{j-1} q P\left\{M_{n-j,k} = x - \max\left(0, j-k\right)\right\} & \text{if } 0 \le x < n-k \end{cases}$$

and

$$P\left\{M_{n,k} = x\right\} = \sum_{i=0}^{n} \sum_{\substack{x_1 + 2x_2 + \cdots + nx_n + i = n \\ \max(0, i-k+1) + \sum_{j=k+1}^{n} (j-k)x_j = x}} \binom{x_1 + x_2 + \cdots + x_n}{x_1, x_2, \ldots, x_n} p^n \left(\frac{q}{p}\right)^{\sum_{i=1}^{n} x_i}.$$

Godbole [42] derived a simpler formula for the distribution of $M_{n,k}$ in the case of i.i.d. trials

$$P\{M_{n,k}=x\}=\begin{cases} p^n & \text{if } x=n-k+1,\\ \sum\limits_{y=\lfloor \frac{n}{k}\rfloor}^{n} q^y p^{n-y} \sum\limits_{j=0}^{\lfloor \frac{n-y}{k}\rfloor} (-1)^j \binom{y+1}{j}\binom{n-jk}{y} & \text{if } x=0,\\ \sum\limits_{y} q^y p^{n-y} \sum\limits_{v} \binom{y}{v} & \\ \quad \times \Bigg\{ \sum\limits_{j} (-1)^j \binom{y-v}{j}\binom{x-j(n-k)-1}{y-v-1} & \\ \qquad \times \sum\limits_{m} (-1)^m \binom{v+1}{m}\binom{n-x-k(y-v+m)}{v} & \text{if } 1\le x\le n-k.\\ \qquad + \sum\limits_{j} (-1)^j \binom{y-v+1}{j}\binom{x-j(n-k)-1}{y-v} & \\ \qquad \times \sum\limits_{m} (-1)^m \binom{v}{m}\binom{n-x-k(y-v+1+m)}{v-1} \Bigg\} & \end{cases}$$

Chryssaphinou, Papastavridis and Tsapelas [17] obtained the distribution of $M_{n,k}$ in the case of i.n.i.d. trials as

$$P\{M_{n,k}=x\}=\begin{cases} \prod\limits_{i=1}^{n} p_i & \text{if } x=n-k+1,\\ (q_1p_n+q_np_1)\prod\limits_{i=2}^{n-1} p_i & \text{if } x=n-k,\\ \sum\limits_{i=0}^{x+k-1} \left(\prod\limits_{m=n-i+1}^{n+1} p_m\right) q_{n-1} & \text{if } 0\le x< n-k.\\ \quad \times P\{M_{n-i-1,k}=x-\max(0,i-k+1)\} & \end{cases}$$

Fu and Lou ([37]) obtained the distribution of $M_{n,k}$ using the imbedded Markov chain

$$Y_t=(M_{t,k},E_t)\,,\ t=1,2,\ldots,n,$$

on the state space

$$\Omega=\{(x,i): x=0,1,\ldots,l_n-1 \text{ and } i=\gamma,0,1,\ldots,k-1\}\bigcup\{(l_n,\gamma)\}-\{(0,\gamma)\}\,,$$

where $l_n=n-k+1$, $M_{t,k}$ is the number of overlapping consecutive k successes in the first t trials, and

$$E_t = \begin{cases} m & \text{if } m = 0, 1, \ldots, k-1, \\ \gamma & \text{if } m \geq k \end{cases}$$

is the ending block variable keeping track of the number of trailing successes. For example, in the sequence of $n = 12$ trials 010011011101, for $k = 2$, the realization of the Markov chain $\{Y_t : t = 1, 2, \ldots, 12\}$ is $\{Y_1 = (0,0), Y_2 = (0,1), Y_3 = (0,0), Y_4 = (0,0), Y_5 = (0,1), Y_6 = (1,\gamma), Y_7 = (1,0), Y_8 = (1,1), Y_9 = (2,\gamma), Y_{10} = (3,\gamma), Y_{11} = (3,0), Y_{12} = (3,1)\}$. They constructed the following partition for Ω:

$$\begin{aligned} C_0 &= \{(0,i) : i = 0, 1, \ldots, k-1\}, \\ C_x &= \{(x,i) : i = \gamma, 0, 1, \ldots, k-1\} \text{ for } x = 1, 2, \ldots, l_n - 1, \\ C_{l_n} &= \{(l_n, \gamma)\}. \end{aligned}$$

The one-step transition probabilities in $M_t(M_{n,k})$ for i.n.i.d. trials are given by

$$p_{(x,i)(y,j)}(t) = \begin{cases} q_t & \begin{array}{l} \text{if } y = x \text{ and } j = 0 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ i = \gamma, 0, 1, \ldots, k-1, \end{array} \\ p_t & \begin{array}{l} \text{if } y = x \text{ and } j = i+1 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ i = 0, 1, \ldots, k-2, \end{array} \\ p_t & \text{if } y = x+1, j = \gamma, \text{ and } i = k-1 \text{ for } x = 0, 1, \ldots, l_n - 1, \\ p_t & \text{if } y = x+1 \text{ and } j = i = \gamma \text{ for } x = 0, 1, \ldots, l_n - 1, \\ 1 & \text{if } y = x = l_n \text{ and } j = i = \gamma, \\ 0 & \text{otherwise} \end{cases}$$

for $t = 1, 2, \ldots, n$. In general, $M_t(M_{n,k})$ are matrices of the form

$$M_t\left(M_{n,k}\right) = \begin{bmatrix} A_t & p_t e_k' & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & q_t e_1 & p_t & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & A_t & p_t e_k' & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & q_t e_1 & p_t & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & q_t e_1 & p_t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & A_t & p_t e_k' \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix}_{d\times d}$$

for $t = 1, 2, \ldots, n$, where A_t is given in (2.2), $e_1 = (1, 0, \ldots, 0)$ and $e_k = (0, \ldots, 0, 1)$ are $1 \times k$ unit row vectors, and $d = l_n(k+1)$.

Fu and Lou ([37]) also obtained the distributions of $G_{n,k}$ and $E_{n,k}$. In the following, we will give the imbedded Markov chains used, the one-step transition probabilities in $M_t(G_{n,k})$ and $M_t(E_{n,k})$, and the form of the matrices $M_t(G_{n,k})$.

The imbedded Markov chain used for $G_{n,k}$ on the state space

$$\Omega = \{(x,i) : x = 0, 1, \ldots, l_n \text{ and } i = \gamma, 0, 1, \ldots, k-1\} - \{(0,\gamma)\}$$

is

$$Y_t = (G_{t,k}, E_t),\ t = 1, 2, \ldots, n,$$

where $l_n = \left\lfloor \frac{n+1}{k+1} \right\rfloor$, $G_{t,k}$ is the number of success runs of length greater than or equal to k in the first t trials, and

$$E_t = \begin{cases} m & \text{if } m = 0, 1, \ldots, k-1, \\ \gamma & \text{if } m \geq k \end{cases}$$

is the ending block variable. For example, in the sequence of $n = 12$ trials 010011011101, for $k = 2$, the realization of the Markov chain $\{Y_t : t = 1, 2, \ldots, 12\}$ is $\{Y_1 = (0,0),\ Y_2 = (0,1),\ Y_3 = (0,0),\ Y_4 = (0,0),\ Y_5 = (0,1),\ Y_6 = (1,\gamma),$

$Y_7 = (1,0)$, $Y_8 = (1,1)$, $Y_9 = (2,\gamma)$, $Y_{10} = (2,\gamma)$, $Y_{11} = (2,0)$, $Y_{12} = (2,1)\}$. They constructed the following partition for Ω:

$$\begin{aligned} C_0 &= \{(0,i) : i = 0,1,\ldots,k-1\}, \\ C_x &= \{(x,i) : i = \gamma,0,1,\ldots,k-1\} \text{ for } x = 1,2,\ldots,l_n. \end{aligned}$$

The one-step transition probabilities in $M_t(G_{n,k})$ for i.n.i.d. trials are given by

$$p_{(x,i)(y,j)}(t) = \begin{cases} q_t & \begin{array}{l}\text{if } y = x \text{ and } j = 0 \text{ for } x = 0,1,\ldots,l_n \text{ and} \\ i = \gamma,0,1,\ldots,k-1,\end{array} \\ p_t & \text{if } y = x \text{ and } j = i = \gamma \text{ for } x = 0,1,\ldots,l_n, \\ p_t & \begin{array}{l}\text{if } y = x \text{ and } j = i+1 \text{ for } x = 0,1,\ldots,l_n \text{ and} \\ i = \gamma,0,1,\ldots,k-2,\end{array} \\ p_t & \text{if } y = x+1,\ i = k-1, \text{ and } j = \gamma \text{ for } x = 0,1,\ldots,l_n-1, \\ 1 & \text{if } y = x = l_n \text{ and } j = i = k-1, \\ 0 & \text{otherwise} \end{cases}$$

for $t = 1,2,\ldots,n$. In general, $M_t(G_{n,k})$ are matrices of the form

$$M_t(G_{n,k}) = \begin{bmatrix} A_t & p_t e_k' & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & p_t & q_t e_1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & A_t & p_t e_k' & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & p_t & q_t e_1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & p_t & q_t e_1 \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & A_t^* \end{bmatrix}_{d\times d}$$

for $t = 1,2,\ldots,n$, where A_t is given in (2.2), $e_1 = (1,0,\ldots,0)$ and $e_k = (0,\ldots,0,1)$ are $1\times k$ unit row vectors, and $d = (l_n+1)(k+1)-1$.

The imbedded Markov chain used for $E_{n,k}$ on the state space

$$\Omega = \{(x,i) : x = 0,1,\ldots,l_n \text{ and } i = \beta,\gamma,0,1,\ldots,k-1\} - \{(0,\gamma)\}$$

is

$$Y_t = (E_{t,k}, E_t)\,,\ t = 1, 2, \ldots, n,$$

where $l_n = \left\lfloor \frac{n+1}{k+1} \right\rfloor$, $E_{t,k}$ is the number of success runs of length exactly k in the first t trials, and

$$E_t = \begin{cases} m & \text{if } m = 0, 1, \ldots, k-1, \\ \gamma & \text{if } m = k, \\ \beta & \text{if } m > k \end{cases}$$

is the ending block variable with two ending block states

(i) waiting state (x, γ), $x = 1, 2, \ldots, l_n$:
$Y_t = (x, \gamma)$ means that $m = k$ and that the x-th success run of size k has occurred at the t-th trial, and

(ii) overflow state (x, β), $x = 1, 2, \ldots, l_n$:
$Y_t = (x, \beta)$ means that $m > k$ and that exactly x success runs of size k have appeared prior to the last $m + 1$ trials.

They constructed the following partition for Ω:

$$\begin{aligned} C_0 &= \{(0, i) : i = \beta, 0, 1, \ldots, k-1\}\,, \\ C_x &= \{(x, i) : i = \gamma, \beta, 0, 1, \ldots, k-1\} \text{ for } x = 1, 2, \ldots, l_n. \end{aligned}$$

The one-step transition probabilities in $M_t\,(E_{n,k})$ for i.n.i.d. trials are given by

$$p_{(x,i)(y,j)}\,(t) = \begin{cases} q_t & \begin{array}{l}\text{if } y = x \text{ and } j = 0 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ i = \gamma, \beta, 0, 1, \ldots, k-1,\end{array} \\ p_t & \begin{array}{l}\text{if } y = x \text{ and } j = i+1 \text{ for } x = 0, 1, \ldots, l_n \text{ and} \\ i = 0, 1, \ldots, k-2,\end{array} \\ p_t & \text{if } y = x+1,\ j = \gamma, \text{ and } i = k-1 \text{ for } x = 0, 1, \ldots, l_n - 1, \\ p_t & \text{if } y = x-1,\ j = \beta, \text{ and } i = \gamma \text{ for } x = 0, 1, \ldots, l_n, \\ p_t & \text{if } y = x \text{ and } j = i = \beta \text{ for } x = 0, 1, \ldots, l_n, \\ 1 & \text{if } y = x = l_n \text{ and } j = i = k-1, \\ 0 & \text{otherwise} \end{cases}$$

for $t = 1, 2, \ldots, n$.

The distribution of L_n, the length of the longest success run, for i.i.d. Bernoulli trials has also attracted the interest of many authors because of its wide range of applications. The most common application of the length of the longest success run is the reliability studies of consecutive k-out-of-n:F system. A consecutive k-out-of-n:F system consists of n components and fails if and only if at least k consecutive components fail. The reliability of the system is then defined by

$$R_{n,k} = P\left\{L_n < k\right\}.$$

It should be noted that, here, we consider the failure of the i-th component as a success and functioning of the i-th component as a failure so that L_n denotes the length of the longest run of failed components.

Burr and Cane [13], Lambiris and Papastavridis [55], and Hwang [48] obtained the distribution of L_n for the case of i.i.d. trials as

$$P\left\{L_n < k\right\} = \sum_{m=0}^{\left\lfloor \frac{n+1}{k+1} \right\rfloor} (-1)^m p^{mk} q^{m-1} \left(\binom{n-mk}{m-1} + q \binom{n-mk}{m} \right).$$

Schilling [77] derived the distribution of L_n in a different way. Let $C_n^{(x)}(k)$ be the number of sequences of length n in which exactly x successes occur, but no more than k of these successes occur consecutively. Then

$$P\left\{L_n \le k\right\} = \begin{cases} q^n & \text{if } k = 0, \\ \sum_{x=0}^{n} C_n^{(x)}(k)\, p^x q^{n-x} & \text{if } 1 \le k \le n, \end{cases}$$

where

$$C_n^{(x)}(k) = \begin{cases} \sum_{j=0}^{k} C_{n-1-j}^{(x-j)}(k) & \text{if } k < x < n, \\ \binom{n}{x} & \text{if } x \le k \le n, \\ 0 & \text{if } k < x = n. \end{cases}$$

The distribution of L_n for the case of i.n.i.d. trials has been obtained by Fu and Lou [37] in the following theorem.

Theorem 2.2 *For* $0 \leq k \leq n$,

$$P\{L_n \leq k\} = \xi \left(\prod_{i=1}^{n} N_t \right) 1'_{1\times(k+1)},$$

where $\xi = (1, 0, \ldots, 0)$ *is a* $1 \times (k+1)$ *unit row vector, and* N_t *is, as indicated below, the* $(k+1)\times(k+1)$ *essential submatrix of the following transition probability matrix:*

$$M_t = \begin{array}{c} 0 \\ 1 \\ \vdots \\ \vdots \\ k \\ \alpha \end{array} \begin{bmatrix} q_t & p_t & 0 & \cdots & 0 & 0 \\ q_t & 0 & p_t & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ q_t & 0 & \cdots & \cdots & 0 & p_t \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}_{(k+2)\times(k+2)} = \begin{bmatrix} N_t & C_t \\ 0 & 1 \end{bmatrix}.$$

Fu and Lou [37] also obtained a recursive formula for $1 \leq k < n$,

$$P\{L_n \leq k\} = q_n P\{L_{n-1} \leq k\} + \sum_{i=1}^{k} q_{n-i} \prod_{j=n-i+1}^{n} p_j P\{L_{n-i-1} \leq k\}$$

with $P\{L_n = 0\} = \prod_{j=1}^{n} q_j$ and $P\{L_n \leq n\} \equiv 1$ for $k = n$. For the length of the longest success run and its applications, we also refer to the works of Philippou and Makri [73][74], Muselli [70], and Makri, Philippou and Psillakis [60].

