AISM (2006) 58: 291–310
DOI 10.1007/s10463-006-0038-8

James C. Fu · W.Y. Wendy Lou

Waiting time distributions of simple and compound patterns in a sequence of r-th order Markov dependent multi-state trials

Received: 24 December 2004 / Revised: 23 May 2005 / Published online: 1 June 2006

Abstract Waiting time distributions of simple and compound runs and patterns have been studied and applied in various areas of statistics and applied probability. Most current results are derived under the assumption that the sequence is either independent and identically distributed or first order Markov dependent. In this manuscript, we provide a comprehensive theoretical framework for obtaining waiting time distributions under the most general structure where the sequence consists of r-th order ($r \geq 1$) Markov dependent multi-state trials. We also show that the waiting time follows a generalized geometric distribution in terms of the essential transition probability matrix of the corresponding imbedded Markov chain. Further, we derive a large deviation approximation for the tail probability of the waiting time distribution based on the largest eigenvalue of the essential transition probability matrix. Numerical examples are given to illustrate the simplicity and efficiency of the theoretical results.

Keywords Runs and patterns · Markov chain imbedding · Transition probability matrix · Large deviation approximation

1 Introduction

Let $\{X_t\} = \{X_{-\infty}, \ldots, X_{-1}, X_0, X_1, \ldots, X_\infty\}$ be a sequence of m-state ($m \geq 2$) trials defined on the set $\mathcal{S} = \{b_1, b_2, \ldots, b_m\}$, and let Λ denote a *simple pattern* of length k if Λ is composed of a specified sequence of k symbols from $\mathcal{S}$, *i.e.*, $\Lambda = b_{i_1}, \ldots b_{i_k}$; the symbols b_i in the pattern Λ may be repeated. For example, in

J.C. Fu (✉)
Department of Statistics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
E-mail: fu@cc.umanitoba.ca

W.Y.W. Lou
Department of Public Health Sciences, University of Toronto, Toronto, ON M5T 3M7, Canada

a sequence of Bernoulli trials ($m = 2$), the success run of size k, $\Lambda = S \dots S$, is a simple pattern of size k.

Let Λ_1 and Λ_2 be two simple patterns of lengths k_1 and k_2, respectively. We say that Λ_1 and Λ_2 are two distinct patterns if neither Λ_1 belongs to Λ_2 nor Λ_2 belongs to Λ_1. We define the union $\{\Lambda_1\} \cup \{\Lambda_2\}$ as the occurrence of either pattern Λ_1 or pattern Λ_2. For simplicity, we denote the union of two patterns by $\Lambda_1 \cup \Lambda_2$ throughout the manuscript. A pattern Λ is referred to as a *compound pattern* if it is a union of l ($l \geq 1$) distinct simple patterns $\Lambda_1, \dots, \Lambda_l$, i.e., $\Lambda = \cup_{i=1}^{l} \Lambda_i$.

Given a simple pattern $\Lambda = b_{i_1} \dots b_{i_k}$, we define a waiting time random variable $W(\Lambda)$ as

$$W(\Lambda) = \inf\{n : n \in J^+, n \geq k, X_{n-k+1} = b_{i_1}, \dots, X_n = b_{i_k}\}, \tag{1}$$

where $J^+ = \{1, 2, \dots, \}$. Throughout this article, we assume that the counting of symbols toward forming the pattern Λ starts at $t = 1$. For a compound pattern $\Lambda = \cup_{i=1}^{l} \Lambda_i$, we define the waiting time random variable $W(\Lambda)$ as

$$W(\Lambda) = \inf\{n : n \in J^+, \text{occurrence of any pattern } \Lambda_1, \dots, \Lambda_l \text{ at the } n\text{-th trial}\}. \tag{2}$$

It follows from the above definition that mathematically, the waiting time $W(\Lambda)$ of a compound pattern $\Lambda = \cup_{i=1}^{l} \Lambda_i$ can always be represented as

$$W(\Lambda) = \inf\{W(\Lambda_1), W(\Lambda_2), \dots, W(\Lambda_l)\}, \tag{3}$$

where $W(\Lambda_i)$ are the waiting time random variables of the simple patterns Λ_i, $i = 1, \dots, l$, defined by Eq. (1). For this reason, the waiting time $W(\Lambda)$ of a compound pattern is often referred to as the *sooner waiting time* of l distinct simple patterns.

Waiting time distributions of runs and patterns, such as geometric, geometric of order k, negative binomial, and sooner and later, have been applied successfully in numerous areas of statistics and applied probability. Over the past several decades, the theory of waiting time distributions has become an indispensable tool for studying applications in fields such as DNA sequence homology and the reliability of engineering systems. A considerable amount of literature treating waiting time distributions was published in the last century, but most formulae were obtained through combinatoric analysis under the restriction that $\{X_t\}$ is a sequence of independent and identically distributed ($i.i.d.$) two- or multi-state trials. The recent book by Balakrishnan and Koutras (2002) provides excellent information on past and current developments in this area. Recently, Fu and Koutras (1994), Koutras and Alexandrou (1995, 1997), Fu (1996), Lou (1996), Koutras (1997), Aki and Hirano (1999), Han and Aki (2002a), Han and Aki (2002b), and Fu and Chang (2002) studied waiting time problems, via various techniques including finite Markov chain imbedding, for the case where $\{X_t\}$ is a sequence of first order Markov dependent two- or multi-state trials. Han and Hirano (2003) treated the waiting times for compound patterns generated by two patterns in a sequence of multi-state first-order Markov dependent trials using the generating function technique. Aki and Hirano (2004) further studied waiting time problems for a two-dimensional pattern. The recent book by Fu and Lou (2003) provides some details and results for this approach. However, there are no general results for the waiting time distributions of simple or compound patterns when $\{X_t\}$ is a sequence of r-th ($r \geq 1$) order Markov dependent multi-state trials.

This manuscript mainly studies the waiting time distributions for simple and compound patterns under the general setting where $\{X_t\}$ consists of a sequence of r-th order homogeneous Markov dependent multi-state trials. We show that the waiting time of a specified run or pattern Λ (simple or compound) has a general geometric distribution in the following sense: there exists an imbedded finite Markov chain $\{Y_t\}$ defined on a state space $\Omega(\Lambda)$ with a transition probability matrix of the form

$$\boldsymbol{M} = \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{O} & \boldsymbol{I} \end{array}\right), \tag{4}$$

such that, for $n \geq r + 1$,

$$P(W(\Lambda) = n) = \boldsymbol{\xi} \boldsymbol{N}^{n-r-1}(\boldsymbol{I} - \boldsymbol{N})\mathbf{1}', \tag{5}$$

or

$$P(W(\Lambda) \geq n) = \boldsymbol{\xi} \boldsymbol{N}^{n-r-1}\mathbf{1}', \tag{6}$$

where $\boldsymbol{\xi}$ is the initial distribution at the r-th trial induced by the ergodic distribution of the r-th order Markov dependent trials $\{X_t\}$, $\boldsymbol{I}$ denotes an identity matrix, and $\mathbf{1}$ is a row vector with each entry equal to one. The matrix $\boldsymbol{N}$ is referred to as the *essential transition probability matrix*, and is defined by Eq. (4). The two equations (5) and (6) are equivalent, but Eq. (6) is easier to deal with mathematically; hence we will focus on the tail probability given by Eq. (6) throughout the manuscript. Some characteristics of the waiting time distribution of $W(\Lambda)$, such as the mean and the probability generating function, are also studied. Further, we provide a large deviation approximation for the tail probability of $W(\Lambda)$ in the sense that

$$\lim_{n\to\infty} \frac{1}{n} \log P(W(\Lambda) \geq n) = -\beta, \tag{7}$$

where $\beta = -\log \lambda_{[1]}$ and $\lambda_{[1]}$ is the largest eigenvalue of the essential transition probability matrix $\boldsymbol{N}$. More precisely, the tail probability $P(W(\Lambda) \geq n)$ can be approximated by its exponential rate β in the sense that there exists a constant c, which is mainly a function of the largest eigenvalue $\lambda_{[1]}$ and its corresponding eigenvector $\boldsymbol{\eta}_{[1]}$, such that

$$P(W(\Lambda) \geq n) \sim c \exp\{-n\beta\}. \tag{8}$$

This manuscript is organized in the following way. Section 2 sets up the notation, and studies the ergodic distribution induced by the higher order Markov dependent multi-state trials. Section 3 treats mainly the exact distribution of the waiting time of a simple pattern, and Sect. 4 considers the waiting time of a compound pattern. Section 5 provides the generating function, the mean and the large deviation approximation. Section 6 gives numerical examples along with some discussion for possible extensions of our approach.

