THE UNIVERSITY OF MANITOBA

Waiting Time Distributions of Runs and Patterns

by

Yung-Ming Chang

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Statistics
Winnipeg, Manitoba

July, 2002

0-612-79841-0

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

WAITING TIME DISTRIBUTIONS OF RUNS AND PATTERNS

BY

YUNG-MING CHANG

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

Doctor of Philosophy

Abstract

Waiting time distributions of runs and patterns have been successfully used in various areas of statistics and applied probability—for example, in reliability, sampling inspection, quality control, DNA sequencing and hypothesis testing. The main goal of this thesis is to give a comprehensive study of waiting time distributions of runs and patterns using the finite Markov Chain imbedding technique. We provide a simple and general method to obtain the exact distributions, means and probability generating functions for waiting time distributions of compound and later patterns. Computational algorithms based on the finite Markov chain imbedding technique are developed for automatically computing the exact distributions, means and probability generating functions of waiting times for compound and later patterns.

To see the applications of waiting time distributions, we introduce a general theoretical framework that leads to the run-length distribution for a multitude of control charts that are based either on a simple rule (e.g., Shewhart, Cusum, EWMA charts) or on a compound set of rules (e.g., Shewhart with runs rules, robust Cusum and robust EWMA charts). It handles both discrete and continuous cases and can incorporate process properties, such as different types of shifts, directly. The framework is simple to apply and is fully automated.

Acknowledgments

My advisor, Dr. James C. Fu, is the first person to whom I would like to express my gratitude. I thank him for offering me a research assistantship, for leading me into the field of applied probability and statistics, and for his patient guidance and invaluable help on this thesis. In particular, I learned the spirit and persistence required to produce outstanding academic research from him.

I would also like to thank my committee members Dr. R. Shivakumar, Dr. A. Thavaneswaran and Dr. Liqun Wang of the University of Manitoba, and Dr. R. A. Johnson from the University of Wisconsin—Madison for their encouragement and useful suggestions. The financial support of University of Manitoba Graduate Fellowship is also appreciated.

Two individuals are acknowledged for enhancing my professional achievements: Dr. Dennis Murphy, who provided me with useful information about statistical computation and gave me suggestions in English writing skill; and Keoagile Thaga, my officemate and future Ph.D., who regularly discussed interesting statistical issues with me. I would also like to acknowledge those professors in the Department of Statistics for imparting much useful knowledge.

The friendship and support of my friends are unforgettable, particularly my good friend, Wen-Jung Hsiao, who has been so helpful in my personal life over the past few years.

Finally, I would like to express my sincerest gratitude and dedicate this thesis to my parents: Hu-Sheng Chang and Jui-Mei Huang. These are the people who brought me up into being what I am. The completion of this work was solely due to their steadfast love, support and encouragement. I would further like to thank my sisters: Shu-Jen Chang, Shu-Hui Chang and Shu-Yuan Chang for their strong support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

The distribution theory of runs and patterns has been widely studied and applied in many fields such as reliability (Chiang and Niu, 1981; Fu, 1985, 1986a, 1986b, 1993; Chao and Fu, 1989, 1991; Chao, Fu and Koutras, 1995), sampling inspection (Shmueli and Cohen, 2000), quality control (Mosteller, 1941; Wolfowitz, 1943; Fu, Spiring and Xie, 2002), hypothesis testing (Wald and Wolfowitz, 1940; Wolfowitz, 1943; Walsh, 1962; Lou, 1996, 1997), DNA sequencing (Waterman, 1995; Fu, Lou and Chen, 1999), psychology (Schwager, 1983; Koutras and Alexandrou, 1997) and ecology (Schwager, 1983). Traditionally, most of the research work focused on the study of runs in a sequence of bistate trials. There are two general types of problems that arise in the study of runs: (i) the distribution of the number of occurrences of k consecutive successes (or failures), and (ii) the distribution of the number of trials (waiting time) to observe the first (or rth) occurrence of k consecutive successes (or failures). The latter case is known as the waiting time problem, which is the main focus of this thesis.

Historically, the distribution theory of runs has been of scientific interest since the time of De Moivre (1667-1754) (see Johnson, Kotz and Kemp, 1992, page 426). From around 1940 to 1970, there were many papers that contributed to this area, but most were concerned with the study of conditional distributions of runs given the total number of successes or of deriving approximate formulas for the distributions of runs (Fu, 1996). In the past two decades, more complicated problems have been proposed and treated based on different random sequences (for example, Markov dependent trials). This work has not only been focused on runs but also on the more general concept of patterns.

Traditionally, a combinatorial approach was adopted to study the distribution theory of runs and patterns. However, from both theoretical and computational points of view, it always entails heavy and tedious task, if not impossible. In addition to the combinatorial approach, two popular approaches have been extensively used in this field over the past decade. One is the *conditional probability generating functions approach* (e.g., Ebneshahrashoob and Sobel, 1990; Aki, 1992, 1997; Aki and Hirano, 1995; Aki, Balakrishnan and Mohanty, 1996; Hirano, Aki and Uchida, 1997), while the other approach is the *finite Markov chain imbedding technique* introduced by Fu and Koutras (1994) and Fu (1996). The finite Markov chain imbedding technique has certain practical advantages; for example, it can be used to deal with very general classes of waiting time problems. Throughout this thesis, our work will be based on this approach.

The remainder of this chapter introduces and gives a brief literature review of several important distributions associated with runs and patterns; in particular, distributions of order k and sooner and later waiting time distributions. These distributions are central to the work in this thesis.

1.2 Distributions of Order k

Distributions of order k are among the most important class set of distributions associated with runs. Philippou and Muwafi (1982) studied the distribution of the waiting time until the first occurrence of k consecutive successes (k fixed) in a sequence of Bernoulli trials with success probability p ($q = 1 - p$). Philippou, Georghiou and Philippou (1983) called this distribution a *Geometric distribution of order k*. They further defined a *Negative Binomial distribution of order k* and a *Poisson distribution of order k*. Subsequently, various distributions of order k have been extensively studied. Since the class of this type of distribution is rather large, we only introduce some important distributions related to our work. To that end, let X denote the random variable of interest, which has some distribution of order k.

Geometric distribution of order k ($G_k(p)$)

The expression of the exact distribution given by Philippou and Muwafi (1982) is

$$P(X=x) = \sum_{x_1,\ldots,x_k} \binom{x_1+\cdots+x_k}{x_1,\ldots,x_k} p^x \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}, \tag{1.1}$$

for $x = k, k+1, \ldots$, where the summation is over all nonnegative integers $x_1, \cdots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x - k$. Clearly, it reduces to the usual Geometric distribution when $k = 1$.

Negative Binomial distribution of order k ($NB_k(r,p)$)

The distribution of the waiting time until the rth occurrence of k consecutive successes is called the *Negative Binomial distribution of order k*. It is clear that a Geometric distribution of order k is a special case of a Negative Binomial distribution of order k when $r = 1$. The exact distribution given by Philippou, Georghiou and Philippou (1983) is

$$P(X=x) = \sum_{x_1,\ldots,x_k} \binom{x_1+\cdots+x_k+r-1}{x_1,\ldots,x_k,r-1} p^x \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}, \tag{1.2}$$

for $x = kr, kr+1, \ldots$, where the summation is over all nonnegative integers $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x - kr$.

Poisson distribution of order k ($P_k(p)$)

This distribution is obtained as a limiting form of a Negative Binomial distribution of order k. Let X_r be a random variable distributed as $NB_k(r,p)$ and assume that

$q \to 0$ and $\lim_{r\to\infty} rq = \lambda$ ($\lambda > 0$). It can be shown (Philippou, Georghiou and Philippou, 1983) that

$$P(X_r - kr = x) \to e^{-k\lambda} \sum_{x_1,\ldots,x_k} \frac{\lambda^{(x_1+\cdots+x_k)}}{x_1!\cdots x_k!},\ x = 0, 1, 2, \ldots, \tag{1.3}$$

as $r \to \infty$, where the summation is over all nonnegative integers $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x$. The limit form in Equation (1.3) is called a *Poisson distribution of order k*. When $k = 1$, it reduces to the usual Poisson distribution.

Logarithmic (series) distribution of order k ($LS_k(p)$)

Like the Poisson distribution of order k, this distribution is also derived as a limiting form of a Negative Binomial distribution of order k. Let X_r be a random variable distributed as $NB_k(r,p)$. Assuming that $r \to 0$, it can be shown (Aki, Kuboki and Hirano, 1984) that

$$P(X_r = x | X_r \geq [kr] + 1) \to \frac{p^x}{-k \log p} \sum_{x_1,\ldots,x_k} \frac{(x_1 + \cdots + x_k - 1)!}{x_1!\cdots x_k!} \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}, \tag{1.4}$$

for $x = 1, 2, \ldots$, where the summation is over all nonnegative integers $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x$. The limit form in Equation (1.4) is called a *Logarithmic (series) distribution of order k.*

Binomial distribution of order k ($B_k(n,p)$)

The distribution of the number of occurrences of k consecutive successes is called the *Binomial distribution of order k*. The exact distribution was derived independently by both Hirano (1986) and Philippou and Makri (1986) as

$$P(X = x) = \sum_{i=0}^{k-1} \sum_{x_1,\ldots,x_k} \binom{x_1 + \cdots + x_k + x}{x_1, \ldots, x_k, x} p^n \left(\frac{q}{p}\right)^{x_1+\cdots+x_k}, \tag{1.5}$$

for $x = 0, 1, 2, \ldots, [n/k]$, where the inner summation is over all nonnegative integers $x_1, \ldots, x_k$ such that $x_1+2x_2+\cdots+kx_k = n-i-kx$. It reduces to the usual Binomial distribution when $k = 1$. Fu and Koutras (1994) gave a different formula by using the finite Markov chain imbedding technique which is introduced in Chapter 2.

Other distributions of order k

In addition to the above, there are many distributions of order k that have been investigated; for example, the Compound Poisson distribution of order k (e.g., Philippou, 1983; Panaretos and Xekalaki, 1986), the hypergeometric and inverse (or negative) hypergeometric distribution of order k (e.g., Panaretos and Xekalaki, 1986; Godbole, 1990b), the Pólya and inverse Pólya distributions of order k (e.g., Philippou, Tripsiannis and Antzoulakos, 1989; Philippou and Tripsiannis, 1991; Tripsiannis, 1993), etc. No further introduction of these distributions is given since they exceed the scope of this thesis.

Remark 1.1 The Negative Binomial and Binomial distributions of order k discussed above assume *nonoverlapping* counting. This type of counting produces *Type I distributions of order k*. Similarly, *at least k* and *overlapping* counting schemes produce *Type II* and *Type III distributions of order k*, respectively (Balakrishnan and Koutras, 2002). These counting schemes are defined more precisely in Chapter 2.

Since the formulas for the exact distributions of order k are quite complicated, alternative formulas have been derived to avoid difficulty in numerical computation. For example, Uppuluri and Patil (1983) obtained a simpler formula for Geometric distribution of order k as

$$\begin{aligned} P(X=x) &= p^k \sum_{j=0}^{\infty} (-1)^j \binom{x-k-jk}{j} (qp^k)^j \\ &- p^{k+1} \sum_{j=0}^{\infty} (-1)^j \binom{x-k-jk-1}{j} (qp^k)^j, \quad x \geq k, \qquad (1.6) \end{aligned}$$

which involves only two single summations. Muselli (1996) gave a more computationally attractive expression for the distribution as

$$P(X = x) = \sum_{j=1}^{\left[\frac{x+1}{k+1}\right]} (-1)^{j-1} p^{jk} q^{j-1} \left\{ \binom{x - jk - 1}{j - 2} + q \binom{x - jk - 1}{j - 1} \right\}, \tag{1.7}$$

which entails only a single summation. Another way to avoid computational difficulties is to derive the probability generating function of a distribution of order k. Probability generating functions are useful not only in finding an exact distribution, but also in studying the characteristics of that distribution.

Replacing different underlying sequences, there were numerous papers that extended the study of distributions of order k. Several authors dealt with these distributions based on Markov dependent trials (e.g., Hirano and Aki, 1993; Mohanty, 1994). Aki (1985) defined a binary sequence of order k as an extension of a sequence of Bernoulli trials. He studied several distributions of order k based on this sequence and called the resulting class of distributions *extended distributions of order k* (see also Hirano and Aki, 1987). Philippou (1988) developed a new class of distributions called *multiparameter distributions of order k* (equivalent to extended distributions of order k by appropriately changing the parameters). Further generalizations of multiparameter distributions of order k have been developed (e.g., Philippou, Antzoulakos and Tripsiannis, 1989; Philippou and Antzoulakos, 1990; Antzoulakos and Philippou, 1997).

In addition to the study of exact distributions, Poisson approximations of these distributions have been derived by several authors. Chen-Stein approximations may be the most popular method of treating such problems. References include: Fu (1985, 1986a, 1986b, 1993), Arratia, Goldstein and Gordon (1989, 1990), Godbole (1990a, 1991), Goldstein (1990) and Wang (1993).

1.3 Sooner and Later Waiting Time Distributions

In many situations, it is necessary to study waiting time problems associated with two or more runs or patterns; for example, an experiment stops or a system fails whenever one of several predefined runs or patterns occurs. Sometimes, order among these runs or patterns is important and must be taken into consideration; for example, the DNA sequence of a virus contains certain patterns that occur in order. In the previous section, two types of distributions of order k that belong to the class of waiting time distributions have been introduced: the Geometric and Negative Binomial distributions of order k. In this section, we introduce another important class of waiting time distributions called *sooner and later waiting time distributions.*

The distribution of the waiting time until the first occurrence of a success run or a failure run (of fixed length) in a sequence of Bernoulli trials was first introduced by Feller (1968). Ebneshahrashoob and Sobel (1990) called this a *sooner waiting time distribution* and referred to this type of problem as a succession quota (SQ) problem or a sooner waiting time problem. They also considered as a dual the SQ later problem—waiting for a success run or a failure run, whichever comes later—and derived explicit formulas for their probability generating functions. The resulting distribution is termed a *later waiting time distribution.* Generally speaking, sooner and later waiting time distributions refer to the waiting times for several simple patterns, whichever comes sooner or later.

From 1980 to 1990, some general results were derived for the waiting time to the first occurrence of one or more specified patterns; see, for example, Li (1980), Blom and Thorburn (1982), Breen, Waterman and Zhang (1985) and Chryssaphinou and Papastavridis (1990). However, most of these results were discussed under some conditions or explained the routes for deriving the probability generating functions of such waiting times but did not provide explicit expressions as Ebneshahrashoob and Sobel did. Since Ebneshahrashoob and Sobel's work, the sooner and later waiting time problems have been extensively studied. Ling and Low (1993) generalized Ebneshahrashoob and Sobel's results in a sequence of independent and identically distributed (i.i.d.) multistate trials. Aki and Hirano (1993) and Balasubramanian, Viveros and Balakrishnan (1993) studied this problem in a sequence of homogeneous Markov bistate trials. Other pertinent references include: Uchida and Aki (1995), Aki, Balakrishnan and Mohanty (1996), Koutras (1997a, 1997b), Aki and Hirano (1999), Antzoulakos (1999), Han and Aki (2000a, 2000b). Most of this research included the study of sooner and later waiting time distributions of a success run and a failure run in a sequence of bistate trials. Koutras and Alexandrou (1997) investigated the sooner waiting time problems in a sequence of trinary trials. Uchida (1998) derived probability generating functions for sooner waiting times of countably many simple patterns, and for later waiting times of two simple patterns in a sequence of i.i.d. multistate trials, but under a very restrictive assumption that each simple pattern was aligned in ascending order. To the best of our knowledge, there are no general results for sooner and later waiting time distributions of l ($l > 2$) simple patterns when the underlying sequence consists of Markov dependent multistate trials. In particular, when we consider the later waiting time problem of several simple patterns, the enumeration schemes (nonoverlapping and overlapping) should be taken into consideration.

One primary goal of the thesis is to develop a simple and general method, both from theoretical and computational points of view, to deal with the sooner and later waiting time problems. Our results are presented in Chapters 3 and 4.

1.4 Summary

In this chapter, we have briefly reviewed and discussed: (i) primary developments in the study of the distribution theory of runs and patterns, (ii) approaches that are often adopted in studying the distribution theory of runs and patterns, and (iii) some well-known distributions associated with runs and patterns.

The rest of this thesis is organized as follows. In Chapter 2, we introduce some basic concepts of runs and patterns and the finite Markov chain imbedding technique. Chapter 3 and Chapter 4 give a comprehensive study of waiting time distributions of compound and later patterns, respectively. Numerical examples are given to illustrate our results. In Chapter 5, computer algorithms based on the finite Markov chain imbedding technique are developed for automatically obtaining the results derived in Chapters 3 and 4. Chapter 6 shows the application of waiting time distributions in quality control. Finally, in Chapter 7, we extend the results obtained in Chapter 3 and list some open problems for future research.

Chapter 2

Finite Markov Chain Imbedding

2.1 Basic Concepts of Runs and Patterns

Let $\{X_i\}$ be a sequence of m-state ($m \geq 2$) random variables defined on the state space $\Gamma = \{b_1, b_2, \ldots, b_m\}$. Traditionally, a run means a finite sequence of consecutive successes or failures. For example, the sequence $SSSSS$ means a success run of length 5. For multistate trials, a run is defined to be a finite sequence of consecutive identical symbols. Due to recent rapid developments in science, this definition has become rather restrictive and is not sufficient for solving more complicated problems. For broader applications, we require a more general definition.

Definition 2.1 We say that Λ is a *simple pattern* if Λ is composed of a specified sequence of k states; i.e. $\Lambda = b_{i_1} \cdots b_{i_k}$ (the length of the pattern k is fixed, and the states in the pattern are allowed to be repeated).

It is clear that a success run (or a failure run) of length k (k fixed) is a special case of a simple pattern. We define a *subpattern* of a simple pattern Λ to be a finite sequence having the general form $b_{i_1} \cdots b_{i_j}$, $1 \leq j \leq k$. It is clear that a simple pattern is a subpattern of itself. The subpattern plays an important role in the finite Markov chain imbedding technique.

Define a *segment* to be any (contiguous) subset of a simple pattern. For example, let $\Lambda = b_1 b_1 b_2 b_2$ be a simple pattern; then, the subpatterns b_1, $b_1 b_1$ and $b_1 b_1 b_2$ are segments of Λ. On the other hand, b_2, $b_1 b_2$, $b_2 b_2$ and $b_1 b_2 b_2$ are segments of Λ, but not subpatterns. Let Λ_1 and Λ_2 be two simple patterns with lengths k_1 and k_2, respectively. We say that Λ_1 and Λ_2 are *distinct* if neither is a segment of the other. We define the union $\Lambda_1 \cup \Lambda_2$ to be the occurrence of either the pattern Λ_1 or the pattern Λ_2.

Definition 2.2 We say that Λ is a *compound pattern* if it is a union of l distinct simple patterns; i.e. $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$ (the lengths of simple patterns do not have to be the same).

It is obvious that a compound pattern reduces to a simple pattern when $l = 1$.

Definition 2.3 Let $\Lambda_1, \ldots, \Lambda_l$ be l distinct simple patterns. We say that $\sigma = \Lambda_1 \circ \Lambda_2 \circ \cdots \circ \Lambda_l$ is an *ordered series pattern* if Λ_1 is the first to occur among the patterns $\Lambda_1, \ldots, \Lambda_l$, Λ_2 is the next to occur among the patterns $\Lambda_2, \ldots, \Lambda_l$, and so on.

From the above definition, we see that an ordered series pattern is formed by observing the first occurrence of the simple patterns Λ_i, $i = 1, \ldots, l$, in the defined ordering. To clarify this definition, we provide the following example.

Example 2.1 Let $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ be an ordered series pattern with $\Lambda_1 = 13$, $\Lambda_2 = 22$ and $\Lambda_3 = 31$. Consider two realizations R_1 and R_2 of a sequence of eighteen three-state trials:

$$R_1 : \; 112\underline{13}3213\underline{22}12213\underline{31},$$

$$R_2 : \; 11\overline{22}\underline{13}323\underline{22}12213\underline{31}.$$

It is easy to see that σ occurs in R_1, however, σ does not occur in R_2 since Λ_1 is not the first one to occur in the ordering as specified by σ. On the other hand, the ordered series pattern $\Lambda_2 \circ \Lambda_1 \circ \Lambda_3$ has occurred in R_2. Also note that in R_1, Λ_1 occurs three times before Λ_3, and Λ_2 occurs twice before Λ_3.

Let $\mathcal{I} = \{1, 2, \ldots, l\}$ and define $\mathcal{P}$ to be the set of all ordered series patterns generated by permuting l distinct simple patterns $\Lambda_1, \ldots, \Lambda_l$; that is,

$$\mathcal{P} = \{\sigma_i : \sigma_i = \Lambda_{i_1} \circ \Lambda_{i_2} \circ \cdots \circ \Lambda_{i_l},\ i = 1, \ldots, l!,\ i_j \in \mathcal{I} \text{ and } i_j \neq i_k \text{ for } j \neq k\}. \quad (2.1)$$

For example, for $l = 2$, we have $\mathcal{P} = \{\sigma_1, \sigma_2 : \sigma_1 = \Lambda_1 \circ \Lambda_2 \text{ and } \sigma_2 = \Lambda_2 \circ \Lambda_1\}$. It is easy to verify that $card(\mathcal{P}) = l!$ ($l!$ permutations).

Definition 2.4 We say that Λ_L is a *later pattern* if it is a union of all $l!$ ordered series patterns in $\mathcal{P}$; i.e. $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$, for each $\sigma_i \in \mathcal{P}$.

Remark 2.1 Given a later pattern $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$, by definition, it should be clear that any two ordered series patterns σ_i and σ_j $(i \neq j)$ cannot occur at the same time. For example, let $\sigma_1 = 11 \circ 22 \circ 33$ and $\sigma_2 = 22 \circ 11 \circ 33$. Consider a realization of a sequence of twelve three-state trials: $\underline{11}3\underline{\overline{22}}3\overline{11}21\underline{\overline{33}}$. It is easy to see that σ_1 occurs on the twelfth trial since the simple patterns '11', '22' and '33' occur in order and '11' is the first one to occur in the sequence; however, σ_2 does not occur on the twelfth trial since the pattern '22' is not the first one to occur in the order.

Next, we consider the waiting time problems. Define the waiting time for a simple pattern $\Lambda = b_{i_1} b_{i_2} \cdots b_{i_k}$ to be

$$\begin{aligned} W(\Lambda) &= \inf\{n : X_{n-k+1} = b_{i_1}, \ldots, X_n = b_{i_k}\} \\ &= \text{Minimum number of trials required to observe the pattern } \Lambda, \end{aligned}$$

and define the waiting time for a compound pattern $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$ as

$$\begin{aligned} W(\Lambda) &= \text{Minimum number of trials required to observe one of the} \\ &\quad\ \text{simple patterns } \Lambda_1, \ldots, \Lambda_l. \end{aligned}$$

For given integer r, $r = 1, 2, \ldots$, we define the waiting time $W(r : \Lambda)$ to the rth occurrence of a pattern Λ (simple or compound) as

$$\begin{aligned} W(r : \Lambda) &= \inf\{n : n \text{ is the number of trials required to observe the } r\text{th} \\ &\quad\ \text{occurrence of the pattern } \Lambda\}. \end{aligned}$$

Similarly, we define the waiting time for an ordered series pattern $\sigma = \Lambda_1 \circ \cdots \circ \Lambda_l$ as

$$W(\sigma) = \text{Minimum number of trials required to observe the pattern } \sigma,$$

and define the waiting time for a later pattern $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$ as

$$W(\Lambda_L) = \begin{array}{l}\text{Minimum number of trials required to observe one of the}\\ \text{ordered series patterns } \sigma_1, \ldots, \sigma_{l!}.\end{array}$$

Clearly, waiting time problems associated with a compound pattern $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$ and a later pattern $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$ ($l \geq 2$) are the sooner and later waiting time problems, respectively. From the above definitions, it is easy to see that the later waiting time distribution of l ($l \geq 2$) simple patterns can be viewed as the waiting time distribution of a compound pattern defined through all the ordered series patterns generated by permuting all the simple patterns. From this point of view, the later waiting time problems can be treated in a similar way as the sooner case.

With these definitions of patterns, we further introduce different counting schemes. Four of the most frequently used counting schemes are:

- Nonoverlapping counting in the sense of Feller (1968): recounting immediately after a given pattern has occurred.

- Overlapping counting in the sense of Ling (1988): when a given pattern (with length k) occurs, start counting backward (up to the last $(k-1)$th trial) to find the overlap of the current pattern and the next occurring pattern, recounting from this subpattern.

- Exactly k counting in the sense of Mood (1940): counting the number of runs of exact length k.

- At least k counting: counting the number of runs of length greater than or equal to k.

The first two counting schemes can be used for any pattern; however, the last two counting schemes can only be used for runs. Given a specified pattern Λ, let $X_n(\Lambda)$ be the number of occurrences of Λ in a sequence of n multistate trials with respect to nonoverlapping or overlapping counting. Similarly, let $E_{n,k}$ and $G_{n,k}$ be the numbers of success runs of length k in a sequence of n bistate trials with respect to exactly k and at least k counting, respectively. We give an example to illustrate these counting schemes.

Example 2.2 Suppose we flip a coin fifteen times with outcomes

$$SSSSFFFSFSFSSFS,$$

where the success event S denotes a head. Consider a simple pattern $\Lambda = SS$; then we have $X_{15}(\Lambda) = 3$ under nonoverlapping counting $(\underline{SS}\overline{SS}FFFSFSF\underline{SS}FS)$ and $X_{15}(\Lambda) = 4$ under overlapping counting $(\overline{S\underline{S}}\underline{\underline{S}S}FFFSFSF\underline{SS}FS)$. Similarly, for $k = 2$, then we have $E_{15,2} = 1$ $(SSSSFFFSFSF\underline{SS}FS)$ and $G_{15,2} = 2$ $(\underline{SSSS}FFFSFSF\underline{SS}FS)$. Consider another simple pattern $\Lambda = SFS$; then we have $X_{15}(\Lambda) = 2$ under nonoverlapping counting $(SSSSFFF\underline{SFS}FS\underline{SFS})$ and $X_{15}(\Lambda) = 3$ under overlapping counting $(SSSSFFF\underline{SF\overline{S}}\overline{FS}S\underline{SFS})$. Obviously, it does not make sense if the exactly k and at least k counting were used for this pattern since it is not a run.

To close this section, we point out that the distributions generated by compound patterns cover many well-known distributions; for example, Binomial distribution, Binomial distribution of order k, geometric distribution, geometric distribution of order k, negative binomial distribution, negative binomial distribution of order k, sooner and later waiting time distributions and distributions of the scan statistic. Table 2.1 summarizes patterns corresponding to different random variables and their distributions.

Remark 2.2 Two things should be noted in Table 2.1: (i) the random variable associated with Binomial distribution of order k is often denoted by $N_{n,k}$ under nonoverlapping counting, and by $M_{n,k}$ under overlapping counting; and (ii) the scan statistic $S_n(r)$ of window size r in a sequence of bistate trials is defined as

$$S_n(r) = \max_{1 \leq t \leq n-r+1} \sum_{k=t}^{t+r-1} X_k. \tag{2.2}$$

We refer to Fu (2002) for the study of the exact distribution of $S_n(r)$.

2.2 Finite Markov Chain Imbedding

2.2.1 Introduction

The finite Markov chain imbedding technique was first employed by Fu (1986b) and successfully used by Chao and Fu (1989, 1991) in studying the reliability of a large series system, such as repairable systems and consecutive-k-out-of-n:F systems. Fu and Koutras (1994) gave a complete introduction to this approach and applied it to the study of distributions of runs in a sequence of Bernoulli trials with respect to different counting schemes. Since their work, the finite Markov chain imbedding technique has become a popular approach to study the distribution theory of runs and patterns and its related applications; see, for example, Lou (1996, 1997), Koutras (1997b), Koutras and Alexandrou (1997), Doi and Yamamoto (1998), Boutsikas and Koutras (2000).

The fundamental idea of the finite Markov chain imbedding technique is to imbed the random variable of interest into a Markov chain. Since the probabilistic

Table 2.1: Some well-known distributions generated by compound patterns.