Another important run statistic is $T_k^{(r)}$, the waiting time for the r-th appearance of a success run of length k. For i.i.d. Bernoulli trials the distributions for $r = 1$ and $r \geq 2$ were first studied respectively by Philippou and Muwafi [75] and Philippou, Georghiou and Philippou [72]. Their distributions are called *geometric distribution of order* k and *negative binomial distribution of order* k, respectively. For an extensive review on distributions of order k, see, for instance [51], [4], and [71].

Philippou and Muwafi [75] showed that

$$P\left\{T_k^{(1)} = x\right\} = \sum_{x_1,\ldots,x_k} \binom{x_1+\cdots+x_k}{x_1,\ldots,x_k} p^x \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}$$

for $x \geq k$, where the summation is over all nonnegative integers $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x - k$.

Uppuluri and Patil [83] derived another formula which involves single summations and binomial coefficients. For $x \geq k$,

$$\begin{aligned} P\left\{T_k^{(1)} = x\right\} &= p^k \sum_{j=0}^{\infty} (-1)^j \binom{x-k-jk}{j} \left(qp^k\right)^j \\ &- p^{k+1} \sum_{j=0}^{\infty} (-1)^j \binom{x-k-jk-1}{j} \left(qp^k\right)^j . \end{aligned}$$

Muselli [69] obtained a simpler formula that contains a single summation

$$P\left\{T_k^{(1)} = x\right\} = \sum_{j=1}^{\left\lfloor \frac{x+1}{k+1} \right\rfloor} (-1)^{j-1} p^{jk} q^{j-1} \left[\binom{x-jk-1}{j-2} - q\binom{x-jk-1}{j-1}\right].$$

Balakrishnan and Koutras [11] obtained an exact formula for the distribution of $T_k^{(1)}$ for the case of i.i.d. trials using the finite Markov chain imbedding technique. Let $\Omega = \{0, 1, \ldots, k\}$ be a finite state space and assume that the chain enters state $i \in \Omega - \{k\}$ at trial t if i consecutive successes have been observed at trials $t - i + 1, \ldots, t$. Then, for $x \geq 1$,

$$P\left\{T_k^{(1)} = x\right\} = pe_1\Lambda^{x-1}e_k',$$

where

$$\Lambda = \begin{bmatrix} q & p & 0 & \cdots & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ q & 0 & \cdots & \cdots & 0 & p \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}_{(k+1)\times(k+1)},$$

$e_1 = (1, 0, \ldots, 0)$, and $e_k = (0, \ldots, 0, 1, 0)$ are $1 \times k + 1$ unit row vectors.

Balakrishnan, Balasubramanian and Viveros [10] obtained the distribution of $T_k^{(1)}$ for the case of i.n.i.d. trials by letting $f(x) = P\left\{T_k^{(1)} = x\right\}$ as

$$f(x) = \begin{cases} 0 & \text{if } x = 0, 1, \ldots, k-1, \\ p_0^k & \text{if } x = k, \\ q_0 p_1^k & \text{if } x = k+1, \\ (1+p_0) f(k+1) - q_0 p_1^{k+1} & \text{if } x = k+2, \\ (1+p_0) f(x-1) - p_0 f(x-2) & \text{if } x = k+3, \ldots, 2k, \\ \begin{array}{l} (1+p_0) f(2k) - p_0 f(2k-1) \\ -q_0 p_0^k p_1^k \end{array} & \text{if } x = 2k+1, \\ \begin{array}{l} (1+p_0) f(2k+1) - p_0 f(2k) \\ -q_1 p_1^k f(k+1) + q_0 p_0^k p_1^{k+1} \end{array} & \text{if } x = 2k+2, \\ \begin{array}{l} (1+p_0) f(x-1) - p_0 f(x-2) \\ -q_1 p_1^k f(x-k-1) \\ +p_0 q_1 p_1^k f(x-k-2) \end{array} & \text{if } x = 2k+3, 2k+4, \ldots. \end{cases}$$

Chapter 3

Runs and Run-related Statistics in Markovian Sequences

In distribution theory of runs and run-related statistics under the assumption of dependent trials, most of the results are based on Markovian type dependence because this type of dependence has been found useful and flexible for modeling stochastic events appearing in many areas of science. Some of these areas in natural sciences are biology and physics and in social sciences are economics and psychology.

Runs of Markovian sequences have been first considered in the paper of Rajarshi [76]. Then, almost 20 years later, Hirano and Aki [47] obtained the distribution of number of success runs of length k (under the counting schemes *overlapping* and *at least*) and Aki and Hirano [2] obtained the distributions of numbers of failures and successes until the first occurrence of consecutive k successes in n trials. Joint distributions of the numbers of failures, successes, and success runs of length less than k until the first occurrence of consecutive k successes have been studied by Aki and Hirano [3] and Lou [57]. Waiting time for the r-th occurrence of a success run of length k ([66], [52], and [5]) and exact distribution and bounds for the distribution of the longest run statistic ([57], [38], and [20]) have also been studied. Lou [57] used the finite Markov chain imbedding technique, introduced in Chapter 2, to obtain the conditional distributions of the two most commonly used run statistics R_n, the number of success runs, and L_n, the length of the longest success run, given the number of successes in n trials. She derived the critical regions under the null hypothesis $H_0 =$ the trials are independent and identically distributed and the powers under the alternative hypothesis $H_A =$ the trials are one-step homogeneous Markov dependent.

More recently, Antzoulakos and Chadjiconstantinidis [7] established some formulae for the probability generating function, probability mass function, and moments of the number of success runs of length k (under the counting schemes *non-overlapping*, *overlapping* and *at least*) in n trials and Eryilmaz [23] considered the mean success run length which is the arithmetic mean of the lengths of the success runs in n trials.

Throughout this chapter, runs and run-related statistics under the assumption of Markov dependent trials is our main interest. In Section 3.1, after presenting some important results from literature, we derive the distributions of the extreme distances between failures whenever the trials are Markov dependent. Then, in Section 3.2, we introduce a new run statistic and obtain its distribution whenever the trials are independent and identically distributed and Markov dependent. Applications such as system reliability and waiting time between extreme events and numerical results are given in Section 3.3.

3.1 Extreme distances between failures

In sequences consisting of independent and identically distributed (i.i.d.) and exchangeable n binary trials, Makri [59] studied the minimum and maximum numbers of successes between two successive failures and called them minimum distance and maximum distance, respectively. As noted in [59] the corresponding run statistics are potentially useful in various areas including hypothesis testing and system reliability. In fact, these extremes are quite useful to elicit information about the minimum and maximum duration between two successive extreme events, e.g. two extreme floods or rainfalls. These statistics could also be useful to test if there is a clustering among the elements of a sequence. The need for this kind of test arises in many fields involving DNA sequence matching and animal learning studies.

Let $\{X_i\}_{i\geq 1}$ be a sequence of trials with two possible outcomes either a success ("1") or a failure ("0") which include among them at least two failures. For $n \geq 2$, denote by $X_n^{(1)}$ and $X_n^{(n)}$ the minimum distance between successive failures in the first n trials of $\{X_i\}_{i\geq 1}$ and the maximum distance between successive failures in the first n trials of $\{X_i\}_{i\geq 1}$, respectively. Let $n = 15$ and the trials be 110111001011001. Then $X_{15}^{(1)} = 0$ and $X_{15}^{(15)} = 3$. Let us first give the result of Makri, Philippou and Psillakis [61] which will be useful in our developments.

Lemma 3.1 ([61]) *The number of allocations of α indistinguishable balls into r distinguishable cells, in such a way that each of m $(0 \leq m \leq r)$ specified cells is occupied by at most k balls, is given by*

$$H_m(\alpha, r, k) = \sum_{j=0}^{\left\lfloor \frac{\alpha}{k+1} \right\rfloor} (-1)^j \binom{m}{j} \binom{\alpha - (k+1)j + r - 1}{\alpha - (k+1)j},$$

for $\alpha \geq 0$, $r > 0$ and $H_m(\alpha, r, k) = 0$, otherwise.

It should be noted that $H_m(\alpha, r, k)$ coincides with the number of integer solutions to the equation $z_1 + \cdots + z_r = \alpha$ such that $0 \leq z_1 \leq k, \ldots, 0 \leq z_m \leq k, z_{m+1} \geq 0, \ldots, z_r \geq 0$.

Makri [59] derived the exact probability mass functions of $X_n^{(1)}$ and $X_n^{(n)}$ when $\{X_i\}_{i\geq 1}$ consists of i.i.d. binary trials with success probability $p = P\{X_i = 1\}$ and failure probability $q = 1 - p = P\{X_i = 0\}$.

Theorem 3.2 ([59]) *For $k = 0, 1, \ldots, n-2$, it holds*

$$P\left\{X_n^{(1)} = k\right\} = \frac{\sum_{y=2}^{\left\lfloor \frac{n+k}{k+1} \right\rfloor} p^{n-y} q^y \left[\binom{n-(y-1)k}{y} - \binom{n-(y-1)(k+1)}{y}\right]}{1 - p^n - nqp^{n-1}}.$$

Theorem 3.3 ([59]) *It holds*

(a) $P\left\{X_n^{(n)} = 0\right\} = \frac{\sum_{y=2}^{n} (n+1-y) p^{n-y} q^y}{1-p^n-nqp^{n-1}}$;

(b) for $k = 1, 2, \ldots, n-2$,

$$P\left\{X_n^{(n)} = k\right\} = \frac{\sum_{y=2}^{n-k} p^{n-y} q^y \sum_{i=1}^{y_1} \binom{y-1}{i} H_{y-i-1}\left(n-y-ik, y-i+1, k-1\right)}{1 - p^n - nqp^{n-1}},$$

where $y_1 = \min\left(y-1, \left\lfloor \frac{n-y}{k} \right\rfloor\right)$.

Makri [59] provided expressions for the mean values of $X_n^{(1)}$ and $X_n^{(n)}$ as

$$E\left(X_n^{(1)}\right) = (n-2) - \sum_{k=0}^{n-3} F_{X_n^{(1)}}(k) \quad \text{and} \quad E\left(X_n^{(n)}\right) = (n-2) - \sum_{l=0}^{n-3} F_{X_n^{(n)}}(l),$$

where

$$F_{X_n^{(1)}}(k) = 1 - \frac{\sum_{y=2}^{\left\lfloor \frac{n+k+1}{k+2} \right\rfloor} p^{n-y} q^y \binom{n-(y-1)(k+1)}{y}}{1 - p^n - nqp^{n-1}}$$

and

$$F_{X_n^{(n)}}(l) = \frac{\sum_{y=2}^{n} p^{n-y} q^y \sum_{j=0}^{\left\lfloor \frac{n-y}{l+1} \right\rfloor} (-1)^j \binom{y-1}{i} \binom{n-(l+1)j}{n-y-(l+1)j}}{1 - p^n - nqp^{n-1}}.$$

Makri [59] also obtained the exact joint probability mass function and joint cumulative distribution of $X_n^{(1)}$ and $X_n^{(n)}$ (see [59]).

3.1.1 Extreme distances in Markov dependent trials

Let $\{X_i\}_{i\geq 1}$ be a time-homogeneous Markov chain with transition probability matrix

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

and initial probabilities $p_0 = P\{X_1 = 0\}$ and $p_1 = P\{X_1 = 1\} = 1 - p_0$. Denote by S_n, R_n and θ_i, $i = 1, 2, \ldots, R_n$ the number of successes, the number of success runs and the length of the i-th success run in $X_1, X_2, \ldots, X_n$, respectively. Assume that Z_n is a success run statistic defined on $X_1, X_2, \ldots, X_n$, that is,

$$Z_n = \phi(\theta_1, \ldots, \theta_{R_n}).$$

Throughout this chapter, we will use the following result of Eryilmaz [23].

Theorem 3.4 ([23]) *Let Z_n be a success run statistic based on a sequence of n Markov-dependent Bernoulli trials with transition probability matrix P and initial probabilities p_0 and p_1. Then*

$$P\{Z_n \in B\} = \sum_r \sum_l |I_\phi(B)|\, g(n, r, l), \tag{3.1}$$

where $I_\phi(B) = \{(i_1, \ldots, i_r) : i_1 + \cdots + i_r = l; \phi(i_1, \ldots, i_r) \in B\}$, B is a Borel set, $|A|$ denotes the cardinality of the set A,

$$g(n, r, l) = \sum_{t=0}^{1} \sum_{s=0}^{1} \binom{n-l-1}{r-t-s} p_{11}^{l-r} p_{01}^{r-t} p_{10}^{r-s} p_{00}^{n-l-r+t+s-1} p_t,$$

and $g(n, 1, n) = p_1 p_{11}^{n-1}$.

As it is seen from Theorem 3.4, it is enough to compute the cardinality of the set $I_\phi(B)$ to derive the distribution of any success run statistic defined in a sequence of Markov-dependent Bernoulli trials. Hence finding the distribution of any success run statistic Z_n in a sequence of Markov-dependent Bernoulli trials is a combinatorial problem which is, specifically, the determination of the total number of integer solutions to the equation $i_1 + \cdots + i_r = l$ such that

$\phi(i_1, \ldots, i_r) \in B$. As a motivational example let us obtain the distribution of L_n in a sequence of Markov dependent trials. Since L_n stands for the longest success run,

$$L_n = \phi(\theta_1, \ldots, \theta_{R_n}) = \max(\theta_1, \ldots, \theta_{R_n}).$$

Then

$$\begin{aligned}
P\{L_n < k\} &= P\left\{\max_{1 \le i \le R_n} \theta_i < k\right\} \\
&= \sum_r P\left\{\max_{1 \le i \le R_n} \theta_i < k \middle| R_n = r\right\} P\{R_n = r\} \\
&= \sum_r P\left\{\max_{1 \le i \le r} \theta_i < k, R_n = r\right\} \\
&= \sum_r \sum_l P\left\{\max_{1 \le i \le r} \theta_i < k, R_n = r, S_n = l\right\} \\
&= \sum_r \sum_l \sum_{\substack{i_1+\cdots+i_r=l \\ 0<i_j<k,\ j=1,\ldots,r}} \cdots \sum P\{\theta_1 = i_1, \ldots, \theta_r = i_r\}
\end{aligned}$$

The occurrence of the event $P\left\{\max\limits_{1 \le i \le r} \theta_i < k, R_n = r, S_n = l\right\}$ has four possible forms:

$$(1)\ \overbrace{11\ldots1}^{0<i_1<k}\ \underbrace{0\ldots0}_{y_1>0}\overbrace{11\ldots1}^{0<i_2<k}\underbrace{0\ldots0}_{y_2>0}\ldots\ \overbrace{11\ldots1}^{0<i_{r-1}<k}\ \underbrace{0\ldots0}_{y_{r-1}>0}\ \overbrace{11\ldots1}^{0<i_r<k}$$

$$(2)\ \overbrace{11\ldots1}^{0<i_1<k}\ \underbrace{0\ldots0}_{y_1>0}\overbrace{11\ldots1}^{0<i_2<k}\underbrace{0\ldots0}_{y_2>0}\ldots\ \overbrace{11\ldots1}^{0<i_{r-1}<k}\ \underbrace{0\ldots0}_{y_{r-1}>0}\ \overbrace{11\ldots1}^{0<i_r<k}\underbrace{0\ldots0}_{y_r>0}$$

$$(3)\ \underbrace{0\ldots0}_{y_1>0}\overbrace{11\ldots1}^{0<i_1<k}\underbrace{0\ldots0}_{y_2>0}\overbrace{11\ldots1}^{0<i_2<k}\ldots\ \overbrace{11\ldots1}^{0<i_{r-1}<k}\ \underbrace{0\ldots0}_{y_r>0}\ \overbrace{11\ldots1}^{0<i_r<k}$$

$$(4)\ \underbrace{0\ldots0}_{y_1>0}\overbrace{11\ldots1}^{0<i_1<k}\underbrace{0\ldots0}_{y_2>0}\overbrace{11\ldots1}^{0<i_2<k}\ldots\ \overbrace{11\ldots1}^{0<i_{r-1}<k}\ \underbrace{0\ldots0}_{y_r>0}\ \overbrace{11\ldots1}^{0<i_r<k}\underbrace{0\ldots0}_{y_{r+1}>0}\ .$$

Using Lemma 3.1, we obtain the total number of integer solutions to the equation

$$\begin{aligned} & i_1 + \cdots + i_r = l \\ & \text{s.t.} \quad 0 < i_j < k,\ j = 1, \ldots, r \end{aligned}$$

as

$$H_r\left(l-r, r, k-2\right) = \sum_{j=0}^{\left\lfloor \frac{l-r}{k-1} \right\rfloor} \left(-1\right)^j \binom{r}{j} \binom{l-\left(k-1\right)j+r-1}{r-1}.$$

The number of arrangements of the form (1) is $\binom{n-l-1}{r-2} H_r\left(l-r, r, k-2\right)$ for $2 \leq r \leq \left\lfloor \frac{n+1}{2} \right\rfloor$ and $r \leq l \leq n-1$ which is the total number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_{r-1} = n-l \\ & \text{s.t.} \quad y_j > 0,\ j = 1, \ldots, r-1 \end{aligned}$$

multiplied by $H_r\left(l-r, r, k-2\right)$ and each sequence in the form of (1) has the probability of occurrence $p_1 p_{11}^{l-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{n-l-r+1}$.

