Ann Inst Stat Math (2013) 65:513–527
DOI 10.1007/s10463-012-0381-x

Recursive equations in finite Markov chain imbedding

Yu-Fei Hsieh · Tung-Lung Wu

Received: 3 October 2011 / Revised: 10 August 2012 / Published online: 15 November 2012

Abstract In this paper, recursive equations for waiting time distributions of r-th occurrence of a compound pattern are studied via the finite Markov chain imbedding technique under overlapping and non-overlapping counting schemes in sequences of independent and identically distributed (i.i.d.) or Markov dependent multi-state trials. Using the relationship between number of patterns and r-th waiting time, distributions of number of patterns can also be obtained. The probability generating functions are also obtained. Examples and numerical results are given to illustrate our theoretical results.

Keywords Recursive equation · Simple and compound patterns · Waiting time · Finite Markov chain imbedding · Probability generating function

1 Introduction

In the past three decades, distribution theory of runs and patterns has been studied widely and extensively (Fu 1996; Hirano and Aki 1993; Koutras and Milienos 2012; Chang 2005; Koutras 1997). In particular, waiting time distributions and distributions of number of occurrences of patterns in a sequence of multi-state trials have been applied in various areas, for example reliability (Cui et al. 2010), quality control (Chang and Wu 2011; Fu et al. 2003), DNA sequence analysis (Nuel 2008), nonparametric test (Lou 1996) and Eulerian and Newcomb numbers in combinatorial analysis (Fu et al. 1999).

Y.-F. Hsieh · T.-L. Wu (✉)
Department of Statistics, University of Manitoba, Winnipeg, MBR3T 2N2, Canada
e-mail: umwu48@cc.umanitoba.ca

Y.-F. Hsieh
e-mail: umhsieh3@cc.umanitoba.ca

Recently, two approaches, the finite Markov chain imbedding (FMCI) technique and probability generating function (or double generating function), have been widely used for obtaining the waiting time distributions and distributions of number of occurrences of patterns. The FMCI technique was first proposed by Fu and Koutras (1994). The idea is to turn a statistic of interest into a Markov chain, avoiding the complexity of dealing with the statistic directly. Then the distribution can be obtained using a unified formula based on transition probability matrices of the imbedded Markov chain (see, e.g., Fu and Lou 2003). The generating function approach for waiting time distributions was introduced by Aki (1992). The probability generating function is obtained as a solution of a system of conditional probability generating functions. Since then, the generating function approach has been extended to double generating function and has been developed for general waiting time distributions by many authors (see, e.g., Inoue and Aki 2007, 2009). One advantage of this approach is the efficiency for symbolic computation, while the FMCI technique is powerful for numerical computation.

Many authors have developed various algorithms to accelerate the computational speed (see, e.g., Zhao and Cui 2009). One important stream is to find the recursive formulae which require less memory space. Han and Hirano (2003) derived recursive equations for sooner and later waiting time distributions of two simple patterns using overlapping indicator function. They also obtained the probability generating function for the sooner waiting time distribution of a compound pattern. Inoue and Aki (2005) studied the generalized Pólya model with $m + 1$ different balls in a urn and obtained recursive equations for the probability generating function of joint distribution of various run statistics. Recently, Chang et al. (2012) gave recursive equations for the distributions of number of occurrences of a compound pattern based on the double generating function which can always be expressed as a ratio of two polynomials.

Although there are many works on the recursive equations for waiting time distributions and distributions of number of occurrences of patterns, none of them covers the general settings, namely compound patterns, Makov dependent trials and overlapping and non-overlapping counting schemes. In this manuscript, we derive recursive equations for waiting time distributions and distributions of number of occurrences of a compound pattern under general settings, based on the FMCI technique. In Sect. 2, notations and basic results of the FMCI technique are introduced. In Sect. 3, the recursive equations for distributions of r-th waiting time are derived. The probability generating functions are also given for $r = 1$. An example and numerical results are given in Sect. 4. Summary and discussion are given in Sect. 5.

2 Notations and preliminary results

Let $X_n, n = 1, 2, \dots,$ be a sequence of first-order homogeneous Markov dependent m-state ($m \geq 2$) trials taking values in the set $\Gamma = \{a_1, a_2, \dots, a_m\}$ with initial distribution $P(X_0 = \emptyset) = 1$, and transition probabilities $P(X_1 = a_j | X_0 = \emptyset) = p_j$ and $P(X_n = a_j | X_{n-1} = a_i) = p_{i;j}$, for $n = 2, 3, \dots,$ and $i, j = 1, \dots, m$. The simple and compound patterns studied here are defined as follows:

Definition 1 Let $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$ and Λ is said to be a compound pattern generated by L distinct simple patterns where $\Lambda_i = b_1^i \cdots b_{l_i}^i$ is a simple pattern with $b_j^i \in \Gamma$ for all $j = 1, \ldots, l_i$. The length l_i of Λ_i is fixed and the symbols in a simple pattern are allowed to be repeated.