Distribution	Compound pattern	Random variable	Standard notation
Binomial	$\Lambda = S$	$X_n(\Lambda)$	$B(n,p)$
Binomial distribution of order k	$\Lambda = S \cdots S$ of length k	$X_n(\Lambda)$	$B_k(n,p)$
Geometric	$\Lambda = S$	$W(\Lambda)$	$G(p)$
Geometric distribution of order k	$\Lambda = S \cdots S$ of length k	$W(\Lambda)$	$G_k(p)$
Negative Binomial	$\Lambda = S$	$W(r:\Lambda)$	$NB(r,p)$
Negative Binomial distribution of order k	$\Lambda = S \cdots S$ of length k	$W(r:\Lambda)$	$NB_k(r,p)$
Sooner waiting time	$\Lambda = \bigcup_{i=1}^{l} \Lambda_i,\ l \geq 2$	$W(\Lambda)$	
Later waiting time	$\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i,.\ l \geq 2$	$W(\Lambda_L)$	
Scan statistic		$S_n(r)$	

behavior of a Markov chain is uniquely characterized by its transition probability matrix, the exact distribution of the random variable can be expressed in a simple form in terms of the transition probability matrix of its imbedded chain. To see how it works, we first introduce the notion *finite Markov chain imbeddable.*

Definition 2.5 We say that a random variable $X_n(\Lambda)$ is *finite Markov chain imbeddable* if: (i) there exists a finite Markov chain $\{Y_t : t = 0, 1, \ldots, n\}$ defined on a finite state space $\Omega = \{a_1, a_2, \ldots, a_m\}$ with initial probability $\boldsymbol{\xi}_0$ and transition probability matrices $\boldsymbol{M}_t$, $t = 1, \ldots, n$, and (ii) there exists a partition $\{C_x : x = 0, 1, \cdots, l\}$ of the state space Ω (l may depend on n), such that

$$P(X_n(\Lambda) = x) = P(Y_n \in C_x | \boldsymbol{\xi}_0)$$

for each $x = 0, 1, \ldots, l$.

The next theorem derived by Fu and Koutras (1994) provides a formula for computing the exact distributions of $X_n(\Lambda)$. We state it without proof.

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

$$P(X_n(\Lambda) = x) = \boldsymbol{\xi}_0 \left(\prod_{t=1}^{n} \boldsymbol{M}_t \right) \boldsymbol{U}'(C_x), \ x = 0, 1, \ldots, l, \tag{2.3}$$

where $\boldsymbol{\xi}_0 = P(Y_0 = a_1, Y_0 = a_2, \ldots, Y_0 = a_m)$, $\boldsymbol{M}_t$, $t = 1, \ldots, n$, are $m \times m$ transition probability matrices associated with $\{Y_t\}_{t=0}^{n}$ and $\boldsymbol{U}(C_x) = \sum_{i:a_i \in C_x} \boldsymbol{e}_i$, and where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ is a unit vector corresponding to a_i.

Equation (2.3) is a matrix version of the Chapman-Kolmogorov equation. It is obvious that Equation (2.3) can be written as

$$P(X_n(\Lambda) = x) = \boldsymbol{\xi}_0 \boldsymbol{M}^n \boldsymbol{U}'(C_x), \ x = 0, 1, \ldots, l,$$

when the imbedded Markov chain is homogeneous. Finding the exact distribution of $X_n(\Lambda)$ may not be trivial since the pattern may be very complicated. According to Fu (1996), there are three essential steps for finding the distribution of a given pattern by using the finite Markov chain technique: (i) construction of a proper state space Ω based on the structure of a specified pattern; (ii) construction of a finite Markov chain and its transition probability matrix; and (iii) construction of a partition $\{C_x\}$ on the state space Ω that has a one-to-one correspondence with the random variable $X_n(\Lambda)$ in the sense that $P(X_n(\Lambda) = x) = P(Y_n \in C_x | \boldsymbol{\xi}_0)$ for all x. We give a simple example to make the above procedure more transparent.

Example 2.3 Let $\{X_i\}_{i=1}^n$ be a sequence of Bernoulli trials with success probability p and failure probability q, respectively, and let $\Lambda = S$. Then the distribution of $X_n(\Lambda)$ is the usual Binomial distribution under nonoverlapping counting. To find such a distribution by using the finite Markov chain imbedding technique, we proceed as follows. Firstly, let $Y_t = \sum_{i=1}^t X_i$, $1 \le t \le n$. Then it is easy to check that $\{Y_t\}$ is a Markov chain with state space $\Omega = \{0, 1, \ldots, n\}$. Secondly, we define the transition probabilities as

$$p_{ij} = P\left(Y_t = j | Y_{t-1} = i\right) = \begin{cases} p & \text{if } j = i+1, \\ q & \text{if } j = i, \end{cases}$$

for $0 \le i \le n-1$, $p_{nn} = 1$, and 0 elsewhere. Then, the transition probability matrix can be written as

$$\boldsymbol{M} = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ n-1 \\ n \end{array} \left[\begin{array}{cccccc} q & p & & & & \\ & q & p & & 0 & \\ & & q & p & & \\ & & & \ddots & \ddots & \\ & 0 & & & q & p \\ & & & & & 1 \end{array} \right].$$

Thirdly, we define $C_x = \{x\}$ for $x = 0, 1, \ldots, n$. Then $\{C_x : x = 0, 1, \ldots, n\}$ forms a partition of the state space Ω. From Theorem 2.1, we have

$$P(X_n(\Lambda) = x) = \boldsymbol{\xi}_0 \boldsymbol{M}^n \boldsymbol{e}_{x+1}', \quad x = 0, 1, \ldots, n, \tag{2.4}$$

where $\boldsymbol{\xi}_0 = (1, 0, \ldots, 0)_{1\times(n+1)}$ is the initial distribution ($P(Y_0 = 0) \equiv 1$) and $\boldsymbol{e}'_{x+1}$ is the transpose of the unit vector $\boldsymbol{e}_{x+1} = (0, \ldots, 1, \ldots, 0)_{1\times(n+1)}$. It can be shown that Equation (2.4) is in fact equivalent to the usual expression for the Binomial distribution; that is,

$$\boldsymbol{\xi}_0 \boldsymbol{M}^n \boldsymbol{e}'_{x+1} = \binom{n}{x} p^x q^{n-x}.$$

The proof can be found in result A1 of the Appendix.

Generally speaking, construction of an imbedded Markov chain $\{Y_t\}$ associated with the random variable of interest may not be as simple as in the above example. For more complicated patterns or random variables, it may not be enough for $\{Y_t\}$ to record information with only one component. In this situation, we can define Y_t with two (or more, if necessary) components (or dimensions) to record enough information. Usually, the first component records the total number of occurrences of a given pattern in the first t trials and the second component records the status of Y_t with respect to different counting schemes or conditions at trial t. Fu (1996) introduced the *forward and backward principle* in studying the distributions of the number of runs and patterns and the waiting time distributions. We give a detailed discussion in the next section.

2.2.2 Forward and Backward Principle

When dealing with problems regarding the distribution theory of runs and patterns via the finite Markov chain imbedding technique, we may encounter some difficulties. For example, how do we imbed a random variable of interest associated with a specified pattern into a Markov chain? The forward and backward principle introduced by Fu (1996) provides a general way to analyze this type of problem and facilitates the study of the finite Markov chain imbedding technique. The idea of the forward and backward principle mainly consists of two parts: (i) understanding of the structure of a specified pattern, and (ii) the counting procedure applied throughout the sequence of n multistate trials. To illustrate this principle, consider a sequence of n i.i.d. m-state trials $\{X_i\}$ defined on the state space $\Gamma = \{b_1, b_2, \ldots, b_m\}$. Given a simple pattern $\Lambda = b_{i_1} b_{i_2} \cdots b_{i_k}$ of length k, suppose we are interested in finding the exact distribution of $X_n(\Lambda)$ with respect to nonoverlapping counting. We proceed as follows:

(i) Decompose the pattern $\Lambda = b_{i_1} b_{i_2} \cdots b_{i_k}$ *forward* into $k-1$ subpatterns labelled $1 = b_{i_1}$, $2 = b_{i_1} b_{i_2}$, …, $k-1 = b_{i_1} b_{i_2} \cdots b_{i_{k-1}}$ and let '0' stand for none of the subpatterns '1',…,'$k-1$'. These subpatterns (including 0) are called *ending blocks.*

(ii) Let $\eta = (x_1, \ldots, x_n)$ be a realization of the sequence $\{X_i\}$, where x_i is the outcome of the ith trial. We define a Markov chain $\{Y_t\}_{i=0}^{n}$ operating on η to be $Y_t(\eta) = (u, v)$ for each $t = 0, \ldots, n$, where u denotes the total number of occurrences of the pattern Λ in the first t trials (counting *forward* from the first trial to the tth trial) and v denotes the subpattern (ending block) (counting *backward* from the tth trial). Based on this construction, the state space Ω associated with $\{Y_t\}$ is defined by

$$\Omega = \{(u, v) : u = 0, 1, \ldots, l \text{ and } v = 0, 1, \ldots, k-1\},$$

where $l = [n/k]$ is the maximum number of occurrences of the pattern Λ in the sequence of n trials. It is easy to verify that $card(\Omega) = (l+1)k$.

(iii) For $t = 1, \ldots, n$, the transition probabilities are determined by the following equations:

$$P\left[Y_t = (u', v') \mid Y_{t-1} = (u, v)\right]$$
$$= \begin{cases} \sum_{v \to v'} p_{i_j} & \text{if } u' = u \text{ for } u = 0, \ldots, l \text{ and } v, v' = 0, \ldots, k-1, \\ p_{i_k} & \text{if } u' = u + 1 \text{ for } u = 0, \ldots, l-1, \;\; v' = 0 \text{ and } v = k-1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\sum_{v \to v'}$ denotes the sum over all states b_{i_j} such that the ending block v is shifted to v', and p_{i_k} is the probability corresponding to the last element b_{i_k} of the pattern Λ.

(iv) Define the partition $\{C_x = [(x, v) : (x, v) \in \Omega,\ v = 0, 1, \ldots, k-1],$ for $x = 0, 1, \ldots, l\}$ on the state space Ω in the sense that

$$P(X_n(\Lambda) = x) = P(Y_n \in C_x | \boldsymbol{\xi}_0),$$

for each $x = 0, 1, \ldots, l$.

From the construction of the transition probabilities in (iii), it is easy to see that $\{Y_t\}$ is a Markov chain. The forward and backward principle provides a classification for a realization η of the sequence $\{X_i\}$ according to the number of patterns and the ending blocks. With some modifications, the forward and backward principle can be extended to a sequence of homogeneous Markov dependent multistate trials, but we will not discuss this case any further. The next example gives an detailed illustration of this principle and catches more ideas about the finite Markov chain imbedding technique.

Example 2.4 Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d. 3-state trials such that each trial has possible outcomes F, S and S^* with corresponding probabilities p_0, p_1 and p_2, respectively. Let Λ be a simple pattern with $\Lambda = SFF$ and assume that nonoverlapping counting is used. To find the exact distribution of $X_n(\Lambda)$, we first decompose the pattern Λ forward into two subpatterns S and SF, and relabel them as $1 = S$ and $2 = SF$. Let '0' stand for neither the subpattern '1' nor the subpattern '2'. Define a Markov chain $Y_t = (u, v)$ for each $t = 0, 1, \ldots, n$, where u represents the total number of occurrences of the pattern Λ in the first t trials and $v = 0, 1, 2$ are ending blocks (counting backward from trial t). The state space based on this imbedded Markov chain is then given by

$$\Omega = \{(u,v) : u = 0,1,\ldots,l \text{ and } v = 0,1,2\},$$

where $l = [n/3]$. Clearly, $card(\Omega) = 3(l+1)$. Then the transition probability matrix associated with the imbedded Markov chain $\{Y_t\}$ is given by

$$
M = \begin{array}{c}
(0,0)\\(0,1)\\(0,2)\\(1,0)\\(1,1)\\(1,2)\\ \cdot\\ \cdot\\ \cdot\\ (l,0)\\(l,1)\\(l,2)
\end{array}
\left[\begin{array}{ccc|ccc|c|ccc}
p_0+p_2 & p_1 & 0 & 0 & 0 & 0 & & & & \\
p_2 & p_1 & p_0 & 0 & 0 & 0 & 0 & & 0 & \\
p_2 & p_1 & 0 & p_0 & 0 & 0 & & & & \\ \hline
0 & 0 & 0 & p_0+p_2 & p_1 & 0 & & & & \\
0 & 0 & 0 & p_2 & p_1 & p_0 & 0 & & 0 & \\
0 & 0 & 0 & p_2 & p_1 & 0 & p_0 & & & \\ \hline
 & & & & & & & & & \\
 & 0 & & & 0 & & \ddots \quad \ddots & & 0 & \\
 & & & & & & & & p_0 & \\ \hline
 & & & & & & & p_0+p_2 & p_1 & 0 \\
 & 0 & & & 0 & & 0 & p_2 & p_1 & p_0 \\
 & & & & & & & 0 & 0 & 1
\end{array}\right].
$$

Finally, we define C_x to be $C_x = \{(x,i) : i = 0,1,2\}$, $x = 0,1,\ldots,l$. Then $\{C_x : x = 0,1,\ldots,l\}$ forms a partition of the state space Ω. Hence, from Theorem 2.1, we have

$$P(X_n(\Lambda) = x) = \boldsymbol{\xi}_0 \boldsymbol{M}^n \boldsymbol{U}'(C_x), \ x = 0,1,\ldots,l,$$

where $\boldsymbol{\xi}_0 = (1,0,\ldots,0)_{1\times 3(l+1)}$ is the initial distribution and $\boldsymbol{U}'(C_x)$ is the transpose of the row vector $(0,\cdots,1,1,1,\ldots,0)_{1\times 3(l+1)}$ such that the locations of 1's correspond to the states (x,i), $i = 0,1,2$, respectively.

The major advantages of the finite Markov chain imbedding technique are that it does not involve heavy mathematics and it is efficient from a computational point of view. Especially with today's high speed computers, computation on large matrices is no longer an impossible task. To close this section, we point out that the finite Markov chain imbedding technique is not only useful in studying the distributions of runs and patterns, but also in solving problems associated with compound statistical decision rules and random permutations. We refer to Fu (1995), Fu, Lou and Wang (1999), Johnson and Fu (2000) and Fu and Lou (2000) for further references.

Chapter 3

Waiting Time Distributions of Compound Patterns

In this chapter, we assume that, unless otherwise stated, $\{X_i\}$ is either a sequence of i.i.d. or first-order homogeneous Markov dependent m-state trials. The main purpose of this chapter is to develop a simple and general method to obtain the exact distributions, means and probability generating functions for the waiting time distributions of compound patterns. Except for Theorem 3.1, all theorems with proofs are new.

3.1 Preliminaries

Let $\{Y_t : t = 0, 1, \ldots\}$ be a homogeneous Markov chain defined on a finite state space Ω with transition probability matrix $\boldsymbol{M}$ having the form

$$\boldsymbol{M} = \begin{array}{c c} & \begin{array}{cc} \Omega \setminus A & A \end{array} \\ \begin{array}{c} \Omega \setminus A \\ A \end{array} & \left[\begin{array}{c|c} \boldsymbol{N}_{k \times k} & \boldsymbol{C}_{k \times l} \\ \hline \mathbf{0}_{l \times k} & \boldsymbol{I}_{l \times l} \end{array}\right] \end{array}, \tag{3.1}$$

where $card(\Omega) = k + l$ is the size of the state space Ω, A is the subset of all absorbing states with $card(A) = l$, and $\Omega \setminus A$ is the subset of all non-absorbing states. Let $\boldsymbol{\xi}_0 = (\boldsymbol{\xi} : \mathbf{0})_{1 \times (k+l)}$ be the initial distribution of $\{Y_t\}$, where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ and $\sum_{i=1}^{k} \xi_i \equiv 1$, and let $(\mathbf{1}_k : \mathbf{0})_{1 \times (k+l)}$ be a row vector, where $\mathbf{1}_k = (1, \ldots, 1)_{1 \times k}$. The following lemma, which is a special case of a result from Fu and Lou (2002), plays an indispensable role in studying waiting time distributions. It yields the probability of the event that the chain enters the set of absorbing states for the first time. We provide details of the result, including the proof.

Lemma 3.1 *For any state $j \in A$, we have*

$$P(Y_n = j, Y_{n-1} \notin A, \ldots, Y_1 \notin A | \boldsymbol{\xi}_0) = \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}_j, \tag{3.2}$$

where $\boldsymbol{C}_j$ is the column vector of the matrix $\boldsymbol{C}$ corresponding to the state j, and more generally,

$$P(Y_n \in A, Y_{n-1} \notin A, \ldots, Y_1 \notin A | \boldsymbol{\xi}_0) = \boldsymbol{\xi} \boldsymbol{N}^{n-1} (\boldsymbol{I} - \boldsymbol{N}) \mathbf{1}_k', \tag{3.3}$$

where $\boldsymbol{I}$ is the $k \times k$ identity matrix, and $\mathbf{1}_k'$ is the transpose of the row vector $\mathbf{1}_k = (1, \ldots, 1)_{1 \times k}$.

Proof. Since $\boldsymbol{M}$ has the form given by (3.1), it follows that

$$\boldsymbol{M}^{n-1} = \left[\begin{array}{c|c} \boldsymbol{N}^{n-1} & \boldsymbol{K}_{n-1} \\ \hline \mathbf{0} & \boldsymbol{I} \end{array} \right],$$

where $\boldsymbol{K}_{n-1} = \boldsymbol{C}^{n-1} + \boldsymbol{N}\boldsymbol{C}^{n-2} + \cdots + \boldsymbol{N}^{n-2}\boldsymbol{C}$. For any state $i \in \Omega \setminus A$, it follows from the Chapman-Kolmogorov equation that

$$P(Y_{n-1} = i, Y_{n-2} \notin A, \ldots, Y_1 \notin A | \boldsymbol{\xi}_0) = (\boldsymbol{\xi} : \mathbf{0}) \boldsymbol{M}^{n-1} (\boldsymbol{e}_i : \mathbf{0})' = \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i',$$

where the dimension of the row vector $(\boldsymbol{e}_i : \mathbf{0})$ is $1 \times (k + l)$, and $\boldsymbol{e}_i'$ is the transpose of the unit vector $\boldsymbol{e}_i = (0, \ldots, 1, \ldots, 0)_{1 \times k}$. From the definition of a Markov chain, we have

$$\begin{aligned} &P(Y_n = j, Y_{n-1} \notin A, \ldots, Y_1 \notin A | \boldsymbol{\xi}_0) \\ &= \sum_{i \in \Omega \setminus A} P(Y_{n-1} = i, Y_{n-2} \notin A, \ldots, Y_1 \notin A | \boldsymbol{\xi}_0) \, P(Y_n = j | Y_{n-1} = i) \\ &= \sum_{i \in \Omega \setminus A} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' p_{ij} = \boldsymbol{\xi} \boldsymbol{N}^{n-1} \sum_{i \in \Omega \setminus A} p_{ij} \, \boldsymbol{e}_i' = \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}_j. \end{aligned}$$

This completes the proof of equality (3.2). Equality (3.3) follows from the definition of a Markov chain and equality (3.2):

$$
\begin{aligned}
&P(Y_n \in A, Y_{n-1} \notin A, \ldots, Y_1 \notin A|\boldsymbol{\xi}_0) \\
&= \sum_{j\in A} P(Y_n = j, Y_{n-1} \notin A, \ldots, Y_1 \notin A|\boldsymbol{\xi}_0) = \sum_{j\in A} \boldsymbol{\xi}\boldsymbol{N}^{n-1}\boldsymbol{C}_j \\
&= \boldsymbol{\xi}\boldsymbol{N}^{n-1}\sum_{j\in A}\boldsymbol{C}_j = \boldsymbol{\xi}\boldsymbol{N}^{n-1}\sum_{j\in A}\sum_{i\in\Omega\backslash A} p_{ij}\,\boldsymbol{e}_i^{'} = \boldsymbol{\xi}\boldsymbol{N}^{n-1}\sum_{i\in\Omega\backslash A}\sum_{j\in A} p_{ij}\,\boldsymbol{e}_i^{'} \\
&= \boldsymbol{\xi}\boldsymbol{N}^{n-1}\sum_{i\in\Omega\backslash A}\left(1 - \sum_{m\in\Omega\backslash A} p_{im}\right)\boldsymbol{e}_i^{'} = \boldsymbol{\xi}\boldsymbol{N}^{n-1}\left(\sum_{i\in\Omega\backslash A}\boldsymbol{e}_i^{'} - \sum_{i\in\Omega\backslash A}\sum_{m\in\Omega\backslash A} p_{im}\boldsymbol{e}_i^{'}\right) \\
&= \boldsymbol{\xi}\boldsymbol{N}^{n-1}(\boldsymbol{1}_k^{'} - \boldsymbol{N}\boldsymbol{1}_k^{'}) = \boldsymbol{\xi}\boldsymbol{N}^{n-1}(\boldsymbol{I} - \boldsymbol{N})\boldsymbol{1}_k^{'}.
\end{aligned}
$$

This completes the proof. □

3.2 Exact Distributions

To derive the exact distribution of the waiting time random variable $W(\Lambda)$, we adopt the forward and backward principle. We start from the i.i.d. case. Given a compound pattern $\Lambda = \bigcup_{i=1}^{l}\Lambda_i$ with each simple pattern having length k_i, then each simple pattern can be decomposed forward into k_i subpatterns. Define the state space Ω as

$$\Omega = \{\emptyset\} \cup \Gamma \cup \bigcup_{i=1}^{l} S(\Lambda_i), \tag{3.4}$$

where $\emptyset$ is the initial state and $S(\Lambda_i) = \{\text{all the subpatterns of } \Lambda_i\}$, $i = 1, \ldots, l$. It is easy to verify that $card(\Omega) \leq 1 + m + \sum_{i=1}^{l}(k_i - 1)$. For example, let $\Gamma = \{b_1, b_2\}$ and $\Lambda = \Lambda_1 \cup \Lambda_2$ with $\Lambda_1 = b_1b_1b_2$ and $\Lambda_2 = b_1b_2b_2$, then $S(\Lambda_1) \cup S(\Lambda_2) = \{b_1, b_1b_1, b_1b_2, b_1b_1b_2, b_1b_2b_2\}$ and $\Omega = \{\emptyset, b_1, b_2, b_1b_1, b_1b_2, b_1b_1b_2, b_1b_2b_2\}$. Hence, the states in the state space Ω can always be relabelled as

$$\Omega = \{1, \ldots, k, \alpha_1, \ldots, \alpha_l\}, \tag{3.5}$$

where $\alpha_1, \ldots, \alpha_l$ are absorbing states corresponding to the patterns $\Lambda_1, \ldots, \Lambda_l$.

It has been shown that the number of patterns (or runs) $X_n(\Lambda)$ and the waiting time $W(\Lambda)$ in a sequence of multistate trials are finite Markov chain imbeddable in the sense of Fu (1986b, 1996), Fu and Koutras (1994), Koutras (1997b) and Koutras and Alexandrou (1997), hence there exists a Markov chain $\{Y_t : t = 0, 1, \ldots\}$ defined on the finite state space Ω whose transition probability matrix $\boldsymbol{M}$ has the form

$$\boldsymbol{M} = \begin{array}{c} 1 \\ \vdots \\ k \\ \alpha_1 \\ \vdots \\ \alpha_l \end{array} \left[\begin{array}{ccc|ccc} p_{11} & \cdots & p_{1k} & p_{1\alpha_1} & \cdots & p_{1\alpha_l} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} & p_{k\alpha_1} & \cdots & p_{k\alpha_l} \\ \hline 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right], \tag{3.6}$$

where the p_{ij}'s are pattern dependent and are determined via the forward and backward principle. From Lemma 3.1, the exact waiting time distribution of $W(\Lambda)$ is given by

$$\begin{aligned} P(W(\Lambda) = n) &= \sum_{j=1}^{l} P(W(\Lambda) = n, W(\Lambda_j) = n) \\ &= \sum_{j=1}^{l} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}(\alpha_j) \qquad (3.7) \\ &= \boldsymbol{\xi} \boldsymbol{N}^{n-1} (\boldsymbol{I} - \boldsymbol{N}) \mathbf{1}'_k, \qquad (3.8) \end{aligned}$$

for $n = 1, 2, \ldots$, where $\boldsymbol{\xi}$ is the initial distribution, $\boldsymbol{C}(\alpha_j)$, $j = 1, \ldots, l$, are the column vectors of the matrix $\boldsymbol{C}$, and $\mathbf{1}'_k$ is the transpose of the row vector $\mathbf{1}_k = (1, 1, \ldots, 1)_{1\times k}$. A different derivation of Equation (3.8) is shown in result A2 of the Appendix.

Remark 3.1 It is important to point out that the waiting time distribution is highly dependent on the initial distribution $\boldsymbol{\xi}$ of the imbedded Markov chain $\{Y_t\}$. Setting up the initial distribution $\boldsymbol{\xi}$ can be very tricky, especially when $\{X_i\}$ is a sequence of Markov dependent multistate trials and Λ is a compound pattern. To avoid this problem, we purposely introduce the initial state $\emptyset$ and define $P(Y_0 = \emptyset) \equiv 1$ ($\boldsymbol{\xi} = (1, 0, \ldots, 0)$).

Remark 3.2 If $\{X_i\}$ is a sequence of i.i.d. multistate trials, then the state space Ω of the imbedded chain $\{Y_t\}$ can be reduced to

$$\Omega = \{\emptyset, \beta\} \cup \bigcup_{i=1}^{l} S(\Lambda_i), \tag{3.9}$$

where the state β means that no subpattern belongs to $\bigcup_{i=1}^{l} S(\Lambda_i)$. The reduction of number of states can be significant especially when the size of Γ is very large.

Equations (3.7) and (3.8) remain applicable to the case when $\{X_i\}$ is a sequence of first-order homogeneous Markov dependent trials by replacing corresponding transition probability matrix in the equalities. If $\{X_i\}$ is a sequence of independent but non-identical multistate trials, then the transition probability matrix at time t associated with the imbedded Markov chain has the form

$$\boldsymbol{M}_t = \left[\begin{array}{c|c} \boldsymbol{N}_t & \boldsymbol{C}_t \\ \hline \mathbf{0} & \boldsymbol{I} \end{array}\right], \tag{3.10}$$

and the exact waiting time distribution of $W(\Lambda)$ is given by

$$\begin{aligned} P(W(\Lambda) = n) &= \sum_{j=1}^{l} \boldsymbol{\xi} \left(\prod_{t=1}^{n-1} \boldsymbol{N}_t \right) \boldsymbol{C}_n(\alpha_j) & (3.11) \\ &= \boldsymbol{\xi} \left(\prod_{t=1}^{n-1} \boldsymbol{N}_t \right) (\boldsymbol{I} - \boldsymbol{N}_n) \mathbf{1}'_k. & (3.12) \end{aligned}$$

The construction procedure for the exact distribution of $W(r : \Lambda)$ is similar to Example 2.4. We only need one more component for each state in the state space Ω to record the total number of occurrences of the pattern Λ in the first t trials, and those subpatterns (excluding Λ itself) are used as ending blocks.

3.3 Means and Probability Generating Functions

3.3.1 Main Results

The probability generating function is an indispensable tool for studying the characteristics of waiting time distributions. Usually, finding the probability generating function for the waiting time entails tedious mathematics and heavy probability theory, even for the case of simple patterns in a sequence of Bernoulli trials. For example, Feller (1968), by using the theory of recurrent events, obtained the probability generating function for the waiting time of a success run, $\Lambda = S \cdots S$, of length k in a sequence of Bernoulli trials as

$$\varphi_{W(\Lambda)}(s) = \frac{(ps)^k(1-ps)}{1-s+qp^k s^{k+1}}. \tag{3.13}$$

In this section, we develop a general method for finding the means and probability generating functions of waiting time distributions of compound patterns in a sequence of i.i.d. or Markov dependent multistate trials.

We first observe that the mean and probability generating function of the waiting time $W(\Lambda)$ can be derived (see Fu, Spiring and Xie, 2002) straightforwardly from Equation (3.8).