The number of arrangements of the forms (2) and (3) are $\binom{n-l-1}{r-1} H_r\left(l-r, r, k-2\right)$ for $1 \leq r \leq \left\lfloor \frac{n}{2} \right\rfloor$ and $r \leq l \leq n-1$ which is the total number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_r = n-l \\ & \text{s.t.} \quad y_j > 0,\ j = 1, \ldots, r \end{aligned}$$

multiplied by $H_r\left(l-r, r, k-2\right)$. Each sequence in the form of (2) has the probability of occurrence $p_1 p_{11}^{l-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{n-l-r}$ and each sequence in the form of (3) has the probability of occurrence $p_0 p_{11}^{l-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{n-l-r}$.

The number of arrangements of the form (4) is $\binom{n-l-1}{r} H_r\left(l-r, r, k-2\right)$ for $1 \leq r \leq \left\lfloor \frac{n}{2} \right\rfloor$ and $r \leq l \leq n-2$ which is the total number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_{r+1} = n-l \\ & \text{s.t.} \quad y_j > 0,\ j = 1, \ldots, r+1 \end{aligned}$$

multiplied by $H_r\left(l-r, r, k-2\right)$ and each sequence in the form of (4) has the probability of occurrence $p_0 p_{11}^{l-r} p_{10}^{r} p_{01}^{r} p_{00}^{n-l-r-1}$.

Hence we obtain

$$
\begin{aligned}
P\left\{L_n<k\right\} = & \sum_{r=2}^{\left\lfloor\frac{n+1}{2}\right\rfloor} \sum_{l=r}^{\min(n-1,n-r+1)} \binom{n-l-1}{r-2} H_r\left(l-r,r,k-2\right) \\
& \times p_1 p_{11}^{l-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{n-l-r+1} \\
& +\sum_{r=1}^{\left\lfloor\frac{n}{2}\right\rfloor} \sum_{l=r}^{\min(n-1,n-r)} \binom{n-l-1}{r-1} H_r\left(l-r,r,k-2\right) \\
& \times p_{11}^{l-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{n-l-r}\left(p_1 p_{10}+p_0 p_{01}\right) \\
& +\sum_{r=1}^{\left\lfloor\frac{n}{2}\right\rfloor} \sum_{l=r}^{\min(n-2,n-r-1)} \binom{n-l-1}{r} H_r\left(l-r,r,k-2\right) \\
& \times p_0 p_{11}^{l-r} p_{10}^{r} p_{01}^{r} p_{00}^{n-l-r-1} \\
& +p_0 p_{00}^{n-1}
\end{aligned}
$$

for $2 \leq k \leq n$. For $k=1$, we have $P\left\{L_n<k\right\}=p_0 p_{00}^{n-1}$.

In the following two theorems we derive respectively the exact distributions of $X_n^{(1)}$ and $X_n^{(n)}$ for Markov dependent trials.

Theorem 3.5 *Let $\left\{X_i\right\}_{i\geq 1}$ be a time-homogeneous Markov chain with transition probability matrix P and initial probabilities p_0 and p_1. Then for $k=1,2,\ldots,n-2$,*

$$
P\left\{X_n^{(1)} \geq k\right\}=\frac{1}{P(n)} \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{y=2}^{u} \binom{n-k(y-1)-2}{y-i-j} p_{11}^{n-2y+i+j-1} p_{10}^{y-i} p_{01}^{y-j}(1-p_i),
$$

where $u=\left\lfloor\frac{n+k-2+i+j}{k+1}\right\rfloor$ and $P(n)=1-p_1 p_{11}^{n-1}-p_0 p_{01} p_{11}^{n-2}-(n-2) p_1 p_{10} p_{01} p_{11}^{n-3}-p_1 p_{11}^{n-2} p_{10}$.

Proof. Let Y_n be the number of failures in a binary sequence of length n consisting at least two failures. Then it is true that

$$
P\left\{X_n^{(1)} \geq k\right\}=\frac{1}{P\left\{Y_n \geq 2\right\}} \sum_{y \geq 2} P\left\{X_n^{(1)} \geq k, Y_n=y\right\}. \tag{3.2}
$$

For $y \geq 2$, the four possible forms for the occurrence of the event

$\left\{X_n^{(1)} \geq k, Y_n = y\right\}$ are:

(A) $\overbrace{11\ldots1}^{z_1>0}0\overbrace{11\ldots1}^{z_2\geq k}0\ldots0\overbrace{11\ldots1}^{z_y\geq k}0\overbrace{11\ldots1}^{z_{y+1}>0}$

(B) $\overbrace{11\ldots1}^{z_1>0}0\overbrace{11\ldots1}^{z_2\geq k}0\ldots0\overbrace{11\ldots1}^{z_y\geq k}0$

(C) $0\overbrace{11\ldots1}^{z_1\geq k}0\overbrace{11\ldots1}^{z_2\geq k}0\ldots0\overbrace{11\ldots1}^{z_{y-1}\geq k}0\overbrace{11\ldots1}^{z_y>0}$

(D) $0\overbrace{11\ldots1}^{z_1\geq k}0\overbrace{11\ldots1}^{z_2\geq k}0\ldots0\overbrace{11\ldots1}^{z_{y-1}\geq k}0.$

The number of arrangements of the form (A) is $\binom{n-k(y-1)-2}{y}$ which is the total number of integer solutions to the equation

$$\begin{aligned}&z_1 + \cdots + z_{y+1} = n - y\\ &\text{s.t.} \quad z_1 > 0, z_2 \geq k, \ldots, z_y \geq k, z_{y+1} > 0\end{aligned}$$

and each sequence in the form of (A) has the probability of occurrence $p_1 p_{11}^{n-2y-1} p_{10}^{y} p_{01}^{y}$.

The number of arrangements of the form (B) is $\binom{n-k(y-1)-2}{y-1}$ which is the total number of integer solutions to the equation

$$\begin{aligned}&z_1 + \cdots + z_y = n - y\\ &\text{s.t.} \quad z_1 > 0, z_2 \geq k, \ldots, z_y \geq k\end{aligned}$$

and each sequence in the form of (B) has the probability of occurrence $p_1 p_{11}^{n-2y} p_{10}^{y} p_{01}^{y-1}$.

The number of arrangements of the form (C) is $\binom{n-k(y-1)-2}{y-1}$ which is the total number of integer solutions to the equation

$$\begin{aligned}&z_1 + \cdots + z_y = n - y\\ &\text{s.t.} \quad z_1 \geq k, \ldots, z_{y-1} \geq k, z_y > 0\end{aligned}$$

and each sequence in the form of (C) has the probability of occurrence $p_0 p_{11}^{n-2y} p_{10}^{y-1} p_{01}^{y}$.

The number of arrangements of the form (D) is $\binom{n-k(y-1)-2}{y-2}$ which is the total number of integer solutions to the equation

$$\begin{aligned} & z_1 + \cdots + z_{y-1} = n - y \\ & \text{s.t.} \quad z_1 \geq k, \ldots, z_{y-1} \geq k \end{aligned}$$

and each sequence in the form of (D) has the probability of occurrence $p_0 p_{11}^{n-2y+1} p_{10}^{y-1} p_{01}^{y-1}$.
We obtain

$$\begin{aligned} \sum_{y\geq 2} P\left\{X_n^{(1)} \geq k, Y_n = y\right\} &= \sum_{y=2}^{\lfloor \frac{n+k-2}{k+1} \rfloor} \binom{n-k(y-1)-2}{y} p_1 p_{11}^{n-2y-1} p_{10}^{y} p_{01}^{y} \\ &+ \sum_{y=2}^{\lfloor \frac{n+k-1}{k+1} \rfloor} \binom{n-k(y-1)-2}{y-1} p_1 p_{11}^{n-2y} p_{10}^{y} p_{01}^{y-1} \\ &+ \sum_{y=2}^{\lfloor \frac{n+k-1}{k+1} \rfloor} \binom{n-k(y-1)-2}{y-1} p_0 p_{11}^{n-2y} p_{10}^{y-1} p_{01}^{y} \\ &+ \sum_{y=2}^{\lfloor \frac{n+k}{k+1} \rfloor} \binom{n-k(y-1)-2}{y-2} p_0 p_{11}^{n-2y+1} p_{10}^{y-1} p_{01}^{y-1}. \end{aligned}$$

On the other hand,

$$\begin{aligned} P(n) &= P\{Y_n \geq 2\} \\ &= 1 - P\{Y_n = 0\} - P\{Y_n = 1\} \\ &= 1 - p_1 p_{11}^{n-1} - p_0 p_{01} p_{11}^{n-2} - (n-2) p_1 p_{10} p_{01} p_{11}^{n-3} - p_1 p_{11}^{n-2} p_{10}. \quad (3.3) \end{aligned}$$

The proof is completed by using the last two equations in (3.2). □

Theorem 3.6 *Let $\{X_i\}_{i\geq 1}$ be a time-homogeneous Markov chain with transition probability matrix P and initial probabilities p_0 and p_1. Then for $k = 2, 3, \ldots, n-1$,*

$$\begin{aligned} P\left\{X_n^{(n)} < k\right\} &= P\left\{X_n^{(n)} = 0\right\} \\ &+ \frac{1}{P(n)} \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{y=2}^{n-i-j-1} \sum_{r=i+j+1}^{\min(n-y, y+i+j-1)} \binom{y-1}{r-i-j} \end{aligned}$$

$$\times H_{r-i-j}(n-y-r,r,k-2)p_{11}^{n-y-r}p_{10}^{r-i}p_{01}^{r-j}p_{00}^{y-r+i+j-1}p_j$$

and

$$\begin{aligned} P\left\{X_n^{(n)}<1\right\} &= P\left\{X_n^{(n)}=0\right\} \\ &= \frac{1}{P(n)}\left[\sum_{y=2}^{n-2}(n-y-1)p_1p_{11}^{n-y-2}p_{10}p_{01}p_{00}^{y-1}\right. \\ &\quad \left.+\sum_{y=2}^{n-1}p_1p_{11}^{n-y-1}p_{10}p_{00}^{y-1}+\sum_{y=2}^{n-1}p_0p_{11}^{n-y-1}p_{01}p_{00}^{y-1}+p_0p_{00}^{n-1}\right], \end{aligned}$$

where $P(n)=1-p_1p_{11}^{n-1}-p_0p_{01}p_{11}^{n-2}-(n-2)p_1p_{10}p_{01}p_{11}^{n-3}-p_1p_{11}^{n-2}p_{10}$.

Proof. First consider the case when $k=2,3,\ldots,n-1$ noting that

$$P\left\{X_n^{(n)}<k\right\}=P\left\{X_n^{(n)}=0\right\}+P\left\{0<X_n^{(n)}<k\right\}. \tag{3.4}$$

Let Y_n be the number of failures and R_n the number of success runs in a binary sequence of length n consisting at least two failures. Then it is true that

$$P\left\{0<X_n^{(n)}<k\right\}=\frac{1}{P\left\{Y_n\geq 2\right\}}\sum_{y\geq 2}\sum_{r\geq 1}P\left\{0<X_n^{(n)}<k,Y_n=y,R_n=r\right\}. \tag{3.5}$$

For $y \geq 2$, the four possible forms for the occurrence of the event $\left\{0<X_n^{(n)}<k,Y_n=y,R_n=r\right\}$ are:

(A′) $\overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{0<z_2<k}\ldots\overbrace{11\ldots1}^{0<z_{r-1}<k}\overbrace{00\ldots0}^{y_{r-1}>0}\overbrace{11\ldots1}^{z_r>0}$

(B′) $\overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{0<z_2<k}\ldots\overbrace{00\ldots0}^{y_{r-1}>0}\overbrace{11\ldots1}^{0<z_r<k}\overbrace{00\ldots0}^{y_r>0}$

(C′) $\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{0<z_1<k}\overbrace{00\ldots0}^{y_2>0}\ldots\overbrace{11\ldots1}^{0<z_{r-1}<k}\overbrace{00\ldots0}^{y_r>0}\overbrace{11\ldots1}^{z_r>0}$

(D′) $\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{0<z_1<k}\overbrace{00\ldots0}^{y_2>0}\ldots\overbrace{00\ldots0}^{y_r>0}\overbrace{11\ldots1}^{0<z_r<k}\overbrace{00\ldots0}^{y_{r+1}>0}.$

The number of arrangements of the form (A′) is $\binom{y-1}{r-2}H_{r-2}(n-y-r,r,k-2)$

for $r \geq 3$ and $y \leq n-3$ which is the total number of integer solutions to the equation

$$\begin{aligned} &z_1 + \cdots + z_r = n - y \\ &\text{s.t.} \quad z_1 > 0, 0 < z_2 < k, \ldots, 0 < z_{r-1} < k, z_r > 0 \end{aligned}$$

multiplied by the total number of integer solutions to the equation

$$\begin{aligned} &y_1 + \cdots + y_{r-1} = y \\ &\text{s.t.} \quad y_i > 0,\ i = 1, \ldots, r-1 \end{aligned}$$

and each sequence in the form of (A′) has the probability of occurrence $p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1}$.

The number of arrangements of the form (B′) is $\binom{y-1}{r-1} H_{r-1}(n-y-r, r, k-2)$ for $r \geq 2$ and $y \leq n-2$ which is the total number of integer solutions to the equation

$$\begin{aligned} &z_1 + \cdots + z_r = n - y \\ &\text{s.t.} \quad z_1 > 0, 0 < z_2 < k, \ldots, 0 < z_r < k \end{aligned}$$

multiplied by the total number of integer solutions to the equation

$$\begin{aligned} &y_1 + \cdots + y_r = y \\ &\text{s.t.} \quad y_i > 0,\ i = 1, \ldots, r \end{aligned}$$

and each sequence in the form of (B′) has the probability of occurrence $p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r}$.