2 Notation and the r-th order markov chain

Let $\{X_t\}$ be a sequence of irreducible, aperiodic and homogeneous r-th order Markov dependent m-state random variables (trials) defined on the state space $\mathcal{S} = \{b_1, \ldots, b_m\}$. For $r \geq 1$, let

$$\mathcal{S}^r = \{\boldsymbol{x} = x_1 \ldots x_r : x_i \in \mathcal{S}, i = 1, \ldots, r\}$$

be the set of all possible outcomes of r trials, with $\mathrm{Card}(\mathcal{S}^r) = m^r$. Given $\boldsymbol{x} \in \mathcal{S}^r$, we define the r-th order transition probabilities for the homogeneous Markov chain $\{X_t\}$:

$$P(X_t = b | X_{t-r} = x_1, \ldots, X_{t-1} = x_r) = p_{\boldsymbol{x}\bullet b} \tag{9}$$

for every $b \in \mathcal{S}$ and $\boldsymbol{x} \in \mathcal{S}^r$, probabilities which are independent of t. This is equivalent to saying that, for every t,

$$P(X_{t-r+1} \ldots X_t = x_2 x_3 \ldots x_r b | X_{t-r} \ldots X_{t-1} = x_1 \ldots x_r) = p_{\boldsymbol{x}\bullet b}. \tag{10}$$

Since $\{X_t\}$ is an irreducible, aperiodic and homogeneous Markov chain, it follows from Eq. (10) that the ergodic probabilities $\pi_{\boldsymbol{x}}$ exist for every $\boldsymbol{x} \in \mathcal{S}^r$, and that

$$\pi_{\boldsymbol{x}} = P(X_{t-r+1} \ldots X_t = \boldsymbol{x}) = P(X_{\tau-r+1} \ldots X_\tau = \boldsymbol{x}) \tag{11}$$

for any t and τ. Denote by $\boldsymbol{\pi} = (\pi_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{S}^r)$ the ergodic distribution of $\boldsymbol{x}$ on $\mathcal{S}^r$. It is well known that the ergodic distribution $\boldsymbol{\pi}$ is the solution of the following equations:

$$\boldsymbol{\pi A} = \boldsymbol{\pi}, \tag{12}$$

and

$$\sum_{\boldsymbol{x} \in \mathcal{S}^r} \pi_{\boldsymbol{x}} = 1, \tag{13}$$

where $\boldsymbol{A} = (p_{\boldsymbol{xy}})$ is an $m^r \times m^r$ matrix, and for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}^r$,

$$p_{\boldsymbol{xy}} = \begin{cases} p_{\boldsymbol{x}\bullet b} & \text{if } \boldsymbol{y} = L_{r-1}(\boldsymbol{x})b, b \in \mathcal{S} \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

with $L_{r-1}(\boldsymbol{x}) = x_2 \ldots x_r$, the subsequence consisting of the last $(r-1)$ symbols of $\boldsymbol{x}$. The ergodic distribution $\boldsymbol{\pi}$ generated by the homogeneous Markov chain will play an important role in forming the initial distribution for the imbedded Markov chain of $W(\Lambda)$.

3 Waiting time distribution of a simple pattern

Let $\Lambda = b_{i_1} \ldots b_{i_k}$ be a specified simple pattern of length k. We study separately the distribution of the waiting time $W(\Lambda)$ for the two cases $k \leq r$ and $k > r$. Before we present general results we would like to provide a simple example with $k \leq r$.

Example 1 Let $\mathcal{S} = \{F, S\}$, $r = 3$, $\mathcal{S}^3 = \{FFF, FFS, FSF, FSS, SFF, SFS, SSF, SSS\}$, and $\{X_t\}_{-\infty}^{\infty}$ be an irreducible, aperiodic and third order homogeneous Markov chain having the transition probabilities

$$\{p_{\boldsymbol{x}\bullet b} : \boldsymbol{x} \in \mathcal{S}^3, b \in \mathcal{S}\} = \{p_{FFF\cdot F}, p_{FFF\cdot S}, \ldots, p_{SSS\cdot F}, p_{SSS\cdot S}\}.$$

Given the transition probabilities $p_{\boldsymbol{x}\cdot b}$ defined above, it follows from Eqs. (12) and (13) that the ergodic distribution $\boldsymbol{\pi} = (\pi_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{S}^r) = (\pi_{FFF}, \pi_{FFS}, \ldots, \pi_{SSF}, \pi_{SSS})$ is the solution of the equations

$$\boldsymbol{\pi} \boldsymbol{A} = \boldsymbol{\pi} \quad \text{and} \quad \sum_{\boldsymbol{x} \in \mathcal{S}^3} \pi_{\boldsymbol{x}} = 1,$$

where the transition probability matrix $\boldsymbol{A}$ is determined via Eq. (14) as

$$\boldsymbol{A} = \begin{array}{c} FFF \\ FFS \\ FSF \\ FSS \\ SFF \\ SFS \\ SSF \\ SSS \end{array} \begin{pmatrix} p_{FFF\cdot F} & p_{FFF\cdot S} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{FFS\cdot F} & p_{FFS\cdot S} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{FSF\cdot F} & p_{FSF\cdot S} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{FSS\cdot F} & p_{FSS\cdot S} \\ p_{SFF\cdot F} & p_{SFF\cdot S} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{SFS\cdot F} & p_{SFS\cdot S} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{SSF\cdot F} & p_{SSF\cdot S} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{SSS\cdot F} & p_{SSS\cdot S} \end{pmatrix}.$$

Given a simple pattern $\Lambda = SS$ with $k = 2$, suppose that we are interested in finding the exact distribution of the waiting time $W(\Lambda)$, where we begin counting symbols toward the pattern Λ at time $t = 1$. Let us consider the first three trials $X_1 X_2 X_3 = \boldsymbol{x} \in \mathcal{S}^3$ and define the subsets

$$\mathcal{S}_2^3(\Lambda) = \{SSF, SSS\} = \{\boldsymbol{x} : W(\Lambda) = 2, \boldsymbol{x} \in \mathcal{S}^3\},$$

$$\mathcal{S}_3^3(\Lambda) = \{FSS\} = \{\boldsymbol{x} : W(\Lambda) = 3, \boldsymbol{x} \in \mathcal{S}^3\},$$

and

$$\mathcal{S}_\alpha^3(\Lambda) = \mathcal{S}_2^3(\Lambda) \cup \mathcal{S}_3^3(\Lambda) = \{SSF, SSS, FSS\}.$$

Since we count toward the pattern starting from $t = 1$, it follows from the definition of $W(\Lambda)$ given by Eq. (1) that $P(W(\Lambda) = 1) \equiv 0$, $P(W(\Lambda) = 2) = \pi_{SSF} + \pi_{SSS}$ and $P(W(\Lambda) = 3) = \pi_{FSS}$.

We construct the imbedded homogeneous Markov chain $\{Y_t\}_4^\infty$ on the state space

$$\Omega(\Lambda) = \{FFF, FFS, FSF, SFF, SFS, \alpha\},$$

where $\{\alpha\} = \{SSF, SSS, FSS\}$ represents an absorbing state. Note that every state in the state space $\Omega(\Lambda)$ has pattern length k, and $\Omega(\Lambda) \subseteq \mathcal{S}^r$. The initial distribution of Y_3 is given by

$$\boldsymbol{\xi}_3 = (P(Y_3 = \boldsymbol{x}) : \boldsymbol{x} \in \Omega(\Lambda)) = (\pi_{FFF}, \pi_{FFS}, \pi_{FSF}, \pi_{SFF}, \pi_{SFS}, \pi_{SSF}+\pi_{SSS}+\pi_{FSS}),$$

and the transition probability matrix of the imbedded Markov chain has the form

$$\boldsymbol{M} = \begin{array}{c} FFF \\ FFS \\ FSF \\ SFF \\ SFS \\ \alpha \end{array} \left(\begin{array}{ccccc|c} p_{FFF\cdot F} & p_{FFF\cdot S} & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{FFS\cdot F} & 0 & 0 & p_{FFS\cdot S} \\ 0 & 0 & 0 & p_{FSF\cdot F} & p_{FSF\cdot S} & 0 \\ p_{SFF\cdot F} & p_{SFF\cdot S} & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{SFS\cdot F} & 0 & 0 & p_{SFS\cdot S} \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & 1 \end{array}\right).$$

For $n \geq 4$, since $\{Y_t\}_4^\infty$ is an imbedded Markov chain for $W(\Lambda)$, it follows from the Chapman-Kolmogorov equation that

$$\begin{aligned} P(W(\Lambda) \geq n|\boldsymbol{\xi}_3) &= P(Y_n \in \Omega(\Lambda) - \{\alpha\}|\boldsymbol{\xi}_3) \\ &= \boldsymbol{\xi}_3 \boldsymbol{M}^{n-4}(1, 1, 1, 1, 1, 0)' \\ &= (\pi_{FFF}, \pi_{FFS}, \pi_{FSF}, \pi_{SFF}, \pi_{SFS})\boldsymbol{N}^{n-4}(1, 1, 1, 1, 1)'. \end{aligned}$$

In view of the above detailed example, we consider that the following general results hold for the initial and waiting time distributions of a simple pattern Λ.