Theorem 3.1 *For a waiting time random variable $W(\Lambda)$, its mean and probability generating function are given by*

$$E\left[W(\Lambda)\right] = \boldsymbol{\xi}(\boldsymbol{I}-\boldsymbol{N})^{-1}\mathbf{1}'_k \tag{3.14}$$

and

$$\varphi_{W(\Lambda)}(s) = 1 + (s-1)\boldsymbol{\xi}(\boldsymbol{I}-s\boldsymbol{N})^{-1}\mathbf{1}'_k, \tag{3.15}$$

respectively.

Although formulas (3.14) and (3.15) are rather simple, from application point of view, Equation (3.15) does not yield an explicit analytical form for the probability generating function in the way that Equation (3.13) does. To find the mathematical form of $\varphi_{W(\Lambda)}(s)$, the major difficulty is that given $\boldsymbol{N}$, we have to find the explicit analytical form of the inverse of the matrix $(\boldsymbol{I} - s\boldsymbol{N})$ which is extremely tedious, if not impossible. It can be done only if $\boldsymbol{N}$ is a very small and simple matrix. This is the primary motivation for developing a general method of finding probability generating functions for any waiting time distribution involving simple or compound patterns.

Traditionally, the mean waiting time $E[W(\Lambda)]$ is obtained by differentiating the probability generating function once and evaluated at $s = 1$; that is, $E[W(\Lambda)] = \varphi^{(1)}_{W(\Lambda)}(s)|_{s=1}$. Contrary to the traditional approach, we develop a simple technique to find $E[W(\Lambda)]$ first and then extend the technique to obtain the probability generating function.

Theorem 3.2 *For a waiting time random variable $W(\Lambda)$, the mean waiting time $E[W(\Lambda)]$ can be expressed as*

$$E[W(\Lambda)] = S_1 + S_2 + \cdots + S_k, \tag{3.16}$$

where $(S_1, \ldots, S_k)$ is the solution of the simultaneous recursive equations

$$S_i = \boldsymbol{\xi}\boldsymbol{e}_i' + (S_1, S_2, \ldots, S_k)\boldsymbol{N}(i), \text{ for } i = 1, \ldots, k, \tag{3.17}$$

and where $\boldsymbol{N}(i)$, $i = 1, \ldots, k$, are the column vectors of the matrix $\boldsymbol{N}$, and $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, $i = 1, \ldots, k$, are unit vectors.

Proof. Since $W(\Lambda)$ is finite Markov chain imbeddable and its imbedded Markov chain $\{Y_t\}$ defined on the state space Ω has transition probability matrix $\boldsymbol{M}$ having the form given by Equation (3.6),

$$P(W(\Lambda) \geq n) = \boldsymbol{\xi}\boldsymbol{N}^{n-1}\mathbf{1}_k'.$$

It follows from the definition that

$$\begin{aligned}
E\left[W(\Lambda)\right] &= \sum_{n=1}^{\infty} P(W(\Lambda) \geq n) \\
&= \sum_{n=1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \mathbf{1}_k' \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' \\
&= \sum_{i=1}^{k} \sum_{n=1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' \\
&= \sum_{i=1}^{k} S_i,
\end{aligned}$$

where $S_i = \sum_{n=1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i'$, for $i = 1, \ldots, k$. Further, last-step analysis yields that, for $i = 1, \ldots, k$,

$$\boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' = \boldsymbol{\xi} \boldsymbol{N}^{n-2} \boldsymbol{N}(i) = \sum_{j=1}^{k} p_{ji}\, \boldsymbol{\xi} \boldsymbol{N}^{n-2} \boldsymbol{e}_j', \tag{3.18}$$

and

$$\begin{aligned}
S_i &= \sum_{n=1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' \\
&= \boldsymbol{\xi} \boldsymbol{e}_i' + \sum_{n=2}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-2} \left(\sum_{j=1}^{k} p_{ji} \boldsymbol{e}_j' \right) \\
&= \boldsymbol{\xi} \boldsymbol{e}_i' + \sum_{j=1}^{k} p_{ji} S_j \\
&= \boldsymbol{\xi} \boldsymbol{e}_i' + (S_1, S_2, \ldots, S_k) \boldsymbol{N}(i).
\end{aligned}$$

This completes the proof. $\square$

Remark 3.3 Since $(S_1, S_2, \ldots, S_k)$ is the solution of the simultaneous recursive equations (3.17), it follows that

$$(S_1, S_2, \ldots, S_k) = \boldsymbol{\xi}(\boldsymbol{I} - \boldsymbol{N})^{-1}.$$

Hence,

$$E[W(\Lambda)] = S_1 + S_2 + \cdots + S_k = \boldsymbol{\xi}(\boldsymbol{I} - \boldsymbol{N})^{-1}\mathbf{1}_k'.$$

This yields Equation (3.14).

In view of Theorem 3.2, it is easy to see that the technique could be extended directly to show that the explicit analytic form of probability generating function $\varphi_{W(\Lambda)}(s)$ can be obtained in terms of $\Phi_{W(\Lambda)}(s)$, the probability generating function of the sequence of cumulative probabilities $\{P(W(\Lambda) \geq n)\}_{n=1}^{\infty}$.

Theorem 3.3 *For a waiting time random variable $W(\Lambda)$, we have*

$$\Phi_{W(\Lambda)}(s) = \phi_1(s) + \cdots + \phi_k(s), \tag{3.19}$$

where $(\phi_1(s), \ldots, \phi_k(s))$ is the solution of the simultaneous recursive equations

$$\phi_i(s) = s\boldsymbol{\xi}\boldsymbol{e}_i' + s(\phi_1(s), \ldots, \phi_k(s))\boldsymbol{N}(i), \ \textit{for } i = 1, 2, \ldots, k, \tag{3.20}$$

and

$$\varphi_{W(\Lambda)}(s) = 1 + \left(1 - \frac{1}{s}\right)\Phi_{W(\Lambda)}(s). \tag{3.21}$$

Proof. The proof of part (3.19) is along the lines of the proof of Theorem 3.2:

$$\begin{aligned}
\Phi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \mathbf{1}_k' \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' \\
&= \sum_{i=1}^{k} \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i' \\
&= \sum_{i=1}^{k} \phi_i(s),
\end{aligned}$$

where $\phi_i(s) = \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i^{'}$ for $i = 1, \ldots, k$. It follows from last-step analysis and Equation (3.18) that

$$\begin{aligned}
\phi_i(s) &= \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i^{'} \\
&= s\boldsymbol{\xi}\boldsymbol{e}_i^{'} + \sum_{n=2}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-2} \left(\sum_{j=1}^{k} p_{ji}\ \boldsymbol{e}_j^{'} \right) \\
&= s\boldsymbol{\xi}\boldsymbol{e}_i^{'} + s \sum_{j=1}^{k} p_{ji} \left(\sum_{n=2}^{\infty} s^{n-1} \boldsymbol{\xi} \boldsymbol{N}^{n-2} \boldsymbol{e}_j^{'} \right) \\
&= s\boldsymbol{\xi}\boldsymbol{e}_i^{'} + s(\phi_1(s), \ldots, \phi_k(s)) \boldsymbol{N}(i).
\end{aligned}$$

This completes the proof of equality (3.19). Equality (3.21) follows from the definition of $\varphi_{W(\Lambda)}(s)$ and $\Phi_{W(\Lambda)}(s)$:

$$\begin{aligned}
\varphi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} (\boldsymbol{I} - \boldsymbol{N}) \boldsymbol{1}_k^{'} \\
&= \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{1}_k^{'} - \frac{1}{s} \sum_{n=0}^{\infty} s^{n+1} \boldsymbol{\xi} \boldsymbol{N}^{n} \boldsymbol{1}_k^{'} + 1 \\
&= 1 + \left(1 - \frac{1}{s}\right) \Phi_{W(\Lambda)}(s).
\end{aligned}$$

This completes the proof. □

In view of the definition of $\phi_i(s)$, it is clear that the $\phi_i(s)$, $i = 1, \ldots, k$, are probability generating functions for the sequences of probabilities $\{P(Y_n = i)\}_{n=1}^{\infty}$, respectively. Since $P(W(\Lambda) \geq n) = \sum_{i=1}^{k} P(Y_n = i)$, the result $\Phi_{W(\Lambda)}(s) = \sum_{i=1}^{k} \phi_i(s)$ comes as no surprise.

Next theorem provides another way to evaluate $\varphi_{W(\Lambda)}(s)$ which is often simpler in computation.

Theorem 3.4 *For a waiting time random variable $W(\Lambda)$, the probability generating function $\varphi_{W(\Lambda)}(s)$ can be expressed as*

$$\varphi_{W(\Lambda)}(s) = \sum_{j=1}^{l} (\phi_1, \phi_2, \ldots, \phi_k) \boldsymbol{C}(\alpha_j), \tag{3.22}$$

where $\boldsymbol{C}(\alpha_j)$, $j = 1, \ldots, l$, *are the column vectors of the matrix* $\boldsymbol{C}$.

Proof. It follows from definition and last-step analysis that

$$\begin{aligned}
\varphi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n P(W(\Lambda) = n) \\
&= \sum_{n=1}^{\infty} s^n \left(\sum_{j=1}^{l} P(W(\Lambda) = n, W(\Lambda_j) = n) \right) \\
&= \sum_{n=1}^{\infty} s^n \left(\sum_{j=1}^{l} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}(\alpha_j) \right) \\
&= \sum_{j=1}^{l} \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}(\alpha_j) \\
&= \sum_{j=1}^{l} \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \left(\sum_{i=1}^{k} p_{i\alpha_j} \boldsymbol{e}_i^{'} \right) \\
&= \sum_{j=1}^{l} \sum_{i=1}^{k} p_{i\alpha_j} \left(\sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{e}_i^{'} \right) \\
&= \sum_{j=1}^{l} \sum_{i=1}^{k} p_{i\alpha_j} \phi_i(s) \\
&= \sum_{j=1}^{l} (\phi_1(s), \phi_2(s), \ldots, \phi_k(s)) \boldsymbol{C}(\alpha_j).
\end{aligned}$$

This completes the proof. □

All the above results in this section are under the assumption that $\{X_i\}$ is a sequence of i.i.d. multistate trials. Moreover, the results are also true if $\{X_i\}$ is a sequence of first-order homogeneous Markov dependent multistate trials. The only difference between i.i.d. and Markov dependent trials is that the transition probability matrices of two imbedded Markov chains are slightly different. We do not repeat the proof for the above results under Markov chain dependency but an example to illustrate this point is provided in the next section.

3.3.2 Examples and Symbolic Computation

To make our results more transparent, we provide the following two examples.

Example 3.1 Let $\{X_i\}$ be a sequence of Bernoulli trials with success and failure probabilities $P(X_i = S) = p$ and $P(X_i = F) = q = 1 - p$, respectively, and let $\Lambda = \Lambda_1 \cup \Lambda_2$ be a compound pattern with $\Lambda_1 = SS$ and $\Lambda_2 = FF$. It is easy to see that the imbedded Markov chain $\{Y_t\}$ associated with the waiting time $W(\Lambda)$ has state space $\Omega = \{\emptyset\} \cup \Gamma \cup \bigcup_{i=1}^{2} S(\Lambda_i) = \{\emptyset, S, F, \alpha_1, \alpha_2\}$ and transition probability matrix

$$\boldsymbol{M} = \begin{array}{c} \emptyset \\ S \\ F \\ \alpha_1 \\ \alpha_2 \end{array} \left[\begin{array}{ccc|cc} 0 & p & q & 0 & 0 \\ 0 & 0 & q & p & 0 \\ 0 & p & 0 & 0 & q \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right]$$

with initial distribution $\boldsymbol{\xi}_0 = (1,0,0,0,0)$ for Y_0 ($\boldsymbol{\xi} = (1,0,0)$). It follows from Theorem 3.2 that (S_1, S_2, S_3) is the solution of the simultaneous recursive equations

$$\begin{aligned} S_1 &= 1, \\ S_2 &= pS_1 + pS_3, \\ S_3 &= qS_1 + qS_2. \end{aligned}$$

This yields

$$E\left[W(\Lambda)\right] = 1 + \frac{p(1+q)}{1-pq} + \frac{q(1+p)}{1-pq} = 2 + \frac{3pq}{1-pq}.$$

Similarly, by Theorem 3.3, $(\phi_1(s), \phi_2(s), \phi_3(s))$ is the solution of the simultaneous recursive equations

$$\begin{aligned} \phi_1(s) &= s, \\ \phi_2(s) &= sp\phi_1(s) + sp\phi_3(s), \\ \phi_3(s) &= sq\phi_1(s) + sq\phi_2(s). \end{aligned}$$

With some simple computation, it follows from Theorem 3.3 that

$$\Phi_{W(\Lambda)}(s) = s + \frac{ps^2 + pqs^3}{1 - pqs^2} + \frac{qs^2 + pqs^3}{1 - pqs^2}$$

and

$$\varphi_{W(\Lambda)}(s) = 1 + \left(1 - \frac{1}{s}\right) \Phi_{W(\Lambda)}(s) = \frac{(p^2 + q^2)s^2 + pqs^3}{1 - pqs^2}.$$

Further, by Theorem 3.4, it also yields

$$\begin{aligned}
\varphi_{W(\Lambda)}(s) &= \sum_{j=1}^{2} (\phi_1(s), \phi_2(s), \phi_3(s)) \boldsymbol{C}(\alpha_j) \\
&= p\phi_2(s) + q\phi_3(s) \\
&= \frac{(p^2 + q^2)s^2 + pqs^3}{1 - pqs^2}.
\end{aligned}$$

In addition to the above, we use this example to illustrate that these results also apply to the case when $\{X_i\}$ is a homogeneous Markov chain having transition probability matrix

$$\boldsymbol{A} = \begin{array}{c} S \\ F \end{array} \begin{bmatrix} p_1 & 1 - p_1 \\ p_2 & 1 - p_2 \end{bmatrix}.$$

Given the initial probabilities $P(X_1 = S) = p$ and $P(X_1 = F) = q = 1 - p$, it is easy to show that the imbedded Markov chain $\{Y_t\}$ associated with the waiting time $W(\Lambda)$ has state space $\Omega = \{\emptyset, S, F, \alpha_1, \alpha_2\}$ and transition probability matrix given by

$$\boldsymbol{M} = \begin{array}{c} \emptyset \\ S \\ F \\ \alpha_1 \\ \alpha_2 \end{array} \left[\begin{array}{ccc|cc} 0 & p & q & 0 & 0 \\ 0 & 0 & 1 - p_1 & p_1 & 0 \\ 0 & p_2 & 0 & 0 & 1 - p_2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right].$$

By the same token, Theorems 3.2, 3.3 and 3.4 yield the mean waiting time and its probability generating function, respectively, as

$$E\left[W(\Lambda)\right] = 1 + \frac{p + qp_2}{1 - (1 - p_1)p_2} + \frac{q + p(1 - p_1)}{1 - (1 - p_1)p_2}$$

and

$$\varphi_{W(\Lambda)}(s) = p_1 \cdot \frac{ps^2 + qp_2s^3}{1-(1-p_1)p_2s^2} + (1-p_2) \cdot \frac{qs^2 + p(1-p_1)s^3}{1-(1-p_1)p_2s^2}.$$

When $p_1 = p_2 = p$ and $1-p_1 = 1-p_2 = q$, this reduces to the i.i.d. case. A general formula for $\varphi_{W(\Lambda)}(s)$ in the case where $\Lambda_1 = S \cdots S$ of length k_1 and $\Lambda_2 = F \cdots F$ of length k_2 are derived as results A3 and A4 in the Appendix.

From the above example, we see that Theorems 3.2–3.4 are applicable both to the case that $\{X_i\}$ is a sequence of i.i.d. multistate trials and to the case that $\{X_i\}$ is a sequence of Markov dependent multistate trials. Moreover, the method is independent of the size of the state space Γ of multistate trials, but the results do depend on the initial distribution $\boldsymbol{\xi}$. For large l (or large k_i, $i = 1, \ldots, l$), it is clear that the $\phi_i(s)$ cannot be readily obtained by hand through the simultaneous recursive equations. We therefore develop symbolic computational algorithms for use in a computer algebra system. Computer programs based on the mathematical software **MAPLE** have been developed to obtain the transition probability matrix $\boldsymbol{N}$, mean $E\left[W(\Lambda)\right]$ and probability generating function $\varphi_{W(\Lambda)}(s)$ automatically. We discuss the algorithms in Chapter 5. The following example is solved as a result of our computer program.

Example 3.2 Let $\{X_i\}$ be a sequence of i.i.d. four-state trials with possible outcomes A, C, G and T, and let $\Lambda = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$ be a compound pattern with $\Lambda_1 = AGTT$, $\Lambda_2 = AAA$ and $\Lambda_3 = GCT$. Assume that $P(X_i = A) = p_a$, $P(X_i = C) = p_c$, $P(X_i = G) = p_g$ and $P(X_i = T) = p_t$, where $p_a + p_c + p_g + p_t = 1$, $i = 1, 2, \cdots$. Note that in our computer program, the letters A, C, G and T are transformed to 1, 2, 3 and 4, respectively. It is easy to see that the imbedded Markov chain $\{Y_t\}$ associated with the waiting time $W(\Lambda)$ has state space

$\Omega = \{\emptyset, A, C, G, T, AG, AA, GC, AGT, \alpha_1, \alpha_2, \alpha_3\}$ and transition probability matrix

$$M = \begin{array}{c} \emptyset \\ A \\ C \\ G \\ T \\ AG \\ AA \\ GC \\ AGT \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{ccccccccc|ccc} 0 & p_a & p_c & p_g & p_t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_c & 0 & p_t & p_g & p_a & 0 & 0 & 0 & 0 & 0 \\ 0 & p_a & p_c & p_g & p_t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_a & 0 & p_g & p_t & 0 & 0 & p_c & 0 & 0 & 0 & 0 \\ 0 & p_a & p_c & p_g & p_t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_a & 0 & p_g & 0 & 0 & 0 & p_c & p_t & 0 & 0 & 0 \\ 0 & 0 & p_c & 0 & p_t & p_g & 0 & 0 & 0 & 0 & p_a & 0 \\ 0 & p_a & p_c & p_g & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_t \\ 0 & p_a & p_c & p_g & 0 & 0 & 0 & 0 & 0 & p_t & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

with initial distribution $\boldsymbol{\xi}_0 = (1,0,0,0,0,0,0,0,0,0,0,0)$ for Y_0 ($\boldsymbol{\xi}$ = (1,0,0,0,0, 0,0,0,0)). From the computer program, we obtain

$$E[W(\Lambda)] = \frac{1 + p_a + p_a^2}{p_a^3 + p_c p_g p_t + p_a p_g p_t (p_c + p_t)(1 + p_a)}$$

and

$$\begin{aligned} \varphi_{W(\Lambda)}(s) &= p_a \cdot \frac{p_a^2 s^3}{\Delta} + p_t \cdot \frac{p_c p_g s^3 (1 + p_a s + p_a^2 s^2)}{\Delta} + p_t \cdot \frac{p_a p_g p_t (1 + p_a s)}{\Delta} \\ &= \frac{s^3 \left[p_a^3 + p_c p_g p_t + p_a p_g p_t (p_c + p_t)(1 + p_a s) s\right]}{\Delta}, \end{aligned}$$

where

$$\begin{aligned} \Delta &= 1 - (1 - p_a)s - p_a(1 - p_a)s^2 + \left[p_c p_g p_t - p_a^2(1 - p_a)\right] s^3 \\ &\quad + p_a p_g p_t (p_c + p_t) s^4 + p_a^2 p_g p_t (p_c + p_t) s^5. \end{aligned}$$

Similarly, the mean and probability generating function can also be obtained from the computer program for the case when $\{X_i\}$ is a sequence of homogeneous Markov dependent trials. In Figure 3.1, we show the probability distributions of $W(\Lambda)$ when $\{X_i\}$ is a sequence of i.i.d. and Markov dependent trials. For the i.i.d. case, we set $p_a = p_c = p_g = p_t = 0.25$. For the Markov dependent case, we set the initial

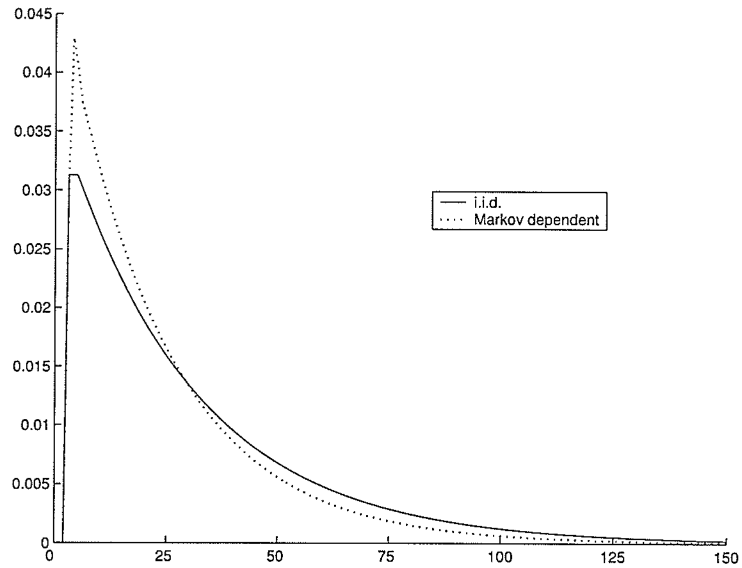

Figure 3.1: Probability distributions of the waiting time $W(\Lambda)$ for Example 3.2.

probabilities to be $p_a = p_c = p_g = p_t = 0.25$, and the transition probability matrix of $\{X_i\}$ as

$$\boldsymbol{A} = \begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{bmatrix} 0.25 & 0.10 & 0.50 & 0.15 \\ 0.35 & 0.25 & 0.10 & 0.30 \\ 0.10 & 0.20 & 0.25 & 0.45 \\ 0.15 & 0.15 & 0.45 & 0.25 \end{bmatrix}.$$

3.3.3 Extensions

We know that the waiting time $W(r:\Lambda)$ to the rth occurrence of the pattern Λ can be written as

$$W(r:\Lambda) = W_1(\Lambda) + \cdots + W_r(\Lambda), \tag{3.23}$$

where $W_i(\Lambda)$, $i = 1, \ldots, r$, are interwaiting times. If $\{X_i\}$ is a sequence of i.i.d. multistate trials and Λ is a simple pattern with a nonoverlapping counting scheme, then the $\{W_i(\Lambda)\}$ are i.i.d. random variables, and the probability generating function of the waiting time $W(r:\Lambda)$ is given by

$$\varphi_{W(r:\Lambda)}(s) = \left(\varphi_{W(\Lambda)}(s)\right)^r. \tag{3.24}$$

If nonoverlapping counting is replaced by overlapping counting, then $W_2(\Lambda), \ldots,$ $W_r(\Lambda)$ are i.i.d. random variables whose common distribution is the distribution of $W_2(\Lambda)$. Note that the $\{W_i(\Lambda)\}_{i=1}^r$ have the same imbedded Markov chain except that the initial distribution for $W_1(\Lambda)$ is $\boldsymbol{\xi}_1 = (1, 0, \ldots, 0) = \boldsymbol{e}_1$ and $\boldsymbol{\xi}_2 = (0, \ldots, 0, 1, 0, \ldots, 0) = \boldsymbol{e}_j$ is the initial distribution for all the waiting times $W_2(\Lambda)$, $\cdots, W_r(\Lambda)$, where the state j is the longest subpattern (excluding Λ itself) in the pattern Λ with respect to overlapping counting (for example, if $\Lambda = SFFSF$ then the state 'j' corresponds to the state 'SF'). Hence, by Equation (3.23), $W(r : \Lambda)$ has a probability generating function given by

$$\varphi_{W(r:\Lambda)}(s) = \varphi_{W_1(\Lambda)}(s) \left(\varphi_{W_2(\Lambda)}(s)\right)^{r-1}. \tag{3.25}$$

Further, if $\{X_i\}$ is a sequence of Markov dependent multistate trials and Λ is a simple pattern, then the probability generating function of $W(r : \Lambda)$ has the same form as Equation (3.25) except that the state 'j' associated with the initial distribution $\boldsymbol{\xi}_2$ for $W_2(\Lambda)$ is the last element of the pattern Λ.

For a compound pattern, Equation (3.24) is valid only if $\{X_i\}$ is a sequence of i.i.d. multistate trials with nonoverlapping counting. In general, things become very complex and the initial distributions $\boldsymbol{\xi}_i$ for $W_i(\Lambda)$, $i = 1, \ldots, r$, may differ a lot. The general form of $\boldsymbol{\xi}_i$ and $\varphi_{W(r:\Lambda)}(s)$ remain unknown for the case of compound patterns, especially when $\{X_i\}$ is a sequence of Markov dependent trials.

Chapter 4

Waiting Time Distributions of Ordered Series and Later Patterns

In this chapter, we assume that, unless otherwise stated, $\{X_i\}$ is a sequence of first-order homogeneous Markov dependent m-state trials. Let σ be an ordered series pattern and let Λ_L be a later pattern as defined in Chapter 2. The aim of this chapter is to investigate the distribution theory of the waiting times $W(\sigma)$ and $W(\Lambda_L)$. All the results in this chapter are new.

Remark 4.1 Given an ordered series pattern $\sigma = \Lambda_1 \circ \cdots \circ \Lambda_l$, it is important to mention that the waiting time distribution of $W(\sigma)$ is an "improper distribution"; that is, the probability $\sum_{n=1}^{\infty} P(W(\sigma) = n) < 1$. The reason for this is rather simple: there always exists a positive probability that at least one of the patterns Λ_j, $j = 2, \ldots, l$, occurs before the pattern Λ_1. Hence, the generating function of $W(\sigma)$,

$$\psi_{W(\sigma)}(s) = \sum_{n=1}^{\infty} s^n P(W(\sigma) = n),$$

is not a probability generating function and $\psi_{W(\sigma)}(1) < 1$.

4.1 Preliminaries

Given a compound pattern $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$, recall the following results obtained in Chapter 3:

(i) The exact distribution of $W(\Lambda)$ is given by

$$\begin{aligned} P(W(\Lambda)=n) &= \sum_{j=1}^{l} P(W(\Lambda)=n, W(\Lambda_j)=n) \\ &= \sum_{j=1}^{l} \boldsymbol{\xi} \boldsymbol{N}^{n-1} \boldsymbol{C}(\alpha_j) = \boldsymbol{\xi} \boldsymbol{N}^{n-1} (\boldsymbol{I}-\boldsymbol{N}) \mathbf{1}_k', \end{aligned} \tag{4.1}$$

for $n = 1, 2, \ldots$, where $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ is the initial distribution with $P(Y_0 = \emptyset) \equiv 1$, and $\boldsymbol{C}(\alpha_j)$, $j = 1, \ldots, l$, are the column vectors of the matrix $\boldsymbol{C}$ corresponding to the absorbing states α_j, $j = 1, \ldots, l$.

(ii) The mean waiting time $E[W(\Lambda)]$ is given by

$$E[W(\Lambda)] = S_1 + S_2 + \cdots + S_k, \tag{4.2}$$

where $(S_1, \ldots, S_k)$ is the solution of the simultaneous recursive equations

$$S_i = \boldsymbol{\xi} \boldsymbol{e}_i' + (S_1, S_2, \ldots, S_k) \boldsymbol{N}(i), \text{ for } i = 1, \ldots, k,$$

and where $\boldsymbol{N}(i)$, $i = 1, \ldots, k$, are the column vectors of the matrix $\boldsymbol{N}$, and $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, $i = 1, \ldots, k$, are unit vectors.