The number of arrangements of the form (C′) is $\binom{y-1}{r-1} H_{r-1}(n-y-r, r, k-2)$ for $r \geq 2$ and $y \leq n-2$ which is the total number of integer solutions to the equation

$$\begin{aligned} &z_1 + \cdots + z_r = n - y \\ &\text{s.t.} \quad 0 < z_1 < k, \ldots, 0 < z_{r-1} < k, z_r > 0 \end{aligned}$$

multiplied by the total number of integer solutions to the equation

$$\begin{aligned} &y_1 + \cdots + y_r = y \\ &\text{s.t.} \quad y_i > 0,\ i = 1, \ldots, r \end{aligned}$$

and each sequence in the form of (C′) has the probability of occurrence

$p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r}$.

The number of arrangements of the form (D′) is $\binom{y-1}{r} H_r(n-y-r, r, k-2)$ for $r \geq 1$ and $y \leq n-1$ which is the total number of integer solutions to the equation

$$\begin{aligned} &z_1 + \cdots + z_r = n - y \\ &\text{s.t.} \quad 0 < z_i < k, \ i = 1, \ldots, r \end{aligned}$$

multiplied by the total number of integer solutions to the equation

$$\begin{aligned} &y_1 + \cdots + y_{r+1} = y \\ &\text{s.t.} \quad y_i > 0, \ i = 1, \ldots, r+1 \end{aligned}$$

and each sequence in the form of (D′) has the probability of occurrence $p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1}$.

We obtain

$$\begin{aligned} &\sum_{y \geq 2} \sum_{r \geq 1} P\left\{0 < X_n^{(n)} < k, Y_n = y, R_n = r\right\} \\ = &\sum_{y=2}^{n-3} \sum_{r=3}^{\min(n-y, y+1)} \binom{y-1}{r-2} H_{r-2}(n-y-r, r, k-2) p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1} \\ &+ \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y, y)} \binom{y-1}{r-1} H_{r-1}(n-y-r, r, k-2) p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r} \\ &+ \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y, y)} \binom{y-1}{r-1} H_{r-1}(n-y-r, r, k-2) p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} \\ &+ \sum_{y=2}^{n-1} \sum_{r=1}^{\min(n-y, y-1)} \binom{y-1}{r} H_r(n-y-r, r, k-2) p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1}. \end{aligned}$$

For $k = 2, 3, \ldots, n-1$, the proof follows by using the last equation and (3.3) in (3.5) and then using the result in (3.4). Now, consider the case when $k = 1$. It is true that

$$\begin{aligned} P\left\{X_n^{(n)} < 1\right\} &= P\left\{X_n^{(n)} = 0\right\} \\ &= \frac{1}{P\{Y_n \geq 2\}} \sum_{y \geq 2} P\left\{X_n^{(n)} = 0, Y_n = y\right\}. \qquad (3.6) \end{aligned}$$

For $y \geq 2$, the four possible forms for the occurrence of the event $\left\{X_n^{(n)} = 0, Y_n = y\right\}$ are:

(A$''$) $\overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{2\leq y\leq n-2}\overbrace{11\ldots1}^{z_2>0}$

(B$''$) $\overbrace{11\ldots1}^{n-y}\overbrace{00\ldots0}^{2\leq y\leq n-1}$

(C$''$) $\overbrace{00\ldots0}^{2\leq y\leq n-1}\overbrace{11\ldots1}^{n-y}$

(D$''$) $\overbrace{00\ldots0}^{y=n}$.

The number of arrangements of the form (A$''$) is $n-y-1$ which is the total number of integer solutions to the equation

$$\begin{aligned} & z_1 + z_2 = n - y \\ & \text{s.t.} \quad z_i > 0, \; i = 1, 2 \end{aligned}$$

and each sequence in the form of (A$''$) has the probability of occurrence $p_1 p_{11}^{n-y-2} p_{10} p_{01} p_{00}^{y-1}$.
There is only one possible arrangement of the form (B$''$) and this sequence has the probability of occurrence $p_1 p_{11}^{n-y-1} p_{10} p_{00}^{y-1}$.
There is only one possible arrangement of the form (C$''$) and this sequence has the probability of occurrence $p_0 p_{11}^{n-y-1} p_{01} p_{00}^{y-1}$.
There is only one possible arrangement of the form (D$''$) and this sequence has the probability of occurrence $p_0 p_{00}^{n-1}$.
We obtain

$$\begin{aligned} \sum_{y\geq 2} P\left\{X_n^{(n)} = 0, Y_n = y\right\} &= \sum_{y=2}^{n-2} (n-y-1) p_1 p_{11}^{n-y-2} p_{10} p_{01} p_{00}^{y-1} + p_0 p_{00}^{n-1} \\ & + \sum_{y=2}^{n-1} p_1 p_{11}^{n-y-1} p_{10} p_{00}^{y-1} + \sum_{y=2}^{n-1} p_0 p_{11}^{n-y-1} p_{01} p_{00}^{y-1} . \end{aligned}$$

Using the last equation and (3.3) in (3.6) completes the proof. □

3.2 Mean distance between failures

In this section we first study the distribution of a different run statistic, namely the mean distance, for a sequence consisting of i.i.d. trials. Then we extend the result to the case of Markov dependent trials. Additionally, we obtain an explicit expression for the expected value of the mean distance for the case of i.i.d. trials.

Let $\{X_i\}_{i\geq 1}$ be a sequence of Bernoulli trials with two possible outcomes either a success ("1") or a failure ("0") which include among them at least two failures. For $n \geq 2$, denote by $\overline{X}_n$ the mean distance between successive failures in the first n trials of $\{X_i\}_{i\geq 1}$. For an illustration, consider the sequence of $n = 15$ trials 110111001011001. Then $\overline{X}_{15} = \frac{3+0+1+2+0}{5} = \frac{6}{5}$. Before we proceed further, it should be noted that the range set of $\overline{X}_n$ is

$$\left\{0, \frac{i}{n-i-1}, \frac{i}{n-i-2}, \ldots, i;\ i = 1, 2, \ldots, n-2\right\}.$$

3.2.1 Independent and identically distributed trials

Theorem 3.7 *Let $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. binary trials with $p = P\{X_i = 1\}$, $i \geq 1$. Then, for $0 \leq x \leq n-2$,*

$$P\left\{\overline{X}_n \leq x\right\} = \frac{1}{Q(n)} \sum_{y=2}^{n} \sum_{a=0}^{n-y} (a+1) \binom{n-a-2}{y-2} p^{n-y} (1-p)^y I\left(\frac{n-y-a}{y-1} \leq x\right),$$

where $Q(n) = 1 - p^n - np^{n-1}(1-p)$, $I(A) = 1$ if A occurs, and $I(A) = 0$, otherwise.

Proof. In a sequence consisting of $y \geq 2$ failures, let Z_i be the number of successes between $(i-1)$-st and i-th failures, that is, the distance between $(i-1)$-st and i-th failures, $i = 2, \ldots, y$. Denote by Z_1 the number of successes before the first failure and Z_{y+1} the number of successes after the y-th failure. Then

$$\begin{aligned} P\left\{\overline{X}_n \le x\right\} &= \frac{1}{P\{Y_n \ge 2\}} \sum_{y\ge 2} P\left\{\overline{X}_n \le x, Y_n = y\right\} \\ &= \frac{1}{P\{Y_n \ge 2\}} \sum_{y\ge 2} P\left\{\frac{Z_2+\cdots+Z_y}{y-1} \le x, Y_n = y\right\} \\ &= \frac{1}{P\{Y_n \ge 2\}} \sum_{y\ge 2} P\left\{\frac{n-y-Z_1-Z_{y+1}}{y-1} \le x, Y_n = y\right\} \\ &= \frac{1}{P\{Y_n \ge 2\}} \sum_{y\ge 2} P\left\{Z_1+Z_{y+1} \ge n-y-\lfloor (y-1)x\rfloor, Y_n = y\right\}. \end{aligned} \tag{3.7}$$

It is clear that

$$\begin{aligned} & P\left\{Z_1+Z_{y+1} \ge n-y-\lfloor (y-1)x\rfloor, Y_n = y\right\} \\ = & \sum_{\substack{z_1\ge 0 \\ n-y-\lfloor (y-1)x\rfloor \le z_1+z_{y+1}\le n-y}} \sum_{z_{y+1}\ge 0} P\left\{Z_1 = z_1, Z_{y+1} = z_{y+1}, Y_n = y\right\} \end{aligned} \tag{3.8}$$

and that

$$P\left\{Z_1 = z_1, Z_{y+1} = z_{y+1}, Y_n = y\right\} = \binom{n-z_1-z_{y+1}-2}{y-2} p^{n-y}(1-p)^y \tag{3.9}$$

for $z_1 \ge 0, z_{y+1} \ge 0$, and $z_1 + z_{y+1} \le n-y$ since $\binom{n-z_1-z_{y+1}-2}{y-2}$ is the number of integer solutions to the equation

$$\begin{aligned} & z_2+\cdots+z_y = n-y-z_1-z_{y+1} \\ & \text{s.t.} \quad z_i \ge 0,\ i = 2,\ldots,y \end{aligned}$$

and $p^{n-y}(1-p)^y$ is the probability of getting y failures and $n-y$ successes in n i.i.d. trials.

Substituting (3.9) into (3.8) and letting $z_1 + z_{y+1} = a$, we have

$$P\left\{Z_1+Z_{y+1} \ge n-y-\lfloor (y-1)x\rfloor, Y_n = y\right\}$$

$$
\begin{aligned}
&= \sum_{\substack{z_1\geq 0 \\ n-y-\lfloor (y-1)x\rfloor \leq z_1+z_{y+1}\leq n-y}} \sum_{z_{y+1}\geq 0} \binom{n-z_1-z_{y+1}-2}{y-2} p^{n-y}(1-p)^y \\
&= \sum_{a=0}^{n-y}(a+1)\binom{n-a-2}{y-2}p^{n-y}(1-p)^y I\left(\frac{n-y-a}{y-1}\leq x\right)
\end{aligned}
$$

since $a+1$ is the number of integer solutions to the equation

$$
\begin{aligned}
& z_1 + z_{y+1} = a \\
& \text{s.t.} \quad z_1 \geq 0, z_{y+1} \geq 0.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
P\{Y_n \geq 2\} &= 1 - P\{Y_n = 0\} - P\{Y_n = 1\} \\
&= 1 - p^n - np^{n-1}(1-p). \qquad (3.10)
\end{aligned}
$$

The required result is obtained using the last two equations in (3.7). □

Proposition 3.8 *Let $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. binary trials with $p = P\{X_i = 1\}$, $i \geq 1$. Then, for $n \geq 2$,*

$$
E(\overline{X}_n) = \frac{p}{1-p}\left[1 - \frac{\binom{n}{2}p^{n-2}(1-p)^2}{1-p^n-np^{n-1}(1-p)}\right].
$$

Proof. By the definition of $\overline{X}_n$,

$$
E(\overline{X}_n) = \frac{1}{P\{Y_n\geq 2\}}\sum_{y=2}^{n} E\left(\frac{n-y-Z_1-Z_{y+1}}{y-1}\,\middle|\, Y_n = y\right)\binom{n}{y}p^{n-y}(1-p)^y. \qquad (3.11)
$$

It is clear that

$$
E\left(\frac{n-y-Z_1-Z_{y+1}}{y-1}\,\middle|\, Y_n = y\right) = \frac{n-y}{y-1} - \frac{1}{y-1}E\left(Z_1+Z_{y+1} \mid Y_n = y\right) \qquad (3.12)
$$

and that

$$
\begin{aligned}
E\left(Z_1 + Z_{y+1} \mid Y_n = y\right) &= \sum_{a=0}^{n-y} aP\left\{Z_1 + Z_{y+1} = a \mid Y_n = y\right\} \\
&= \sum_{a=0}^{n-y} a\frac{P\left\{Z_1 + Z_{y+1} = a, Y_n = y\right\}}{P\left\{Y_n = y\right\}} \\
&= \sum_{a=0}^{n-y} a\frac{(a+1)\binom{n-a-2}{y-2}p^{n-y}(1-p)^y}{\binom{n}{y}p^{n-y}(1-p)^y} \\
&= \frac{1}{\binom{n}{y}}\sum_{a=0}^{n-y} a(a+1)\binom{n-a-2}{y-2} \\
&= 2\frac{(n-y)}{y+1}. \qquad (3.13)
\end{aligned}
$$

Using (3.13) in (3.12), we obtain

$$
\begin{aligned}
E\left(\frac{n-y-Z_1-Z_{y+1}}{y-1}\middle| Y_n = y\right) &= \frac{n-y}{y-1} - \frac{1}{y-1}\frac{2(n-y)}{y+1} \\
&= \frac{(n-y)}{y+1} \qquad (3.14)
\end{aligned}
$$

Thus the proof is completed by using the last equation and (3.10) in (3.11). □

From Proposition 3.8, it is obvious that

$$E(\overline{X}_n) \to \frac{p}{1-p} \text{ as } n \to \infty.$$

3.2.2 Markov dependent trials

Theorem 3.9 *Let $\{X_i\}_{i\geq 1}$ be a time-homogeneous Markov chain with transition probability matrix P and initial probabilities p_0 and p_1. Then, for $0 \leq x \leq n-2$,*

$$P\left\{\overline{X}_n \leq x\right\} = \frac{1}{P(n)}\left[\sum_{y=2}^{n-2}(n-y-1)p_1 p_{11}^{n-y-2}p_{10}p_{01}p_{00}^{y-1} + \sum_{y=2}^{n-1} p_1 p_{11}^{n-y-1}p_{10}p_{00}^{y-1}\right.$$

$$+\sum_{y=2}^{n-1} p_0 p_{11}^{n-y-1} p_{01} p_{00}^{y-1} + p_0 p_{00}^{n-1} + U_1 + U_2 + U_3 + U_4 \Bigg],$$

where

$U_1 = \sum_{y=2}^{n-3} \sum_{r=3}^{\min(n-y,y+1)} \sum_{a=2}^{n-y-r+2} (a-1) \binom{n-y-a-1}{r-3} \binom{y-1}{r-2} p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1}$

$\times I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_2 = \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y,y)} \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2} \binom{y-1}{r-1} p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r} I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_3 = \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y,y)} \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2} \binom{y-1}{r-1} p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_4 = \sum_{y=2}^{n-1} \sum_{r=1}^{\min(n-y,y-1)} \binom{n-y-1}{r-1} \binom{y-1}{r} p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1} I\left(\frac{n-y}{y-1} \leq x\right)$,

$P(n) = 1 - p_1 p_{11}^{n-1} - p_0 p_{01} p_{11}^{n-2} - (n-2) p_1 p_{10} p_{01} p_{11}^{n-3} - p_1 p_{11}^{n-2} p_{10}$, $I(A) = 1$ *if* A *occurs, and* $I(A) = 0$, *otherwise.*

Proof. It is clear that

$$\begin{aligned} P\left\{\overline{X}_n \leq x\right\} &= P\left\{\overline{X}_n = 0\right\} + P\left\{0 < \overline{X}_n \leq x\right\} \\ &= \frac{1}{P\{Y_n \geq 2\}} \sum_{y \geq 2} P\left\{\overline{X}_n = 0, Y_n = y\right\} \\ &\quad + \frac{1}{P\{Y_n \geq 2\}} \sum_{y \geq 2} P\left\{0 < \overline{X}_n \leq x, Y_n = y\right\}. \end{aligned} \tag{3.15}$$

For $y \geq 2$, the four possible forms for the occurrence of the event $\left\{\overline{X}_n = 0, Y_n = y\right\}$ are given by (A″)-(D″) in the proof of Theorem 3.6. Thus

$$\begin{aligned} P\left\{\overline{X}_n = 0\right\} &= P\left\{X_n^{(n)} = 0\right\} \\ &= P\left\{X_n^{(n)} < 1\right\} \\ &= \frac{1}{P(n)} \Bigg[\sum_{y=2}^{n-2} (n-y-1) p_1 p_{11}^{n-y-2} p_{10} p_{01} p_{00}^{y-1} + p_0 p_{00}^{n-1} \\ &\quad + \sum_{y=2}^{n-1} p_1 p_{11}^{n-y-1} p_{10} p_{00}^{y-1} + \sum_{y=2}^{n-1} p_0 p_{11}^{n-y-1} p_{01} p_{00}^{y-1}\Bigg], \end{aligned} \tag{3.16}$$

where $P(n) = 1 - p_1 p_{11}^{n-1} - p_0 p_{01} p_{11}^{n-2} - (n-2) p_1 p_{10} p_{01} p_{11}^{n-3} - p_1 p_{11}^{n-2} p_{10}$.