Lemma 3.1 *Assume $\{X_t\}_{-\infty}^\infty$ is a sequence of irreducible, aperiodic, and r-th order homogeneous Markov dependent m-state trials with the transition probabilities $p_{\boldsymbol{x}\cdot b}$, $\boldsymbol{x} \in \mathcal{S}^r$ and $b \in \mathcal{S}$. Given a specified simple pattern $\Lambda = b_{i_1} b_{i_2} \dots b_{i_k}$ of length $k \leq r$, and assuming that counting toward the pattern starts from $t = 1$, then*

(i) for every $\boldsymbol{x} \in \mathcal{S}^r$, the ergodic probability $\pi_{\boldsymbol{x}}$ exists,
(ii) $P(W(\Lambda) \geq r+1, X_1 X_2 \dots X_r = \boldsymbol{x}) = \pi_{\boldsymbol{x}}$, for all $\boldsymbol{x} \in \{\mathcal{S}^r - \mathcal{S}^r_\alpha\}$, and
(iii)

$$P(W(\Lambda) = j) = \sum_{\boldsymbol{x} \in \mathcal{S}^r_j} \pi_{\boldsymbol{x}}, \tag{15}$$

where

$$\mathcal{S}^r_j = \{\boldsymbol{x} : W(\Lambda) = j \text{ and } \boldsymbol{x} \in \mathcal{S}^r\}, \tag{16}$$

for all $j = k, \ldots, r$, *and*

$$P(W(\Lambda) \leq r) = \sum_{x \in \mathcal{S}^r_\alpha} \pi_x,$$

where

$$\mathcal{S}^r_\alpha = \cup^r_{j=k} \mathcal{S}^r_j.$$

Proof Since $\{X_t\}$ is an irreducible, aperiodic and homogeneous Markov chain, the existence of the ergodic probabilities $\pi_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{S}^r$ is guaranteed, and the ergodic distribution $\boldsymbol{\pi} = (\pi_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{S}^r)$ is the solution of Eqs. (12) and (13) (see Feller 1968).

It follows from the Definition (16) that for every $k \leq j \leq r$,

$$P(W(\Lambda) = j \text{ and } X_1 \ldots X_r = \boldsymbol{x}) = \pi_{\boldsymbol{x}}, \qquad \text{for every } \boldsymbol{x} \in \mathcal{S}^r_j. \tag{17}$$

Results (ii) and (iii) are then immediate consequences of Eq. (17). This completes the proof. □

Let us group *all* the $\boldsymbol{x} \in \mathcal{S}^r_\alpha$ as an *absorbing state* α, and define the state space for the imbedded Markov chain $\{Y_t\}^\infty_r$ of the waiting time random variable $W(\Lambda)$ as

$$\Omega(\Lambda) = \{\mathcal{S}^r - \mathcal{S}^r_\alpha\} \cup \{\alpha\}. \tag{18}$$

We define a mapping $\langle \cdot, \cdot \rangle_{\Omega(\Lambda)}\colon \Omega(\Lambda) \times \mathcal{S} \to \Omega(\Lambda)$ as, for every given $\boldsymbol{x} \in \Omega(\Lambda)$ and $b \in \mathcal{S}$,

$$\boldsymbol{y} = \langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)} = \begin{cases} \alpha & \text{if } \boldsymbol{x} = \alpha \text{ and } b \in \mathcal{S} \\ L_{r-1}(\boldsymbol{x})b & \text{if } \boldsymbol{x} \neq \alpha \text{ and } L_{r-1}(\boldsymbol{x})b \notin \mathcal{S}^r_\alpha \\ \alpha & \text{if } \boldsymbol{x} \neq \alpha \text{ and } L_{r-1}(\boldsymbol{x})b \in \mathcal{S}^r_\alpha. \end{cases} \tag{19}$$

In view of our Example 3.1 and Lemma 3.1, the following results hold.

Theorem 3.1 *Assuming* Λ *is a simple pattern with length* $k \leq r$, *the waiting time random variable* $W(\Lambda)$ *is Markov chain imbeddable. The imbedded Markov chain* $\{Y_t\}^\infty_r$ *defined on the state space* $\Omega(\Lambda)$ *given by Eq. (18) has*

(i) the initial distribution, at $t = r$,

$$\boldsymbol{\xi}_r = (\boldsymbol{\xi} : \boldsymbol{\xi}_\alpha), \tag{20}$$

where $\boldsymbol{\xi} = (\pi_{\boldsymbol{x}}; \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\})$ *and* $\boldsymbol{\xi}_\alpha = \sum_{\boldsymbol{x} \in \mathcal{S}^r_\alpha} \pi_{\boldsymbol{x}}$,

(ii) the transition probability matrix

$$\boldsymbol{M} = (p_{\boldsymbol{xy}}) = \begin{array}{c} \Omega(\Lambda) - \alpha \\ \alpha \end{array} \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & 1 \end{array} \right), \tag{21}$$

where the transition probabilities are given by, for every $\boldsymbol{x}, \boldsymbol{y} \in \Omega(\Lambda)$,

$$p_{\boldsymbol{xy}} = \begin{cases} p_{\boldsymbol{x} \bullet b} & \text{if } \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\},\ \boldsymbol{y} = \langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)}, \text{ and } b \in \mathcal{S} \\ 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} = \alpha \\ 0 & \text{otherwise}, \end{cases} \tag{22}$$

with $\langle \cdot, \cdot \rangle_{\Omega(\Lambda)}$ *given by Eq. (19), and* *(iii) for $n < k$, $P(W(\Lambda) = n) = 0$, for $k \le n \le r$, $P(W(\Lambda) = n) = \sum_{\boldsymbol{x} \in \mathcal{S}_n^r} \boldsymbol{\pi}_{\boldsymbol{x}}$, and for $n \ge r + 1$,*

$$P(W(\Lambda) \ge n) = \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathbf{1}'. \tag{23}$$

Proof Since $\{X_t\}$ is a sequence of irreducible, aperiodic and r-th order homogeneous Markov dependent trials, it follows from Lemma 3.1 that, for $t = r$ and for every $\boldsymbol{x} \in \{\mathcal{S}^r - \mathcal{S}_\alpha^r\}$, we have $P(W(\Lambda) \ge r + 1,\ X_1 \ldots X_r = \boldsymbol{x}) = \pi_{\boldsymbol{x}}$. For $\boldsymbol{x} \in \mathcal{S}_\alpha^r$, it is easy to see that

$$\boldsymbol{\xi}_\alpha = P(W(\Lambda) \le r) = P(\boldsymbol{x} \in \mathcal{S}_\alpha^r) = \sum_{\boldsymbol{x} \in \mathcal{S}_\alpha^r} \pi_{\boldsymbol{x}}.$$

It follows that Y_r has a distribution $\boldsymbol{\xi}_r = (\boldsymbol{\xi} : \boldsymbol{\xi}_\alpha)$ on $\Omega(\Lambda)$, where $\boldsymbol{\xi} = (\pi_{\boldsymbol{x}} : \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\})$ and $\boldsymbol{\xi}_\alpha = \sum_{\boldsymbol{x} \in \mathcal{S}_\alpha^r} \pi_{\boldsymbol{x}}$. This completes the proof of Result (i).

For $t \ge r + 1$, we define a homogeneous Markov chain $\{Y_t\}$ on the state space $\Omega(\Lambda)$ with the transition matrix $\boldsymbol{M} = (p_{\boldsymbol{x}\boldsymbol{y}})$. Given $Y_{t-1} = \boldsymbol{x}$, $\boldsymbol{x} \in \Omega(\Lambda)$ and $X_t = b$, $b \in \mathcal{S}$, we define

$$Y_t = \langle Y_{t-1}, X_t \rangle_{\Omega(\Lambda)}, \tag{24}$$

where $\langle \cdot, \cdot \rangle_{\Omega(\Lambda)}$ is defined by Eq. (19). It follows from the above definition that, for every $t = r, r + 1, \ldots, Y_t = \boldsymbol{x}$ contains the necessary information about the r-th order Markov dependency and the subpattern of Λ. For every $\boldsymbol{x} \in \Omega(\Lambda)$, we define a subset in $\Omega(\Lambda)$:

$$[\boldsymbol{x}, \mathcal{S}] = \{\boldsymbol{y} : \boldsymbol{y} \in \Omega(\Lambda), \boldsymbol{y} = \langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)}, b \in \mathcal{S}\}. \tag{25}$$

Note that $[\alpha, \mathcal{S}] = \{\alpha\}$. It clearly follows from the definitions of Eqs. (19) and (24) that the transition probabilities are defined as: for $\boldsymbol{x}$, $\boldsymbol{y} \in \Omega(\Lambda)$,

$$p_{\boldsymbol{x}\boldsymbol{y}} = \begin{cases} p_{\boldsymbol{x}\bullet b} & \text{if } \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\},\ \boldsymbol{y} = \langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)} \text{ and } b \in \mathcal{S} \\ 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} = \alpha \\ 0 & \text{if } \boldsymbol{y} \notin [\boldsymbol{x}, \mathcal{S}]. \end{cases}$$