(iii) The probability generating function of $W(\Lambda)$ is given by

$$\begin{aligned} \varphi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n \left(\sum_{j=1}^{l} P(W(\Lambda)=n, W(\Lambda_j)=n) \right) \\ &= \sum_{j=1}^{l} (\phi_1, \phi_2, \ldots, \phi_k) \boldsymbol{C}(\alpha_j), \end{aligned} \tag{4.3}$$

where $(\phi_1(s), \ldots, \phi_k(s))$ is the solution of the simultaneous recursive equations

$$\phi_i(s) = s \boldsymbol{\xi} \boldsymbol{e}_i' + s(\phi_1(s), \ldots, \phi_k(s)) \boldsymbol{N}(i), \quad i = 1, 2, \ldots, k.$$

For convenience, let $W(\Lambda_j | \Lambda_1, \ldots, \Lambda_l)$ be the waiting time to the first occurrence of Λ_j, and that Λ_j occurs first among all the patterns $\Lambda_1, \ldots, \Lambda_l$. For the same reason as stated in Remark 4.1, the random variables $W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)$, $j=1,\ldots,l$, are improper random variables. It follows from the definition of $W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)$ that

$$P(W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)=n)=P(W(\Lambda)=n, W(\Lambda_j)=n)$$

and

$$P(W(\Lambda)=n)=\sum_{j=1}^{l}P(W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)=n). \tag{4.4}$$

Further, from Equation (4.3) the generating function of $W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)$ is given by

$$\begin{aligned}\psi_{W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)}(s) &= \sum_{n=1}^{\infty}s^nP(W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)=n)\\ &= \sum_{n=1}^{\infty}s^nP(W(\Lambda)=n, W(\Lambda_j)=n)\\ &= (\phi_1,\phi_2,\ldots,\phi_k)\boldsymbol{C}(\alpha_j),\ j=1,\ldots,l, \end{aligned} \tag{4.5}$$

and

$$\varphi_{W(\Lambda)}(s)=\sum_{j=1}^{l}\psi_{W(\Lambda_j|\Lambda_1,\ldots,\Lambda_l)}(s). \tag{4.6}$$

Lemma 3.1 and Equations (4.1)–(4.6) lay the foundation for studying the exact distributions, means and probability generating functions for the waiting times of order series and later patterns.

4.2 Waiting Time Distribution of an Ordered Series Pattern

Let $\sigma=\Lambda_1\circ\cdots\circ\Lambda_l$ be an ordered series pattern generated by l distinct simple patterns $\Lambda_1,\ldots,\Lambda_l$ with lengths $k_1,\ldots,k_l$, respectively. Our main interest in this section pertains to finding the probabilities $\{P(W(\sigma)=n): n=1,2,\ldots\}$ and the generating function of $W(\sigma)$. The construction of the state space Ω for the imbedded Markov chain $\{Y_t\}$ associated with $W(\sigma)$ can be divided into two cases: (i) $k_i\geq 2$ for all $i=1,\ldots,l$, and (ii) $k_i=1$ for some $i=1,\ldots,l$. Firstly, consider case (i): $k_i \geq 2$ for all $i = 1, \ldots, l$ under nonoverlapping counting. Let $S'(\Lambda_i)$ be the set of all subpatterns of Λ_i excluding Λ_i itself (e.g., if $\Gamma = \{b_1, b_2\}$ and $\Lambda_1 = b_1b_2b_1$, then $S'(\Lambda_1) = \{b_1, b_1b_2\}$), and let $\Upsilon(\Lambda_i) = \{0\} \cup \Gamma \cup S'(\Lambda_i)$, $i = 1, \ldots, l$. We define a Markov chain $\{Y_t : t = 0, 1, \ldots\}$ on the state space Ω having the form

$$\Omega = \Omega_0 \cup \bigcup_{i=1}^{l-1} \Omega_i \cup \bigcup_{i=1}^{l} \{\alpha_i\}, \tag{4.7}$$

where Ω_0 and Ω_i, $i = 1, \ldots, l-1$, are defined as

$$\begin{aligned} \Omega_0 &= \left\{ (0, v) : v \in \bigcup_{j=1}^{l} \Upsilon(\Lambda_j) \right\}, \\ \Omega_i &= \left\{ (u, v) : u = 1 \cdots i \text{ and } v \in \bigcup_{j=i+1}^{l} \Upsilon(\Lambda_j) \right\}, \end{aligned}$$

α_1 is the absorbing state corresponding to the pattern σ and $\alpha_2, \ldots, \alpha_l$ are the absorbing states corresponding to the patterns $\Lambda_2, \ldots, \Lambda_l$, respectively. For example, let $\Gamma = \{b_1, b_2\}$ and $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ with $\Lambda_1 = b_1b_2$, $\Lambda_2 = b_2b_1b_1$ and $\Lambda_3 = b_2b_2$; then the state space is $\Omega = \Omega_0 \cup \Omega_1 \cup \Omega_2 \cup \{\alpha_1, \alpha_2, \alpha_3\}$, where

$$\begin{aligned} \Omega_0 &= \{(0,0), (0,b_1), (0,b_2), (0,b_2b_1)\}, \\ \Omega_1 &= \{(1,0), (1,b_1), (1,b_2), (1,b_2b_1)\}, \\ \Omega_2 &= \{(12,0), (12,b_1), (12,b_2)\}, \end{aligned}$$

and the absorbing states α_1, α_2 and α_3 correspond to the patterns $\sigma = b_1b_2 \circ b_2b_1b_1 \circ b_2b_2$, $\Lambda_2 = b_2b_1b_1$ and $\Lambda_3 = b_2b_2$, respectively. Note that: (i) the state $(0,0)$ is the initial state; (ii) the state $(1,0)$ means that the simple pattern Λ_1 has just occurred; and (iii) the state $(12, b_1)$ means that Λ_1 and Λ_2 have occurred in order with ending block b_1. With this construction, it is clear that the state space Ω can always be relabelled as $\Omega = \{1, \ldots, k, \alpha_1, \ldots, \alpha_l\}$ and hence the transition probability matrix $\boldsymbol{M}_{(N)}$ of the imbedded Markov chain $\{Y_t\}$ associated with $W(\sigma)$ with respect to nonoverlapping counting can be obtained and always has the form

$$M_{(N)} = \left[\begin{array}{c|c} N_{(N)} & C_{(N)} \\ \hline 0 & I \end{array}\right] = (p_{ij}), \tag{4.8}$$

where the transition probabilities p_{ij} depend on the structure of the simple patterns $\Lambda_1, \ldots, \Lambda_l$ and the order of these patterns in σ.

For case (ii): $k_i = 1$ for some $i = 1, \ldots, l$, we require some minor modifications on the state space Ω. If $k_i = 1$ for some i; i.e., $\Lambda_i \in \Gamma$, then the state space Ω still has the same form as Equation (4.7) except that the set $\bigcup_{j=1}^{l} \Upsilon(\Lambda_j)$ in Ω_0 is replaced by the set $\bigcup_{j=1}^{l} \Upsilon(\Lambda_j) \setminus \bigcup_j \{\Lambda_j : k_j = 1\}$ and the set $\bigcup_{j=i+1}^{l} \Upsilon(\Lambda_j)$ in Ω_i is replaced by the set $\bigcup_{j=i+1}^{l} \Upsilon(\Lambda_j) \setminus \bigcup_j \{\Lambda_j : k_j = 1 \text{ and } j \geq i+1\}$. For example, let $\Gamma = \{b_1, b_2, b_3\}$ and $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ with $\Lambda_1 = b_1$, $\Lambda_2 = b_2 b_1 b_2$, and $\Lambda_3 = b_3$; then the state space is $\Omega = \Omega_0 \cup \Omega_1 \cup \Omega_2 \cup \{\alpha_1, \alpha_2, \alpha_3\}$, where

$$\begin{aligned} \Omega_0 &= \{(0,0), (0,b_2), (0,b_2b_1)\}, \\ \Omega_1 &= \{(1,0), (1,b_1), (1,b_2), (1,b_2b_1)\}, \\ \Omega_2 &= \{(12,0), (12,b_1), (12,b_2)\}, \end{aligned}$$

and the absorbing states α_1, α_2 and α_3 correspond to the patterns $\sigma = b_1 \circ b_2 b_1 b_2 \circ b_3$, $\Lambda_2 = b_2 b_1 b_2$ and $\Lambda_3 = b_3$, respectively.

The construction procedure for the case of overlapping counting is the same as the case of nonoverlapping counting except that the transition probabilities are defined with respect to overlapping counting. In order to make the entire imbedding procedure more transparent, we provide the following example, focusing especially on constructing the transition probabilities of the matrix M.

Example 4.1 Let $\{X_i\}$ be a sequence of three-state Markov dependent trials with possible outcomes 1, 2 and 3, and let $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ be an ordered series pattern with $\Lambda_1 = 13$, $\Lambda_2 = 31$ and $\Lambda_3 = 22$. Given the initial probabilities p_1, p_2 and p_3 and the transition probability matrix of $\{X_i\}$,

$$\boldsymbol{A} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix},$$

then the imbedded Markov chain $\{Y_t\}$ associated with the waiting time $W(\sigma)$ has state space

$$\begin{aligned} \Omega &= \{(0,0),(0,1),(0,2),(0,3),(1,0),(1,1),(1,2),(1,3),(12,0),(12,1), \\ &\quad (12,2),(12,3),\alpha_1,\alpha_2,\alpha_3\}, \end{aligned}$$

where the absorbing states α_1, α_2 and α_3 correspond to σ, Λ_2 and Λ_3, respectively. The transition probability matrix $\boldsymbol{M}_{(N)}$ with respect to nonoverlapping counting is given by

$$\begin{aligned} \boldsymbol{M}_{(N)} &= \begin{array}{c} (0,0) \\ (0,1) \\ (0,2) \\ (0,3) \\ (1,0) \\ (1,1) \\ (1,2) \\ (1,3) \\ (12,0) \\ (12,1) \\ (12,2) \\ (12,3) \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{cccc|cccc|cccc|ccc} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & p_{12} & 0 & p_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{21} & 0 & p_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{22} \\ 0 & 0 & p_{32} & p_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{11} & p_{12} & p_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{21} & 0 & p_{23} & 0 & 0 & 0 & 0 & 0 & 0 & p_{22} \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{32} & p_{33} & p_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & 0 & p_{23} & p_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \\ &= \left[\begin{array}{c|c} \boldsymbol{N}_{(N)} & \boldsymbol{C}_{(N)} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right]. \end{aligned}$$

Since neither the pattern '31' nor the pattern '22' could occur until the first occurrence of the pattern '13', the transition probabilities $P[Y_t = \alpha_3 | Y_{t-1} = (0,2)] = p_{22}$ and $P[Y_t = \alpha_2 | Y_{t-1} = (0,3)] = p_{31}$. Similarly, $P[Y_t = \alpha_3 | Y_{t-1} = (1,2)] = p_{22}$. The transition probability matrix $\boldsymbol{M}_{(O)}$ with respect to overlapping counting is almost the same except for the transition probabilities $P[Y_t = (1,1)|Y_{t-1} = (1,0)] = 0$ and $P[Y_t = (12,0)|Y_{t-1} = (1,0)] = p_{31}$. The reason of this is due to the fact that the difference between overlapping and nonoverlapping counting only occurs when $\Lambda_1 = 13$ has occurred at time $t-1$ and the outcome of the trial is '1' at time t.

From the above example, it is clear that the idea of our construction procedure is based on the character of the ordered series pattern σ and the counting procedure. Now the probability of $W(\sigma)$ and its generating function can be obtained.

Theorem 4.1 *Let $\sigma = \Lambda_1 \circ \cdots \circ \Lambda_l$ be an ordered series pattern generated by l simple patterns $\Lambda_1, \ldots, \Lambda_l$. With respect to nonoverlapping counting, we have*

(i) the probability mass function of $W(\sigma)$ is given by

$$P(W(\sigma) = n) = \boldsymbol{\xi} \boldsymbol{N}_{(N)}^{n-1} \boldsymbol{C}_{(N)}(\alpha_1), \ n = 1, 2, \ldots, \tag{4.9}$$

where $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ is the initial distribution with $P(Y_0 = (0,0)) \equiv 1$, and $\boldsymbol{C}_{(N)}(\alpha_1)$ is the first column vector of the matrix $\boldsymbol{C}_{(N)}$;

(ii) the generating function of $W(\sigma)$ is given by

$$\psi_{W(\sigma)}^{(N)}(s) = (\phi_1, \phi_2, \ldots, \phi_k) \boldsymbol{C}_{(N)}(\alpha_1), \tag{4.10}$$

where $(\phi_1(s), \ldots, \phi_k(s))$ is the solution of the simultaneous recursive equations

$$\phi_i(s) = s \boldsymbol{\xi} \boldsymbol{e}_i' + s(\phi_1(s), \ldots, \phi_k(s)) \boldsymbol{N}_{(N)}(i), \ i = 1, 2, \ldots, k,$$

and where $\boldsymbol{N}_{(N)}(i)$, $i = 1, \ldots, k$, are the column vectors of the matrix $\boldsymbol{N}_{(N)}$, and $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, $i = 1, \ldots, k$, are unit vectors.

Proof. Let A denote the set of all absorbing states. Since the absorbing state α_1 corresponds to the pattern σ, it follows from Lemma 3.1 that

$$P(W(\sigma) = n) = P(Y_n = \alpha_1, Y_{n-1} \notin A, \ldots, Y_1 \notin A) = \boldsymbol{\xi} \boldsymbol{N}_{(N)}^{n-1} \boldsymbol{C}_{(N)}(\alpha_1).$$

This completes the proof of part (i). For the proof of part (ii), we let $\Lambda^* = \sigma \cup (\bigcup_{i=2}^{l} \Lambda_i)$. Then it follows from Equation (4.5) that

$$\psi_{W(\sigma)}^{(N)}(s) = \sum_{n=1}^{\infty} s^n P(W(\Lambda^*) = n, W(\sigma) = n) = (\phi_1, \phi_2, \ldots, \phi_k)\boldsymbol{C}_{(N)}(\alpha_1).$$

This completes the proof. $\square$

The results in Theorem 4.1 remain applicable to the case of overlapping counting, except that the matrices $\boldsymbol{N}_{(N)}$ and $\boldsymbol{C}_{(N)}$ are replaced by $\boldsymbol{N}_{(O)}$ and $\boldsymbol{C}_{(O)}$, respectively.

There is yet another way to obtain the generating function of $W(\sigma)$. Let $\Lambda_j^* = \bigcup_{i=j}^{l} \Lambda_i$, $j = 1, \ldots, l$, and let $W(\Lambda_j|\Lambda_j, \ldots, \Lambda_l)$ be the waiting time to the first occurrence of the pattern Λ_j, and that Λ_j occurs first among all the patterns $\Lambda_j, \ldots, \Lambda_l$. For the same reason as stated in Remark 4.1, the random variables $W(\Lambda_j|\Lambda_j, \ldots, \Lambda_l)$, $j = 1, \ldots, l-1$, are improper random variables. However, the random variable $W(\Lambda_l|\Lambda_l) = W(\Lambda_l)$ is a proper random variable. For convenience, we still denote the probability generating function of $W(\Lambda_l)$ by $\psi_{W(\Lambda_l)}(s)$ and use the term "generating function" instead of "probability generating function". For each $j = 1, \ldots, l$, we imbed the random variable $W(\Lambda_j^*)$ with the same arguments as in Chapter 3. Then the corresponding transition probability matrices of the imbedded Markov chains associated with $W(\Lambda_j^*)$ have the form

$$\boldsymbol{M}_{W(\Lambda_j^*)} = \left[\begin{array}{c|c} \boldsymbol{N}_{W(\Lambda_j^*)} & \boldsymbol{C}_{W(\Lambda_j^*)} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right], \; j = 1, \ldots, l.$$

Applying Equation (4.5), we can obtain the generating functions $\psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}(s)$, $j = 1, \ldots, l$. The next theorem shows that the generating function $\psi_{W(\sigma)}(s)$ can be expressed in terms of the product of the generating functions $\psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}(s)$, $j = 1, \ldots, l$.

Remark 4.2 When we study the generating function of $W(\sigma)$ in terms of the generating functions $\psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}(s)$, $j = 1,\ldots,l$, we need to consider the order among all the simple patterns $\Lambda_1,\ldots,\Lambda_l$. The initial distributions of the imbedded Markov chains associated with $W(\Lambda_j^*)$, $j = 1,\ldots,l$, may vary a lot according to underlying sequence or counting procedure. Therefore, in the sequel, we use the notation $\psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}(s|\boldsymbol{\xi})$ to mean that such a generating function depends on the initial distribution of its imbedded Markov chain.

Theorem 4.2 *The generating function of $W(\sigma)$ with respect to nonoverlapping counting is given by*

$$\psi^{(N)}_{W(\sigma)}(s) \;=\; \prod_{j=1}^{l} \psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{j-1})\right), \tag{4.11}$$

where $\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{j-1}) = (1,0,\ldots,0)$, $j = 1,\ldots,l$, are the initial distributions with initial state $\emptyset_{j^\star} = j^\star$ ($P(Y_0 = \emptyset_{j^\star}) \equiv 1$), $j^\star$ is the last element of the pattern Λ_{j-1}, and $\boldsymbol{\xi}^{(N)}_{\emptyset}(\Lambda_0) = \boldsymbol{\xi} = (1,0,\ldots,0)$ (with usual initial state $\emptyset_{j^\star} = \emptyset$) by convention.

Proof. For the ordered series pattern $\sigma = \Lambda_1 \circ \cdots \circ \Lambda_l$, it follows from its definition that Λ_1 must occur first among the patterns $\Lambda_1,\ldots,\Lambda_l$, Λ_2 must occur first among the patterns $\Lambda_2,\ldots,\Lambda_l$ given that Λ_1 has occurred, and so on. Hence, we have

$$W(\sigma) = \sum_{j=1}^{l} W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l).$$

Since Λ_1 is the first manifest pattern, it is clear that the initial distribution of the imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_1^*)$ is $\boldsymbol{\xi} = (1,0,\ldots,0) = \boldsymbol{\xi}^{(N)}_{\emptyset}(\Lambda_0)$ with initial probability $P(Y_0 = \emptyset) \equiv 1$. Thus, the generating function $\psi_{W(\Lambda_1|\Lambda_1\cdots\Lambda_l)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset}(\Lambda_0)\right)$ can be obtained via imbedding the random variable $W(\Lambda_1^*)$. Similarly, since Λ_j, $j = 2,\ldots,l$, occur in sequential order, the initial states $\emptyset_{j^\star}$ of the imbedded Markov chains associated with $W(\Lambda_j^*)$, $j = 2,\ldots,l$, are the last elements $j^\star$ of Λ_{j-1}, $j = 2, \ldots, l$, respectively. Hence, the initial distributions are $\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{j-1}) = (1, 0, \ldots, 0)$ with initial state $\emptyset_{j^\star} = j^\star$ and $P(Y_0 = \emptyset_{j^\star}) \equiv 1$. The generating functions $\psi_{W(\Lambda_j|\Lambda_j\cdots\Lambda_l)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{j-1})\right)$, $j = 2, \ldots, l$, can be obtained via imbedding the random variables $W(\Lambda^*_j)$, $j = 2, \ldots, l$, respectively. Now, since the random variables $W(\Lambda_j|\Lambda_j\cdots\Lambda_l)$, $j = 1, \ldots, l$, are conditionally independent, the generating function $\psi^{(N)}_{W(\sigma)}(s)$ of $W(\sigma)$ is the product of the generating functions $\psi_{W(\Lambda_j|\Lambda_j\cdots\Lambda_l)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{j-1})\right)$, $j = 1, \ldots, l$. This establishes equality (4.11). $\square$

We provide a detailed example to make our result more transparent.

Example 4.2 Let $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ be an ordered series pattern with $\Lambda_1 = 13$, $\Lambda_2 = 31$ and $\Lambda_3 = 22$ as in Example 4.1. We first imbed the random variable $W(\Lambda^*_1)$, where $\Lambda^*_1 = \bigcup_{i=1}^3 \Lambda_i$. Then the imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda^*_1)$ has state space $\Omega = \{\emptyset, 1, 2, 3, \alpha_1, \alpha_2, \alpha_3\}$ and transition probability matrix

$$\boldsymbol{M}_{W(\Lambda^*_1)} = \begin{array}{c} \emptyset \\ 1 \\ 2 \\ 3 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{cccc|ccc} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 \\ 0 & p_{11} & p_{12} & 0 & p_{13} & 0 & 0 \\ 0 & p_{21} & 0 & p_{23} & 0 & 0 & p_{22} \\ 0 & 0 & p_{32} & p_{33} & 0 & p_{31} & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{N}_{W(\Lambda^*_1)} & \boldsymbol{C}_{W(\Lambda^*_1)} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right].$$

Applying Equation (4.5) and using our computer program, we obtain

$$\psi_{W(\Lambda_1|\Lambda_1\Lambda_2\Lambda_3)}(s|\boldsymbol{\xi}) = \frac{p_{13}s^2 \times \Delta_1}{\Delta_2},$$

where

$$\begin{aligned} \Delta_1 &= p_1 - p_1p_{33}s + p_2p_{21}s - p_1p_{23}p_{32}s^2 - p_2p_{21}p_{33}s^2 + p_3p_{21}p_{32}s^2, \\ \Delta_2 &= 1 - p_{11}s - p_{33}s - p_{23}p_{32}s^2 - p_{12}p_{21}s^2 + p_{11}p_{33}s^2 + p_{11}p_{23}p_{32}s^3 + p_{12}p_{21}p_{33}s^3. \end{aligned}$$

Given the occurrence of the pattern Λ_1, we imbed the random variable $W(\Lambda_2^*)$, where $\Lambda_2^* = \bigcup_{i=2}^{3} \Lambda_i$. The imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_2^*)$ has state space $\Omega = \{\emptyset_3 = 3, 1, 2, 3, \alpha_2, \alpha_3\}$ and transition probability matrix

$$\boldsymbol{M}_{W(\Lambda_2^*)} = \begin{array}{c} \emptyset_3 \\ 1 \\ 2 \\ 3 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{cccc|cc} 0 & p_{31} & p_{32} & p_{33} & 0 & 0 \\ 0 & p_{11} & p_{12} & p_{13} & 0 & 0 \\ 0 & p_{21} & 0 & p_{23} & 0 & p_{22} \\ 0 & 0 & p_{32} & p_{33} & p_{31} & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{N}_{W(\Lambda_2^*)} & \boldsymbol{C}_{W(\Lambda_2^*)} \\ \hline \boldsymbol{0} & \boldsymbol{I} \end{array}\right].$$

Since Λ_1 occurs before Λ_2, the initial state $\emptyset_{j^\star} = j^\star$ of this imbedded Markov chain is the last element of Λ_1; i.e. $\emptyset_3 = 3$. Hence, the transition probabilities $P(Y_1 = j | Y_0 = \emptyset_3 = 3) = p_{3j}$, $j = 1, 2, 3$. Using our computer program, it yields

$$\psi_{W(\Lambda_2|\Lambda_2\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_3}^{(N)}(\Lambda_1)\right) = \frac{p_{31}s^2 \times \Delta_3}{\Delta_4},$$

where

$$\begin{aligned} \Delta_3 &= p_{33} - p_{11}p_{33}s + p_{23}p_{32}s + p_{13}p_{31}s - p_{12}p_{21}p_{33}s^2 - p_{11}p_{23}p_{32}s^2 + p_{12}p_{23}p_{31}s^2 \\ &\quad + p_{13}p_{21}p_{32}s^2, \\ \Delta_4 &= 1 - p_{11}s - p_{33}s - p_{23}p_{32}s^2 - p_{12}p_{21}s^2 + p_{11}p_{33}s^2 - p_{13}p_{21}p_{32}s^3 + p_{11}p_{23}p_{32}s^3 \\ &\quad + p_{12}p_{21}p_{33}s^3. \end{aligned}$$

By the same token, we obtain from our computer program that

$$\psi_{W(\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_1}^{(N)}(\Lambda_2)\right) = \frac{p_{22}s^2 \times \Delta_5}{\Delta_6},$$

where

$$\begin{aligned} \Delta_5 &= p_{12} - p_{12}p_{33}s + p_{13}p_{32}s, \\ \Delta_6 &= 1 - p_{11}s - p_{33}s - p_{23}p_{32}s^2 - p_{12}p_{21}s^2 - p_{13}p_{31}s^2 + p_{11}p_{33}s^2 - p_{13}p_{21}p_{32}s^3 \\ &\quad - p_{12}p_{23}p_{31}s^3 + p_{11}p_{23}p_{32}s^3 + p_{12}p_{21}p_{33}s^3. \end{aligned}$$

By Theorem 4.2, we can obtain the generating function $\psi^{(N)}_{W(\sigma)}(s)$ by taking the product of the generating functions $\psi_{W(\Lambda_1|\Lambda_1\Lambda_2\Lambda_3)}(s|\boldsymbol{\xi})$, $\psi_{W(\Lambda_2|\Lambda_2\Lambda_3)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_3}(\Lambda_1)\right)$ and $\psi_{W(\Lambda_3)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_1}(\Lambda_2)\right)$.

When $\{X_i\}$ is a sequence of i.i.d. multistate trials, it is easy to see that $\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_{j-1}) = \boldsymbol{\xi}^{(N)}_{\emptyset}(\Lambda_{j-1}) = \boldsymbol{\xi}$ for all $j = 1, \ldots, l$. In the case of overlapping counting, the result in Theorem 4.2 still holds except that the initial distributions in Equation (4.11) are replaced by $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_{j-1}) = (0, \ldots, 1, \ldots, 0) = \boldsymbol{e}_{j^\circ}$, $j = 1, \ldots, l$, where the state j° is the longest overlap of the patterns Λ_{j-1} with Λ_k, $k = j, \ldots, l$, in the sense of overlapping counting and $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_0) = \boldsymbol{\xi} = (1, 0, \ldots, 0)$ by convention (the state j° corresponding to the usual initial state $\emptyset$). For example, let $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ with $\Lambda_1 = 133$, $\Lambda_2 = 32$ and $\Lambda_3 = 331$; then the state 'j°' for $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_1)$ corresponds to the state '33' (the longest overlap of Λ_1 with Λ_2 and Λ_1 with Λ_3). If there are no overlaps, then it is trivial that $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_{j-1}) = \boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_{j-1})$.

Corollary 4.1 *The generating function of $W(\sigma)$ with respect to overlapping counting is given by*

$$\psi^{(O)}_{W(\sigma)}(s) = \prod_{j=1}^{l} \psi_{W(\Lambda_j|\Lambda_j,\ldots,\Lambda_l)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_{j-1})\right), \tag{4.12}$$

where $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_{j-1}) = (0, \ldots, 1, \ldots, 0) = \boldsymbol{e}_{j^\circ}$, $j = 1, \ldots, l$, are the initial distributions such that the state j° is the longest overlap of the patterns Λ_{j-1} with Λ_k, $k = j, \ldots, l$, in the sense of overlapping counting, and $\boldsymbol{\xi}^{(O)}_{\emptyset_{j\star}}(\Lambda_0) = \boldsymbol{\xi} = (1, 0, \ldots, 0)$ by convention (the state j° corresponding to the usual initial state $\emptyset$).