On the other hand, for $y \geq 2$, the four possible forms for the occurrence of the event $\left\{0 < \overline{X}_n \leq x, Y_n = y\right\}$ are:

$$\text{(A}'''\text{)}\ \overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{z_2>0}\ldots\overbrace{11\ldots1}^{z_{r-1}>0}\overbrace{00\ldots0}^{y_{r-1}>0}\overbrace{11\ldots1}^{z_r>0}$$

$$\text{(B}'''\text{)}\ \overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{z_2>0}\ldots\overbrace{00\ldots0}^{y_{r-1}>0}\overbrace{11\ldots1}^{z_r>0}\overbrace{00\ldots0}^{y_r>0}$$

$$\text{(C}'''\text{)}\ \overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_2>0}\ldots\overbrace{11\ldots1}^{z_{r-1}>0}\overbrace{00\ldots0}^{y_r>0}\overbrace{11\ldots1}^{z_r>0}$$

$$\text{(D}'''\text{)}\ \overbrace{00\ldots0}^{y_1>0}\overbrace{11\ldots1}^{z_1>0}\overbrace{00\ldots0}^{y_2>0}\ldots\overbrace{00\ldots0}^{y_r>0}\overbrace{11\ldots1}^{z_r>0}\overbrace{00\ldots0}^{y_{r+1}>0}.$$

Considering the forms (A‴)-(D‴), we have

$$\begin{aligned}
& \frac{1}{P\left\{Y_n \geq 2\right\}} \sum_{y\geq 2} P\left\{0 < \overline{X}_n \leq x, Y_n = y\right\} \\
= \ & \frac{1}{P\left\{Y_n \geq 2\right\}} \left[\sum_{y\geq 2}\sum_{r\geq 3} P\left\{\frac{Z_2+\cdots+Z_{r-1}}{y-1} \leq x, Y_n = y, R_n = r\right\}\right. \\
& + \sum_{y\geq 2}\sum_{r\geq 2} P\left\{\frac{Z_2+\cdots+Z_r}{y-1} \leq x, Y_n = y, R_n = r\right\} \\
& + \sum_{y\geq 2}\sum_{r\geq 2} P\left\{\frac{Z_1+\cdots+Z_{r-1}}{y-1} \leq x, Y_n = y, R_n = r\right\} \\
& \left. + \sum_{y\geq 2}\sum_{r\geq 1} P\left\{\frac{Z_1+\cdots+Z_r}{y-1} \leq x, Y_n = y, R_n = r\right\}\right]. \qquad (3.17)
\end{aligned}$$

For the form (A‴):

$$\begin{aligned}
& P\left\{\frac{Z_2+\cdots+Z_{r-1}}{y-1} \leq x, Y_n = y, R_n = r\right\} \\
= \ & P\left\{Z_1 + Z_r \geq n - y - \lfloor (y-1)\,x\rfloor, Y_n = y, R_n = r\right\} \\
= \ & \sum_{\substack{z_1>0 \\ n-y-\lfloor (y-1)x\rfloor \leq z_1+z_r \leq n-y-r+2}}\ \sum_{z_r>0} P\left\{Z_1 = z_1, Z_r = z_r, Y_n = y, R_n = r\right\}. \qquad (3.18)
\end{aligned}$$

It is clear that

$$P\{Z_1 = z_1, Z_r = z_r, Y_n = y, R_n = r\} = \binom{n-y-z_1-z_r-1}{r-3}\binom{y-1}{r-2} \times p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1} \quad (3.19)$$

for $z_1 > 0, z_r > 0$, and $z_1 + z_r \leq n - y - r + 2$ since $\binom{n-y-z_1-z_r-1}{r-3}$ is the number of integer solutions to the equation

$$\begin{aligned} & z_2 + \cdots + z_{r-1} = n - y - z_1 - z_r \\ & \text{s.t.} \quad z_i > 0,\ i = 2, \ldots, r-1, \end{aligned}$$

$\binom{y-1}{r-2}$ is the number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_{r-1} = y \\ & \text{s.t.} \quad y_i > 0,\ i = 1, \ldots, r-1, \end{aligned}$$

and each sequence in the form of (A‴) has the probability of occurrence $p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1}$.

Substituting (3.19) into (3.18), we have

$$\begin{aligned} & P\left\{\frac{Z_2 + \cdots + Z_{r-1}}{y-1} \leq x, Y_n = y, R_n = r\right\} \\ = & \sum_{\substack{z_1>0 \\ n-y-\lfloor (y-1)x \rfloor \leq z_1+z_r \leq n-y-r+2}} \sum_{z_r>0} \binom{n-y-z_1-z_r-1}{r-3}\binom{y-1}{r-2} p_1 p_{11}^{n-y-r} p_{10}^{r-1} \\ & \times p_{01}^{r-1} p_{00}^{y-r+1} \\ = & \sum_{a=2}^{n-y-r+2} (a-1) \binom{n-y-a-1}{r-3}\binom{y-1}{r-2} p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1} \\ & \times I\left(\frac{n-y-a}{y-1} \leq x\right). \end{aligned} \quad (3.20)$$

Note that, here we let $z_1 + z_r = a$, then $a - 1$ is the number of integer solutions to the equation

$$\begin{aligned} & z_1 + z_r = a \\ & \text{s.t.} \quad z_1 > 0, z_r > 0. \end{aligned}$$

For the form (B‴):

$$\begin{aligned} & P\left\{\frac{Z_2+\cdots+Z_r}{y-1} \le x, Y_n = y, R_n = r\right\} \\ = \;& P\left\{Z_1 \ge n-y-\lfloor (y-1)\,x \rfloor, Y_n = y, R_n = r\right\} \\ = \;& \sum_{\substack{z_1>0 \\ n-y-\lfloor (y-1)x\rfloor \le z_1 \le n-y-r+1}} P\left\{Z_1 = z_1, Y_n = y, R_n = r\right\}. \end{aligned} \tag{3.21}$$

It is clear that

$$P\left\{Z_1 = z_1, Y_n = y, R_n = r\right\} = \binom{n-y-z_1-1}{r-2}\binom{y-1}{r-1} p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r} \tag{3.22}$$

for $z_1 > 0$ and $z_1 \le n-y-r+1$ since $\binom{n-y-z_1-1}{r-2}$ is the number of integer solutions to the equation

$$\begin{aligned} & z_2+\cdots+z_r = n-y-z_1 \\ & \text{s.t.} \quad z_i > 0,\; i = 2,\ldots,r, \end{aligned}$$

$\binom{y-1}{r-1}$ is the number of integer solutions to the equation

$$\begin{aligned} & y_1+\cdots+y_r = y \\ & \text{s.t.} \quad y_i > 0,\; i = 1,\ldots,r, \end{aligned}$$

and each sequence in the form of (B‴) has the probability of occurrence $p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r}$.

Substituting (3.22) into (3.21), we have

$$\begin{aligned} & P\left\{\frac{Z_2+\cdots+Z_r}{y-1} \le x, Y_n = y, R_n = r\right\} \\ = \;& \sum_{\substack{z_1>0 \\ n-y-\lfloor (y-1)x\rfloor \le z_1 \le n-y-r+1}} \binom{n-y-z_1-1}{r-2}\binom{y-1}{r-1} p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r} \\ = \;& \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2}\binom{y-1}{r-1} p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r}\, I\left(\frac{n-y-a}{y-1} \le x\right). \end{aligned} \tag{3.23}$$

Note that, here we let $z_1 = a$ and there is only one integer solution to the equation

$$\begin{aligned} & z_1 = a \\ & \text{s.t.} \quad z_1 > 0. \end{aligned}$$

For the form (C‴):

$$\begin{aligned} & P\left\{\frac{Z_1 + \cdots + Z_{r-1}}{y-1} \le x, Y_n = y, R_n = r\right\} \\ = \; & P\left\{Z_r \ge n - y - \lfloor (y-1)\, x \rfloor, Y_n = y, R_n = r\right\} \\ = \; & \sum_{\substack{z_r > 0 \\ n-y-\lfloor (y-1)x \rfloor \le z_r \le n-y-r+1}} P\left\{Z_r = z_r, Y_n = y, R_n = r\right\}. \end{aligned} \tag{3.24}$$

It is clear that

$$P\left\{Z_r = z_r, Y_n = y, R_n = r\right\} = \binom{n-y-z_r-1}{r-2}\binom{y-1}{r-1} p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} \tag{3.25}$$

for $z_r > 0$ and $z_r \le n-y-r+1$ since $\binom{n-y-z_r-1}{r-2}$ is the number of integer solutions to the equation

$$\begin{aligned} & z_1 + \cdots + z_{r-1} = n - y - z_r \\ & \text{s.t.} \quad z_i > 0, \; i = 1, \ldots, r-1, \end{aligned}$$

$\binom{y-1}{r-1}$ is the number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_r = y \\ & \text{s.t.} \quad y_i > 0, \; i = 1, \ldots, r, \end{aligned}$$

and each sequence in the form of (C‴) has the probability of occurrence $p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r}$.

Substituting (3.25) into (3.24), we have

$$\begin{aligned} & P\left\{\frac{Z_1 + \cdots + Z_{r-1}}{y-1} \le x, Y_n = y, R_n = r\right\} \\ = \; & \sum_{\substack{z_r > 0 \\ n-y-\lfloor (y-1)x \rfloor \le z_r \le n-y-r+1}} \binom{n-y-z_r-1}{r-2}\binom{y-1}{r-1} p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} \end{aligned}$$

$$= \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2}\binom{y-1}{r-1} p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} I\left(\frac{n-y-a}{y-1} \le x\right). \tag{3.26}$$

Note that, here we let $z_r = a$ and there is only one integer solution to the equation

$$\begin{aligned} & z_r = a \\ & \text{s.t.} \quad z_r > 0. \end{aligned}$$

For the form (D‴):

$$\begin{aligned} & P\left\{\frac{Z_1 + \cdots + Z_r}{y-1} \le x, Y_n = y, R_n = r\right\} \\ = & \binom{n-y-1}{r-1}\binom{y-1}{r} p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1} I\left(\frac{n-y}{y-1} \le x\right) \end{aligned} \tag{3.27}$$

since $\binom{n-y-1}{r-1}$ is the number of integer solutions to the equation

$$\begin{aligned} & z_1 + \cdots + z_r = n - y \\ & \text{s.t.} \quad z_i > 0, \; i = 1, \ldots, r, \end{aligned}$$

$\binom{y-1}{r}$ is the number of integer solutions to the equation

$$\begin{aligned} & y_1 + \cdots + y_{r+1} = y \\ & \text{s.t.} \quad y_i > 0, \; i = 1, \ldots, r+1, \end{aligned}$$

and each sequence in the form of (D‴) has the probability of occurrence $p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1}$.

Using the four results (3.20), (3.23), (3.26), and (3.27) in (3.17), we have

$$\begin{aligned} P\left\{0 < \overline{X}_n \le x\right\} &= \frac{1}{P\left\{Y_n \ge 2\right\}} \sum_{y \ge 2} P\left\{0 < \overline{X}_n \le x, Y_n = y\right\} \\ &= \frac{1}{P(n)}\left[U_1 + U_2 + U_3 + U_4\right], \end{aligned} \tag{3.28}$$

where

$U_1 = \sum_{y=2}^{n-3} \sum_{r=3}^{\min(n-y,y+1)} \sum_{a=2}^{n-y-r+2} (a-1)\binom{n-y-a-1}{r-3}\binom{y-1}{r-2} p_1 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r-1} p_{00}^{y-r+1}$

$\times I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_2 = \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y,y)} \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2}\binom{y-1}{r-1} p_1 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r-1} p_{00}^{y-r} I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_3 = \sum_{y=2}^{n-2} \sum_{r=2}^{\min(n-y,y)} \sum_{a=1}^{n-y-r+1} \binom{n-y-a-1}{r-2}\binom{y-1}{r-1} p_0 p_{11}^{n-y-r} p_{10}^{r-1} p_{01}^{r} p_{00}^{y-r} I\left(\frac{n-y-a}{y-1} \leq x\right)$,

$U_4 = \sum_{y=2}^{n-1} \sum_{r=1}^{\min(n-y,y-1)} \binom{n-y-1}{r-1}\binom{y-1}{r} p_0 p_{11}^{n-y-r} p_{10}^{r} p_{01}^{r} p_{00}^{y-r-1} I\left(\frac{n-y}{y-1} \leq x\right)$,

and $P(n) = 1 - p_1 p_{11}^{n-1} - p_0 p_{01} p_{11}^{n-2} - (n-2) p_1 p_{10} p_{01} p_{11}^{n-3} - p_1 p_{11}^{n-2} p_{10}$.

Substituting (3.16) and (3.28) into (3.15) we have the desired result. □

3.3 Applications and numerical results

3.3.1 System reliability

A consecutive k-within-m-out-of-n:F system consists of n components and it fails if and only if there are at least k failed components among any m consecutive components ($k \leq m \leq n$). For $m = n$ and $m = k$, consecutive k-within-m-out-of-n:F system coincides with the well-known k-out-of-n:F and consecutive k-out-of-n:F systems, respectively.

Let X_i denote the state of component i as either working ("1") or failed ("0") and $R_{m,n}$ the reliability of consecutive 2-within-m-out-of-n:F system. As stated in [59] the reliability of a consecutive 2-within-m-out-of-n:F system is closely related to the random variable $X_n^{(1)}$, which is given by

$$R_{m,n} = [1 - P\{Y_n \geq 2\}] + P\left\{X_n^{(1)} \geq m-1\right\} P\{Y_n \geq 2\}.$$

Assume that the components of the consecutive 2-within-m-out-of-n:F system are dependent in a Markovian fashion. That is, the probability that i-th component fails (operates) depends upon, and only upon, the state of $(i-1)$-st component. In Table 3.1 we compute $R_{m,n}$, the exact reliability of consecutive 2-within-m-out-of-n:F system, (using Theorem 3.5) for $n = 10, 20$ and $m = 2, 3, 5$ for the following two cases:

Case I: The components are independent with common success probability $p = 0.9$,

Case II: The components are Markov dependent with $p_1 = 0.9$, $p_{11} = 0.85$, and $p_{01} = 0.5$.