This completes the proof of Result (ii). The first two parts of Result (iii) are concluded directly from Lemma 3.1 (iii). For $n \ge r + 1$, since $W(\Lambda) \ge n$ if and only if $Y_{n-1} \notin \{\alpha\}$, the last part of Result (iii) is a direct consequence of the identity

$$P(W(\Lambda) \ge n) = P(Y_{n-1} \in \Omega(\Lambda) - \{\alpha\}),$$

the Chapman–Kolmogorov equation, and Eq. (21). □

For the case of $k > r$, we define a set of essential subsequences for the pattern $\Lambda = b_{i_1} \ldots b_{i_r} b_{i_{r+1}} \ldots b_{i_k}$ with length longer than r but less than or equal to $k - 1$,

$$\mathcal{S}_+^r = \{b_{i_1} \ldots b_{i_{r+1}}, b_{i_1} \ldots b_{i_{r+2}}, \ldots, b_{i_1} \ldots b_{i_{k-1}}\} \quad \text{with } \alpha = b_{i_1} \ldots b_{i_k} = \Lambda, \tag{26}$$

and define the state space for the imbedded Markov chain $\{Y_t\}$ as

$$\Omega(\Lambda) = \mathcal{S}^r \cup \mathcal{S}_+^r(\Lambda) \cup \{\alpha\}. \tag{27}$$

Note that every state $\boldsymbol{x}$ in $\Omega(\Lambda)$ defined by Eq. (27) has to have length l greater than or equal to r but less than or equal to k (i.e. $r \leq l \leq k$). Here we need to extend our definition of $\langle \boldsymbol{x}, b\rangle_{\Omega(\Lambda)}$ given by Eq. (19) to cover the cases where the state space $\Omega(\Lambda)$ contains states longer than r. For $\boldsymbol{x} = x_1 \dots x_l$, $r \leq l \leq k$, with $\boldsymbol{x} \in \Omega(\Lambda)$ and $b \in \mathcal{S}$, we define

$$\langle \boldsymbol{x}, b\rangle_{\Omega(\Lambda)} = \begin{cases} \alpha & \text{if } \boldsymbol{x} = \alpha \text{ and for all } b \in \mathcal{S} \\ \boldsymbol{y} & \text{if } \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\}, \end{cases} \tag{28}$$

where $\boldsymbol{y}$ is the longest subsequence of $\{x_1x_2 \dots x_l b, x_2x_3 \dots x_l b, \dots, x_{l-r+2}x_{l-r+3} \dots x_l b\}$ in $\Omega(\Lambda)$. This operation $\langle \cdot, \cdot\rangle_{\Omega(\Lambda)}$ is based on the forward and backward counting principle given by Fu (1996), and we define $Y_t = \langle Y_{t-1}, X_t\rangle_{\Omega(\Lambda)}$. Following from the definition of $\langle \cdot, \cdot\rangle_{\Omega(\Lambda)}$, if $Y_t = \alpha$, it means Λ has occurred before or at time t.

Remark 1 Theorem 3.1 (iii) can be viewed as the conditional probability distribution in the following sense: for $n \geq r+1$ and $\boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\}$,

$$P(W(\Lambda) \geq n | (x_1, \dots, x_r) = \boldsymbol{x}) = (0, \dots, 0, 1, 0, \dots, 0)\boldsymbol{N}^{n-r-1}\mathbf{1},$$

where $(0, \dots, 0, 1, 0, \dots, 0)$ is the unit vector corresponding to $\boldsymbol{x}$. The conditional probability is useful when studying more complicated problems. The proof of this result is an immediate consequence of Theorem 3.1 (iii) with given $(x_1, \dots, x_r) = \boldsymbol{x}$ and $(0, \dots, 0, 1, 0, \dots, 0)$ as the initial probability.

Theorem 3.2 *For $k > r$, the waiting time random variable $W(\Lambda)$ is finite Markov chain imbeddable. The imbedded Markov chain $\{Y_t\}$ defined on the state space $\Omega(\Lambda)$ given by Eq. (27) has*

(i) the initial distribution, at $t = r$,

$$\boldsymbol{\xi}_r = (\boldsymbol{\pi} : \mathbf{0}),$$

where $\boldsymbol{\xi}_r$ is a $1 \times (m^r + k - r)$ vector, $\boldsymbol{\pi} = (\pi_{\boldsymbol{x}}; \boldsymbol{x} \in \mathcal{S}^r)$ is a $1 \times m^r$ row vector, and $\mathbf{0} = (0, \dots, 0)$ is a $1 \times (k-r)$ row vector,

(ii) the imbedded Markov chain $Y_t = \langle Y_{t-1}, X_t\rangle_{\Omega(\Lambda)}$, $t = r+1, \dots$, has the transition probability matrix

$$\boldsymbol{M} = (p_{\boldsymbol{xy}}) = \begin{array}{c} \Omega(\Lambda) - \alpha \\ \alpha \end{array} \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \mathbf{0} & 1 \end{array}\right),$$

where the transition probabilities of the Markov chain $\{Y_t\}$ are given by

$$p_{\boldsymbol{xy}} = \begin{cases} p_{L_r(\boldsymbol{x})\bullet b} & \text{if } \boldsymbol{x} \neq \alpha,\ \boldsymbol{y} = \langle \boldsymbol{x}, b\rangle_{\Omega(\Lambda)} \text{ and } b \in \mathcal{S} \\ 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} = \alpha \\ 0 & \text{otherwise,} \end{cases} \tag{29}$$

with $L_r(\boldsymbol{x})$ representing the last r symbols of $\boldsymbol{x}$, and

(iii) for $n \geq k$,

$$P(W(\Lambda) \geq n) = \boldsymbol{\xi}\boldsymbol{N}^{n-r-1}\mathbf{1}', \tag{30}$$

where $\boldsymbol{\xi}$ is equivalent to $\boldsymbol{\xi}_r$ without the last coordinate, i.e. $\boldsymbol{\xi}_r = (\boldsymbol{\xi} : 0)$.

Proof Since the counting toward pattern Λ starts at $t = 1$, then at the time $t = r$, $P(Y_r = \boldsymbol{x}) = \pi_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{S}^r$, where $\pi_{\boldsymbol{x}}$ are the ergodic probabilities on $\mathcal{S}^r$ induced by the r-th order Markov chain $\{X_t\}$. Further, $P(Y_r = \boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathcal{S}_+$, as every state in $\mathcal{S}^r_+$ has length longer than r. Therefore Y_r has the distribution $\boldsymbol{\xi}_r = (\boldsymbol{\pi} : \boldsymbol{0})$ on $\Omega(\Lambda)$. This proves Result (i).

Note that α is an absorbing state, *i.e.* if $Y_i = \alpha$ then $Y_j = \alpha$ for all $j \geq i$ and $p_{\alpha\alpha} = 1$. Since $\{X_t\}$ is a sequence of r-th order Markov dependent trials, and since every $\boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\}$ has length longer than or equal to r, then for $X_t = b$, $b \in \mathcal{S}$ and $\boldsymbol{y} = \langle \boldsymbol{x}, b\rangle_{\Omega(\Lambda)}$, the transition probability from $\boldsymbol{x}$ to $\boldsymbol{y}$ has to be $p_{L_r(\boldsymbol{x})\bullet b}$. Further, if $\boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\}$ and if $\boldsymbol{y} \neq \langle \boldsymbol{x}, b\rangle_{\Omega(\Lambda)}$ for any $b \in \mathcal{S}$, then there is zero probability to transition from state $\boldsymbol{x}$ to state $\boldsymbol{y}$. This proves Result (ii). Since $\{Y_t\}$ forms a homogeneous Markov chain with transition probability matrix

$$\boldsymbol{M} = (p_{\boldsymbol{x}\boldsymbol{y}}) = \begin{array}{c} \Omega(\Lambda) - \alpha \\ \alpha \end{array} \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & 1 \end{array}\right) \tag{31}$$

and initial distribution $\boldsymbol{\xi}_r = (\boldsymbol{\xi} : 0)$ at $t = r$, Result (iii) is an immediate consequence of the identity

$$P(W(\Lambda) \geq n) = P(Y_{n-1} \in \Omega(\Lambda) - \{\alpha\}),$$

the Chapman–Kolmogorov equation, and the structure of Eq. (31). This completes the proof. □

4 Waiting time of a compound pattern

The results in Theorems 3.1 and 3.2 apply to the waiting time of a simple pattern, and can be extended to the waiting time of a compound pattern $\Lambda = \cup_{i=1}^{l} \Lambda_i$, a union of l distinct simple patterns. Before we state our general results for the waiting time distribution of a compound pattern, we again would like to first illustrate the construction of imbedded Markov chain $\{Y_t\}$ by providing the following example.