Example 4.3 Under the same setup as in Example 4.1, consider the ordered series pattern $\sigma = \Lambda_1 \circ \Lambda_2 \circ \Lambda_3$ with $\Lambda_1 = 133$, $\Lambda_2 = 32$ and $\Lambda_3 = 331$ with respect to overlapping counting. We first imbed the random variable $W(\Lambda_1^*)$, where $\Lambda_1^* =$

$\cup_{i=1}^{3} \Lambda_i$. The imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_1^*)$ has state space $\Omega = \{\emptyset, 1, 2, 3, 13, 33, \alpha_1, \alpha_2, \alpha_3\}$ and transition probability matrix

$$M_{W(\Lambda_1^*)} = \begin{array}{c} \emptyset \\ 1 \\ 2 \\ 3 \\ 13 \\ 33 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{cccccc|ccc} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & p_{12} & 0 & p_{13} & 0 & 0 & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{31} & 0 & 0 & 0 & p_{33} & 0 & p_{32} & 0 \\ 0 & p_{31} & 0 & 0 & 0 & 0 & p_{33} & p_{32} & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{33} & 0 & p_{32} & p_{31} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} N_{W(\Lambda_1^*)} & C_{W(\Lambda_1^*)} \\ \hline \mathbf{0} & I \end{array}\right].$$

The initial distribution of the imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_1^*)$ is $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ $(P(Y_0 = \emptyset) \equiv 1)$. Applying Equation (4.5) and using our computer program, we obtain

$$\psi_{W(\Lambda_1|\Lambda_1\Lambda_2\Lambda_3)}(s|\boldsymbol{\xi}) = \frac{p_{13}p_{33}s^3 \times \Delta_1}{\Delta_2},$$

where

$$\begin{aligned} \Delta_1 &= p_1 - p_1p_{22}s + p_2p_{21}s + p_3p_{31}s - p_3p_{22}p_{31}s^2 + p_2p_{23}p_{31}s^2, \\ \Delta_2 &= 1 - p_{11}s - p_{22}s - p_{12}p_{21}s^2 - p_{13}p_{31}s^2 + p_{11}p_{22}s^2 - p_{12}p_{23}p_{31}s^3 + p_{13}p_{22}p_{31}s^3. \end{aligned}$$

Given the occurrence of the pattern Λ_1, we imbed the random variable $W(\Lambda_2^*)$, where $\Lambda_2^* = \cup_{i=2}^{3} \Lambda_i$. The imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_2^*)$ has state space $\Omega = \{\emptyset_3 = 3, 1, 2, 3, 33, \alpha_2, \alpha_3\}$ and transition probability matrix

$$M_{W(\Lambda_2^*)} = \begin{array}{c} \emptyset_3 \\ 1 \\ 2 \\ 3 \\ 33 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{ccccc|cc} 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 \\ 0 & p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & 0 & 0 \\ 0 & p_{31} & 0 & 0 & p_{33} & p_{32} & 0 \\ 0 & 0 & 0 & 0 & p_{33} & p_{32} & p_{31} \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} N_{W(\Lambda_2^*)} & C_{W(\Lambda_2^*)} \\ \hline \mathbf{0} & I \end{array}\right].$$

For the same reason as stated in Example 4.2, the initial state $\emptyset_{j^\star} = j^\star$ of this imbedded chain is the last element of Λ_1; i.e. $\emptyset_3 = 3$. But note that, since the longest overlap of the patterns Λ_1 with Λ_2 and Λ_1 with Λ_3 is '33', the initial distribution of the imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda_2^*)$ is $\boldsymbol{\xi}_{\emptyset_3}^{(O)}(\Lambda_1) = (0,0,0,0,1) = e_{j^\circ}$ (the state 'j°' corresponds to the state '33'). Using our computer program, we find that

$$\psi_{W(\Lambda_2|\Lambda_2\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_3}^{(O)}(\Lambda_1)\right) = \frac{p_{32}s}{1-p_{33}s}.$$

By the same token, we obtain from our computer program that

$$\psi_{W(\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_2}^{(O)}(\Lambda_2)\right) = \frac{p_{31}p_{33}s^3 \times \Delta_3}{\Delta_4},$$

where

$$\begin{aligned}
\Delta_3 &= p_{23} - p_{11}p_{23}s + p_{13}p_{21}s,\\
\Delta_4 &= 1 - p_{11}s - p_{22}s - p_{33}s - p_{12}p_{21}s^2 - p_{13}p_{31}s^2 - p_{23}p_{32}s^2 + p_{11}p_{22}s^2\\
&\quad + p_{11}p_{33}s^2 + p_{22}p_{33}s^2 - p_{11}p_{22}p_{33}s^3 - p_{12}p_{23}p_{31}s^3 - p_{13}p_{21}p_{32}s^3 + p_{11}p_{23}p_{32}s^3\\
&\quad + p_{12}p_{21}p_{33}s^3 + p_{13}p_{22}p_{31}s^3 + p_{13}p_{31}p_{33}s^3 - p_{13}p_{22}p_{31}p_{33}s^4 + p_{12}p_{23}p_{31}p_{33}s^4.
\end{aligned}$$

By Corollary 4.1, we can obtain the generating function $\psi_{W(\sigma)}^{(O)}(s)$ by taking the product of the generating functions $\psi_{W(\Lambda_1|\Lambda_1\Lambda_2\Lambda_3)}(s|\boldsymbol{\xi})$, $\psi_{W(\Lambda_2|\Lambda_2\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_3}^{(O)}(\Lambda_1)\right)$ and $\psi_{W(\Lambda_3)}\left(s|\boldsymbol{\xi}_{\emptyset_2}^{(O)}(\Lambda_2)\right)$.

4.3 Later Waiting Time Distributions of Two Simple Patterns

Let Λ_1 and Λ_2 be two simple patterns and $\Lambda = \Lambda_1 \cup \Lambda_2$. Then we have

$$\mathcal{P} = \{\sigma_1, \sigma_2 : \sigma_1 = \Lambda_1 \circ \Lambda_2 \text{ and } \sigma_2 = \Lambda_2 \circ \Lambda_1\},$$

and the later pattern $\Lambda_L = \sigma_1 \cup \sigma_2$. We introduce three different ways to obtain the exact distribution and probability generating function of $W(\Lambda_L)$. Firstly, following the construction procedure described in Section 4.2, we can imbed the random variables $W(\sigma_1)$ and $W(\sigma_2)$ into two separate Markov chains. With respect to nonoverlapping counting, let $\boldsymbol{M}_{1(N)}$ and $\boldsymbol{M}_{2(N)}$ denote the transition probability matrices associated with the imbedded Markov chains of $W(\sigma_1)$ and $W(\sigma_2)$, respectively.

Theorem 4.3 *Let $\Lambda_L = \sigma_1 \cup \sigma_2$ be a later pattern with $\sigma_1 = \Lambda_1 \circ \Lambda_2$ and $\sigma_2 = \Lambda_2 \circ \Lambda_1$. With respect to nonoverlapping counting, we have*

(i) the exact distribution of $W(\Lambda_L)$ is given by

$$\begin{aligned} P(W(\Lambda_L) = n) &= P(W(\sigma_1) = n) + P(W(\sigma_2) = n) \\ &= \boldsymbol{\xi} \boldsymbol{N}_{1(N)}^{n-1} \boldsymbol{C}_{1(N)}(\alpha_1) + \boldsymbol{\xi} \boldsymbol{N}_{2(N)}^{n-1} \boldsymbol{C}_{2(N)}(\alpha_1), \end{aligned} \tag{4.13}$$

for $n = 1, 2, \ldots$, where $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ is the initial distribution with $P(Y_0 = (0,0)) \equiv 1$, and $\boldsymbol{C}_{i(N)}(\alpha_1)$, $i = 1, 2$, are the first column vectors of the matrices $\boldsymbol{C}_{i(N)}$, $i = 1, 2$, respectively;

(ii) the probability generating function of $W(\Lambda_L)$ is given by

$$\varphi_{W(\Lambda_L)}^{(N)}(s) = \psi_{W(\sigma_1)}^{(N)}(s) + \psi_{W(\sigma_2)}^{(N)}(s). \tag{4.14}$$

Proof. From Theorem 4.1(i), we have

$$P(W(\sigma_i) = n) = \boldsymbol{\xi} \boldsymbol{N}_{i(N)}^{n-1} \boldsymbol{C}_{i(N)}(\alpha_1), \ i = 1, 2.$$

It follows from the definitions of the random variables $W(\sigma_1)$, $W(\sigma_2)$ and $W(\Lambda_L)$ that the event $\{W(\Lambda_L) = n\}$ is the union of the events $\{W(\sigma_1) = n\}$ and $\{W(\sigma_2) = n\}$, where $\{W(\sigma_1) = n\}$ and $\{W(\sigma_2) = n\}$ are mutually exclusive. Hence, we have

$$\begin{aligned} P(W(\Lambda_L) = n) &= P(W(\sigma_1) = n) + P(W(\sigma_2) = n) \\ &= \boldsymbol{\xi} \boldsymbol{N}_{1(N)}^{n-1} \boldsymbol{C}_{1(N)}(\alpha_1) + \boldsymbol{\xi} \boldsymbol{N}_{2(N)}^{n-1} \boldsymbol{C}_{2(N)}(\alpha_1). \end{aligned}$$

This completes the proof of part (i). Part (ii) follows immediately from the definition of $\varphi^{(N)}_{W(\Lambda_L)}(s)$. $\square$

Note that the generating functions $\psi^{(N)}_{W(\sigma_1)}(s)$ and $\psi^{(N)}_{W(\sigma_2)}(s)$ can be obtained through Theorem 4.1(ii). In the case of overlapping counting, the above results still hold except that the transition probability matrices are replaced by $\boldsymbol{M}_{1(O)}$ and $\boldsymbol{M}_{2(O)}$, respectively.

Without imbedding the random variables $W(\sigma_1)$ and $W(\sigma_2)$ into two Markov chains separately, we can also imbed the random variable $W(\Lambda_L)$ into a Markov chain directly. Define a Markov chain $\{Y_t\}$ on the state space Ω having the form

$$\Omega = \Omega_0 \cup \Omega_1 \cup \{\alpha_1, \alpha_2\}, \tag{4.15}$$

where Ω_0 and Ω_1 are defined as

$$\begin{aligned} \Omega_0 &= \left\{ (0, v) : v \in \bigcup_{j=1}^{2} \Upsilon(\Lambda_j) \right\}, \\ \Omega_1 &= \{(u, v) : u = 1, 2 \text{ and } v \in \Upsilon(\Lambda_j),\ j \neq u\}, \end{aligned}$$

and α_1 and α_2 are absorbing states corresponding to the ordered series patterns σ_1 and σ_2, respectively. Note that the imbedded Markov chains for nonoverlapping and overlapping counting cases are defined on the same state space Ω, but their transition probabilities are slightly different according to the counting procedures and structures of Λ_1 and Λ_2. However, both transition probability matrices defined on the state space Ω have the same form as Equation (3.1). Applying Equations (4.1)$-$(4.3) to the pattern Λ_L, this imbedding procedure provides a second way to obtain the exact distribution, mean and probability generating function of $W(\Lambda_L)$.

The next theorem provides a third way to obtain the probability generating function of $W(\Lambda_L)$ under nonoverlapping counting.

Theorem 4.4 *Let $\Lambda_L = \sigma_1 \cup \sigma_2$ be a later pattern with $\sigma_1 = \Lambda_1 \circ \Lambda_2$ and $\sigma_2 = \Lambda_2 \circ \Lambda_1$. With respect to nonoverlapping counting, we have*

(i) *the generating functions of $W(\sigma_1)$ and $W(\sigma_2)$ are given by*

$$\psi^{(N)}_{W(\sigma_1)}(s) = \psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_1)\right), \tag{4.16}$$

$$\psi^{(N)}_{W(\sigma_2)}(s) = \psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_2)\right), \tag{4.17}$$

where $\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_i)$, $i = 1, 2$, are the initial distributions as defined in Theorem 4.2;

(ii) *the probability generating function of $W(\Lambda_L)$ is given by*

$$\begin{aligned}\varphi^{(N)}_{W(\Lambda_L)}(s) &= \psi^{(N)}_{W(\sigma_1)}(s) + \psi^{(N)}_{W(\sigma_2)}(s) \\ &= \psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_1)\right) \\ &\quad + \psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_2)\right). \end{aligned} \tag{4.18}$$

Proof. Equalities (4.16) and (4.17) are special cases of Equation (4.11). Since the events $\{W(\sigma_1) = n\}$ and $\{W(\sigma_2) = n\}$ are mutually exclusive, it follows from the definition and the results of part (i) that

$$\begin{aligned}&\varphi^{(N)}_{W(\Lambda_L)}(s) = \sum_{n=1}^{\infty} s^n \left(\sum_{i=1}^{2} P(W(\sigma_i) = n)\right) = \psi^{(N)}_{W(\sigma_1)}(s) + \psi^{(N)}_{W(\sigma_2)}(s) \\ &= \psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_1)\right) + \psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\, \psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j\star}}(\Lambda_2)\right).\end{aligned}$$

This completes the proof. □

Corollary 4.2 *Let $\Lambda_L = \sigma_1 \cup \sigma_2$ be a later pattern with $\sigma_1 = \Lambda_1 \circ \Lambda_2$ and $\sigma_2 = \Lambda_2 \circ \Lambda_1$. With respect to overlapping counting, we have* *(i) the generating functions of $W(\sigma_1)$ and $W(\sigma_2)$ are given by*

$$\psi^{(O)}_{W(\sigma_1)}(s) = \psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\,\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j*}}(\Lambda_1)\right), \qquad (4.19)$$

$$\psi^{(O)}_{W(\sigma_2)}(s) = \psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\,\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j*}}(\Lambda_2)\right), \qquad (4.20)$$

where $\boldsymbol{\xi}^{(O)}_{\emptyset_{j*}}(\Lambda_i)$, $i = 1, 2$, are the initial distributions as defined in Corollary 4.1;

(ii) the probability generating function of $W(\Lambda_L)$ is given by

$$\begin{aligned}\varphi^{(O)}_{W(\Lambda_L)}(s) &= \psi^{(O)}_{W(\sigma_1)}(s) + \psi^{(O)}_{W(\sigma_2)}(s)\\ &= \psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\,\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j*}}(\Lambda_1)\right)\\ &\quad + \psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi})\,\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j*}}(\Lambda_2)\right). \qquad (4.21)\end{aligned}$$

In order to demonstrate various ways of obtaining the probability generating functions and to make our theoretical results more transparent, we provide the following example.

Example 4.4 Let $\{X_i\}$ be a sequence of (i.i.d. or first-order homogeneous Markov dependent) three-state trials with possible outcomes 1, 2 and 3, and let $\Lambda_1 = 13$ and $\Lambda_2 = 31$ be two simple patterns. We are interested in finding the probability generating function for the later pattern $\Lambda_L = \sigma_1 \cup \sigma_2$, where $\sigma_1 = 13 \circ 31$ and $\sigma_2 = 31 \circ 13$. We show the three different ways of obtaining the probability generating function of $W(\Lambda_L)$. We first consider the i.i.d. case. The imbedded Markov chain $\{Y_t\}$ associated with the random variable $W(\sigma_1)$ has state space Ω given by

$$\Omega = \{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (1,3), \alpha_1, \alpha_2\},$$

where α_1 and α_2 correspond to the patterns σ_1 and Λ_2, respectively, and transition probability matrix with respect to nonoverlapping counting given by

$$
M_{(N)} = \begin{array}{c} (0,0) \\ (0,1) \\ (0,2) \\ (0,3) \\ (1,0) \\ (1,1) \\ (1,2) \\ (1,3) \\ \alpha_1 \\ \alpha_2 \end{array}
\left[\begin{array}{cccccccc|cc}
0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p_1 & p_2 & 0 & p_3 & 0 & 0 & 0 & 0 & 0 \\
0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & p_1 \\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p_2 & p_3 & p_1 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
= \left[\begin{array}{c|c} N_{(N)} & C_{(N)} \\ \hline \mathbf{0} & I \end{array}\right].
$$

Theorem 4.1(ii) and our computer program yield

$$\psi^{(N)}_{W(\sigma_1)}(s) = \frac{p_1^2 p_3^2 s^4 (1 - p_3 s)}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}.$$

Following the same procedure for $W(\sigma_2)$, we obtain

$$\psi^{(N)}_{W(\sigma_2)}(s) = \frac{p_1^2 p_3^2 s^4 (1 - p_1 s)}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}.$$

Hence, summing over $\psi^{(N)}_{W(\sigma_1)}(s)$ and $\psi^{(N)}_{W(\sigma_2)}(s)$, we have

$$\varphi^{(N)}_{W(\Lambda_L)}(s) = \frac{p_1^2 p_3^2 s^4 (2 - p_1 s - p_3 s)}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}. \tag{4.22}$$

By the same token, we obtain from our computer program that

$$
\begin{aligned}
\psi^{(O)}_{W(\sigma_1)}(s) &= \frac{p_1^2 p_3 s^3 (1 - p_3 s)(1 - p_1 s - p_2 s)}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}, \\
\psi^{(O)}_{W(\sigma_2)}(s) &= \frac{p_1 p_3^2 s^3 (1 - p_1 s)(1 - p_2 s - p_3 s)}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}, \\
\varphi^{(O)}_{W(\Lambda_L)}(s) &= \frac{\Delta}{(1 - s + p_1 p_3 s^2)(1 - s + p_1 p_3 s^2 + p_1 p_2 p_3 s^3)}, \qquad (4.23) \\
\Delta &= p_1 p_3 s^3 (p_1 + p_3 - p_1 p_2 s - p_2 p_3 s - p_1^2 s - p_3^2 s \\
&\quad - 2 p_1 p_3 s + p_1 p_3^2 s^2 + p_1^2 p_3 s^2 + 2 p_1 p_2 p_3 s^2).
\end{aligned}
$$

Now, let us imbed the random variable $W(\Lambda_L)$ directly. The imbedded Markov chain $\{Y_t\}$ has state space

$$\begin{aligned}\Omega &= \{(0,0),(0,1),(0,2),(0,3),(1,0),(1,1),(1,2),(1,3),\\ &\qquad (2,0),(2,1),(2,2),(2,3),\alpha_1,\alpha_2\},\end{aligned}$$

where α_1 and α_2 correspond to the pattern σ_1 and σ_2, respectively, and transition probability matrix with respect to nonoverlapping counting given by

$$\begin{aligned}\boldsymbol{M}_{(N)} &= \begin{array}{c} (0,0)\\(0,1)\\(0,2)\\(0,3)\\(1,0)\\(1,1)\\(1,2)\\(1,3)\\(2,0)\\(2,1)\\(2,2)\\(2,3)\\ \alpha_1\\ \alpha_2 \end{array}\left[\begin{array}{cccccccccccc|cc}
0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & p_1 & p_2 & 0 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & p_2 & p_3 & 0 & 0 & 0 & 0 & p_1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & p_2 & p_3 & 0 & 0 & 0 & 0 & p_1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & 0 & 0 & p_3\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0\\ \hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]\\
&= \left[\begin{array}{c|c} \boldsymbol{N}_{(N)} & \boldsymbol{C}_{(N)}\\ \hline \mathbf{0} & \boldsymbol{I}\end{array}\right].\end{aligned}$$

The transition probability matrix $\boldsymbol{M}_{(O)}$ with respect to overlapping counting can be obtained by the same arguments. Applying Equation (4.3) yields the same results given by Equations (4.22) and (4.23). Further, we compute the following generating functions:

$$\psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi}) = \frac{p_1p_3s^2(1-p_3s)}{1-s+p_1p_3s^2+p_1p_2p_3s^3},$$

$$\psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi}) = \frac{p_1p_3s^2(1-p_1s)}{1-s+p_1p_3s^2+p_1p_2p_3s^3},$$

$$\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}_{\emptyset}^{(N)}(\Lambda_2)\right) = \frac{p_1p_3s^2}{1-s+p_1p_3s^2},$$

$$\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}_{\emptyset}^{(N)}(\Lambda_1)\right) = \frac{p_1p_3s^2}{1-s+p_1p_3s^2},$$

$$\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}_{\emptyset}^{(O)}(\Lambda_2)\right) = \frac{p_3s(1-p_2s-p_3s)}{1-s+p_1p_3s^2},$$

$$\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}_{\emptyset}^{(O)}(\Lambda_1)\right) = \frac{p_1s(1-p_1s-p_2s)}{1-s+p_1p_3s^2}.$$

Theorem 4.4 and Corollary 4.2 yield the same results given by Equations (4.22) and (4.23) after some simple algebra. The probability generating functions for the Markov dependent case can be obtained as the i.i.d. case. Since the analytic form of the probability generating functions is quite complicated, we set the initial probabilities to be $p_1 = 0.3$, $p_2 = 0.3$ and $p_3 = 0.4$, and the transition probability matrix of $\{X_i\}$ as

$$\boldsymbol{A} = \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.25 & 0.30 & 0.45 \\ 0.35 & 0.35 & 0.30 \end{bmatrix}.$$

Imbedding the random variables $W(\sigma_1)$ and $W(\sigma_2)$ separately or imbedding the random variable $W(\Lambda_L)$ directly produces $\varphi^{(N)}_{W(\Lambda_L)}(s) = \Delta_1/\Delta_2$, where

$$\begin{aligned} \Delta_1 &= -0.525s^4(15687s^7 + 33003s^6 + 974590s^5 + 8278200s^4 + 27928000s^3 \\ &\quad + 144240000s^2 - 1390400000s + 1344000000), \\ \Delta_2 &= (201s^3 + 50s^2 - 3600s + 4000)(48s^3 + 25s^2 - 1800s + 2000) \\ &\quad (51s^3 - 50s^2 + 3600s - 4000), \end{aligned}$$

and $\varphi^{(O)}_{W(\Lambda_L)}(s) = \Delta_3/\Delta_4$, where

$$\begin{aligned} \Delta_3 &= 10.5s^3(5241s^7 + 14855s^6 - 326970s^5 + 3317400s^4 + 26556000s^3 \\ &\quad - 215520000s^2 + 403200000s - 224000000), \\ \Delta_4 &= (201s^3 + 50s^2 - 3600s + 4000)(48s^3 + 25s^2 - 1800s + 2000) \\ &\quad (51s^3 - 50s^2 + 3600s - 4000). \end{aligned}$$

Further, we compute the following generating functions:

$$\begin{aligned}
\psi_{W(\Lambda_1|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi}) &= \frac{-0.3s^2\,(31s^2+420s-1200)}{201s^3+50s^2-3600s+4000},\\
\psi_{W(\Lambda_2|\Lambda_1\Lambda_2)}(s|\boldsymbol{\xi}) &= \frac{0.70s^2\,(19s^2-210s+800)}{201s^3+50s^2-3600s+4000},\\
\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_1}(\Lambda_2)\right) &= \frac{-3s^2\,(3s^2+10s+120)}{51s^3-50s^2+3600s-4000},\\
\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_3}(\Lambda_1)\right) &= \frac{5.25s^2\,(s^2+11s+40)}{48s^3+25s^2-1800s+2000},\\
\psi_{W(\Lambda_1)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_1}(\Lambda_2)\right) &= \frac{3s(27s^2+240s-400)}{51s^3-50s^2+3600s-4000},\\
\psi_{W(\Lambda_2)}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_3}(\Lambda_1)\right) &= \frac{-7s(s^2+60s-100)}{48s^3+25s^2-1800s+2000}.
\end{aligned}$$

From Theorem 4.4 and Corollary 4.2, we obtain the same results for the probability generating functions $\varphi^{(N)}_{W(\Lambda_L)}(s)$ and $\varphi^{(O)}_{W(\Lambda_L)}(s)$ after simplification.

4.4 Later Waiting Time Distributions of l ($l \geq 2$) Simple Patterns

In this section, we extend the results from the previous section to l ($l \geq 2$) simple patterns. Let $\Lambda_1, \ldots, \Lambda_l$ be l simple patterns. Then we have the set $\mathcal{P}$ of all ordered series patterns generated by these l simple patterns as defined in (2.1) and the later pattern $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$. Firstly, following the construction procedure described in Section 4.2, we can imbed the random variables $W(\sigma_i)$, $i = 1, \ldots, l!$, into $l!$ separate Markov chains. With respect to nonoverlapping counting, let $\boldsymbol{M}_{i(N)}$, $i = 1, \ldots, l!$, be the transition probability matrices associated with the imbedded Markov chains of $W(\sigma_i)$, $i = 1, \ldots, l!$, respectively. The next theorem is an extension of Theorem 4.3.

Theorem 4.5 *Let $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$ ($l \geq 2$) be a later pattern, where $\sigma_i \in \mathcal{P}$ for each $i = 1, \ldots, l!$. With respect to nonoverlapping counting, we have* (*i*) *the exact distribution of* $W(\Lambda_L)$ *is given by*

$$P(W(\Lambda_L)=n)=\sum_{\sigma_i\in\mathcal{P}}P(W(\sigma_i)=n)=\sum_{i=1}^{l!}\boldsymbol{\xi}\boldsymbol{N}_{i(N)}^{n-1}\boldsymbol{C}_{i(N)}(\alpha_1), \tag{4.24}$$

where $\boldsymbol{\xi}=(1,0,\ldots,0)$ *is the initial distribution with* $P(Y_0=(0,0))\equiv 1$, *and* $\boldsymbol{C}_{i(N)}(\alpha_1)$, $i=1,\ldots,l!$, *are the first column vectors of the matrices* $\boldsymbol{C}_{i(N)}$, $i=1,\ldots,l!$, *respectively;*

(*ii*) *the probability generating function of* $W(\Lambda_L)$ *is given by*

$$\varphi^{(N)}_{W(\Lambda_L)}(s)=\sum_{i=1}^{l!}\psi^{(N)}_{W(\sigma_i)}(s). \tag{4.25}$$

Note that the generating functions $\psi^{(N)}_{W(\sigma_i)}(s)$, $i=1,\ldots,l!$, can be obtained through Theorem 4.1(ii). In the case of overlapping counting, the above results still hold except that the transition probability matrices are replaced by $\boldsymbol{M}_{i(O)}$, $i=1,\ldots,l!$, respectively.

In addition to the above imbedding procedure, we can imbed the random variable $W(\Lambda_L)$ into a Markov chain directly. Let $\mathcal{P}_j$, $j=1,\ldots,l-1$, be the sets of all permutations of the elements of $\mathcal{I}=\{1,2,\ldots,l\}$ taken j at a time; that is,

$$\mathcal{P}_j=\{i_1i_2\cdots i_j : i_k\in\mathcal{I} \text{ for } 1\le k\le j \text{ and } i_k\ne i_m \text{ for } k\ne m\},$$

and $card(\mathcal{P}_j)=C^l_j\times j!=\frac{l!}{(l-j)!}$. For example, let $\mathcal{I}=\{1,2,3\}$; then $\mathcal{P}_2=\{12,21,13,31,23,32\}$. We define a Markov chain $\{Y_t : t=0,1,\ldots\}$ on the state space Ω having the form

$$\Omega=\Omega_0\cup\bigcup_{j=1}^{l-1}\Omega_j\cup\bigcup_{j=1}^{l!}\{\alpha_j\}, \tag{4.26}$$

where Ω_0 and Ω_j, $j=1,\ldots,l-1$, are defined as

$$\begin{aligned}\Omega_0 &= \left\{(0,v) : v\in\bigcup_{j=1}^{l}\Upsilon(\Lambda_j)\right\},\\ \Omega_j &= \left\{(u,v) : u=i_1i_2\cdots i_j\in\mathcal{P}_j \text{ and } v\in\bigcup_{k\ne i_1,\ldots,i_j}\Upsilon(\Lambda_k)\right\},\end{aligned}$$

and $\alpha_1, \ldots, \alpha_{l!}$ are absorbing states corresponding to the ordered series patterns $\sigma_1, \ldots, \sigma_{l!}$, respectively. The above construction procedure is the same for both nonoverlapping and overlapping cases except that their transition probabilities are slightly different according to the counting procedures and structures of $\Lambda_1, \ldots, \Lambda_l$. Applying Equations (4.1)$-$(4.3), the exact distribution, mean and probability generating function of $W(\Lambda_L)$ can be obtained.

The next theorem, an extension of Theorem 4.4, provides a third way to obtain the probability generating function of $W(\Lambda_L)$.

Theorem 4.6 *Let $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$ $(l \geq 2)$ be a later pattern, where $\sigma_i \in \mathcal{P}$ for each $i = 1, \ldots, l!$. With respect to nonoverlapping counting, we have*

(i) for any ordered series pattern $\sigma_i = \Lambda_{i_1} \circ \cdots \circ \Lambda_{i_l} \in \mathcal{P}$, the generating function of $W(\sigma_i)$ is given by

$$\psi^{(N)}_{W(\sigma_i)}(s) = \prod_{j=1}^{l} \psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{i_{(j-1)}})\right), \ i = 1, \ldots, l!, \tag{4.27}$$

where $\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{i_{(j-1)}}) = (1, 0, \ldots, 0)$, $j = 1, \ldots, l$, are the initial distributions with initial state $\emptyset_{j^\star} = j^\star$ $(P(Y_0 = \emptyset_{j^\star}) \equiv 1)$, $j^\star$ is the last element of the pattern $\Lambda_{i_{(j-1)}}$, and $\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{i_0}) = \boldsymbol{\xi} = (1, 0, \ldots, 0)$ (with usual initial state $\emptyset_{j^\star} = \emptyset$) by convention;

(ii) the probability generating function of $W(\Lambda_L)$ is given by

$$\varphi^{(N)}_{W(\Lambda_L)}(s) = \sum_{\sigma_i \in \mathcal{P}} \psi^{(N)}_{W(\sigma_i)}(s) = \sum_{i=1}^{l!} \prod_{j=1}^{l} \psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^\star}}(\Lambda_{i_{(j-1)}})\right). \tag{4.28}$$

Proof. The proof of part (i) is similar to the proof of Theorem 4.2. The proof of part (ii) follows from the definition of $\varphi^{(N)}_{W(\Lambda_L)}(s)$ and part (i). $\square$

Equations (4.27) and (4.28) still hold when $\{X_i\}$ is a sequence of i.i.d. multistate trials, except that the initial distributions $\boldsymbol{\xi}^{(N)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}})$, $j = 1, \ldots, l$, are replaced by $\boldsymbol{\xi}$.