		Case I	Case II
n	m	$R_{m,n}$	$R_{m,n}$
10	2	0.9197	0.5335
	3	0.8662	0.5047
	5	0.8007	0.4688
20	2	0.8388	0.2598
	3	0.7350	0.2304
	5	0.6098	0.1940

Table 3.1: Reliability of a consecutive 2-within-m-out-of-n:F system

3.3.2 Waiting time between extreme events

The occurrences of extreme events are highly important in the context of risk management. For example, extreme floods and rainfalls are important extreme events in insurance, and stock market crashes are important extreme events in finance. For analyzing a certain stochastic process, the distance (or the time) between two successive critical events may be a good indicator. Assume that associated with a certain stochastic process $\{Y_i\}_{i\geq 1}$, if the value of this process exceeds a critical level c at any time, then a failure ("0") occurs, which means that an extreme event occurs. That is,

$$X_i = \begin{cases} 0 & \text{if } Y_i \geq c \text{ (or } Y_i \leq c \text{ depending on the problem)}, \\ 1 & \text{otherwise.} \end{cases}$$

Thus the random variables $X_n^{(1)}$, $X_n^{(n)}$, and $\overline{X}_n$ represent respectively the minimum distance, the maximum distance, and the mean distance between the occurrences of successive critical events. Obviously, the distributional properties of these statistics are helpful for understanding the behavior of the corresponding stochastic process.

In Table 3.2 we compute the distributions of $X_5^{(1)}$ and $X_5^{(5)}$ for Markov dependent trials with $p_1 = 0.9$, $p_{11} = 0.85$, and $p_{01} = 0.5$. Table 3.3 includes the distribution of $\overline{X}_5$ for Case I and Case II defined in 3.3.1.

	$k=0$	$k=1$	$k=2$	$k=3$	$E(.)$
$P\left\{X_5^{(1)}=k\right\}$	0.8719	0.0644	0.0437	0.0200	0.2118
$P\left\{X_5^{(5)}=k\right\}$	0.7761	0.1367	0.0672	0.0200	0.3311

Table 3.2: Distributions of the extremes for Markov dependent trials

	$x=0$	$x=\frac{1}{3}$	$x=\frac{1}{2}$	$x=1$	$x=2$	$x=3$	$E(\overline{X}_5)$
Case I	0.3901	0.0033	0.0398	0.2983	0.1790	0.0895	0.9458
Case II	0.7761	0.0208	0.0513	0.0880	0.0436	0.0202	0.2681

Table 3.3: Distribution of $\overline{X}_5$ for i.i.d. and Markov dependent trials

In Table 3.4 we present exact (E) and simulated (S) values of $P\left\{\overline{X}_{10} \leq x\right\}$ for Case II considering all possible values of x. The exact results are consistent with the simulation results. Table 3.5 includes the expected values of $X_n^{(1)}$, $\overline{X}_n$, and $X_n^{(n)}$ for Case I and Case II.

x	E	S	x	E	S
0	0.4304	0.4323	1	0.7276	0.7258
1/8	0.4312	0.4375	5/4	0.7354	0.7340
1/7	0.4339	0.4390	4/3	0.7567	0.7602
1/6	0.4395	0.4418	3/2	0.7859	0.7903
1/5	0.4497	0.4475	5/3	0.8028	0.8080
1/4	0.4669	0.4635	2	0.8660	0.8602
2/7	0.4688	0.4702	5/2	0.8869	0.8869
1/3	0.5004	0.4989	3	0.9276	0.9252
2/5	0.5110	0.5132	7/2	0.9352	0.9342
1/2	0.5663	0.5696	4	0.9563	0.9581
3/5	0.5761	0.5760	5	0.9734	0.9709
2/3	0.6018	0.6024	6	0.9866	0.9851
3/4	0.6185	0.6223	7	0.9958	0.9961
4/5	0.6242	0.6251	8	1.0000	1.0000

Table 3.4: Exact and simulated values of $P\left\{\overline{X}_{10} \leq x\right\}$ for Case II

	Case I			Case II		
n	$E(X_n^{(1)})$	$E(\overline{X}_n)$	$E(X_n^{(n)})$	$E(X_n^{(1)})$	$E(\overline{X}_n)$	$E(X_n^{(n)})$
5	0.9048	0.9458	0.9877	0.2118	0.2681	0.3311
10	2.1097	2.3937	2.6943	0.5511	0.9790	1.6216
20	3.5404	4.7803	6.1921	0.7502	2.2277	5.0731
50	3.1446	8.2740	15.6502	0.1532	3.2783	12.8481

Table 3.5: Expected values of $X_n^{(1)}$, $\overline{X}_n$, and $X_n^{(n)}$

Chapter 4

Previous-Sum Dependent Model

Let $\{X_i\}_{i\geq 1}$ be a sequence of Bernoulli trials with two possible outcomes either a success ("1") or a failure ("0") satisfying

$$P\left\{X_n = 1 \middle| X_1, X_2, \ldots, X_{n-1}\right\} = a_n + b_n S_{n-1}, \tag{4.1}$$

for $n \geq 2$; and a_n and b_n satisfy $0 < a_n < 1$ and $0 < a_n + (n-1)\, b_n < 1$ for $n \geq 1$, where S_n, $n \geq 1$, denotes the total number of successes in $X_1, X_2, \ldots, X_n$, and for convenience $P\{S_0 = 0\} = 1$. The model given by (4.1) was first considered by Vellaisamy [85] to get new generalized binomial distributions. Vellaisamy and Sankar [86] called the model (4.1) the *previous-sum dependent model* and used it for modeling a dependent production process. When $a_i = p_i$ and $b_i = 0$, for $i \geq 2$, the model (4.1) corresponds to independent but nonidentically distributed Bernoulli trials with probability of success p_i.

As a direct consequence of (4.1), we have

$$P\left\{X_n = 1 \middle| S_{n-1}\right\} = a_n + b_n S_{n-1} \tag{4.2}$$

(see [87]).

A start-up demonstration test is a procedure by which a vendor elicits information about the reliability of a power generation equipment (unit) before purchasing it. Lawn mowers, water pumps, car batteries, and outboard motors are examples of units that are placed on start-up demonstration tests. A start-up demonstration test consists of attempting to start-up the unit under test several times, and observing the outcomes, either a success ("1") or a failure ("0"). It should be noted that observing a success means the unit turned on and a failure means it did not.

Start up demonstration testing was first discussed by Hahn and Gage [44] in 1983. They considered the CS (Consecutive Successes) type start-up demonstration test in which the testing continued until k consecutive successes are observed and assumed that the start-ups are independent and identically distributed. Viveros and Balakrishnan [88], Balakrishnan, Balasubramanian and Viveros [10], and Balakrishnan, Mohanty and Aki [12] studied the same type of test assuming that the start-ups are Markov dependent. CS type test has been found to be impractical because in this type of test only consecutive number of successful start-ups is considered and the number of unsuccessful start-ups is not taken into account. Koutras and Balakrishnan [54] suggested an alternative start-up demonstration test using a simple scan-based statistic. This type of test has two rejection criteria. One of them is based on observation of early unsuccessful start-ups and the other one is based on unsuccessful start-ups that are close to each other. If neither of them occurs, then the unit is accepted.

In 2000, Balakrishnan and Chan [9] proposed a new type of start-up demonstration test, CSTF (Consecutive Successes Total Failures), and studied this type of test under the assumption of i.i.d. start-ups. According to CSTF type test, the unit is accepted if k consecutive successful start-ups are observed before a total of d failures and rejected if d failures are observed before k consecutive successes. Martin [64] and Smith and Griffith [80] considered CSTF type test in the case of first order Markov dependent trials. Smith and Griffith [80] and Chan, Ng and Balakrishnan [14] studied estimation of the success probability p in CSTF type test with i.i.d. start-ups using different methods. Eryilmaz and Chakraborti [27] studied CSTF type test assuming that the start-ups are exchangeable.

Smith and Griffith [81] proposed two different types of start-up demonstration test, which are TSTF (Total Successes Total Failures) and CSCF (Consecutive Successes Consecutive Failures). According to TSTF (CSCF) type test, the unit is accepted if k (consecutive) successful start-ups are observed before d (consecutive) failures and rejected if d (consecutive) failures are observed before k (consecutive) successes. They also compared various types of start-up demonstration tests and showed that TSTF type test performs very well and in many cases it is more desirable than other types of test. Martin [65] studied TSTF, CSCF, CSTF, and TSCF (Total Successes Consecutive Failures) types of start-up demonstration tests assuming that the start-ups are Markov dependent of general order. Antzoulakos, Koutras and Rakitzis [8] introduced another type of test CSDF (Consecutive Successes Distance Failures), which is based on the number of consecutive successes and the distance between failures. According to CSDF type test, the unit is accepted if k consecutive successes are observed before two failures separated by at most $r-2$ successes and rejected if two failures separated by at most $r-2$ successes are observed before k consecutive successes. It should be noted that this rejection criteria is similar to the second criteria of Koutras and Balakrishnan [54].

Recently, Eryilmaz [25] studied TSTF and CSTF types of tests in the case of first order Markov dependent start-ups, obtained the distributions of the test lengths in a nonrecursive form, and estimated the expected test lengths. Gera [39][40] improved the CSTF type test to two new types of start-up demonstration tests TSCSTF (Total Successes Consecutive Successes Total Failures) and TSCSTFCF (Total Successes Consecutive Successes Total Failures Consecutive Failures) and studied them in case of i.i.d. start-ups. According to TSCSTF (TSCSTFCF) test, if either k_{cs} consecutive or a total of k_s successful start-ups are observed before (either k_{cf} consecutive or) a total of k_f failures, then the unit is accepted, otherwise it is rejected. Gera [39][40] proposed TSCSTF and TSCSTFCF because these types of tests will significantly reduce the test length which is beneficial from practical and financial aspects.

So far, in the literature, only the Markovian type dependence has been used to model dependent start-ups. In this chapter, we study TSTF type test under the model (4.1), the *previous-sum dependent model*. In view of (4.1), the outcome of the present start-up depends on the total number of successful start-ups so far. This model exhibits a stronger form of dependence than the Markovian type dependence, and hence it may be more appropriate when the outcome of the present start-up depends on all previous start-ups instead of only the immediately previous one. The magnitude of the dependence can be adjusted using the parameters of the model. The reason behind the choice of TSTF type test is twofold. First, based on the detailed analysis and comparison of various types of start-up demonstration tests in [81], this type of test has been shown to be more desirable than the others. Second, the TSTF type tests is easy to apply because it requires the vendor to keep track of only two things while conducting the test, which are total number of successes and total number of failures.

4.1 Characteristics of TSTF type test under previous-sum dependent model

In this section, we will derive the characteristics of TSTF type test using the recurrence

$$P\left\{S_n = x\right\} = \left(a_n + b_n\left(x-1\right)\right)P\left\{S_{n-1} = x-1\right\} + \left(1 - a_n - b_n x\right)P\left\{S_{n-1} = x\right\}.$$

See Appendix A for the proofs of the results for previous-sum dependent trials. Appendix A also includes the results for the case of i.i.d. trials.

Let $Z_{k,d}$ denote the total number of start-ups until termination of the experiment, that is, the test length. The distribution and the expected value of $Z_{k,d}$ are important characteristics in the context of start-up demonstration testing. As it is mentioned before, the unit under TSTF type test is accepted if k successful start-ups are observed before d failures and rejected if d failures are observed before k successes. Therefore

$$Z_{k,d} = \min\left(W_k^{(1)}, W_d^{(0)}\right),$$

where $W_k^{(1)}$ is the total number of start-ups until the first k successes (waiting time for the first k successes) and $W_d^{(0)}$ is the total number of start-ups until the first d failures (waiting time for the first d failures).

For $z = min(k, d), \ldots, k + d - 1$, the probability mass function of $Z_{k,d}$ can be computed from

$$P\left\{Z_{k,d} = z\right\} = P\left\{Z_{k,d} \geq z\right\} - P\left\{Z_{k,d} \geq z + 1\right\},$$

where

$$P\left\{Z_{k,d} \geq z\right\} = \sum_{i=max(0,z-d)}^{k-1} P\left\{S_{z-1} = i\right\}.$$

The expected test length can be found using

$$E\left(Z_{k,d}\right) = \sum_{z=\min(k,d)}^{k+d-1} z \; P\left\{Z_{k,d} = z\right\}.$$

The probability of acceptance of a unit is given by

$$P\left\{\text{Acceptance}\right\} = \sum_{i=k}^{k+d-1} \left[a_i + b_i\left(k-1\right)\right] P\left\{S_{i-1} = k-1\right\}.$$

The conditional probability of the test length $Z_{k,d}$ given that the unit is accepted in the end is

$$P\left\{Z_{k,d} = z \middle| \text{Acceptance}\right\} = \begin{cases} \frac{[a_z + b_z(k-1)]P\{S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k \leq z \leq k+d-1, \\ 0 & \text{otherwise} \end{cases}$$

and the conditional probability of the test length $Z_{k,d}$ given that the unit is rejected in the end is

$$P\left\{Z_{k,d} = z \middle| \text{Rejection}\right\} = \begin{cases} \frac{[1 - a_z - b_z(z-d)]P\{S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d \leq z \leq k+d-1, \\ 0 & \text{otherwise.} \end{cases}$$

4.2 Illustrations and numerical computations

Suppose that a vendor wishes to demonstrate the reliability of a power generation equipment. This demonstration is done by starting-up the unit several times and using TSTF type test on the basis of the observed successes and failures. Consider the model (4.1) with $a_i = (1-\theta_i)\, p_i$ and $b_i = \theta_i / (i-1)$ for $i \geq 2$ so that $P\{X_1 = 1\} = p_1$ and for $i \geq 2$

$$P\left\{X_i = 1 \middle| X_1, X_2, \ldots, X_{i-1}\right\} = (1-\theta_i)\, p_i + \frac{\theta_i}{(i-1)} S_{i-1}, \tag{4.3}$$

where

$$max\left(\frac{p_1 - 1}{p_1}, \frac{p_1}{p_1 - 1}\right) \leq \theta_i \leq 1. \tag{4.4}$$

Under the model (4.3), the random variables $X_1, X_2, \ldots$ are identically distributed with success probability $P\{X_i = 1\} = p_1$, $i \geq 1$ (see [85]). It should be noted that, if $\theta_i = 0$, $i \geq 1$, then $X_1, X_2, \ldots$ are independent and identically distributed.

For $\theta_i > 0$ ($\theta_i < 0$), X_i is stochastically increasing (decreasing) in S_{i-1}, $i \geq 2$, which implies a positive (negative) dependence between X_i and S_{i-1}. The correlation between X_i and S_{i-1} can be computed from

$$Cor\,(X_i, S_{i-1}) = \frac{\frac{\theta_i}{i-1} E\left(S_{i-1}^2\right) - p_1^2 \theta_i\, (i-1)}{\sqrt{p_1\,(1-p_1)\left[E\left(S_{i-1}^2\right) - (i-1)^2\, p_1^2\right]}}$$

for $i \geq 2$ (see [85]).