Example 2 Let $\{X_t\}$ be a sequence of 3rd ($r = 3$) order Markov dependent two-state trials. Consider a compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ generated by two distinct simple patterns $\Lambda_1 = SS$ and $\Lambda_2 = FFFFF$ with lengths $k_1 = 2$ and $k_2 = 5$, respectively. Since $k_1 < r < k_2$, following Theorems 3.1 and 3.2, we define

$$\begin{aligned} &\mathcal{S}^3 = \{FFF, FFS, FSF, FSS, SFF, SFS, SSF, SSS\} \\ &\mathcal{S}^3_2(\Lambda_1) = \{SSS, SSF\}, \quad \mathcal{S}^3_3(\Lambda_1) = \{FSS\}, \\ &\text{and } \mathcal{S}^3_+(\Lambda_2) = \{FFFF\}. \end{aligned} \tag{32}$$

Further, we define

$$\alpha_1 \equiv \{SSS, SSF, FSS\} \text{ and } \alpha_2 \equiv \{FFFFF\} \tag{33}$$

as absorbing states for the state space $\Omega(\Lambda)$ defined by

$$\begin{aligned} \Omega(\Lambda) &= \{\mathcal{S}^3 - \cup_{j=2}^{3} \mathcal{S}^3_j(\Lambda_1)\} \cup \{\alpha_1, \alpha_2\} \cup \mathcal{S}^3_+(\Lambda_2) \\ &= \{FFF, FFS, FSF, SFF, SFS, FFFF, \alpha_1, \alpha_2\}. \end{aligned} \tag{34}$$

Again, for each $\boldsymbol{x} \in \Omega(\Lambda)$ and $b \in \mathcal{S}$, we define, similarly to Eq. (28), a mapping $\langle ., . \rangle_{\Omega(\Lambda)}$: $\Omega(\Lambda) \times \mathcal{S} \to \Omega(\Lambda)$ as

$$\langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)} = \begin{cases} \alpha_i & \text{if } \boldsymbol{x} = \alpha_i, i = 1, 2, \text{ and } b \in \mathcal{S} \\ \boldsymbol{y} & \text{if } \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha_1, \alpha_2\}, \end{cases} \tag{35}$$

where $\boldsymbol{y}$ is the longest subsequence of $\{x_1 x_2 \ldots x_l b, x_2 x_3 \ldots x_l b, \ldots, x_{l-r+2} \ldots x_l b\}$ in $\Omega(\Lambda)$ and $r \leq l \leq \max(k_1, k_2)$.

The imbedded Markov chain $\{Y_t\}$ associated with the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ is defined as

$$Y_t = \langle Y_{t-1}, X_t \rangle_{\Omega(\Lambda)}, \quad \text{for } t \geq r + 1, \tag{36}$$

with transition probability matrix

$$\boldsymbol{M} = \begin{array}{r} FFF \\ FFS \\ FSF \\ SFF \\ SFS \\ FFFF \\ \alpha_1 \\ \alpha_2 \end{array} \left(\begin{array}{cccccc|cc} 0 & p_{FFF\cdot S} & 0 & 0 & 0 & p_{FFF\cdot F} & 0 & 0 \\ 0 & 0 & p_{FFS\cdot F} & 0 & 0 & 0 & p_{FFS\cdot S} & 0 \\ 0 & 0 & 0 & p_{FSF\cdot F} & p_{FSF\cdot S} & 0 & 0 & 0 \\ p_{SFF\cdot F} & p_{SFF\cdot S} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{SFS\cdot F} & 0 & 0 & 0 & p_{SFS\cdot S} & 0 \\ 0 & p_{FFF\cdot S} & 0 & 0 & 0 & 0 & 0 & p_{FFF\cdot F} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

$$= \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{O} & \boldsymbol{I} \end{array} \right). \tag{37}$$

Given the parameters $\{p_{\boldsymbol{x} \bullet b} : \boldsymbol{x} \in \mathcal{S}^r \text{ and } b \in \mathcal{S}\}$, the initial distribution is

$$\boldsymbol{\xi}_3 = (\pi_{FFF}, \pi_{FFS}, \pi_{FSF}, \pi_{SFF}, \pi_{SFS}, 0, \pi_{SSS} + \pi_{SSF} + \pi_{FSS}, 0) \tag{38}$$

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

$$P(W(\Lambda) \geq n) = (\pi_{FFF}, \pi_{FFS}, \pi_{FSF}, \pi_{SFF}, \pi_{SFS}, 0) \boldsymbol{N}^{n-4} (1, 1, 1, 1, 1, 1)'. \tag{39}$$

Remark 2 The two absorbing states α_1 and α_2 could be grouped together as one single absorbing state α. This will not affect the computation of the distribution of the waiting time random variable $W(\Lambda)$. In this case, the system first enters the state α when the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ has occurred.

Let $\Lambda_1, \ldots, \Lambda_m$ be m distinct simple patterns with lengths $k_1, \ldots, k_m$, respectively, and let $\Lambda = \cup_{i=1}^m \Lambda_i$ be the compound pattern generated by the distinct simple patterns Λ_i. We assume that $k_1, \ldots, k_g \leq r < k_{g+1}, \ldots, k_m$. We define

(i) for $i = 1, \ldots, g$,

$$\begin{aligned} \mathcal{S}_-^r(\Lambda_i) &= \{\boldsymbol{x} : \boldsymbol{x} \in \mathcal{S}^r \text{ and pattern } \Lambda_i \text{ occurred in the first } r \text{ trials }\} \\ &= \cup_{j=k_i}^r \mathcal{S}_j^r(\Lambda_i), \end{aligned} \tag{40}$$

and

$$\mathcal{S}_-^r(\Lambda) = \cup_{i=1}^g \mathcal{S}_-^r(\Lambda_i), \tag{41}$$

(ii) for $i = g+1, \dots, m$,

$$
\begin{aligned}
\mathcal{S}_+^r(\Lambda_i) &= \{\text{all the sequential subpatterns of } \Lambda_i \text{ with} \\
&\qquad \text{length greater than } r \text{ but less than } k_i\} \\
&= \{b_{i_1} \dots b_{i_{r+1}}, \dots, b_{i_1} \dots b_{i_{k_i-1}}\},
\end{aligned} \tag{42}
$$

and

$$
\mathcal{S}_+^r(\Lambda) = \cup_{i=g+1}^m \mathcal{S}_+^r(\Lambda_i), \tag{43}
$$

(iii)

$$
\mathcal{S}_\alpha = \mathcal{S}_-^r(\Lambda) \cup \{\Lambda_{g+1}, \dots, \Lambda_m\}, \tag{44}
$$

and the state space

$$
\Omega(\Lambda) = \{\mathcal{S}^r - \mathcal{S}_-^r(\Lambda)\} \cup \mathcal{S}_+^r(\Lambda) \cup \{\alpha\}, \tag{45}
$$

where all $\boldsymbol{x} \in \mathcal{S}_\alpha$ are grouped as an absorbing state α.

Further, we define the mapping $\langle ., . \rangle_{\Omega_\Lambda} : \Omega(\Lambda) \times \mathcal{S} \to \Omega(\Lambda)$ for the compound pattern Λ, analogous to Eqs. (28) and (35), as

$$
\langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)} = \begin{cases} \alpha & \text{if } \boldsymbol{x} = \alpha \text{ and all } b \in \mathcal{S} \\ \boldsymbol{y} & \text{if } \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\}, \end{cases} \tag{46}
$$

where $\boldsymbol{y}$ is the longest subsequence of $\{x_1 x_2 \dots x_l b, x_2 x_3 \dots x_l b, \dots, x_{l-r+2} \dots x_l b\}$ in $\Omega(\Lambda)$ and $r \leq l \leq \max(k_1, \dots, k_m)$. It follows from Eqs. (40) to (46) that the imbedded homogeneous Markov chain $\{Y_t\}_r^\infty$ may be defined as

$$
Y_t = \langle Y_{t-1}, X_t \rangle_{\Omega(\Lambda)}
$$

on the state space $\Omega(\Lambda)$ with the transition probability matrix

$$
\boldsymbol{M} = (p_{\boldsymbol{xy}}) = \begin{array}{c} \Omega(\Lambda) - \alpha \\ \alpha \end{array} \left(\begin{array}{c|c} \boldsymbol{N} & \boldsymbol{C} \\ \hline \boldsymbol{0} & 1 \end{array} \right), \tag{47}
$$

where the transition probabilities $P(Y_t = \boldsymbol{y} | Y_{t-1} = \boldsymbol{x}) = p_{\boldsymbol{xy}}$ are defined by the following equation: for $\boldsymbol{x}, \boldsymbol{y} \in \Omega(\Lambda)$,

$$
p_{\boldsymbol{xy}} = \begin{cases} p_{L_r(\boldsymbol{x}) \bullet b} & \text{if } \boldsymbol{x} \neq \alpha, \, \boldsymbol{y} = \langle \boldsymbol{x}, b \rangle_{\Omega(\Lambda)}, \, b \in \mathcal{S} \\ 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} = \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{48}
$$

Note again that the state Y_t contains implicitly two pieces of essential information: (1) the longest sequential subpattern of Λ at time t, and (2) the r-th order Markov dependent sequence at time t. In view of our construction of the imbedded Markov chain and Eqs. (40, 41, 42, 43, 44, 45, 46, 47, 48), the following theorem holds.