Corollary 4.3 *Let $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$ $(l \geq 2)$ be a later pattern, where $\sigma_i \in \mathcal{P}$ for each $i = 1, \ldots, l!$. With respect to overlapping counting, we have*

(i) for any ordered series pattern $\sigma_i = \Lambda_{i_1} \circ \cdots \circ \Lambda_{i_l} \in \mathcal{P}$, the generating function of $W(\sigma_i)$ is given by

$$\psi^{(O)}_{W(\sigma_i)}(s) = \prod_{j=1}^{l} \psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}})\right), \quad i = 1, \ldots, l!, \tag{4.29}$$

where $\boldsymbol{\xi}^{(O)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}}) = (0, \ldots, 1, \ldots, 0) = \boldsymbol{e}_{j^\circ}$, $j = 1, \ldots, l$, are the initial distributions such that the state j° is the longest overlap of the patterns $\Lambda_{i_{(j-1)}}$ with Λ_{i_k}, $k = j, \ldots, l$, in the sense of overlapping counting, and $\boldsymbol{\xi}^{(O)}_{\emptyset_{j^*}}(\Lambda_{i_0}) = \boldsymbol{\xi} = (1, 0, \ldots, 0)$ by convention (the state j° corresponding to the usual initial state $\emptyset$);

(ii) the probability generating function of $W(\Lambda_L)$ is given by

$$\varphi^{(O)}_{W(\Lambda_L)}(s) = \sum_{\sigma_i \in \mathcal{P}} \psi^{(O)}_{W(\sigma_i)}(s) = \sum_{i=1}^{l!} \prod_{j=1}^{l} \psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}})\right). \tag{4.30}$$

Clearly, it is impossible to compute the probability and generating function of $W(\sigma)$ and the exact distribution and probability generating function of $W(\Lambda_L)$ by hand. Computer programs based on the Mathematical software **MAPLE** have been developed to make our work fully automated. We discuss the computational issues in the next chapter.

Example 4.5 Let $\{X_i\}_{i=1}^n$ be a sequence of four-state trials with possible outcomes 1, 2, 3 and 4 (or A, C, G and T), and let $\Lambda_1 = 1414$, $\Lambda_2 = 4141$ and $\Lambda_3 = 2323$. We consider the following two cases:

1. Set $p_1 = p_2 = p_3 = p_4 = 0.25$ for $\{X_i\}$ being a sequence of i.i.d. trials.

2. Set the initial probabilities to be $p_1 = p_2 = p_3 = p_4 = 0.25$ and the transition probability matrix of $\{X_i\}$ as

$$\boldsymbol{A} = \begin{bmatrix} 0.25 & 0.35 & 0.20 & 0.20 \\ 0.15 & 0.25 & 0.25 & 0.35 \\ 0.35 & 0.30 & 0.25 & 0.10 \\ 0.25 & 0.30 & 0.20 & 0.25 \end{bmatrix}$$

   for $\{X_i\}$ being a sequence of first-order homogeneous Markov dependent trials.

The state spaces Ω and the transition probability matrices of the imbedded Markov chains for both cases are generated by our computer programs automatically. Figures 4.1 and 4.2 show the probability distributions of $W(\Lambda_L)$ for both cases with respect to nonoverlapping and overlapping counting.

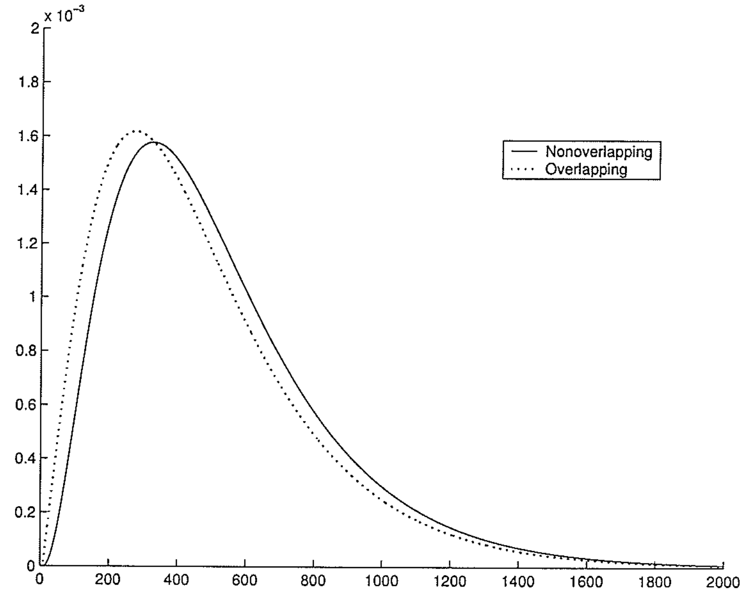

Figure 4.1: Probability distributions of the later waiting time $W(\Lambda_L)$ for Case 1, Example 4.5 with respect to nonoverlapping and overlapping counting.

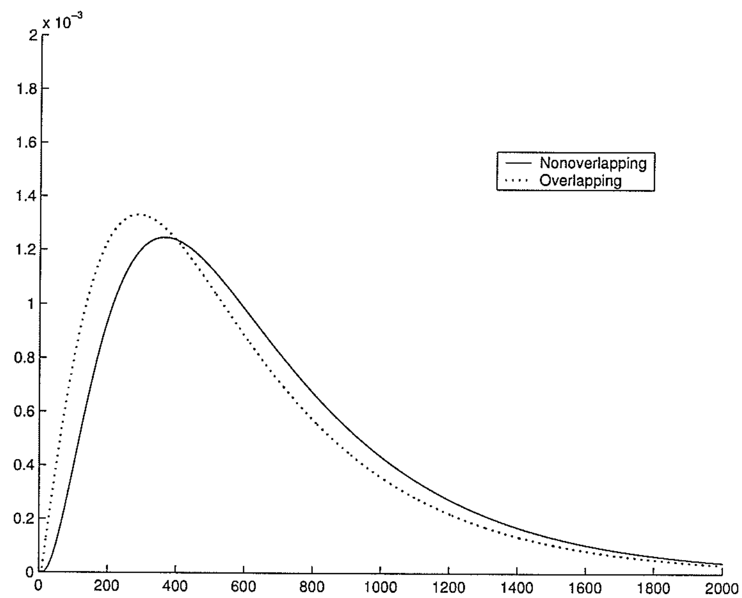

Figure 4.2: Probability distributions of the later waiting time $W(\Lambda_L)$ for Case 2, Example 4.5 with respect to nonoverlapping and overlapping counting.

Chapter 5

Algorithms for Waiting Time Distributions

In this chapter, we develop computational algorithms for use in a computer algebra system to implement the results in Chapters 3 and 4. For convenience, relabel the possible outcomes $b_1, b_2, \ldots, b_m$ as $1, 2, \ldots, m$ with probabilities $p_1, p_2, \cdots, p_m$, respectively. Given a compound pattern $\Lambda = \bigcup_{i=1}^{l} \Lambda_i$, we first discuss the algorithms for computing the exact distribution, mean and probability generating function of the waiting time $W(\Lambda)$.

Algorithm I: $\{X_i\}$ is assumed to be a sequence of i.i.d. multistate trials.

Step 1. Decompose each simple pattern forward and create the state space Ω having the form given by (3.4) for the imbedded Markov chain $\{Y_t\}$. Let

$$\Omega_1 = \Omega \setminus \{\text{absorbing states}\} = \Omega \setminus \{\alpha_1, \ldots, \alpha_l\}$$

and $k = card(\Omega_1)$. We define a one-to-one index map f as

$$f(\omega) = \begin{cases} i & \forall\, \omega \in \Omega_1 \text{ and } i = 1, 2, \ldots, k, \\ j & \forall\, \omega \in \{\alpha_1, \ldots, \alpha_l\} \text{ and } j = k+1, \ldots, k+l. \end{cases}$$

Step 2. Generate the transition probability matrix $\boldsymbol{M}$ having the form given by (3.6). Initially, we set $\boldsymbol{M}$ to be a $(k+l) \times (k+l)$ $\mathbf{0}$ matrix. For each $\omega \in \{\alpha_1, \cdots, \alpha_l\}$, replace $p_{f(w),f(w)}$ by 1. For each $\omega \in \Omega_1$ and $j = 1, \ldots, m$, let ω_1 be a pattern which is composed of states ω and j such that j follows ω.

Find the longest subpattern (counting backward) of ω_1 that belongs to Ω and denote it by ω_2. Then we replace $p_{f(\omega),f(\omega_2)}$ by p_j.

Step 3. Extract the submatrix $\boldsymbol{N}$ and $\boldsymbol{C}$ from the transition probability matrix $\boldsymbol{M}$ obtained in Step 2. From Equation (3.8) and Theorems 3.2–3.4, we can easily set up the simultaneous recursive equations and compute the distribution, mean and probability generating function.

It is easy to implement Step 1 in the computer program. For Step 2, we provide an example to illustrate the idea of setting up the transition probability matrix.

Example 5.1 Let $\{X_i\}$ be a sequence of i.i.d. three-state trials with possible outcomes 1, 2 and 3 ($m = 3$) and let $\Lambda = \Lambda_1 \cup \Lambda_2$ be a compound pattern with $\Lambda_1 = 112$ and $\Lambda_2 = 1332$. Then we have

$$\begin{aligned} \Omega &= \{\emptyset, 1, 2, 3, 11, 13, 133, \alpha_1, \alpha_2\}, \\ \Omega_1 &= \{\emptyset, 1, 2, 3, 11, 13, 133\}, \end{aligned}$$

where α_1 and α_2 correspond to $\Lambda_1 = 112$ and $\Lambda_2 = 1332$, respectively. We define a one-to-one index map f as described in Step 1: $f(\emptyset) = 1$, $f(1) = 2$, $f(2) = 3$, $f(3) = 4$, $f(11) = 5$, $f(13) = 6$, $f(133) = 7$, $f(\alpha_1) = 8$, $f(\alpha_2) = 9$. Suppose the imbedded Markov chain is in state $\omega = 11$ at trial $t-1$; then ω_1 is composed of ω and $j = 1, 2, 3$ (possible outcomes at trial t) such that j follows ω; that is,

$$\omega_1 = \begin{cases} 111 & \text{if } j = 1, \\ 112 & \text{if } j = 2, \\ 113 & \text{if } j = 3. \end{cases}$$

For $\omega_1 = 111$, the longest subpattern is $\omega_2 = 11$ ($1\underline{\underline{11}}$: counting backward $\longleftarrow$). Note that '111' is not the longest subpattern since it does not belong to Ω. Hence, the transition probability $P(Y_t = 11 | Y_{t-1} = 11) = p_{f(\omega),f(\omega_2)} = p_{f(11),f(11)} = p_{5,5} = p_1$. Similarly, the transition probabilities $P(Y_t = \alpha_1 | Y_{t-1} = 11) = p_2$ and $P(Y_t =$

$13|Y_{t-1} = 11) = p_3$. This illustrates the idea of setting up the transition probability matrix in our computer program.

As stated in Remark 3.2, when $\{X_i\}$ is a sequence of i.i.d. multistate trials, the state space Ω can be reduced to a smaller one. However, when the state space of the imbedded chain is changed, the corresponding transition matrix is changed, too. Therefore, the way of assigning such a matrix in the algorithm is different from Algorithm I. With a simple modification in Step 1 and Step 2, we have the following algorithm.

Algorithm II: $\{X_i\}$ is assumed to be a sequence of i.i.d. multistate trials and the imbedded Markov chain $\{Y_t\}$ has a smaller state space Ω.

Step 1. Decompose each simple pattern forward and create the state space Ω having the form given by (3.9) for the imbedded Markov chain $\{Y_t\}$ as described in Remark 3.2. Let

$$\begin{aligned} \Omega_1 &= \Omega \setminus \{\text{absorbing states}\} = \Omega \setminus \{\alpha_1, \ldots, \alpha_l\}, \\ \Gamma_1 &= \{j : j \text{ is the first element of the simple pattern } \Lambda_i,\ i = 1, \ldots, l\}, \\ \Gamma_2 &= \Gamma \setminus \Gamma_1, \end{aligned}$$

and $k = card(\Omega_1)$. We define a one-to-one index map f as

$$f(\omega) = \begin{cases} i & \forall\, \omega \in \Omega_1 \text{ and } i = 1, 2, \ldots, k, \\ j & \forall\, \omega \in \{\alpha_1, \ldots, \alpha_l\} \text{ and } j = k+1, \ldots, k+l. \end{cases}$$

Step 2 Generate the transition probability matrix $\boldsymbol{M}$ having the form given by (3.6). Initially, we set $\boldsymbol{M}$ to be a $(k+l) \times (k+l)$ $\mathbf{0}$ matrix. Then we replace the transition probabilities as follows:

(1) For each $\omega \in \{\alpha_1, \cdots, \alpha_l\}$, replace $p_{f(w),f(w)}$ by 1.

(2) For each $\omega \in \Omega_1$ and $j \in \Gamma_1$, let ω_1 be a pattern which is composed of states ω and j such that j follows ω. Find the longest subpattern (counting backward) of ω_1 that belongs to Ω and denote it by ω_2. Then replace $p_{f(\omega),f(\omega_2)}$ by p_j.

(3) For each $\omega \in \Omega_1$ and $j \in \Gamma_2$, let ω_3 be a pattern which is composed of states ω and j such that j follows ω. Find the longest subpattern (counting backward) of ω_3 and denote it by ω_4. If $\omega_4 \in \Omega$ then replace $p_{f(\omega),f(\omega_4)}$ by p_j. Otherwise, replace $p_{f(\omega),f(\beta)}$ by $\sum_{j \in \Gamma_3} p_j$, where $\Gamma_3 = \{j : j \in \Gamma_2 \text{ and } w_4 \notin \Omega\}$.

Step 3. Same as Step 3 in Algorithm I.

The following example shows the difference between Algorithm I and Algorithm II.

Example 5.2 Let $\{X_i\}$ be a sequence of i.i.d. five-state trials with possible outcomes 1, 2, 3, 4 and 5, and let $\Lambda = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$ be a compound pattern with $\Lambda_1 = 112$, $\Lambda_2 = 13$, and $\Lambda_3 = 15$. Using our computer program based on Algorithm I, we obtain the state space $\Omega = \{\emptyset, 1, 2, 3, 4, 5, 11, \alpha_1, \alpha_2, \alpha_3\}$ for the imbedded Markov chain $\{Y_t\}$ and transition probability matrix

$$\boldsymbol{M} = \begin{array}{c} \emptyset \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 11 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{ccccccc|ccc} 0 & p_1 & p_2 & p_3 & p_4 & p_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_2 & 0 & p_4 & 0 & p_1 & 0 & p_3 & p_5 \\ 0 & p_1 & p_2 & p_3 & p_4 & p_5 & 0 & 0 & 0 & 0 \\ 0 & p_1 & p_2 & p_3 & p_4 & p_5 & 0 & 0 & 0 & 0 \\ 0 & p_1 & p_2 & p_3 & p_4 & p_5 & 0 & 0 & 0 & 0 \\ 0 & p_1 & p_2 & p_3 & p_4 & p_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_4 & 0 & p_1 & p_2 & p_3 & p_5 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

with initial distribution $\boldsymbol{\xi}_0 = (1,0,0,0,0,0,0,0,0,0)$ for Y_0 ($\boldsymbol{\xi} = (1,0,0,0,0,0,0)$). Similarly, from our computer program based on Algorithm II, we obtain the state space $\Omega = \{\emptyset, 1, 6, 11, \alpha_1, \alpha_2, \alpha_3\}$ for the imbedded Markov chain $\{Y_t\}$ (the state '6' represents the state 'β': no subpattern belongs to $\bigcup_{i=1}^{3} S(\Lambda_i) = \{1, 11, \alpha_1, \alpha_2, \alpha_3\}$) and transition probability matrix

$$\boldsymbol{M} = \begin{array}{c} \emptyset \\ 1 \\ 6 \\ 11 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \left[\begin{array}{cccc|ccc} 0 & p_1 & p_2+\cdots+p_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_2+p_4 & p_1 & 0 & p_3 & p_5 \\ 0 & p_1 & p_2+\cdots+p_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_4 & p_1 & p_2 & p_3 & p_5 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

with initial distribution $\boldsymbol{\xi}_0 = (1, 0, 0, 0, 0, 0, 0)$ for Y_0 ($\boldsymbol{\xi}$ = (1,0,0,0)). Both computer programs yield

$$E\left[W(\Lambda)\right] = \frac{1}{p_1p_3 + p_1p_5 + p_1^2p_2}$$

and

$$\varphi_W(s) = \frac{p_1s^2(p_3 + p_5 + p_1p_2s)}{1 - s + (p_1p_3 + p_1p_5)s^2 + p_1^2p_2s^3}.$$

Next, we discuss the algorithm for the case when $\{X_i\}$ is a sequence of first-order homogeneous Markov dependent trials. We assume that $\{X_i\}$ has initial probabilities $p_1, p_2, \ldots, p_m$ and transition probability matrix

$$\boldsymbol{A} = \begin{array}{c} 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ m \end{array} \left[\begin{array}{ccccc} p_{11} & p_{12} & \cdot\ \cdot\ \cdot & p_{1m} \\ p_{21} & p_{22} & \cdot\ \cdot\ \cdot & p_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{m1} & p_{m2} & \cdot\ \cdot\ \cdot & p_{mm} \end{array}\right].$$

Algorithm III: $\{X_i\}$ is assumed to be a sequence of first-order homogeneous Markov dependent trials.

Step 1. Decompose each simple pattern forward and create the state space Ω having the form given by (3.4) for the imbedded Markov chain $\{Y_t\}$. Let

$$\begin{aligned} \Omega_1 &= \Omega \setminus \{\text{absorbing states}\} = \Omega \setminus \{\alpha_1, \ldots, \alpha_l\}, \\ \Omega_2 &= \Omega_1 \setminus \{\emptyset\}, \end{aligned}$$

and $k = card(\Omega_1)$. We define a one-to-one index map f as

$$f(\omega) = \begin{cases} i & \forall\, \omega \in \Omega_1 \text{ and } i = 1, 2, \ldots, k, \\ j & \forall\, \omega \in \{\alpha_1, \ldots, \alpha_l\} \text{ and } j = k+1, \ldots, k+l. \end{cases}$$

Step 2. Generate the transition probability matrix $\boldsymbol{M}$ having the form given by (3.6). Initially, we set $\boldsymbol{M}$ to be a $(k+l) \times (k+l)$ $\mathbf{0}$ matrix. Then we replace the transition probabilities as follows:

(1) For each $\omega \in \{\alpha_1, \cdots, \alpha_l\}$, replace $p_{f(\omega),f(\omega)}$ by 1.

(2) For each $j = 1, \ldots, m$, replace $p_{f(\emptyset),f(j)}$ by p_j.

(3) For each $\omega \in \Omega_2$ and $j = 1, \ldots, m$, let i denote the last element of ω and let ω_1 be a pattern which is composed of states ω and j such that j follows ω. Find the longest subpattern (counting backward) of ω_1 that belongs to Ω and denote it by ω_2. Then replace $p_{f(\omega),f(\omega_2)}$ by p_{ij}.

Step 3. Same as Step 3 in Algorithm I.

Given a later pattern $\Lambda_L = \bigcup_{i=1}^{l!} \sigma_i$, we have introduced three different ways in Chapter 4 to obtain the exact distribution and probability generating function of the waiting time $W(\Lambda_L)$: (i) imbedding the random variables $W(\sigma_i)$, $i = 1, \ldots, l!$, separately, (ii) imbedding the random variable $W(\Lambda_L)$ directly, and (iii) using Equation (4.28) (or Equation (4.30) for overlapping counting) which is expressed in terms of the generating functions $\psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j*}}(\Lambda_{i_{(j-1)}})\right)$ for $i = 1, \ldots, l!$ and $j = 1, \ldots, l$. When l is large, the transition probability matrices associated with the imbedded Markov chains of $W(\sigma_i)$, $i = 1, \ldots, l!$, and $W(\Lambda_L)$ are quite large. Hence, it may not be efficient to obtain the analytic form for the probability generating function of $W(\Lambda_L)$ through (i) or (ii). Therefore, we do not provide further discussion for these algorithms. However, computer programs based on (i) and (ii) are still useful in numerical computation. To compute the generating function

$\psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(N)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}})\right)$ or $\psi_{W(\Lambda_{i_j}|\Lambda_{i_j},\ldots,\Lambda_{i_l})}\left(s|\boldsymbol{\xi}^{(O)}_{\emptyset_{j^*}}(\Lambda_{i_{(j-1)}})\right)$ for fixed i and j, we can use our computer programs based on Algorithms I and III for the compound pattern $\bigcup_{k=j}^{l}\Lambda_{i_k}$ with minor modifications in setting up the initial distribution and transition probability matrix. The transition probability matrix associated with the imbedded Markov chain of the waiting time $W(\bigcup_{k=j}^{l}\Lambda_{i_k})$ is much smaller. This provides an efficient way to obtain the probability generating function of $W(\Lambda_L)$.

We see that all the algorithms discussed in this chapter are similar. The key is how to correctly set up the transition probability matrix. In summary, the forward and backward principle lays the foundation for constructing the state space and setting up the transition probability matrix. Our algorithms are very easy to implement. With today's computers, it takes no more than one hour to compute the exact distributions, means and probability generating functions for waiting times of reasonably large compound and later patterns.

Chapter 6

Applications of Waiting Time Distributions in Quality Control

6.1 Introduction

The multitude of control charts for monitoring various process parameters (such as the mean, variance, and proportion) exists due to the multiple types of shifts that can occur in that parameter over time. No single chart is optimal for detecting all types of shifts. Sometimes several charts are used simultaneously, while in other cases new combined charts are used (e.g., robust Cusum charts and Shewhart charts with runs rules).

Various control charts have been investigated in terms of the run length distributions based on a given pattern (usually a step function) by various methods. Comparisons among several of these charts have been done using different methods: by simulation (Roberts, 1959), through numerical analysis (e.g., solving integrals numerically, such as Robinson and Ho, 1978; Luceno and Puig-Pey, 2000; and Rao, Disney and Pignatiello Jr., 2001) to theoretical approximations and exact derivations.

Among the theoretical derivations, many authors have used the Markov chain approach (Champ and Woodall, 1987; Lucas and Saccucci, 1990; Lucas and Crosier, 2000) introduced by Brook and Evans (1972). Since each author focused on one or more charts, the different Markov chain applications were tailored to each case. For instance, the state space of the Markov chain has been formulated differently by different authors (Lucas and Crosier, 1982; Champ and Woodall, 1987), and in several cases it was not specified at all.

We introduce a general unified framework that is based on discretization and the use of a finite Markov chain imbedding technique for run statistics (Fu, 1996). The method can be applied to any control scheme that is based either on a simple boundary crossing rule, or on a compound rule based on run or scan statistics that include several criteria. Some known results can be viewed as special cases of our general method.

Our unified approach sheds light on the relation between different types of monitoring schemes, and their performance in the presence of different data structures. It also enables a straightforward performance comparison of various schemes, thus being important from an applied point of view as well.

The rest of this chapter is laid out as follows: Section 6.2 describes the general framework, which is based on a Markov chain imbedding formulation. We show how some well known charts can be formulated in this framework. In Section 6.3, we provide a detailed numerical example of a compound rule based on a run statistic and Cusum to demonstrate how to imbed it as a Markov chain. In Section 6.4, we discuss the implications and possible extensions of this general approach.

6.2 The Markov Chain Approach

Brook and Evans (1972) introduced a Markov chain representation for computing the run-length distribution of a Cusum chart. Their basic idea for a discrete monitoring statistic (e.g., a count) is to treat the m values that the monitoring statistic can obtain within the control limits as states of a Markov chain, and all the values that exceed the limits as an absorbing state. If the monitoring statistic is continuous, the same method is used, after discretizing the area of the control chart into m regions within the control limits and one region that exceeds the limits (the absorbing state).

We formalize and generalize this method as follows: for a control scheme that involves a compound decision rule, the run length is imbedded into a finite Markov chain $\{Y_t\}$, according to the rules applied (denoted by $\phi_1, \ldots \phi_l$). $\{Y_t\}$ then has a state space Ω and transition probability matrix $\boldsymbol{M}_t$ having the form given by (3.10). The run length probability distribution is then given by

$$P(RL = n) = \boldsymbol{\xi} \left(\prod_{t=1}^{n-1} \boldsymbol{N}_t \right) (\boldsymbol{I} - \boldsymbol{N}_n) \mathbf{1}_k', \; n = 1, 2, \ldots, \tag{6.1}$$

where $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ and $\mathbf{1}_k = (1, 1, \ldots, 1)$.

We denote a simple rule by ϕ and the monitoring statistic by

$$W_t(k) = [W_{t-k+1}, \ldots, W_t],$$

where k is the length of the "history" that is retained in order to reach a decision at time t. A compound rule dictates that the chart signals an alarm at time t if any one rule $\phi_i(W_t(k))$ based on the monitoring statistic $W_t(k) = [W_{t-k+1}, \ldots, W_t]$ exceeds some limit at time t (or falls within a certain area on the chart).

For this general pattern, we define the imbedded Markov chain $\{Y_t\}$ as

$$Y_t(W_t(k); \phi_1, \ldots, \phi_l) = H(\phi_i(W_t(k)); i = 1, \ldots, l), \tag{6.2}$$

where H is a function that combines the information from the l different rules. In general, the state space Ω of the imbedded Markov chain $\{Y_t\}$ is induced by ϕ_i and the vector $W_t(k)$.

6.2.1 Discretizing W_t

Although all the different uses of the Markov chain approach rely on discretizing a continuous variable, we make an important distinction between two types of discretization: natural vs. artificial. Discretizing a random variable can either arise naturally from the monitoring scheme, or else it is artificially imposed. Examples where discretization arises naturally are charts with discrete monitoring statistics (such as counts) and charts with a continuous monitoring statistic which effectively divide the values of the statistic into two or more regions (e.g., Shewhart charts with or without runs rules).

Examples where discretization is carried out artificially are Cusum and EWMA charts (Brook and Evans, 1972; Fu, Spiring and Xie, 2002). The reason for imposing an artificial discretization is to simplify the calculation of a complicated probability. For example, in Cusum and EWMA schemes it is hard to calculate the probability that the monitoring statistic exceeds the boundary. The alternative, which is based on discretizing the continuous measurements, is described in Section 6.2.2.

For W_t, the monitoring statistic at time t, we use $R(W_t)$ to denote a natural discretization of W_t and $D(W_t)$ an artificial discretization of W_t. A rule $\phi(W_t(k))$ determines which type of discretization is used, and the values of the last k monitoring statistics that should be retained.

6.2.2 Imbedding Well-Known Monitoring Schemes

The Markov chain approach introduced by Brook and Evans (1972) was used by several authors to derive the average run length (ARL) or the entire run length distribution for various control charts. We show how the different results can be formalized as described above, and how the run length is imbedded into a finite Markov chain. We describe rules that lead to natural discretization, rules that require artificial discretization, and rules that involve both types of discretization.

Class 1: Rules that lead to natural discretization

The class of monitoring schemes where the decision rule leads to a natural discretization of W_t includes schemes that are based on a discrete statistic and schemes where the control chart is divided into discrete regions.

The "history" that is retained in this case has the form $[R(W_{t-k+1}), \ldots, R(W_t)]$; that is, the values of the last k monitoring statistics must be tracked in order to reach a decision. Examples of charts with naturally discretizing rules and different values of k are:

- A Shewhart chart, where we only retain information on the statistic X_t at time t, i.e. $R(X_t)$.

- A discrete Cusum chart such as a Poisson Cusum, where the monitoring statistic is a function of the observation at time t (X_t) and of the cumulative sum of the observations until time $t-1$ (S_{t-1}). In such cases we retain information on times $t-1$ and t, i.e. $[R(S_{t-1}), R(X_t)]$.