Assume that a unit is started-up ten times and the following two realizations are obtained $X = (1011101111)$ and $Y = (1001001101)$. Intuitively, under the positive dependence among the start-ups, a successful start-up is more likely to occur in X than in Y in the future. According to the model (4.3), if $\theta_i > 0$, then the probability of observing a success in the next start-up in X is greater than that in Y. This result is a direct consequence of the positive dependence between X_i and S_{i-1} when $\theta_i > 0$. Since the case $\theta_i < 0$ seems irrational in the context of start-up demonstration testing (because $\theta_i < 0$ implies that the probability of observing a success in the next start-up in Y is greater than that in X), we consider only the case $\theta_i > 0$.

Let probability of observing a successful start-up be $p_1 = 0.7$. Then from (4.4) we have $-3/7 \leq \theta_i \leq 1$. In Table 4.1 we compute the correlation between X_i and S_{i-1} for $\theta_i \in \{-0.4, 0.1, 0.5, 0.9\}$, $i \geq 2$.

i	$\theta_i = -0.4$	$\theta_i = 0.1$	$\theta_i = 0.5$	$\theta_i = 0.9$
2	-0.4000	0.1000	0.5000	0.9000
3	-0.0838	0.0363	0.2372	0.5203
4	-0.0508	0.0242	0.1864	0.4756
5	-0.0370	0.0185	0.1582	0.4542
6	-0.0293	0.0150	0.1390	0.4400
7	-0.0243	0.0126	0.1248	0.4291
8	-0.0208	0.0109	0.1136	0.4203
9	-0.0181	0.0096	0.1046	0.4128
10	-0.0161	0.0086	0.0971	0.4062

Table 4.1: Correlation between X_i and S_{i-1}

From Table 4.1 we observe that the dependence between X_i and S_{i-1} decreases in i for $\theta_i > 0$, which means that in the long term the effect of S_{i-1} on X_i decreases.

In Table 4.2 and Table 4.3 we compute and present respectively the probability of acceptance of a particular unit and the expected test length for various values of k and d when $\theta_i \in \{0, 0.1, 0.5, 0.9\}$, $i \geq 2$. For all cases, success probabilities of individual start-ups are the same with $P\{X_i = 1\} = 0.7$, $i \geq 1$.

We observe that the probability of acceptance of the unit is decreasing in k and increasing in d, whereas the expected test length is increasing in both k and d. As it can be seen from Table 4.3, expected test length gets smaller as θ_i, $i \geq 2$ get larger, i.e., the start-ups get more dependent.

k	d	$\theta_i = 0$	$\theta_i = 0.1$	$\theta_i = 0.5$	$\theta_i = 0.9$
8	4	0.5696	0.5704	0.5968	0.6727
	5	0.7237	0.7085	0.6661	0.6835
	6	0.8346	0.8117	0.7221	0.6919
	7	0.9067	0.8832	0.7684	0.6989
9	4	0.4925	0.5012	0.5600	0.6662
	5	0.6543	0.6456	0.6325	0.6778
	6	0.7805	0.7605	0.6914	0.6866
	7	0.8689	0.8448	0.7403	0.6938

Table 4.2: Probability of acceptance of a particular unit

k	d	$\theta_i = 0$	$\theta_i = 0.1$	$\theta_i = 0.5$	$\theta_i = 0.9$
8	4	9.3133	9.1396	8.2369	7.1096
	5	10.2345	10.0503	8.9956	7.5141
	6	10.7858	10.6358	9.6046	7.8991
	7	11.0968	10.9954	10.0985	8.2701
9	4	10.0169	9.8386	8.9481	7.8178
	5	11.1692	10.9616	9.8129	8.2396
	6	11.9008	11.7131	10.5115	8.6386
	7	12.3380	12.1955	11.0823	9.0220

Table 4.3: Expected length of TSTF type test

Additionally, in Table 4.4 and Table 4.5 we compute respectively the probability of acceptance of a particular unit and the expected test length for two non-stationary models: M1 when $a_i = \frac{1}{i}$ and $b_i = \frac{1}{i+1}$, and M2 when $a_i = \frac{1}{i}$ and $b_i = \frac{1}{2i}$ in (4.2). For both models, success probabilities of individual start-ups are the same with $P\{X_i = 1\} = 0.7$, $i \geq 1$.

k	d	M1	M2
8	4	0.5679	0.1566
	5	0.6216	0.2021
	6	0.6638	0.2451
	7	0.6978	0.2851
9	4	0.5374	0.1067
	5	0.5920	0.1434
	6	0.6354	0.1797
	7	0.6707	0.2148

Table 4.4: Probability of acceptance of a particular unit for models M1 and M2

k	d	M1	M2
8	4	8.0751	7.7307
	5	8.7756	8.8982
	6	9.3531	9.9568
	7	9.8430	10.9267
9	4	8.7253	7.9620
	5	9.5100	9.2249
	6	10.1593	10.3852
	7	10.7116	11.4605

Table 4.5: Expected length of TSTF type test for models M1 and M2

The probability of acceptance of the unit for the model M1 is substantially greater than that for M2 because the conditional probability of a successful start-up in M1 is larger than that in M2.

Chapter 5

Conclusion

In this thesis, distributions of runs and run-related statistics have been studied under various dependence structures of random binary sequences. The distributions of extreme distances between successive failures have been obtained for Markov dependent trials. These results extend the results in Makri [59]. The distribution of mean distance between successive failures has also been obtained for both independent and identically distributed trials and Markov dependent trials.

Start-up demonstration testing, an application of the theory of runs, has been considered under a different dependence model given by

$$P\left\{X_n = 1 \middle| X_1, X_2, \ldots, X_{n-1}\right\} = a_n + b_n S_{n-1}, \tag{5.1}$$

for $n \geq 2$; such that a_n and b_n satisfy $0 < a_n < 1$ and $0 < a_n + (n-1)\, b_n < 1$ for $n \geq 1$, where S_n, $n \geq 1$, denotes the total number of successes in $X_1, X_2, \ldots, X_n$, and for convenience $P\{S_0 = 0\} = 1$.

One can easily see that the outcome of the present start-up depends on the total number of successful start-ups so far and hence this model is called the *previous-sum dependent model*. The model (5.1) exhibits a stronger form of dependence than the Markov dependence so it may be more appropriate to use in real life applications.

As a future work, we plan to study the distributions of various run statistics under the model given by (5.1). The characteristics of these run statistics under the same model can also be derived.

Appendix A

Calculations of characteristics of TSTF type test for previous-sum dependent and i.i.d. trials

If S_z and $S_z^{(0)}$ denote respectively the number of successes and failures in z start-ups, then we have

$$
\begin{aligned}
P\{Z_{k,d} \geq z\} &= P\left\{\min\left(W_d^{(0)}, W_k^{(1)}\right) \geq z\right\} \\
&= P\left\{W_d^{(0)} \geq z, W_k^{(1)} \geq z\right\} \\
&= P\left\{S_{z-1}^{(0)} \leq d-1, S_{z-1} \leq k-1\right\} \\
&= P\{z-1-S_{z-1} \leq d-1, S_{z-1} \leq k-1\} \\
&= P\{S_{z-1} \geq z-d, S_{z-1} \leq k-1\} \\
&= P\{z-d \leq S_{z-1} \leq k-1\} \\
&= \sum_{i=\max(0,z-d)}^{k-1} P\{S_{z-1} = i\}
\end{aligned}
$$

It should be noted that, for i.i.d. trials, we have

$$P\{S_n = x\} = \binom{n}{x} p^x (1-p)^{n-x}.$$

Hence

$$
\begin{aligned}
P\left\{Z_{k,d} \geq z\right\} &= \sum_{i=\max(0,z-d)}^{k-1} P\left\{S_{z-1} = i\right\} \\
&= \sum_{i=\max(0,z-d)}^{k-1} \binom{z-1}{i} p^{i} \left(1-p\right)^{z-i-1}.
\end{aligned}
$$

The probability of acceptance of the unit can be computed from

$$
\begin{aligned}
P\left\{\text{Acceptance}\right\} &= P\left\{W_k^{(1)} < W_d^{(0)}\right\} \\
&= \sum_i P\left\{W_k^{(1)} < W_d^{(0)} \middle| W_k^{(1)} = i\right\} P\left\{W_k^{(1)} = i\right\} \\
&= \sum_i P\left\{W_d^{(0)} > i, W_k^{(1)} = i\right\} \\
&= \sum_i P\left\{S_i^{(0)} \leq d-1, W_k^{(1)} = i\right\} \\
&= \sum_i P\left\{i - S_i \leq d-1, W_k^{(1)} = i\right\} \\
&= \sum_i P\left\{S_i \geq i-d+1, W_k^{(1)} = i\right\} \\
&= \sum_i P\left\{X_i = 1, S_{i-1} = k-1, S_i \geq i-d+1\right\} \\
&= \sum_{i=k}^{k+d-1} P\left\{X_i = 1, S_{i-1} = k-1\right\}.
\end{aligned}
$$

We have, for previous-sum dependent trials

$$
\begin{aligned}
P\left\{\text{Acceptance}\right\} &= \sum_{i=k}^{k+d-1} P\left\{X_i = 1, S_{i-1} = k-1\right\} \\
&= \sum_{i=k}^{k+d-1} P\left\{X_i = 1 \middle| S_{i-1} = k-1\right\} P\left\{S_{i-1} = k-1\right\} \\
&= \sum_{i=k}^{k+d-1} \left[a_i + b_i\left(k-1\right)\right] P\left\{S_{i-1} = k-1\right\}
\end{aligned}
$$

and for i.i.d. trials

$$\begin{aligned}
P\left\{\text{Acceptance}\right\} &= \sum_{i=k}^{k+d-1} P\left\{X_i = 1, S_{i-1} = k-1\right\} \\
&= \sum_{i=k}^{k+d-1} \binom{i-1}{k-1} p^{k-1}(1-p)^{i-1-(k-1)} p \\
&= \sum_{i=0}^{d-1} \binom{i+k-1}{k-1} p^{k}(1-p)^{i}.
\end{aligned}$$

The conditional probabilities of $Z_{k,d}$ given acceptance and rejection of the unit can be computed, respectively, from

$$\begin{aligned}
P\left\{Z_{k,d} = z \middle| \text{Acceptance}\right\} &= P\left\{Z_{k,d} = z \middle| W_k^{(1)} < W_d^{(0)}\right\} \\
&= \frac{P\left\{Z_{k,d} = z, W_k^{(1)} < W_d^{(0)}\right\}}{P\left\{W_k^{(1)} < W_d^{(0)}\right\}} \\
&= \frac{P\left\{\min\left(W_d^{(0)}, W_k^{(1)}\right) = z, W_k^{(1)} < W_d^{(0)}\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{W_k^{(1)} = z, W_d^{(0)} > z\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{W_k^{(1)} = z, W_d^{(0)} \geq z+1\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{S_z^{(0)} \leq d-1, W_k^{(1)} = z\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{z - S_z \leq d-1, W_k^{(1)} = z\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{S_z \geq z-d+1, W_k^{(1)} = z\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \frac{P\left\{X_z = 1, S_{z-1} = k-1, S_z \geq z-d+1\right\}}{P\left\{\text{Acceptance}\right\}} \\
&= \begin{cases} \frac{P\{X_z=1, S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k \leq z \leq k+d-1, \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

and from

$$
\begin{aligned}
P\left\{Z_{k,d}=z\middle|\text{Rejection}\right\} &= P\left\{Z_{k,d}=z\middle|W_d^{(0)}<W_k^{(1)}\right\}\\
&= \frac{P\left\{Z_{k,d}=z,W_d^{(0)}<W_k^{(1)}\right\}}{P\left\{W_d^{(0)}<W_k^{(1)}\right\}}\\
&= \frac{P\left\{\min\left(W_d^{(0)},W_k^{(1)}\right)=z,W_d^{(0)}<W_k^{(1)}\right\}}{P\left\{\text{Rejection}\right\}}\\
&= \frac{P\left\{W_d^{(0)}=z,W_k^{(1)}>z\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \frac{P\left\{W_d^{(0)}=z,W_k^{(1)}\geq z+1\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \frac{P\left\{S_z\leq k-1,W_d^{(0)}=z\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \frac{P\left\{X_z=0,S_{z-1}^{(0)}=d-1,S_z\leq k-1\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \frac{P\left\{X_z=0,S_{z-1}=z-1-(d-1),S_z\leq k-1\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \frac{P\left\{X_z=0,S_{z-1}=z-d,S_z\leq k-1\right\}}{1-P\left\{\text{Acceptance}\right\}}\\
&= \begin{cases} \frac{P\{X_z=0,S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d\leq z\leq k+d-1,\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}
$$

If the trials are previous-sum dependent, then

$$
\begin{aligned}
&P\left\{Z_{k,d}=z\middle|\text{Acceptance}\right\}\\
&= \begin{cases} \frac{P\{X_z=1,S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k\leq z\leq k+d-1,\\ 0 & \text{otherwise,}\end{cases}\\
&= \begin{cases} \frac{P\left\{X_z=1\middle|S_{z-1}=k-1\right\}P\{S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k\leq z\leq k+d-1,\\ 0 & \text{otherwise,}\end{cases}\\
&= \begin{cases} \frac{[a_z+b_z(k-1)]P\{S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k\leq z\leq k+d-1,\\ 0 & \text{otherwise}\end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
&P\left\{Z_{k,d} = z \middle| \text{Rejection}\right\} \\
= &\begin{cases} \frac{P\{X_z=0, S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d \le z \le k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
= &\begin{cases} \frac{P\left\{X_z=0 \middle| S_{z-1}=z-d\right\} P\{S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d \le z \le k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
= &\begin{cases} \frac{[1-a_z-b_z(z-d)]P\{S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d \le z \le k+d-1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

If the trials are i.i.d., then

$$
\begin{aligned}
&P\left\{Z_{k,d} = z \middle| \text{Acceptance}\right\} \\
= &\begin{cases} \frac{P\{X_z=1, S_{z-1}=k-1\}}{P\{\text{Acceptance}\}} & \text{if } k \le z \le k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
= &\begin{cases} \frac{\binom{z-1}{k-1} p^{k-1}(1-p)^{z-1-(k-1)} p}{\sum_{i=0}^{d-1} \binom{i+k-1}{k-1} p^k (1-p)^i} & \text{if } k \le z \le k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
= &\begin{cases} \frac{\binom{z-1}{k-1} p^k (1-p)^{z-k}}{\sum_{i=0}^{d-1} \binom{i+k-1}{k-1} p^k (1-p)^i} & \text{if } k \le z \le k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
= &\begin{cases} \frac{\binom{z-1}{k-1} (1-p)^{z-k}}{\sum_{i=0}^{d-1} \binom{i+k-1}{k-1} (1-p)^i} & \text{if } k \le z \le k+d-1, \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

and

$$
P\left\{Z_{k,d} = z \middle| \text{Rejection}\right\}
$$

$$
\begin{aligned}
&= \begin{cases} \frac{P\{X_z=0, S_{z-1}=z-d\}}{1-P\{\text{Acceptance}\}} & \text{if } d \leq z \leq k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} \dfrac{\binom{z-1}{z-d} p^{z-d}(1-p)^{z-1-(z-d)}(1-p)}{1-\sum_{i=0}^{d-1}\binom{i+k-1}{k-1} p^k(1-p)^i} & \text{if } d \leq z \leq k+d-1, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} \dfrac{\binom{z-1}{d-1} p^{z-d}(1-p)^{d}}{\sum_{i=0}^{d-1}\binom{i+k-1}{k-1} p^k(1-p)^i} & \text{if } d \leq z \leq k+d-1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

BIBLIOGRAPHY

[1] S. Aki, *Waiting time problems for a sequence of discrete random variables*, Annals of the Institute of Statistical Mathematics **44** (1992), no. 2, 363–378.