Theorem 4.1 *If $\{X_t\}_{-\infty}^{\infty}$ is a sequence of irreducible, aperiodic and homogeneous r-th order Markov dependent m-state trials, and $\Lambda = \cup_{i=1}^{m}\Lambda_i$ is a compound pattern generated by m simple patterns, then the imbedded Markov chain $\{Y_t\}_r^{\infty}$ of the waiting time $W(\Lambda)$, defined on the state space $\Omega(\Lambda)$ given by Eq. (45), has*

(i) the initial distribution, at $t = r$,

$$\begin{aligned}\boldsymbol{\xi}_r &= (\xi(\boldsymbol{x}), \boldsymbol{x} \in \Omega(\Lambda))\\ &= ((\xi(\boldsymbol{x}), \boldsymbol{x} \in \{\mathcal{S}^r - \mathcal{S}_-^r(\Lambda)\}) : (\xi(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_+^r(\Lambda)) : (\xi(\boldsymbol{x}), \boldsymbol{x} = \alpha)),\end{aligned}$$

where

$$\xi(\boldsymbol{x}) = \begin{cases} \pi_{\boldsymbol{x}} & \text{if } \boldsymbol{x} \in \{\mathcal{S}^r - \mathcal{S}_-^r(\Lambda)\} \\ 0 & \text{if } \boldsymbol{x} \in \mathcal{S}_+^r(\Lambda) \\ \sum_{\boldsymbol{x}\in\mathcal{S}_-^r(\Lambda)} \pi_{\boldsymbol{x}} & \text{if } \boldsymbol{x} = \alpha, \end{cases} \tag{49}$$

with the transition probability matrix $\boldsymbol{M}$ given by Eqs. (47) and (48), and

(ii) for $n \geq r + 1$,

$$P(W(\Lambda) \geq n) = \boldsymbol{\xi}\boldsymbol{N}^{n-r-1}\mathbf{1}', \tag{50}$$

where $\boldsymbol{\xi} = (\xi(\boldsymbol{x}), \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha\})$.

Proof The proof is very similar to the proofs of Theorems 3.1 and 3.2. The results are immediate consequences of our construction of Eqs. (40, 41, 42, 43, 44, 45, 46, 47, 48). We leave the details for the reader. □

Further, if we are also interested in knowing the individual probabilities $P(W(\Lambda) = W(\Lambda_i) = n)$ for $i = 1, \ldots, m$, we have to set $\Lambda_1 = \alpha_1, \ldots, \Lambda_m = \alpha_m$ as absorbing states, as in Example 4.1. We expect the following result holds.

Theorem 4.2 *Given $\Lambda = \cup_{i=1}^{m}\Lambda_i$, then for $n \geq r + 1$,*

$$P(W(\Lambda) = W(\Lambda_i) = n) = \boldsymbol{\xi}\boldsymbol{N}^{n-r-1}\boldsymbol{C}'(\alpha_i), \tag{51}$$

where $\boldsymbol{N}$ is defined by Eq. (47) and $\boldsymbol{C}'(\alpha_i) = (p_{\boldsymbol{x}\alpha_i} : \boldsymbol{x} \in \Omega(\Lambda) - \{\alpha_1, \ldots, \alpha_m\})'$, the i-th column of matrix $\boldsymbol{C}$.

Proof For every $1 \leq i \leq m$, it follows from Theorem 4.1 that

$$\begin{aligned}P(W(\Lambda) = W(\Lambda_i) = n) &= \sum_{\mathrm{x}\in\Omega(\Lambda)-\{\alpha_1,\ldots,\alpha_l\}} P(Y_{n-1} = \mathrm{x})P(Y_n = \alpha_i | Y_{n-1} = \mathrm{x})\\ &= \sum_{\mathrm{x}\in\Omega(\Lambda)-\{\alpha_1,\ldots,\alpha_l\}} \boldsymbol{\xi}\boldsymbol{N}^{n-r-1} p_{\mathrm{x}\alpha_i}\mathrm{e}_{\mathrm{x}}'\\ &= \boldsymbol{\xi}\boldsymbol{N}^{n-r-1}\boldsymbol{C}'(\alpha_i),\end{aligned}$$

where $\mathrm{e}_{\mathrm{x}} = (0, \ldots, 0, 1, 0, \ldots, 0)$ is a unit vector associated with $\boldsymbol{x}$. □

5 Generating functions, spectrum analysis, and large deviation approximation

Let $\lambda_{[1]}, \ldots \lambda_{[d]}$ be the ordered eigenvalues of the essential transition probability matrix $\boldsymbol{N}$, such that $|\lambda_{[1]}| \geq |\lambda_{[2]}| \geq \cdots \geq |\lambda_{[d]}|$, and let $\boldsymbol{\eta}'_{[1]}, \ldots, \boldsymbol{\eta}'_{[d]}$ be the column eigenvectors corresponding to the eigenvalues $\lambda_{[i]}$, $i = 1, \ldots, d$, respectively. Note that since $\boldsymbol{N}$ is a substochastic matrix and since its eigenspace has the same dimension as $\boldsymbol{NB}>$, it follows from the Perron–Frobenius Theorem (see Seneta 1981) that the largest eigenvalue $\lambda_{[1]}$ is unique and $1 > \lambda_{[1]} > |\lambda_{[2]}| \geq 0$. Define

$$\Phi_{W(\Lambda)}(s) = \sum_{n=r+1} s^n P(W(\Lambda) \geq n), \quad \text{for } |s| < 1/\lambda_{[1]}.$$

The following general results hold for the probability generating function of the waiting time distribution of a compound pattern $\Lambda = \cup_{i=1}^m \Lambda_i$.

Theorem 5.1 *Assume that $\{X_t\}_{-\infty}^{\infty}$ is a sequence of irreducible, aperiodic and homogeneous r-th order ($r \geq 1$) Markov dependent multi-state trials with transition probabilities $p_{\boldsymbol{x}\bullet b}$, $\boldsymbol{x} \in \mathcal{S}^r$ and $b \in \mathcal{S}$, and that $\Lambda = \cup_{i=1}^m \Lambda_i$ is a compound pattern generated by m distinct simple patterns Λ_i with lengths $k_1, \ldots, k_g \leq r < k_{g+1}, \ldots, k_m$. Then the generating function of the waiting time distribution of $W(\Lambda)$ is*

$$\varphi_{W(\Lambda)}(s) = \sum_{j=1}^{r} s^j \sum_{\boldsymbol{x} \in \cup_{i=1}^g \mathcal{S}_j^r(\Lambda_i)} \pi_{\boldsymbol{x}} + s^r \boldsymbol{\xi} \mathbf{1}' + \left(1 - \frac{1}{s}\right) \Phi_{W(\Lambda)}(s), \tag{52}$$

where $\pi_{\boldsymbol{x}}$, $\boldsymbol{x} \in \mathcal{S}^r$, are ergodic probabilities defined by Lemma 3.1, $\boldsymbol{\xi}$ is defined by Eq. (20),

$$\Phi_{W(\Lambda)}(s) = \sum_{i=1}^{d} \frac{c_i s^{r+1} \boldsymbol{\xi} \boldsymbol{\eta}'_{[i]}}{1 - \lambda_{[i]} s}, \quad \textit{for } |s| < 1/\lambda_{[1]}, \tag{53}$$

and c_i, for $i = 1, 2, \ldots, d$, are the coefficients of $\mathbf{1}' = \sum_{i=1}^d c_i \boldsymbol{\eta}'_{[i]}$.