- A Shewhart chart with Western Electric rules, where the decision rule is based on information of the locations of the k previous values of the statistic. The following set of Western Electric rules (see Montgomery, 2001) are widely applied: signal an alarm if

  1. one or more points exceed the 3-sigma control limits;
  2. two of three consecutive points fall beyond the 2-sigma limits;
  3. four of five consecutive points fall beyond the 1-sigma limits;
  4. eight consecutive points fall on one side of the center line.

  If all four rules are combined, then we must retain the locations of the last 8 points, i.e. $[R(X_{t-7}), R(X_{t-6}), \ldots, R(X_t)]$.

According to the rules ϕ_i, $i = 1, \ldots, l$ (with or without runs rules), the statistic $W_t(k)$ is imbedded into a Markov chain $Y_t(W_t(k); \phi_1, \ldots, \phi_l)$. A simple example is a Shewhart chart where X_t is naturally discretized into an indicator taking two possible levels: α (outside the limits) and R (within the limits). The state space for $Y_t(W_t(1);\phi) = R(X_t)$ is $\Omega = \{\emptyset, R, \alpha\}$, and the corresponding transition matrix is given by

$$\boldsymbol{M} = \begin{bmatrix} 0 & p_R & 1-p_R \\ 0 & p_R & 1-p_R \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.3}$$

where $\emptyset$ is the dummy initial state for the imbedded Markov chain and $p_R = P(LCL < X_t < UCL)$, under a given value of the monitored parameter.

Class 2: Rules that require artificial discretization

When the monitoring statistic is continuous there are cases where it is too complicated to compute the run-length distribution directly. Two such examples are the Cusum and EWMA charts. An alternative is to discretize X_t artificially. This leads to a discrete Cusum/EWMA statistic W_t, and we retain the last k discrete values $[D(W_{t-k+1}), \ldots, D(W_t)]$.

For example, a one-sided Cusum chart can be discretized such that the Cusum statistic S_t obtains $m+2$ values (or $2(m+1)+1$ values for a two-sided chart): the in-control area $[0, h)$ is divided into m equally-sized regions and the out-of-control area $[h, \infty)$ is the $m+1$ region (as described in Brook and Evans, 1972; Fu, Spiring and Xie, 2002). For both the Cusum and EWMA schemes we retain information on the accumulating statistic (denoted by S) at time $t-1$ and on the accumulated statistic (denoted by X) at time t, i.e. $k = 2$ and we retain 2 last discrete values $[D(S_{t-1}), D(X_t)]$. This information is imbedded into a Markov chain $Y_t(W_t(2);\phi) = H(\phi(W_t(2))) = D(S_t)$.

The transition matrix includes $m+3$ states. By selecting m to be large enough, the run-length of the discretized statistic will approach that of the continuous one.

Class 3: Rules that involve natural and artificial discretization

This class is the most general. It includes decision rules that require monitoring a naturally discrete statistic and a continuous one. Two examples are the "robust Cusum chart" (Lucas and Crosier, 1982) and the "robust EWMA chart" (Lucas and Saccucci, 1990) which are combinations of a Shewhart chart (with 4 or 5-sigma limits) and a Cusum or EWMA chart. Two statistics are tracked simultaneously: the Cusum/EWMA continuous statistic (S_t) and the naturally-discretized Shewhart statistic (X_t).

The decision rule is compound: signal an alarm if the continuous (Cusum or EWMA) statistic exceeds the (Cusum/EWMA) limits, or if two consecutive Shewhart statistics exceed the Shewhart chart limits. The information that is needed in order to reach a decision is based on the last Cusum/EWMA value and the last two Shewhart statistics: $W_t(2) = [S_{t-1}, X_t]$. Here the Cusum/EWMA statistic is artificially discretized, and the Shewhart statistic is naturally binary:

$$\phi_1(W_t(2)) = D(S_t), \tag{6.4}$$

$$\phi_2(W_t(2)) = R(X_t). \tag{6.5}$$

These two rules are then combined and imbedded into a Markov chain $\{Y_t\}$.

Other hypothetical decision rules that fall into this category would be Cusum or EWMA schemes with runs rules.

6.3 An Example of a Compound Rule that Involves Natural and Artificial Discretization

To illustrate this general case, we consider a hypothetical one-sided Cusum chart, which in addition to the upper control limit has a "warning limit". The compound decision rule is to signal at time t if the Cusum statistic, given by $S_t = \max(0, S_{t-1} + X_t)$ exceeds the upper control limit at time t; or if two of three consecutive Cusum statistics fall in the interval between the warning and control limits. Without loss of generality, we assume that the X_t are i.i.d. $N(0,1)$.

As in the ordinary Cusum chart, for computational reasons the accumulated statistic X is artificially discretized with step Δ:

$$D(X) = i\Delta, \quad i = 0, \pm 1, \pm 2, \ldots, \pm(m+1). \tag{6.6}$$

We define $p_i = P(D(X) = i\Delta)$ and $F(i) = P(D(X) \le i\Delta)$ as follows:

$$\begin{aligned}
p_i &= \int_{(i-.5)\Delta}^{(i+.5)\Delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \; i = 0, \pm 1, \cdots, \pm m, \\
p_{-(m+1)} &= \int_{-\infty}^{[-(m+1)+.5]\Delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \\
p_{m+1} &= \int_{[(m+1)-.5]\Delta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \\
F(i) &= \sum_{j=-(m+1)}^{i} p_j .
\end{aligned}$$

This results in a discrete Cusum statistic S. We denote the upper control limit by $h = (m+1)\Delta$ and the warning limit by $h^* = m^*\Delta$, $m^* \le m$. Then S_t can assume the values $i\Delta$, $i = 0, 1, \ldots, m+1$ in the interval $[0, h]$.

In addition, the chart in this example is naturally divided by the second rule into three regions (as illustrated in Figure 6.1):

$$R(S_t) = \begin{cases} r_1 & \text{if } 0 \le S_t < h^*, \\ r_2 & \text{if } h^* \le S_t < h, \\ r_3 & \text{if } S_t \ge h. \end{cases} \tag{6.7}$$

The "history" that is required in this scheme in order to reach a decision is $k = 3$, with $W_t(3) = [S_{t-2}, S_{t-1}, S_t]$. We can thus write the two rules as:

$$\phi_1(W_t(3)) = S_t, \tag{6.8}$$

$$\phi_2(W_t(3)) = [R(S_{t-1}), R(S_t)]. \tag{6.9}$$

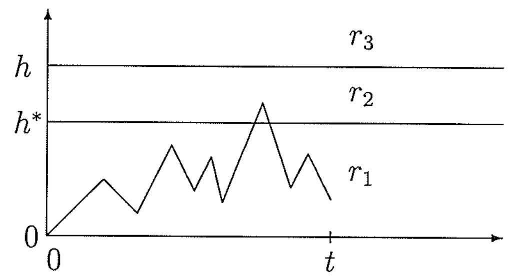

Figure 6.1: An illustration of the three-region chart from a one-sided Cusum with a warning limit.

The combined rules induce a state space of a Markov chain of the form:

$$Y_t(W_t(3);\phi_1,\phi_2) = H(\phi_1(W_t(3)),\phi_2(W_t(3))) = [R(S_{t-1}),S_t], \tag{6.10}$$

where $R(S_t) = r_1, r_2$ or r_3 and $S_t = i\Delta, i = 0, 1, \ldots, m+1$. For simplicity we write $S_t = i$ to denote $S_t = i\Delta$. States that include r_3 or $m+1$ as one of their coordinates can be combined into α, the absorbing state. Hence, the state space is:

$$\Omega = \{(\emptyset,\emptyset),(\emptyset,0),\ldots,(\emptyset,m),(r_1,0),\ldots,(r_1,m),(r_2,0),\ldots,(r_2,m^*-1),\alpha\} \tag{6.11}$$

with $2+2(m+1)+m^*$ states.

Remark 6.1 Here we assumed that X_0 is in the dummy state $\emptyset$ with probability one, that is, $P(X_0 = \emptyset) \equiv 1$. Hence the states $(\emptyset,\emptyset)$, $(\emptyset,0)$, $\cdots$, $(\emptyset,m)$ are required so that Y_0 and Y_1 of the imbedded Markov chain $\{Y_t\}$ can be properly defined with the initial distribution given by $P(Y_0 = (\emptyset,\emptyset)) \equiv 1$.

If $\{Y_t\}$ is not in the absorbing state, it will *not* move into α if: (i) $S_{t+1} < h^*$, (ii) $\{h^* \le S_{t+1} < h,\ R(S_{t-1}) = \emptyset$ and $R(S_t) = r_1\}$, and (iii) $\{h^* \le S_{t+1} < h,\ R(S_{t-1}) = R(S_t) = \emptyset$ or $r_1\}$. If one of the three conditions is met, then

$$P\{Y_{t+1} = [R(S_t),S_{t+1}] \mid Y_t = [R(S_{t-1}),S_t]\}$$
$$= \begin{cases} F(X_{t+1} = -S_t) & \text{if } S_{t+1} = 0, \\ P(X_{t+1} = S_{t+1} - S_t) & \text{if } 0 < S_{t+1} < h^*, \\ P(X_{t+1} = S_{t+1} - S_t) & \text{if condition (ii) or (iii) hold,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.12}$$

Since the rows of the transition matrix add to 1, the probability of moving into the absorbing state can be obtained by subtracting all the positive probabilities in the row from 1.

To illustrate such a transition matrix, we choose $m = 3$, $h = 4$ ($\Delta = \frac{h}{m+1} = 1$) and $h^* = 2$ ($m^* = \frac{h^*}{\Delta} = 2$). Then, $D(X_t) = i\Delta = i$, $i = 0, \pm1, \pm2, ..., \pm4$, and S_t obtains the values $i = 0, 1, 2, 3, 4$ (with 4 denoting the absorbing state α). The transition matrix $\boldsymbol{M}$ is given by

	$(\emptyset,\emptyset)$	$(\emptyset,0)$	$(\emptyset,1)$	$(\emptyset,2)$	$(\emptyset,3)$	$(r_1,0)$	$(r_1,1)$	$(r_1,2)$	$(r_1,3)$	$(r_2,0)$	$(r_2,1)$	α
$(\emptyset,\emptyset)$		$F(0)$	p_1	p_2	p_3							$1-F(3)$
$(\emptyset,0)$						$F(0)$	p_1	p_2	p_3			$1-F(3)$
$(\emptyset,1)$						$F(-1)$	p_0	p_1	p_2			$1-F(2)$
$(\emptyset,2)$										$F(-2)$	p_{-1}	$1-F(-1)$
$(\emptyset,3)$										$F(-3)$	p_{-2}	$1-F(-2)$
$(r_1,0)$						$F(0)$	p_1	p_2	p_3			$1-F(3)$
$(r_1,1)$						$F(-1)$	p_0	p_1	p_2			$1-F(2)$
$(r_1,2)$										$F(-2)$	p_{-1}	$1-F(-1)$
$(r_1,3)$										$F(-3)$	p_{-2}	$1-F(-2)$
$(r_2,0)$						$F(0)$	p_1					$1-F(1)$
$(r_2,1)$						$F(-1)$	p_0					$1-F(0)$
α												1

and it can always be written in the form

$$\boldsymbol{M} = \left[\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & 1 \end{array}\right].$$

Figure 6.2 gives the run length probability distributions for the case when $h = 3$ and $h^* = 2$ with different values of m. The mean and standard deviation of run length are computed by using the following formulas:

$$\begin{aligned} E[RL] &= \boldsymbol{\xi}(\boldsymbol{I}-\boldsymbol{N})^{-1}\mathbf{1}_k', \\ E\left[RL^2\right] &= \boldsymbol{\xi}(\boldsymbol{I}+\boldsymbol{N})(\boldsymbol{I}-\boldsymbol{N})^{-2}\mathbf{1}_K', \end{aligned}$$

from Fu, Spiring and Xie (2002), Theorem 1(iii). Numerical results are given in Table 6.1.

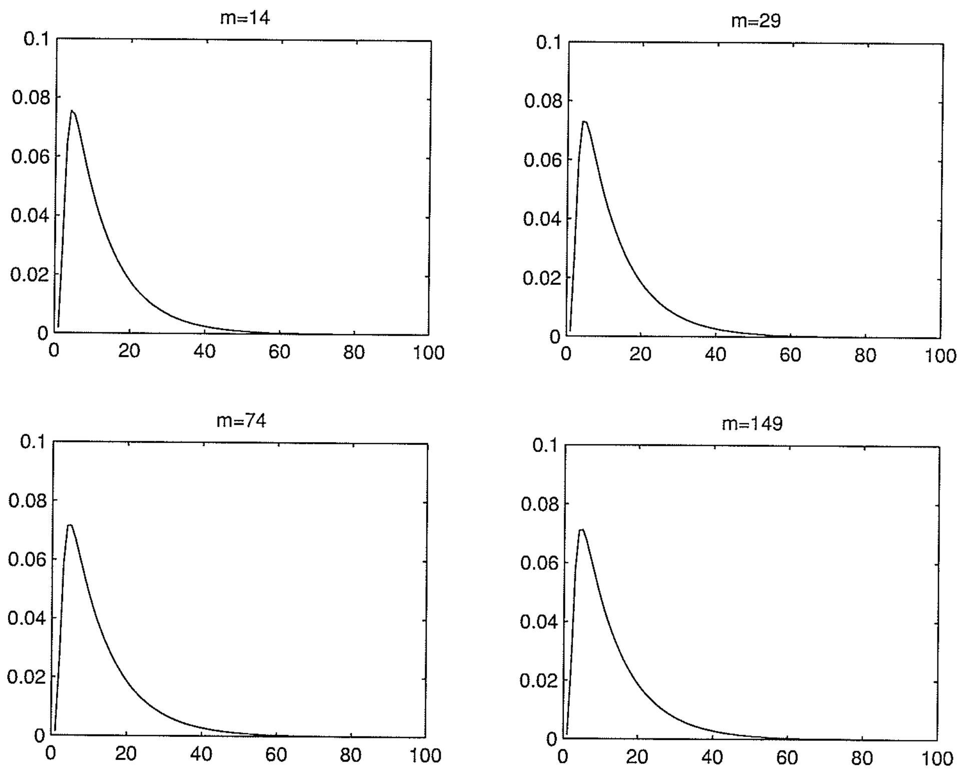

Figure 6.2: Probability distributions of the run length for a one sided Cusum with 3-sigma control limit, 2-sigma warning limit and a compound decision rule, at $m = 14, 29, 74, 149$.

Table 6.1: Quartiles, mean and standard deviation of the run length for a one-sided Cusum with 3-sigma control limit, 2-sigma warning limit and a compound decision rule at different levels of discretization (m).

m	Q_1	Q_2	Q_3	Mean	StdDev.
5	4.626	8.444	14.904	11.739	9.386
14	5.066	9.213	16.232	12.749	10.187
29	5.223	9.486	16.701	13.103	10.473
74	5.316	9.649	16.976	13.319	10.649
149	5.348	9.703	17.075	13.392	10.709
299	5.364	9.730	17.124	13.428	10.738
749	5.373	9.747	17.154	13.450	10.756
1499	5.376	9.753	17.164	13.457	10.762
1874	5.376	9.753	17.164	13.459	10.763

Remark 6.2 Traditionally, the mean and standard deviation of the run-length distribution are used for comparing the performance of control charts. However, in view of our numerical result that the run length distribution for a compound control rule is rather right skewed, we feel that displaying the quantiles of the distribution is more adequate. The quantiles can be derived directly from the distribution. In general, the distribution resulting from a compound control rule is always highly skewed to the right, especially when it involves several control charts. This is a direct consequence of the fact that the waiting time T resulting from the compound rule is the minimum waiting time among the individual rules $T_1, T_2, \ldots, T_k$.

6.4 Discussion

In many cases two or more control charts are used simultaneously, either to monitor several parameters (e.g., the mean and the standard deviation), or to be able to detect different sizes or types of shifts (e.g., a Cusum and an X-bar simultaneously). In such cases, it is important to know the signaling behavior of the joint charts.

The combined run-length is the time until the first signal is raised by any one of the charts. Mathematically, it is a waiting time problem with

$$T = \min(T_1, \ldots, T_r),$$

where T_i, $i = 1, \ldots, r$, are waiting times of r charts (or decision rules). When the monitoring charts are independent, the distribution of the combined run-length can be easily derived. In the dependent case, the combined run-length distribution of a compound rule is more complicated. Even in this situation, it still can be incorporated into the general Markov chain imbedding framework. For example, the additional Western Electric rule ϕ_4 "eight consecutive points fall on one side of the centerline" can be easily incorporated into the example of the compound rule of Section 6.3 by using an additional coordinate with states $0, 1, \ldots, 7$ and state 8 as absorbing state. We leave the details to the reader. Further, with a simple modification of the transition matrix, the results in Section 6.3 can be extended to the case when the sequence $\{X_t\}$ has a Markovian dependence structure.

The unified approach can be used to learn about a chart's ability to detect different types of signals. In order to study the performance of some monitoring scheme for a particular parameter pattern (e.g., a step-shift or linear trend from the target value), we can integrate a given pattern into the Markov chain and compute the run length distribution. In comparison to the case of a constant parameter, the Markov chain is no longer time homogeneous. This means that the transition probabilities will now depend on the value of the parameter at time t. A simple example would be to incorporate a simulated pattern of the process mean μ_t into a Shewhart scheme. The transition matrix, $\boldsymbol{M}_t$, would be the same as (6.3), except that p_R would depend on time, i.e. $p_{R,t}$.

In conclusion, the general framework described here can be used for framing a multitude of control charts into a Markov chain imbedding setting for the purpose of computing the run length distribution. It extends to combinations of charts, to dependence, and to many other monitoring schemes, both hypothetical and ones suggested in the control chart literature.

Chapter 7

Future Directions

7.1 Asymptotics of Runs and Patterns

In this section, we develop some general results which may be useful for studying the asymptotics of runs and patterns. Let Λ be a simple pattern and $\{X_i\}$ be a sequence of i.i.d. multistate trials. Suppose nonoverlapping counting is used. Recall the following results obtained in Chapter 3:

(a) $\varphi_{W(r:\Lambda)}(s) = \left[\varphi_{W(\Lambda)}(s)\right]^r$.

(b) Since $\varphi_{W(\Lambda)}(s) = 1 + \left(1 - \frac{1}{s}\right)\Phi_{W(\Lambda)}(s)$, it follows that

$$\Phi_{W(\Lambda)}(s) = \frac{s}{1-s}\left[1 - \varphi_{W(\Lambda)}(s)\right].$$

(c) From (b), we have $\Phi_{W(r:\Lambda)}(s) = \frac{s}{1-s}\left[1 - \varphi_{W(r:\Lambda)}(s)\right]$.

(d) Since $P(X_n(\Lambda) < r) = P(W(r:\Lambda) > n)$ (Feller, 1968), we have

$$\begin{aligned} P(X_n(\Lambda) = r) &= P(X_n(\Lambda) < r+1) - P(X_n(\Lambda) < r) \\ &= P(W(r+1:\Lambda) > n) - P(W(r:\Lambda) > n). \end{aligned}$$

Based on these results, we have the following theorem.

Theorem 7.1 *Given a simple pattern Λ, the double probability generating function of the number of occurrences $X_n(\Lambda)$ is given by*

$$G(s,t) = \sum_{n=0}^{\infty} \varphi_{X_n(\Lambda)}(t)s^n = \frac{1}{1-s}\left[\frac{1 - \varphi_{W(\Lambda)}(s)}{1 - \varphi_{W(\Lambda)}(s)t}\right]. \tag{7.1}$$

Proof. It follows from the definition that

$$\begin{aligned}
G(s,t) &= \sum_{n=0}^{\infty} \varphi_{X_n(\Lambda)}(t) s^n \\
&= \sum_{n=0}^{\infty} \left(\sum_{r=0}^{\infty} P(X_n(\Lambda) = r) t^r \right) s^n \\
&= \sum_{r=0}^{\infty} \sum_{n=0}^{\infty} \left[P(W(r+1:\Lambda) > n) - P(W(r:\Lambda) > n) \right] s^n t^r \\
&= \sum_{r=0}^{\infty} \frac{1}{s} \sum_{n=0}^{\infty} \left[P(W(r+1:\Lambda) \geq n+1) - P(W(r:\Lambda) \geq n+1) \right] s^{n+1} t^r \\
&= \sum_{r=0}^{\infty} \frac{1}{s} \left[\Phi_{W(r+1:\Lambda)}(s) - \Phi_{W(r:\Lambda)}(s) \right] t^r \\
&= \frac{1}{s} \sum_{r=0}^{\infty} \frac{s}{1-s} \left[\varphi_{W(r:\Lambda)}(s) - \varphi_{W(r+1:\Lambda)}(s) \right] t^r \\
&= \frac{1}{1-s} \sum_{r=0}^{\infty} \varphi_{W(r:\Lambda)}(s) t^r - \frac{1}{1-s} \sum_{r=0}^{\infty} \varphi_{W(r+1:\Lambda)}(s) t^r \\
&= \frac{1}{1-s} \left[\frac{1}{1-\varphi_{W(\Lambda)}(s)t} - \frac{\varphi_{W(\Lambda)}(s)}{1-\varphi_{W(\Lambda)}(s)t} \right] \\
&= \frac{1}{1-s} \left[\frac{1-\varphi_{W(\Lambda)}(s)}{1-\varphi_{W(\Lambda)}(s)t} \right].
\end{aligned}$$

This completes the proof. □

Corollary 7.1 *We have from Theorem 7.1 that*

$$\varphi_{X_n(\Lambda)}(t) = \frac{1}{n!} D_s^n G(s,t) \Big|_{s=0}. \tag{7.2}$$

Theorem 7.1 states that the probability generating function $\varphi_{X_n(\Lambda)}(t)$ of $X_n(\Lambda)$ can be obtained through the double probability generating function $G(s,t)$ of $X_n(\Lambda)$. We give an example to illustrate this result.

Example 7.1 Let a simple pattern be $\Lambda = S$ in a sequence of n Bernoulli trials. Then we can obtain the probability generating function of $W(\Lambda)$ as

$$\varphi_{W(\Lambda)}(s) = \frac{ps}{1-qs}.$$

From Theorem 7.1 and Corollary 7.1, we have

$$\begin{aligned} G(s,t) &= \frac{1}{1-s}\left[\frac{1-\varphi_{W(\Lambda)}(s)}{1-\varphi_{W(\Lambda)}(s)t}\right] = \frac{1}{1-qs-pst} \\ &= 1 + (q+pt)\,s + (q+pt)^2\,s^2 + (q+pt)^3\,s^3 \\ &\quad + (q+pt)^4\,s^4 + (q+pt)^5\,s^5 + \cdots. \end{aligned}$$

Similarly, if $\Lambda = SS$, then we have

$$\varphi_{W(\Lambda)}(s) = \frac{p^2s^2}{1-qs-pqs^2},$$

and hence

$$\begin{aligned} G(s,t) &= \frac{1}{1-s}\left[\frac{1-\varphi_{W(\Lambda)}(s)}{1-\varphi_{W(\Lambda)}(s)t}\right] = \frac{1-qs-p^2s^2-pqs^2}{(1-s)\,(1-qs-pqs^2-p^2s^2t)} \\ &= 1 + s + \left(1-p^2+p^2t\right)s^2 + \left(1-p^2-p^2q+p^2t+qp^2t\right)s^3 \\ &\quad + (1-p^2-p^2q-p^2q^2-p^3q+p^2t-p^4t+p^2qt+p^3qt \\ &\quad + p^2q^2t + p^4t^2)s^4 + \cdots. \end{aligned}$$

However, differentiating the double generating function $G(s,t)$ n times may be troublesome. It is necessary to look for another way to solve this problem. In view of Equation (7.1), the double generating function $G(s,t)$ always has a rational form. Stanley (1986) proposed a method for computing the coefficients of a rational function $R(x)/Q(x) = \sum_{n\geq 0} f_n x^n$, which we briefly describe. Without loss of generality, suppose that $R(x) = \beta_0 + \beta_1 x + \cdots + \beta_e x^e$ and $Q(x) = 1 + \gamma_1 x + \cdots + \gamma_d x^d$ (possibly $e \geq d$). Then, equating the coefficients of x^n in $R(x) = Q(x)\sum_{n\geq 0} f_n x^n$ yields the recursive formula

$$f_n = -\gamma_1 f_{n-1} - \cdots - \gamma_d f_{n-d} + \beta_n, \tag{7.3}$$

where $f_n = 0$ for $n < 0$ and $\beta_n = 0$ for $n > e$. Equation (7.3) provides an easy way to obtain the probability generating function of $X_n(\Lambda)$. For brevity, we denote $\varphi_{X_n(\Lambda)}(t)$ as $\varphi_n(t)$.

Theorem 7.2 *The probability generating function of $X_n(\Lambda)$ with $\Lambda = S\cdots S$ of length k in a sequence of Bernoulli trials satisfies the following set of recurrence relations:*

$$\varphi_n(t) = \begin{cases} \varphi_{n-1}(t) + p^k t\varphi_{n-k}(t) - p^k(q+pt)\varphi_{n-k-1}(t) & \textit{for } n > k, \\ 1 - p^k + p^k t & \textit{for } n = k, \\ 1 & \textit{for } n < k. \end{cases} \tag{7.4}$$

Proof. Since

$$\varphi_{W(\Lambda)}(s) = \frac{(ps)^k(1-ps)}{1-s+qp^k s^{k+1}},$$

it follows from Theorem 7.1 that

$$\begin{aligned} G(s,t) &= \frac{1}{1-s}\left[\frac{1-\varphi_{W(\Lambda)}(s)}{1-\varphi_{W(\Lambda)}(s)t}\right] \\ &= \frac{1}{1-s}\left[\frac{1-s+qp^k s^{k+1} - p^k s^k(1-ps)}{1-s-p^k t s^k + p^k(q+pt)s^{k+1}}\right] \\ &= \frac{1}{1-s}\left[\frac{(1-s)-p^k s^k(1-s)}{1-s-p^k t s^k + p^k(q+pt)s^{k+1}}\right] \\ &= \frac{1-p^k s^k}{1-s-p^k t s^k + p^k(q+pt)s^{k+1}}. \end{aligned}$$

Let $R(s) = 1 - p^k s^k$ and $Q(s) = 1 - s - p^k t s^k + p^k(q+pt)s^{k+1}$. Setting $\beta_0 = 1$, $\beta_k = -p^k$, $\gamma_1 = -1$, $\gamma_k = -p^k t$, and $\gamma_{k+1} = p^k(q+pt)$ and applying Equation (7.3), we establish Equations (7.4). □

When $\Lambda = S$, it is well-known that the random variable $X_n(\Lambda)$ (binomial random variable) in a sequence of n Bernoulli trials with success probability $p = p_n$ tends to a Poisson random variable with parameter $\lambda = \lim_{n\to\infty} np_n$; that is,

$$\lim_{n\to\infty} P(X_n(\Lambda) = x) = e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, \ldots. \tag{7.5}$$

This classical result is known in the literature as "Poisson theorem" (see Balakrishnan and Koutras, 2002, page 175). Furthermore, several asymptotic results for $X_n(\Lambda)$ with $\Lambda = S \cdots S$ of length k have been constructed that are parallel to the Poisson theorem under different assumptions on p and k. For example, assume that the success probability $0 < p < 1$ ($q = 1 - p$) is fixed and the length $k = k_n$ depends on n, so that $nqp^{k_n} \to \lambda$ as both $k_n \to \infty$ and $n \to \infty$. Feller (1968), by using the asymptotic form of the probability generating function of $X_n(\Lambda)$, showed that the limiting distribution of $X_n(\Lambda)$ is Poisson. In addition, he also established the asymptotic normality of $X_n(\Lambda)$ via renewal theoretic arguments.

Most of the aforementioned asymptotic results in the literature were focused on runs in a sequence of bistate trials. For any pattern Λ, it is challenging to study these asymptotic problems, especially when $\{X_i\}$ is a sequence of i.i.d. (or homogeneous Markov dependent) m-state ($m > 2$) trials. Equations (7.1) and (7.3) provide a direction for studying these asymptotic problems. We leave this unfinished work to the interested reader.

7.2 Other Issues

In addition to the study of asymptotics of runs and patterns, some interesting issues that were not addressed in this thesis are listed below for future research.