[2] S. Aki and K. Hirano, *Distributions of numbers of failures and successes until the first consecutive k successes*, Annals of the Institute of Statistical Mathematics **46** (1994), no. 1, 193–202.

[3] ———, *Joint distributions of numbers of success-runs and failures until the first consecutive k successes*, Annals of the Institute of Statistical Mathematics **47** (1995), no. 2, 225–235.

[4] S. Aki, H. Kuboki, and K. Hirano, *On discrete distributions of order k*, Annals of the Institute of Statistical Mathematics **36** (1984), no. 3, 431–440.

[5] D.L. Antzoulakos, *On waiting time problems associated with runs in markov dependent trials*, Annals of the Institute of Statistical Mathematics **51** (1999), no. 2, 323–330.

[6] ———, *Waiting times for patterns in a sequence of multistate trials*, Journal of Applied Probability **38** (2001), no. 2, 508–518.

[7] D.L. Antzoulakos and S. Chadjiconstantinidis, *Distributions of numbers of success runs of fixed length in markov dependent trials*, Annals of the Institute of Statistical Mathematics **53** (2001), no. 3, 599–619.

[8] D.L. Antzoulakos, M.V. Koutras, and A.C. Rakitzis, *Start-up demonstration tests based on run and scan statistics*, Journal of Quality Technology **41** (2009), no. 1, 48–59.

[9] N. Balakrishnan and P.S. Chan, *Start-up demonstration tests with rejection of units upon observing d failures*, Annals of the Institute of Statistical Mathematics **52** (2000), no. 1, 184–196.

[10] N. Balakrishnan, Balasubramanian K., and R. Viveros, *Start-up demonstration tests under correlation and corrective action*, Naval Research Logistics **42** (1995), 1271–1276.

[11] N. Balakrishnan and M.V. Koutras, *Runs and scans with applications*, Wiley, New York, 2001.

[12] N. Balakrishnan, S.G. Mohanty, and S. Aki, *Start-up demonstration tests under markov dependence model with corrective actions*, Annals of the Institute of Statistical Mathematics **49** (1997), no. 1, 155–169.

[13] E.J. Burr and G. Cane, *Longest run of consecutive observations having a specified attribute*, Biometrika **48** (1961), no. 3, 461–465.

[14] P.S. Chan, H.K.T. Ng, and N. Balakrishnan, *Statistical inference for start-up demonstration tests with rejection of units upon observing d failures*, Journal of Applied Statistics **35** (2008), no. 8, 867–878.

[15] M.T. Chao, J.C. Fu, and M.V. Koutras, *Survey of reliability studies of consecutive-k-out-of-n:f and related systems*, IEEE Transactions on Reliability **44** (1995), no. 1, 120–127.

[16] D.T. Chiang and S. Niu, *Reliability of consecutive-k-out-of-n:f system*, IEEE Transactions on Reliability **30** (1981), no. 1, 87–89.

[17] O. Chryssaphinou, S. Papastavridis, and T. Tsapelas, *On the number of overlapping success runs in a sequence of independent bernoulli trials*, Applications of Fibonacci Numbers, Vol. 5 (eds. G.E. Bergum, A.N. Philippou and A.F. Horadam) (1993), 103–112.

[18] S. Demir and S. Eryilmaz, *Run statistics in a sequence of arbitrarily dependent binary trial*, Statistical Papers **51** (2010), 959–973.

[19] S. Eryilmaz, *On the distribution and expectation of success runs in nonhomogeneous markov dependent trials*, Statistical Papers **46** (2005), 117–128.

[20] ———, *Some results associated with the longest run statistic in a sequence of markov dependent trials*, Applied Mathematics and Computation **175** (2006), 119–130.

[21] ———, *Extension of runs to the continuous-valued sequences*, Statistics & Probability Letters **77** (2007), 383–388.

[22] ———, *Distribution of runs in a sequence of exchangeable multi-state trials*, Statistics & Probability Letters **78** (2008), 1505–1513.

[23] ———, *Mean success run length*, Journal of the Korean Statistical Society **38** (2009), 65–71.

[24] ———, *Review of recent advances in reliability of consecutive k-out-of-n and related systems*, Proceedings of the Institution of Mechanical Engineers, Part O, Journal of Risk and Reliability **224** (2010), 225–237.

[25] ———, *Start-up demonstration tests under markov dependence*, Pakistan Journal of Statistics **26** (2010), no. 4, 637–647.

[26] ———, *Joint distribution of run statistics in partially exchangeable processes*, Statistics & Probability Letters **81** (2011), 163–168.

[27] S. Eryilmaz and S. Chakraborti, *On start-up demonstration tests under exchangeability*, IEEE Transactions on Reliability **57** (2008), no. 4, 627–632.

[28] S. Eryilmaz and S. Demir, *Success runs in a sequence of exchangeable binary trials*, Journal of Statistical Planning and Inference **137** (2007), 2954–2963.

[29] S. Eryilmaz and J.C. Fu, *Runs in continuous-valued sequences*, Statistics & Probability Letters **78** (2008), 759–765.

[30] S. Eryilmaz and A. Stepanov, *Runs in an ordered sequence of random variables*, Metrika **67** (2008), 299–313.

[31] S. Eryilmaz and F. Yalcin, *Distribution of run statistics in partially exchangeable processes*, Metrika **73** (2011), 293–304.

[32] A. Fan, Z. Wang, and F. Ding, *Some limit theorems of runs to the continuous-valued sequence*, Kybernetes **37** (2008), 1279–1286.

[33] W. Feller, *An introduction to probability theory and its applications, vol. i. 3rd ed.*, Wiley, New York, 1968.

[34] J.C. Fu, *Reliability of consecutive k-out-of-n:f system with $k-1$ step markov dependence*, IEEE Transactions on Reliability **R-35** (1986), no. 5, 602–606.

[35] ———, *Distribution theory of runs and patterns associated with a sequence of multi-state trials*, Statistica Sinica **6** (1996), 957–974.

[36] J.C. Fu and M.V. Koutras, *Distribution theory of runs: A markov chain approach*, Journal of the American Statistical Association **89** (1994), 1050–1058.

[37] J.C. Fu and W.Y.W. Lou, *Distribution theory of runs and patterns and its applications, a finite markov chain imbedding approach*, World Scientific Publishing, Singapore, 2003.

[38] J.C. Fu, L. Wang, and W.Y.W. Lou, *On exact and large deviation approximation for the distribution of the longest run in a sequence of two-state markov dependent trials*, Journal of Applied Probability **40** (2003), no. 2, 346–360.

[39] A.E. Gera, *A new start-up demonstration test*, IEEE Transactions on Reliability **59** (2010), no. 1, 128–131.

[40] ———, *A general model for start-up demonstration tests*, IEEE Transactions on Reliability **60** (2011), no. 1, 295–304.

[41] A.P. Godbole, *Specific formulae for some success run distributions*, Statistics & Probability Letters **10** (1990), 119–124.

[42] ———, *The exact and asymptotic distribution of overlapping success runs*, Communications in Statistics - Theory and Methods **21** (1992), no. 4, 953–967.

[43] L. Goldstein, *Poisson approximation and dna sequence matching*, Communications in Statistics - Theory and Methods **19** (1990), no. 4, 4167–4179.

[44] G.J. Hahn and J.B. Gage, *Evaluation of a start-up demonstration test*, Journal of Quality Technology **15** (1983), no. 3, 103–106.

[45] Q. Han and S. Aki, *Joint distributions of runs in a sequence of multi-state trials*, Annals of the Institute of Statistical Mathematics **51** (1999), no. 3, 419–447.

[46] K. Hirano, *Some properties of the distributions of order k*, Fibonacci Numbers and Their Applications (eds. G.E. Bergum, A.N. Philippou and A.F. Horadam) (1986), 43–53.

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

[48] F.K. Hwang, *Simplified reliabilities for consecutive k-out-of-n:f systems*, SIAM Journal on Algebraic and Discrete Methods **7** (1986), 258–264.

[49] K. Inoue and S. Aki, *Joint distributions of numbers of runs of specified lengths in a sequence of markov dependent multistate trials*, Annals of the Institute of Statistical Mathematics **59** (2007), no. 3, 577–595.

[50] K. Inoue, S. Aki, and Hirano K., *Distributions of simple patterns in some kinds of exchageable sequences*, Journal of Statistical Planning and Inference **141** (2011), 2532–2544.

[51] N.L. Johnson, A.W. Kemp, and S. Kotz, *Univariate discrete distributions, 3rd ed.*, Wiley, New York, 1992.

[52] M.V. Koutras, *Waiting time distributions associated with runs of fixed length in two-state markov chains*, Annals of the Institute of Statistical Mathematics **49** (1997), no. 1, 123–139.

[53] M.V. Koutras and V.A. Alexandrou, *Runs, scans and urn model distributions: A unified markov chain approach*, Annals of the Institute of Statistical Mathematics **47** (1995), no. 4, 743–766.

[54] M.V. Koutras and N. Balakrishnan, *A start-up demonstration test using a simple scan-based statistics*, Scan Statistics and Applications, (eds. J. Glaz and N. Balakrishnan) (1999), 251–267.

[55] M. Lambiris and S. Papastavridis, *Exact reliability formulas for linear and circular consecutive k-out-of-n:f systems*, IEEE Transactions on Reliability **R-34** (1985), no. 2, 124–126.

[56] K.D. Ling, *On binomial distribution of order k*, Statistics & Probability Letters **6** (1988), 247–250.

[57] W.Y.W. Lou, *On runs and longest run tests: A method of finite markov chain imbedding*, Journal of the American Statistical Association **91** (1996), 1595–1601.

[58] F.S. Makri, *On occurrences of f-s strings in linearly and circularly ordered binary sequeces*, Journal of Applied Probability **47** (2010), no. 1, 157–178.

[59] ______, *Minimum and maximum distances between failures in binary sequences*, Statistics & Probability Letters **81** (2011), 402–410.

[60] F.S. Makri, A.N. Philippou, and Z.M. Psillakis, *Shortest and longest length of success runs in binary sequences*, Journal of Statistical Planning and Inference **137** (2007), 2226–2239.

[61] ______, *Success run statistics defined on an urn model*, Advances in Applied Probability **39** (2007), 991–1019.

[62] F.S. Makri and Z.M. Psillakis, *On success runs of length exceeded a threshold*, Methodology and Computing in Applied Probability **13** (2011), 269–305.

[63] ______, *Counting certain binary strings*, Journal of Statistical Planning and Inference **142** (2012), 908–924.

[64] D.E.K. Martin, *Markovian start-up demonstration tests with rejection of units upon observing d failures*, European Journal of Operation Research **155** (2004), 474–486.

[65] ______, *Application of auxiliary markov chains to start-up demonstration tests*, European Journal of Operation Research **184** (2008), 574–583.

[66] S.G. Mohanty, *Success runs of length k in markov dependent trials*, Annals of the Institute of Statistical Mathematics **46** (1994), no. 4, 777–796.

[67] A.M. Mood, *The distribution theory of runs*, Annals of Mathematical Statistics **11** (1940), no. 4, 367–392.

[68] F. Mosteller, *Note on an application of runs to quality control charts*, Annals of Mathematical Statistics **12** (1941), no. 2, 228–232.

[69] M. Muselli, *Simple expressions for success run distributions in bernoulli trials*, Statistics & Probability Letters **31** (1996), 121–128.

[70] ______, *Useful inequalities for the longest run distribution*, Statistics & Probability Letters **46** (2000), 239–249.

[71] A.N. Philippou and D.L. Antzoulakos, *Distributions of order k*, International Encyclopedia of Statistical Sciences (ed. M. Lovric) (2011), 400–402.

[72] A.N. Philippou, C. Georghiou, and G.N. Philippou, *A generalized geometric distribution and some of its properties*, Statistics & Probability Letters **1** (1983), 171–175.

[73] A.N. Philippou and F.S. Makri, *Longest success runs and fibonacci type polynomials*, The Fibonacci Quarterly **23** (1985), 338–346.

[74] ______, *Successes, runs, and longest runs*, Statistics & Probability Letters **4** (1986), 101–105.

[75] A.N. Philippou and A.A. Muwafi, *Waiting for the k-th consecutive success and the fibonacci sequence of order k*, The Fibonacci Quarterly **20** (1982), 28–32.

[76] M.B. Rajarshi, *Success runs in a two-state markov chain*, Journal of Applied Probability **11** (1974), no. 1, 190–192.

[77] M.F. Schilling, *The longest run of heads*, The College Mathematics Journal **21** (1990), 196–207.

[78] E.F. Schuster, *Distribution theory of runs via exchangeable random variables*, Statistics & Probability Letters **11** (1991), 379–386.

[79] S. Schwager, *Run probabilities in sequences of markov-dependent trials*, Journal of the American Statistical Association **78** (1983), 168–175.

[80] M.L. Smith and W.S. Griffith, *Start-up demonstration tests based on consecutive successes and total failures*, Journal of Quality Technology **37** (2005), no. 3, 186–198.

[81] ______, *The analysis and comparison of start-up demonstration tests*, European Journal of Operation Research **186** (2008), 1029–1045.

[82] A. Stepanov, *Runs based on the ratios of consecutive order statistics*, Communications in Statistics - Theory and Methods **40** (2011), no. 18, 3252–3268.

[83] V.R.R. Uppuluri and S.A. Patil, *Waiting times and generalized fibonacci sequences*, The Fibonacci Quarterly **21** (1983), 242–249.

[84] E. Vaggelatou, *On the length of the longest run in a multi-state markov chain*, Statistics & Probability Letters **62** (2003), 211–221.

[85] P. Vellaisamy, *On the number of successes in dependent trials*, Communications in Statistics - Theory and Methods **25** (1996), no. 8, 1745–1756.

[86] P. Vellaisamy and S. Sankar, *A unified approach for modeling and designing attribute sampling plans for monitoring dependent production processes*, Methodology and Computing in Applied Probability **7** (2005), 307–323.

[87] P. Vellaisamy and N.S. Upadhye, *On the negative binomial distribution and its generalizations*, Statistics & Probability Letters **77** (2007), 173–180.

[88] R. Viveros and Balakrishnan N., *Statistical inference from start-up demonstration test data*, Journal of Quality Technology **25** (1993), no. 2, 119–130.

[89] A. Wald and J. Wolfowitz, *On a test whether two samples are from the same population*, Annals of Mathematical Statistics **11** (1940), no. 2, 147–162.

[90] J. Wolfowitz, *On the theory of runs with some applications to quality control*, Annals of Mathematical Statistics **14** (1943), no. 3, 280–288.

VITA

Femin Yalçın was born in 1987 in İzmir, Turkey. She got her bachelor's degree in Mathematics as a high honour student in June 2009 and right after that she started her PhD education in Applied Probability and Statistics program in İzmir University of Economics, Turkey. She has been working as a research assistant at Department of Mathematics of the same university since September 2009.

Her research areas include applied probability, stochastic processes, and reliability theory. She has three papers jointly written with Prof. Dr. Serkan Eryılmaz published in journals Metrika, Journal of Computational and Applied Mathematics, and IEEE Transactions on Reliability.

She has been awarded by The Scientific and Technological Research Council of Turkey (TÜBİTAK) with a PhD scholarship since 2009.