Note that, if $\{X_t\}$ is a sequence of $i.i.d.$ trials ($r = 0$), it is easy to see that

$$\sum_{j=1}^{r} s^j \sum_{\boldsymbol{x} \in \cup_{i=1}^g \mathcal{S}_j^r(\Lambda_i)} \pi_{\boldsymbol{x}} \equiv 0, \tag{54}$$

and

$$s^r \boldsymbol{\xi} \mathbf{1}' \equiv 1. \tag{55}$$

The generating function $\varphi_{W(\Lambda)}(s)$ in Eq. (52) then reduces to the well-known formula (see Fu and Lou 2003)

$$\varphi_{W(\Lambda)}(s) = 1 + \left(1 - \frac{1}{s}\right) \Phi_{W(\Lambda)}(s). \tag{56}$$

Proof Since $\mathbf{1}'$ can be represented as a linear combination of eigenvectors, i.e. $\mathbf{1}' = \sum_{i=1}^{d} c_i \boldsymbol{\eta}'_{[i]}$, it follows from the definition of $\Phi_{W(\Lambda)}(s)$, Eq. (20), and Theorem 4.1 that

$$\begin{aligned}
\Phi_{W(\Lambda)}(s) &= \sum_{n=r+1}^{\infty} s^n P(W(\Lambda) \geq n) \\
&= s^{r+1} \sum_{n=r+1}^{\infty} s^{n-r-1} \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \left(\sum_{i=1}^{d} c_i \boldsymbol{\eta}'_{[i]} \right) \\
&= \sum_{i=1}^{d} s^{r+1} \sum_{n=r+1}^{\infty} c_i (s\lambda_{[i]})^{n-r-1} \boldsymbol{\xi} \boldsymbol{\eta}'_{[i]} \\
&= \sum_{i=1}^{d} \frac{c_i s^{r+1} \boldsymbol{\xi} \boldsymbol{\eta}'_{[i]}}{1 - \lambda_{[i]} s}.
\end{aligned} \tag{57}$$

The existence of Eq. (57) requires the convergence of all power series $\sum_{n=r+1}^{\infty} (s\lambda_{[i]})^{n-r-1}$, $i = 1, \ldots, d$, and that requires $|s| < 1/\lambda_{[1]}$. This completes the proof of Eq. (53). It follows from the definition of $\varphi_{W(\Lambda)}(s)$ and Theorem 4.1 that

$$\begin{aligned}
\varphi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n P(W(\Lambda) = n) \\
&= \sum_{n=1}^{r} s^n P(W(\Lambda) = n) + \sum_{n=r+1}^{\infty} s^n P(W(\Lambda) = n),
\end{aligned} \tag{58}$$

where

$$\sum_{n=1}^{r} s^n P(W(\Lambda) = n) = \sum_{n=1}^{r} s^n \sum_{\boldsymbol{x} \in \cup_{i=1}^{g} \mathcal{S}_n^r(\Lambda_i)} \pi_{\boldsymbol{x}}, \tag{59}$$

and

$$\begin{aligned}
\sum_{n=r+1}^{\infty} s^n P(W(\Lambda) = n) &= \sum_{n=r+1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} (\boldsymbol{I} - \boldsymbol{N}) \mathbf{1}' \\
&= \sum_{n=r+1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathbf{1}' - \sum_{n=r+1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-r} \mathbf{1}' \\
&= \sum_{n=r+1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathbf{1}' - \frac{1}{s} \sum_{n=r+1}^{\infty} s^n \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathbf{1}' + s^r \boldsymbol{\xi} \mathbf{1}'.
\end{aligned} \tag{60}$$

The result in Eq. (52) follows immediately from Eqs. (58), (59), and (60). This completes the proof. □

Theorem 5.2 *The mean waiting time of $W(\Lambda)$ is given by*

$$EW(\Lambda) = \sum_{j=1}^{r} P(W(\Lambda) \geq j) + \sum_{i=1}^{w} S_i, \tag{61}$$

where $P(W(\Lambda) \geq 1) \equiv 1$,

$$P(W(\Lambda) \geq j) = 1 - \sum_{\mathbf{x} \in \cup_{i=1}^{j-1} \mathcal{S}_j^r} \pi_{\mathbf{x}}, \quad j = 2, \ldots, r, \tag{62}$$

and $(S_1, \ldots, S_w)$ is the solution of the simultaneous equations

$$S_i = \boldsymbol{\xi} e_i' + (S_1, \ldots, S_w)\boldsymbol{N}\mathrm{e}_i', \quad i = 1, \ldots, w, \tag{63}$$

where w is the size of the matrix $\boldsymbol{N}$.

Proof Since $EW(\Lambda) = \sum_{n=1}^{r} P(W(\Lambda) \geq n) + \sum_{n=r+1}^{\infty} P(W(\Lambda) \geq n)$, Eq. (62) follows immediately from the definitions of $P(W(\Lambda) \geq n)$ and $\pi_{\mathbf{x}}$. Note that

$$\sum_{n=r+1}^{\infty} P(W(\Lambda \geq n) = \sum_{i=1}^{w} \sum_{n=r+1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathrm{e}_i' = \sum_{i=1}^{w} S_i,$$

where $S_i = \sum_{n=r+1}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-r-1} \mathrm{e}_i'$, which can be expressed as, for $i = 1, \ldots, w$,

$$\begin{aligned} S_i &= \boldsymbol{\xi} \mathrm{e}_i' + \sum_{n=r+2}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-r-2} (\boldsymbol{N} \mathrm{e}_i') \\ &= \boldsymbol{\xi} \mathrm{e}_i' + \sum_{j=1}^{w} p_{ji} \sum_{n=r+2}^{\infty} \boldsymbol{\xi} \boldsymbol{N}^{n-r-2} \mathrm{e}_j' \\ &= \boldsymbol{\xi} \mathrm{e}_i' + (S_1, \ldots, S_w) \boldsymbol{N} e_i'. \end{aligned}$$

This completes the proof. □

Remark 3 The function $\Phi_{W(\Lambda)}(s)$ can also be obtained directly via the essential transition probability matrix $\boldsymbol{N}$ by using the method of Fu and Chang (2002):

$$\Phi_{W(\Lambda)}(s) = \phi_1(s) + \ldots + \phi_w(s),$$

where $(\phi_1(s), \phi_2(s), \ldots, \phi_w(s))$ is the solution of the simultaneous equations

$$\phi_i(s) = s^{r+1} \boldsymbol{\xi} e_i' + s(\phi_1(s), \phi_2(s), \ldots, \phi_w(s)) \boldsymbol{N} \mathrm{e}_i',$$

for $i = 1, \ldots, w$.

The tail probability $P(W(\Lambda) \geq n)$ can also be approximated via the probability of large deviation in terms of the largest eigenvalue and its corresponding eigenvector.

Theorem 5.3 *Under the assumptions of Theorem 5.1, we have*

$$(i) \lim_{n\to\infty} \frac{1}{n} \log P(W(\Lambda) \geq n) = -\beta, \tag{64}$$

where $\beta = -\log \lambda_{[1]}$, *and*

$$(ii) \lim_{n\to\infty} \frac{P(W(\Lambda) \geq n)}{C^* \lambda_{[1]}^n} = 1, \tag{65}$$

where $C^* = c_1(\boldsymbol{\xi}\boldsymbol{\eta}'_{[1]})\lambda_{[1]}^{-r-1}$.

In view of Theorems 5.1 and 5.3, the waiting time random variable $W(\Lambda)$ has a general geometric distribution characterized by the essential transition probability matrix $\boldsymbol{N}$, and for large n, its tail probability $P(W(\Lambda) \geq n)$ depends mainly on the largest eigenvalue $\lambda_{[1]}$ and its corresponding eigenvector $\boldsymbol{\eta}'_{[1]}$:

$$P(W(\Lambda) \geq n) \sim C^* \exp\{-n\beta\}, \tag{66}$$

where $\beta = -\log \lambda_{[1]}$ and $C^* = c_1(\boldsymbol{\xi}\boldsymbol{\eta}'_{[1]})\lambda_{[1]}^{-r-1}$. Hence the tail probability $P(W(\Lambda) \geq n)$ converges to zero exponentially with a rate constant $\beta = -\log \lambda_{[1]}$. In view of Theorem 5.3 (ii), numerically we expect the large deviation approximation in Eq. (66) to perform well for moderate and large n; this can be seen in our numerical examples presented in Sect. 6.

Proof (of Theorem 5.3) Since the essential transition probability matrix $\boldsymbol{N}$ is a substochastic matrix, it follows from the Perron-Frobenius Theorem that the largest eigenvalue, $\lambda_{[1]} > |\lambda_{[2]}| \geq \ldots \geq |\lambda_{[d]}|$, is unique with $1\rangle\lambda_{[1]}\rangle 0$. Note that $\mathbf{1}' = \sum_{i=1}^d c_i \boldsymbol{\eta}'_{[i]}$. For $n \geq r+1$, we have

$$\begin{aligned} P(W(\Lambda) \geq n) &= \boldsymbol{\xi}\boldsymbol{N}^{n-r-1}\mathbf{1}' \\ &= \sum_{i=1}^{d} c_i \lambda_{[i]}^{n-r-1} \boldsymbol{\xi}\boldsymbol{\eta}'_{[i]} \\ &= C^* \lambda_{[1]}^n \{1 + O((\frac{|\lambda_{[2]}|}{\lambda_{[1]}})^n)\}, \end{aligned} \tag{67}$$

where $C^* = c_1(\boldsymbol{\xi}\boldsymbol{\eta}'_{[1]})\lambda_{[1]}^{-r-1}$. Since d is fixed, and $(\frac{|\lambda_{[2]}|}{\lambda_{[1]}})^n \to 0$ exponentially, Result (i) is obtained directly by taking logarithms and dividing by n on both sides of Eq. (67), followed by letting $n \to \infty$. Result (ii) is also an immediate consequence of Eq. (67). This completes the proof. □

Remark 4 The coefficients c_i, $i = 1, \ldots, d$, can be computed by

$$c_i = e_i[\boldsymbol{\eta}'_{[1]}, \ldots, \boldsymbol{\eta}'_{[d]}]^{-1}\mathbf{1}', \tag{68}$$

and can be used directly in Eq. (67) to obtain the exact probabilities for $P(W(\Lambda) \geq n)$.