1. In Chapters 3 and 4, we developed a simple and general method for computing the probability generating functions of compound and later patterns. The probability generating function can also be used to obtain the variance; however, it may be a tedious task when the form of the probability generating function is very complicated. It is necessary to develop a simple and general method for computing the variance.

2. All the results in Chapters 3 and 4 can be extended to the case when $\{X_i\}$ is a sequence of higher order homogeneous Markov dependent trials. The key to solving this problem is the appropriate setup of the initial distributions, and we need two or more components for each state in the state space Ω.

3. The unified approach for computing the run length probability distributions in Chapter 6 can be extended to the case when the observations $\{X_t\}$ have $AR(1)$ or $AR(2)$ dependent structure.

4. There is not much literature regarding statistical inference of parameters in waiting time distributions. Since the finite Markov chain imbedding technique has advantages in numerical computation, a computer-based approach may be useful in treating such problems.

Appendix

A1. *Given the transition probability matrix*

$$\boldsymbol{M} = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ n-1 \\ n \end{array} \begin{bmatrix} q & p & & & & \\ & q & p & & 0 & \\ & & q & p & & \\ & & & \ddots & \ddots & \\ & 0 & & & q & p \\ & & & & & 1 \end{bmatrix},$$

we have, for $0 \le x \le n$,

$$\boldsymbol{\xi}_0 \boldsymbol{M}^n \boldsymbol{e}'_{x+1} = \binom{n}{x} p^x q^{n-x}, \tag{1}$$

where $\boldsymbol{\xi}_0 = (1, 0, \ldots, 0)_{1\times(n+1)}$ *and* $\boldsymbol{e}_{x+1} = (0, \ldots, 1, \ldots, 0)_{1\times(n+1)}$ *with 1 at the* $(x+1)$*th component.*

Proof. We prove Equation (1) by induction. For $n = 1$, we have $x = 0$ or 1. Thus, we have for $x = 0$,

$$\boldsymbol{\xi}_0 \boldsymbol{M} \boldsymbol{e}'_1 = (1, 0) \begin{bmatrix} q & p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = q = \binom{1}{0} p^0 q^1$$

and for $x = 1$,

$$\boldsymbol{\xi}_0 \boldsymbol{M} \boldsymbol{e}'_2 = (1, 0) \begin{bmatrix} q & p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = p = \binom{1}{0} p^1 q^0.$$

This implies that Equation (1) holds for $n = 1$. Now suppose Equation (1) holds for $n - 1$; that is,

$$\boldsymbol{\xi}_0 \boldsymbol{M}^{n-1} \boldsymbol{e}'_{x+1} = \binom{n-1}{x} p^x q^{(n-1)-x}, \ 0 \le x \le n-1.$$

Note that

$$\boldsymbol{M}\boldsymbol{e}_{x+1}^{'} = \begin{array}{c} 0 \\ 1 \\ \vdots \\ x \\ x+1 \\ \vdots \\ n-1 \\ n \end{array} \begin{bmatrix} q & p & & & & & & \\ & q & p & & & 0 & & \\ & & \ddots & \ddots & & & & \\ & & & q & p & & & \\ & & & & q & p & & \\ & & 0 & & & \ddots & \ddots & \\ & & & & & & q & p \\ & & & & & & & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$= p\boldsymbol{e}_{x}^{'} + q\boldsymbol{e}_{x+1}^{'}.$$

Hence, from our assumption for $n-1$ and the above equation, we have

$$\begin{aligned}
\boldsymbol{\xi}_0\boldsymbol{M}^n\boldsymbol{e}_{x+1}^{'} &= \boldsymbol{\xi}_0\boldsymbol{M}^{n-1}\left(\boldsymbol{M}\boldsymbol{e}_{x+1}^{'}\right) \\
&= \boldsymbol{\xi}_0\boldsymbol{M}^{n-1}\left(p\boldsymbol{e}_{x}^{'} + q\boldsymbol{e}_{x+1}^{'}\right) \\
&= p\boldsymbol{\xi}_0\boldsymbol{M}^{n-1}\boldsymbol{e}_{x}^{'} + q\boldsymbol{\xi}_0\boldsymbol{M}^{n-1}\mathbf{e}_{x+1}^{'} \\
&= p\binom{n-1}{x-1}p^{x-1}q^{(n-1)-(x-1)} + q\binom{n-1}{x}p^{x}q^{(n-1)-x} \\
&= \left[\binom{n-1}{x-1} + \binom{n-1}{x}\right]p^x q^{n-x} \\
&= \binom{n}{x}p^x q^{n-x},\ 0 \le x \le n.
\end{aligned}$$

This implies that Equation (1) holds for any n. This completes the proof. $\square$

A2. *Given a pattern Λ (simple or compound), the exact distribution of the waiting time random variable $W(\Lambda)$ is given by*

$$P(W(\Lambda) = n) = \boldsymbol{\xi}\boldsymbol{N}^{n-1}(\boldsymbol{I} - \boldsymbol{N})\mathbf{1}_k^{\prime},$$

where $\boldsymbol{\xi} = (1, 0, \ldots, 0)$ is the initial distribution, $\boldsymbol{I}$ is the $k \times k$ identity matrix, and $\mathbf{1}_k^{\prime}$ is the transpose of the row vector $\mathbf{1}_k = (1, 1, \ldots, 1)_{1\times k}$.

Proof. For the same reason as stated in Theorem 3.2, we note that the tail probability of $W(\Lambda)$ is given by

$$P(W(\Lambda) \geq n) = \boldsymbol{\xi} \boldsymbol{N}^{n-1} \mathbf{1}'_k.$$

Hence, we have

$$\begin{aligned} P(W(\Lambda) = n) &= P(W(\Lambda) \geq n) - P(W(\Lambda) \geq n+1) \\ &= \boldsymbol{\xi} \boldsymbol{N}^{n-1} \mathbf{1}'_k - \boldsymbol{\xi} \boldsymbol{N}^{n} \mathbf{1}'_k \\ &= \boldsymbol{\xi} \boldsymbol{N}^{n-1} (\boldsymbol{I} - \boldsymbol{N}) \mathbf{1}'_k. \end{aligned}$$

This completes the proof. $\square$

A3. *Let $\Lambda = \Lambda_1 \cup \Lambda_2$ be a compound pattern with $\Lambda_1 = S \cdots S$ of length k_1 and $\Lambda_2 = F \cdots F$ of length k_2 in a sequence of Bernoulli trials with success probability p and failure probability $q = 1 - p$, respectively. The probability generating function of the waiting time random variable $W(\Lambda)$ is given by*

$$\varphi_{W(\Lambda)}(s) = \frac{(1-ps)(ps)^{k_1}\left[1-(qs)^{k_2}\right] + (1-qs)(qs)^{k_2}\left[1-(ps)^{k_1}\right]}{(1-ps)(1-qs) - pqs^2\left[1-(ps)^{k_1-1}\right]\left[1-(qs)^{k_2-1}\right]}. \tag{2}$$

Proof. First, we note that the imbedded Markov chain $\{Y_t\}$ associated with $W(\Lambda)$ has state space

$$\begin{aligned} \Omega &= \{\emptyset, S, SS, \ldots, \overbrace{S \cdots S}^{k_1-1}, F, FF, \ldots, \overbrace{F \cdots F}^{k_2-1}, \alpha_1, \alpha_2\} \\ &= \{1, 2, 3, \ldots, k_1, k_1+1, k_1+2, \ldots, k_1+k_2-1, \alpha_1, \alpha_2\} \end{aligned}$$

and transition probability matrix

$$M = \begin{array}{r} 1 \\ 2 \\ \vdots \\ k_1-1 \\ k_1 \\ k_1+1 \\ k_1+2 \\ \vdots \\ k_1+k_2-2 \\ k_1+k_2-1 \\ \alpha_1 \\ \alpha_2 \end{array} \left[\begin{array}{ccccc|ccccc|cc} 0 & p & 0 & \cdots & 0 & q & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & p & \cdots & 0 & q & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p & q & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & q & 0 & 0 & \cdots & 0 & p & 0 \\ \hline 0 & p & 0 & \cdots & 0 & 0 & q & 0 & \cdots & 0 & 0 & 0 \\ 0 & p & 0 & \cdots & 0 & 0 & 0 & q & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & q & 0 & 0 \\ 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & q \\ \hline \multicolumn{5}{c|}{0} & \multicolumn{5}{c|}{0} & 1 & 0 \\ \multicolumn{5}{c|}{0} & \multicolumn{5}{c|}{0} & 0 & 1 \end{array}\right].$$

It follows from Theorem 3.3 that $(\phi_1(s), \phi_2(s), \ldots, \phi_{k_1+k_2-1}(s))$ is the solution of the simultaneous recursive equations

$$\begin{aligned} \phi_1(s) &= s, \\ \phi_2(s) &= ps\left[\phi_1(s) + \phi_{k_1+1}(s) + \cdots + \phi_{k_1+k_2-1}(s)\right], \\ \phi_3(s) &= ps\phi_2(s), \\ \phi_4(s) &= ps\phi_3(s), \\ &\vdots \\ \phi_{k_1}(s) &= ps\phi_{k_1-1}(s), \\ \phi_{k_1+1}(s) &= qs\left[\phi_1(s) + \cdots + \phi_{k_1}(s)\right], \\ \phi_{k_1+2}(s) &= qs\phi_{k_1+1}(s), \\ \phi_{k_1+3}(s) &= qs\phi_{k_1+2}(s), \\ &\vdots \\ \phi_{k_1+k_2-1}(s) &= qs\phi_{k_1+k_2-2}(s). \end{aligned}$$

From the above simultaneous recursive equations, note that $\phi_3(s), \ldots, \phi_{k_1}(s)$ can be expressed in terms of $\phi_2(s)$; that is, $\phi_3(s) = ps\phi_2(s), \ldots, \phi_{k_1}(s) = (ps)^{k_1-2}\phi_2(s)$.

Hence, $\phi_2(s)$ can be written as

$$\begin{aligned}
\phi_2(s) &= ps\left[\phi_1(s) + \phi_{k_1+1}(s) + \cdots + \phi_{k_1+k_2-1}(s)\right] \\
&= ps^2 + ps\left[\phi_{k_1+1}(s) + \cdots + \phi_{k_1+k_2-1}(s)\right] \\
&= ps^2 + ps\Big\{qs^2 + qs\left[\phi_2(s) + ps\phi_2(s) + \cdots + (ps)^{k_1-2}\phi_2(s)\right] \\
&\quad + q^2s^3 + (qs)^2\left[\phi_2(s) + ps\phi_2(s) + \cdots + (ps)^{k_1-2}\phi_2(s)\right] + \cdots \\
&\quad + q^{k_2-1}s^{k_2} + (qs)^{k_2-1}\left[\phi_2(s) + ps\phi_2(s) + \cdots + (ps)^{k_1-2}\phi_2(s)\right]\Big\} \\
&= ps^2 + ps\left\{\frac{qs^2\left[1-(qs)^{k_2-1}\right]}{1-qs} + \frac{qs\left[1-(qs)^{k_2-1}\right]\left[1-(ps)^{k_1-1}\right]}{(1-qs)(1-ps)}\phi_2(s)\right\}.
\end{aligned}$$

Solving the above equation, we obtain

$$\phi_2(s) = \frac{ps^2(1-ps)\left[1-(qs)^{k_2}\right]}{(1-ps)(1-qs) - pqs^2\left[1-(ps)^{k_1-1}\right]\left[1-(qs)^{k_2-1}\right]}.$$

Thus, $\phi_{k_1}(s) = (ps)^{k_1-2}\phi_2(s)$ can be obtained by substituting the solution of $\phi_2(s)$. Similarly,

$$\begin{aligned}
\phi_{k_1+k_2-1}(s) &= (qs)^{k_2-2}\phi_{k_1+1}(s) \\
&= (qs)^{k_2-2}\left\{qs^2 + qs\left[\phi_2(s) + ps\phi_2(s) + \cdots + (ps)^{k_1-2}\phi_2(s)\right]\right\} \\
&= (qs)^{k_2-1}\left\{s + \frac{\left[1-(ps)^{k_1-1}\right]}{1-ps}\phi_2(s)\right\}
\end{aligned}$$

can be obtained by substituting the solution of $\phi_2(s)$ into the above expression. By Theorem 3.4 and some simple algebra, we get

$$\begin{aligned}
\varphi_{W(\Lambda)}(s) &= p\phi_{k_1}(s) + q\phi_{k_1+k_2-1}(s) \\
&= \frac{(1-ps)(ps)^{k_1}\left[1-(qs)^{k_2}\right] + (1-qs)(qs)^{k_2}\left[1-(ps)^{k_1}\right]}{(1-ps)(1-qs) - pqs^2\left[1-(ps)^{k_1-1}\right]\left[1-(qs)^{k_2-1}\right]}.
\end{aligned}$$

This completes the proof. □

The result in A3 matches the formula derived by Feller (1968) and Ebneshahrashoob and Sobel (1990). It also shows that the finite Markov chain imbedding technique can be used in deriving a general formula. Under the same assumption as in A3, we can obtain a general formula for the case of first-order homogeneous Markov dependent bistate trials. We state it without proof.

A4. *Let $\Lambda = \Lambda_1 \cup \Lambda_2$ be a compound pattern with $\Lambda_1 = S \cdots S$ of length k_1 and $\Lambda_2 = F \cdots F$ of length k_2 in a sequence of first-order homogeneous Markov dependent bistate trials with initial probabilities $P(X_1 = S) = p$ and $P(X_1 = F) = q = 1 - p$, respectively, and transition probability matrix*

$$\boldsymbol{A} = \begin{array}{c} S \\ F \end{array} \begin{bmatrix} p_1 & q_1 \\ p_2 & q_2 \end{bmatrix},$$

where $q_i = 1 - p_i$, $i = 1, 2$. Then the probability generating function of the waiting time random variable $W(\Lambda)$ is given by

$$\varphi_{W(\Lambda)}(s) = \frac{\Delta_1 + \Delta_2}{\Delta_3}, \tag{3}$$

where

$$\begin{aligned} \Delta_1 &= (1 - p_1 s)(p_1 s)^{k_1 - 1} \left\{ ps - pq_2 s^2 + qp_2 s^2 \left[1 - (q_2 s)^{k_2 - 1}\right] \right\}, \\ \Delta_2 &= (1 - q_2 s)(q_2 s)^{k_2 - 1} \left\{ qs - qp_1 s^2 + pq_1 s^2 \left[1 - (p_1 s)^{k_1 - 1}\right] \right\}, \\ \Delta_3 &= (1 - p_1 s)(1 - q_2 s) - p_2 q_1 s^2 \left[1 - (p_1 s)^{k_1 - 1}\right] \left[1 - (q_2 s)^{k_2 - 1}\right]. \end{aligned}$$

Bibliography

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

[2] Aki, S. (1992). Waiting time problems for a sequence of discrete random variables. *Annals of the Institute of Statistical Mathematics*, **44**, 363-378.

[3] Aki, S. (1997). On sooner and later problems between success and failure runs. In *Advances in Combinatorial Methods and Applications to Probability and Statistics* (Ed., N. Balakrishnan), pp. 385-400, Birkhäuser, Boston.

[4] Aki, S. and Hirano, K. (1993). Discrete distributions related to succession events in a two-state Markov chain. In *Statistical Science and Data Analysis* (Eds., K. Matusita, M. L. Puri and T. Hayakawa), pp. 467-474, VSP International Science Publishers, Zeist.

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

[6] Aki, S. and Hirano, K. (1999). Sooner and later waiting time problems for runs in Markov dependent bivariate trials. *Annals of the Institute of Statistical Mathematics*, **51**, 17-29.

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

[8] Aki, S., Kuboki, H. and Hirano, K. (1984). On discrete distributions of order k. *Annals of the Institute of Statistical Mathematics*, **36**, 431-440.

[9] Antzoulakos, D. L. (1999). On waiting time problems associated with runs in Markov dependent trials. *Annals of the Institute of Statistical Mathematics*, **51**, 323-330.

[10] Antzoulakos, D. L. and Philippou, A. N.(1997). On multivariate distributions of various orders obtained by waiting for the r-th success run of length k in trials with multiple outcomes. In *Advances in Combinatorial Methods and Applications to Probability and Statistics* (Ed., N. Balakrishnan), pp. 411-426, Birkhäuser, Boston.

[11] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *Annals of Probability*, **17**, 9-25.

[12] Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statistical Science*, **5**, 403-434.

[13] Balakrishnan, N. and Koutras, M. V. (2002). *Runs and scans with applications.* John Wiley & Sons, New York.

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

[15] Blom, G. and Thorburn, D. (1982). How many random digits are required until given sequences are obtained? *Journal of Applied Probability*, **19**, 518-531.

[16] Boutsikas, M. V. and Koutras, M. V. (2000). Reliability approximation for Markov chain imbeddable systems. *Methodology and Computing in Applied Probability*, **2**, 393-411.

[17] Breen, S., Waterman, M. S. and Zhang, N. (1985). Renewal theory for several patterns. *Journal of Applied Probability*, **22**, 228-234.

[18] Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of CUSUM run length. *Biometrika*, **59**, 539-549.

[19] Champ, C. W. and Woodall, W. H. (1987). Exact results for Shewhart control charts with supplementary runs rules. *Technometrics*, **29**, 393-399.

[20] Chao, M. T. and Fu, J. C. (1989). A limit theorem of certain repairable systems. *Annals of the Institute of Statistical Mathematics*, **41**, 809-818.

[21] Chao, M. T. and Fu, J. C. (1991). The reliability of a large series system under Markov structure. *Advances in Applied Probability*, **23**, 894-908.

[22] Chao, M. T., Fu, J. C. and Koutras, M. V. (1995). Survey of reliability studies of consecutive-k-out-of-n:F & related systems. *IEEE Transactions on Reliability*, **44**, 120-127.

[23] Chiang, D. and Niu, S. C. (1981). Reliability of consecutive-k-out-of-n:F system. *IEEE Transactions on Reliability*, **30**, 87-89.

[24] Chryssaphinou, O. and Papastavridis, S. G. (1990). The occurrence of sequence patterns in repeated dependent experiments. *Theory of Probability and its Applications*, **35**, 145-152.

[25] Doi, M. and Yamamoto, E. (1998). On the joint distribution of runs in a sequence of multistate trials. *Statistics & Probability Letters*, **39**, 133-141.

[26] Ebneshahrashoob, M. and Sobel, M. (1990). Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas. *Statistics & Probability Letters*, **9**, 5-11.

[27] Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Vol. I, Third edition, John Wiley & Sons, New York.

[28] Fu, J. C. (1985). Reliability of a large consecutive-k-out-of-n:F system. *IEEE Transactions on Reliability*, **34**, 127-130.

[29] Fu, J. C. (1986a). Bounds for reliability of large consecutive-k-out-of-n:F systems with unequal component reliability. *IEEE Transactions on Reliability*, **35**, 316-319.

[30] Fu, J. C. (1986b). Reliability of consecutive-k-out-of-n:F systems with $(k-1)$-step Markov dependence. *IEEE Transactions on Reliability*, **35**, 602-606.

[31] Fu, J. C. (1993). Poisson convergence in reliability of a large linearly connected system as related to coin tossing. *Statistica Sinica*, **3**, 261-275.

[32] Fu, J. C. (1995). Exact and limiting distributions of the number of successions in a random permutation. *Annals of the Institute of Statistical Mathematics*, **47**, 435-446.

[33] Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials. *Statistica Sinica*, **6**, 957-974.

[34] Fu, J. C. (2002). Distribution of the scan statistic for a sequence of bistate trials. *Journal of Applied Probability*, **38**, 908-916.

[35] Fu, J. C. and Koutras, M. V. (1994). Distribution theory of runs: a Markov chain approach. *Journal of the American Statistical Association*, **89**, 1050-1058.

[36] Fu, J. C. and Lou, W.Y. W. (2000). Joint distribution of rises and falls. *Annals of the Institute of Statistical Mathematics*, **52**, 415-425.

[37] Fu, J. C. and Lou, W.Y. W. (2002). *Distribution Theory of Runs and Patterns and its Applications: A Finite Markov Chain Approach.* World Scientific Publishing Co., to appear.

[38] Fu, J. C., Lou, W.Y. W. and Chen, S.C. (1999). On the probability of pattern matching in nonaligned DNA sequences: a finite Markov chain imbedding approach. In *Scan Statistics and Applications* (Eds., J. Glaz and N. Balakrichnan), pp. 287-302, Birkhäuser, Boston.

[39] Fu, J. C., Lou, W.Y. W. and Wang, Y.J. (1999). On the exact distributions of Eulerian and Simon Newcomb numbers associated with random permutations. *Statistics & Probability Letters*, **42**, 115-125.

[40] Fu, J. C., Spiring, F. A. and Xie, H. (2002). On the average run lengths of quality control schemes using a Markov Chain approach. *Statistics & Probability Letters*, **56**, 369-380.

[41] Godbole, A. P. (1990a). Degenerate and Poisson convergence criteria for success runs. *Statistics & Probability Letters*, **10**, 247-255.

[42] Godbole, A. P. (1990b). On hypergeometric and related distributions of order k. *Communications in Statistics — Theory and Methods*, **19**, 1291-1301.

[43] Godbole, A. P. (1991). Poisson approximations for runs and patterns of rare events. *Advances in Applied Probability*, **23**, 851-865.

[44] Goldstein, L. (1990). Poisson approximation and DNA sequence matching. *Communications in Statistics — Theory and Methods*, **19**, 4167-4179.

[45] Han, Q. and Aki, S. (2000a). Sooner and later waiting time problems based on a dependent sequence. *Annals of the Institute of Statistical Mathematics*, **52**, 407-414.

[46] Han, Q. and Aki, S. (2000b). Waiting time problems in a two-state Markov chain. *Annals of the Institute of Statistical Mathematics*, **52**, 778-789.

[47] Hirano, K. (1986). Some properties of the distributions of order k. In *Fibonacci Numbers and Their Applications* (Eds., A. N. Philippou, A. F. Horadam and G. E. Bergum), pp. 43-53, D. Reidel Publishing Company, Dordrecht, The Neterhlands.

[48] Hirano, K. and Aki, S. (1987). Properties of the extended distributions of order k. *Statistics & Probability Letters*, **6**, 67-69.

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

[50] Hirano, K., Aki, S. and Uchida, M. (1997). Distributions of numbers of success-runs until the first consecutive k successes in higher order Markov dependent trials. In *Advances in Combinatorial Methods and Applications to Probability and Statistics* (Ed., N. Balakrishnan), pp. 401-410, Birkhäuser, Boston.

[51] Johnson, B. C. and Fu, J. C. (2000). The distribution of increasing l-sequences in random permutations: a Markov chain approach. *Statistics & Probability Letters*, **49**, 337-344.

[52] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions.* Second edition, John Wiley & Sons, New York.

[53] Koutras, M. V. (1997a). Waiting times and number of appearances of events in a sequence of discrete random variables. In *Advances in Combinatorial Methods*

and Applications to Probability and Statistics (Ed., N. Balakrishnan), pp. 363-384, Birkhäuser, Boston.

[54] Koutras, M. V. (1997b). Waiting time distributions associated with runs of fixed length in two-state Markov chains. *Annals of the Institute of Statistical Mathematics*, **49**, 123-139.

[55] Koutras, M. V. and Alexandrou, V. A. (1997). Sooner waiting time problems in a sequence of trinary trials. *Journal of Applied Probability*, **34**, 593-609.

[56] Li, S.-Y. R. (1980). A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Annals of Probability*, **8**, 1171-1176.

[57] Ling, K. D. (1988). On binomial distributions of order k. *Statistics & Probability Letters*, **6**, 247-250.

[58] Ling, K. D. and Low, T. Y. (1993). On the soonest and latest waiting time distributions: succession quotas. *Communications in Statistics — Theory and Methods*, **22**, 2207-2221.

[59] Lou, W.Y. W. (1996). On runs and longest run tests: a method of finite Markov chain imbedding. *Journal of the American Statistical Association*, **91**, 1595-1601.

[60] Lou, W.Y. W. (1997). An application of the method of finite Markov chain imbedding to runs tests. *Statistics & Probability Letters*, **31**, 155-161.

[61] Lucas, J. M. and Crosier, R. B. (1982). Robust CUSUM: a robustness study for Cusum quality control schemes. *Communications in Statistics — Theory and Methods*, **11**, 2669-2687.

[62] Lucas, J. M. and Crosier, R. B. (2000). Fast initial response for CUSUM quality-control schemes: give your CUSUM a head start. *Technometrics*, **42**, 102-107.

[63] Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, **32**, 1-12.

[64] Luceno, A. and Puig-Pey, J. (2000). Evaluation of the run-length probability distribution for CUSUM charts: assessing chart performance. *Technometrics*, **42**, 411-416.

[65] Mohanty, S. G. (1994). Success runs of length k in Markov dependent trials. *Annals of the Institute of Statistical Mathematics*, **46**, 777-796.

[66] Montgomery, D. C. (2001). *Introduction to Statistical Quality Control.* Fourth edition, John Wiley & Sons, New York.

[67] Mood, A. M. (1940). The distribution theory of runs. *Annals of Mathematical Statistics*, **11**, 367-392.

[68] Mosteller, F. (1941). Note on an application of runs to quality control charts. *Annals of Mathematical Statistics*, **12**, 228-232.

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

[70] Panaretos, J. and Xekalaki, E. (1986). On some distributions arising from certain generalized sampling schemes. *Communications in Statistics — Theory and Methods*, **15**, 873-891.

[71] Philippou, A. N. (1983). The Poisson and compound Poisson distributions of order k and some of its properties. *Zapiski Nauchnyka Seminarov Lenigrand, Math. Inst. Steklova*, **130**, 175-180.

[72] Philippou, A. N. (1988). On multiparameter distributions of order k. *Annals of the Institute of Statistical Mathematics*, **40**, 467-475.

[73] Philippou, A. N. and Antzoulakos, D. L. (1990). Multivariate distributions of order k on a generalized sequence. *Statistics & Probability Letters*, **9**, 453-463.

[74] Philippou, A. N. and Makri, F. S. (1986). Successes, runs and longest runs. *Statistics & Probability Letters*, **4**, 101-105.

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

[76] Philippou, A. N. and Tripsiannis, G. A. (1991). Multivariate Polya and inverse Polya distributions of order k. *Biometrical Journal*, **33**, 225-236.

[77] Philippou, A. N., Antzoulakos, D. L. and Tripsiannis, G. A. (1989). Multivariate distributions of order k. *Statistics & Probability Letters*, **7**, 207-216.

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

[79] Philippou, A. N., Tripsiannis, G. A and Antzoulakos, D. L. (1989). New Polya and inverse Polya distributions of order k. *Communications in Statistics — Theory and Methods*, **18**, 2125-2137.

[80] Rao, B. V., Disney, R. L. and Pignatiello Jr., J. J. (2001). Uniqueness and convergence of solutions to average run length integral equation for cumulative sum and other control charts. *IIE Transactions*, **33**, 463-469.

[81] Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, **1**, 239-250.

[82] Robinson, P. B. and Ho, T. Y. (1978). Average run lengths of geometric moving average charts by numerical methods. *Technometrics*, **20**, 85-94.

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

[84] Shmueli, G. and Cohen, A. (2000). Run-related probability functions applied to sampling inspection. *Technometrics*, **42**, 188-202.

[85] Stanley, R. P. (1986). *Enumerative Combinatorics.* Vol. I, Wadsworth & Brooks/Cole, California.

[86] Tripsiannis, G. A (1993). Modified multivariate Polya and inverse Polya distributions of order k. *Journal of the Indian Society of Statistics and Operations Research*, **14**, 1-14.

[87] Uchida, M. (1998). On generating functions of waiting time problems for a sequence patterns of discrete random variables. *Annals of the Institute of Statistical Mathematics*, **50**, 655-671.

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

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

[90] Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics*, **11**, 147-162.

[91] Walsh, J. E. (1962). *Handbook of Nonparametric Statistics.* Van Nostrand Co., Princeton, New Jersey.

[92] Wang, Y. H. (1993). On the number of successes in dependent trials. *Statistica Sinica*, **3**, 295-312.

[93] Waterman, M. S. (1995). *Introduction to Computational Biology: Maps, Sequences and Genomes.* Chapman & Hall, London.

[94] Wolfowitz, J. (1943). On the theory of runs with some applications to quality control. *Annals of Mathematical Statistics*, **14**, 280-288.