6 Numerical example and discussion

We will use Example 4.1 to illustrate our theoretical results, to check the efficiency of the computations, and to confirm the accuracy of the large deviation approximation.

Let $\{X_t\}$ be a sequence of 3rd order Markov dependent two-state trials with transition probabilities $p_{FFF\cdot F} = 0.1$, $p_{FFF\cdot S} = 0.9$, $p_{FFS\cdot F} = 0.15$, $p_{FFS\cdot S} = 0.85$, $p_{FSF\cdot F} = 0.2$, $p_{FSF\cdot S} = 0.8$, $p_{FSS\cdot F} = 0.25$, $p_{FSS\cdot S} = 0.75$, $p_{SFF\cdot F} = 0.3$, $p_{SFF\cdot S} = 0.7$, $p_{SFS\cdot F} = 0.35$, $p_{SFS\cdot S} = 0.65$, $p_{SSF\cdot F} = 0.4$, $p_{SSF\cdot S} = 0.6$, $p_{SSS\cdot F} = 0.45$, and $p_{SSS\cdot S} = 0.55$. Using the equations $\boldsymbol{\pi A} = \boldsymbol{\pi}$ and $\pi_{FFF} + \cdots + \pi_{SSS} = 1$ gives the ergodic probabilities $\pi_{FFF} = 0.0262$, $\pi_{FFS} = 0.0786$, $\pi_{FSF} = 0.0643$, $\pi_{FSS} = 0.1643$, $\pi_{SFF} = 0.0786$, $\pi_{SFS} = 0.15$, $\pi_{SSF} = 0.1643$, and $\pi_{SSS} = 0.2738$.

For the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$, with $\Lambda_1 = SS$ and $\Lambda_2 = FFFFF$, the tail probabilities of the waiting time of the compound pattern Λ are computed through the equation $P(W(\Lambda) \geq n) = \boldsymbol{\xi} \boldsymbol{N}^{n-4} \mathbf{1}'$ for $n = 4, 5, \cdots$, using the essential transition probability matrix $\boldsymbol{N}$ and the initial distribution $\boldsymbol{\xi}$ defined in Example 4.1. For $n = 1, 2$, and 3, the probabilities $P(W \geq n)$ are calculated directly from the ergodic distribution. The numerical results are provided in Fig. 1.

The largest eigenvalue and the corresponding eigenvector of $\boldsymbol{N}$ are $\lambda_{[1]} = 0.5860$ and $\boldsymbol{\eta}_{[1]} = (0.3189, 0.1774, 0.6931, 0.3752, 0.4140, 0.2725)$, respectively. This yields the large deviation approximation

$$P(W(\Lambda) \geq n) \cong 3.0916 \exp\{n \log \lambda_{[1]}\}$$

for large n. Table 1 provides a comparison of exact probabilities versus the large deviation approximation for moderate and large n.

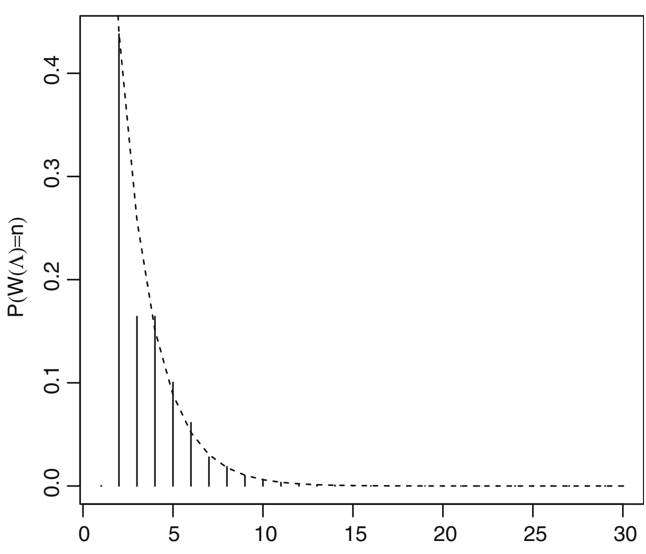

Fig. 1 The exact probabilities (*solid lines*) and the large deviation approximations (*dashed line*) for the waiting time, $P(W(\Lambda) = n)$, of the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ in Example 4.1, where $\Lambda_1 = SS$ and $\Lambda_2 = FFFFF$

Table 1 The exact probabilities and the large deviation approximations for $P(W(\Lambda) \geq n)$ of Example 4.1

n	Exact probability	Large deviation Approximation
10	1.49681e-02	1.47725e-02
20	7.08312e-0.5	7.05871e-05
30	3.37598e-007	3.37286e-007
50	7.70143e-012	7.70092e-012
100	1.91825e-023	1.91825e-023
500	2.84318e-116	2.84318e-116
1000	2.61473e-232	2.61473e-232

The numerical computation of the exact probabilities $P(W(\Lambda) \geq n)$ via the finite Markov chain imbedding approach is rather simple and efficient. In Example 4.1, for example, for the case $n = 1{,}000$, the CPU time is only a fraction of a second on a current PC. The large deviation approximation performs extremely well even when n is moderate ($n = 30$). In the limit of large n, the ratio of the exact and the approximate probabilities tends to one.

For applications where the dimension of the transition probability matrix is large, then, in the numerical evaluations for moderate and large values of n, the computational time can be significantly reduced by taking advantage of the sparse structure of the transition probability matrix (see, for example, the recursive equations presented in Fu and Lou 2003). Alternatively, various numerical methods for obtaining several or all of the eigenvalues of the transition probability matrix, and for calculating high powers of the matrix, can be applied effectively for moderate values of n; for large n, our large deviation approximation should perform well.

Further, for a given pattern Λ (simple or compound) and a complete specification of the transition probabilities $p_{\boldsymbol{x}\bullet b}$, the construction of the state space $\Omega(\Lambda)$ and the transition probability matrix $\boldsymbol{M}$ of the imbedded Markov chain $\{Y_t\}$ can be fully automated, based on Eqs. (22) and (29) or (48). Hence, obtaining the distribution of patterns given by Eqs. (30) and (50) also can be fully automated. Loosely speaking, all the results may be viewed as an extension of the application of the forward and backward principle (Fu 1996) for r-th order Markov dependent multi-state trials.

The extension of the results in Sects. 3, 4 and 5, for obtaining the distribution of the random variable $W(l, \Lambda)$, the waiting time of l patterns Λ (under non-overlap counting), requires only the simple modification of adding an additional coordinate to the state for the purpose of recording the number of patterns Λ that have occurred. However, to find the probability generating function $\varphi_{W(l,\Lambda)}(s)$ for higher order Markov dependent sequences is somewhat complicated and tedious. We shall not pursue this here.

Acknowledgements This work was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs Program.

References

Aki, S., 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.

Aki, S., Hirano, K. (2004). Waiting time problems for a two-dimensional pattern. *Annals of the Institute of Statistical Mathematics*, *56*, 169–182.
Balakrishnan, N., Koutras, M. V. (2002). *Runs and scans with applications*. New York: John Wiley & Sons.
Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. I, 3rd ed.). New York: Wiley
Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials. *Statistica Sinica*, *6*, 957–974.
Fu, J. C., Koutras, M. V. (1994). Distribution theory of runs: A Markov chain approach. *Journal of the American Statistical Association*, *89*, 1050–1058.
Fu, J. C., Chang, Y. M. (2002). On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials. *Journal of Applied Probability*, *39*, 70–80.
Fu, J. C., Lou, W. Y. W. (2003). *Distribution theory of runs and patterns and its applications: a finite Markov chain imbedding approach.* World Scientific Publishing Co.
Han, Q., Aki, S. (2000a). Sooner and later waiting time problems based on a dependent sequence. *Annals of the Institute of Statistical Mathematics*, *52*, 407–414.
Han, Q., Aki, S. (2000b). Waiting time problems in a two-state Markov chain. *Annals of the Institute of Statistical Mathematics*, *52*, 778–789.
Han, Q., Hirano, K. (2003). Sooner and later waiting time problems for patterns in Markov dependent trials. *Journal of Applied Probability*, *40*, 73–86.
Koutras, M. V. (1997). Waiting time distributions associated with runs of fixed length in two-state Markov chains, *Annals of the Institute of Statistical Mathematics*, *49*, 123–139.
Koutras, M. V., Alexandrou, V. (1995). Runs, scans and urn model distributions: a unified Markov chain approach, *Annals of the Institute of Statistical Mathematics*, *47*, 743–766.
Koutras, M. V., Alexandrou, V. A. (1997). Sooner waiting time problems in a sequence of trinary trials. *Journal of Applied Probability*, *34*, 593–609.
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.
Seneta, E. (1981). *Non-negative matrices and Markov chains*, 2nd edition. Berlin Heidelberg New York: Springer-Verlag