Let $W(\Lambda_i)$ be the waiting time until the first occurrence of Λ_i, and $W(\Lambda) = \min\{W(\Lambda_i) : i = 1, \ldots, L\}$ be the waiting time until the first occurrence of $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$. For a given integer $r \geq 1$, the random variable $W(r, \Lambda)$ denotes the waiting time until the r-th occurrence of Λ. Let $\Gamma(\Lambda_i)$ denote the set of all subpatterns of Λ_i, excluding Λ_i itself, $\Gamma(\Lambda) = \bigcup_{i=1}^{L} \Gamma(\Lambda_i)$, $\Gamma^+(\Lambda_i) = \Gamma(\Lambda_i) \cup \{\Lambda_i\}$ and $\Gamma^+(\Lambda) = \bigcup_{i=1}^{L} \Gamma^+(\Lambda_i)$. It has been shown that the waiting time variable $W(r, \Lambda)$ is finite Markov chain imbeddable (see, e.g., Fu 1996), and there exists an imbedded Markov chain $\{Y_n\}$ defined on the state space

$$\Omega = \{\emptyset\} \cup \{(\ell, \omega) : \omega \in \Gamma \cup \Gamma^+(\Lambda), \ell = 0, \ldots, r-1\} \cup \{\alpha\}, \tag{1}$$

having transition probability matrix $\boldsymbol{M}_r$ of the form

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

It follows from the FMCI technique (Fu and Lou 2003) that the distribution of $W(r, \Lambda)$ is given by

$$P(W(r, \Lambda) > n) = \boldsymbol{\xi}_0 \boldsymbol{N}_r^n \boldsymbol{1}^{'}, \tag{2}$$

where $\boldsymbol{\xi}_0$ is the initial distribution with $P(Y_0 = \emptyset) = 1$ and $\boldsymbol{1}^{'}$ is a column vector with all elements 1. With minor modification in the transition probability matrix, Eq. (2) holds for both overlapping and non-overlapping counting schemes (see, e.g., Chang 2005).

Given a compound pattern Λ, let $X_n(\Lambda)$ be the number of occurrences of Λ in $X_1, X_2, \ldots, X_n$. It is well known that $X_n(\Lambda)$ and $W(r, \Lambda)$ have the following relationship

$$X_n(\Lambda) < r \quad \text{if and only if} \quad W(r, \Lambda) > n, \tag{3}$$

and the probability $P(X_n(\Lambda) = r)$ can be computed by

$$P(X_n(\Lambda) = r) = P(W(r+1, \Lambda) > n) - P(W(r, \Lambda) > n). \tag{4}$$

3 Recursive equations for distributions of waiting time *W(r, Λ)*

We study the recursive equations for distributions of $W(r, \Lambda)$ under overlapping and non-overlapping counting schemes in this section. Given a compound pattern $\Lambda = \cup_{i=1}^{L} \Lambda_i$, we define $\Lambda_i \setminus j = b_1^i \cdots b_{l_i - j}^i$ by removing the last j elements of Λ_i, and that Λ_i is concatenated to Λ_k is denoted by $\Lambda_i \cdot \Lambda_k$. For example, let $\Lambda_1 = ACT$ and

$\Lambda_2 = GT$, then $\Lambda_1 \setminus 1 = AC$, $\Lambda_1 \setminus 2 = A$ and $\Lambda_1 \cdot \Lambda_2 = ACTGT$. An indicator function is defined by

$$I_{\Lambda_k}(\Lambda_i \setminus j) = \begin{cases} 1 & \text{if } \omega \cdot (\Lambda_i \setminus j) = \Lambda_k,\ \omega \in \Gamma \cup \Gamma(\Lambda), \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Note that the indicator function determines whether one pattern overlaps another. If $I_{\Lambda_k}(\Lambda_i \setminus j) = 1$ or $I_{\Lambda_i}(\Lambda_k \setminus j) = 1$ for some j, then patterns Λ_i and Λ_k have an overlap of $l_i - j$ or $l_k - j$ elements, respectively, and if $I_{\Lambda_k}(\Lambda_i \setminus j) = 0$ and $I_{\Lambda_i}(\Lambda_k \setminus j) = 0$ for all j, then they do not overlap. For example, let $\Lambda_1 = ACT$, $\Lambda_2 = GT$ and $\Lambda_3 = CTTG$, then $I_{\Lambda_1}(\Lambda_3 \setminus 2) = 1$ and $I_{\Lambda_3}(\Lambda_2 \setminus 1) = 1$ show both Λ_1 and Λ_2 overlap Λ_3, but Λ_1 and Λ_2 do not overlap since $I_{\Lambda_1}(\Lambda_2 \setminus j) = 0$ and $I_{\Lambda_2}(\Lambda_1 \setminus j) = 0$ for all possible j.

3.1 Non-overlapping counting scheme

Given a compound pattern $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$, let $\{W(r, \Lambda) = n, LO(\Lambda_i)\}$ denote the event that $W(r, \Lambda) = n$ and the last (r-th) pattern occurring is Λ_i. It follows that the probability $P(W(r, \Lambda) = n)$ can be expressed as

$$P(W(r, \Lambda) = n) = \sum_{i=1}^{L} P(W(r, \Lambda) = n, LO(\Lambda_i)).$$

Under non-overlapping counting, the event $\{W(r, \Lambda) = n, LO(\Lambda_i)\}$ can be considered in the following sense: let us fix Λ_i to be the r-th pattern observed at time n ($X_{n-l_i+1} = b_1^i, \ldots, X_n = b_{l_i}^i$) and the $(r-1)$-th pattern must occur before time $n - l_i + 1$. However, due to the existence of overlapping between and within the simple patterns, the event that the r-th pattern occurs between times $n - l_i + 1$ and $n - 1$ may exist and should be excluded. We give an example to illustrate the above idea.

Example 1 Consider a binary sequence with possible outcomes $\{S, F\}$ and let $\Lambda = \Lambda_1 \cup \Lambda_2$ be a compound pattern generated by $\Lambda_1 = SFS$ and $\Lambda_2 = SS$. For $r = 2$, let us consider the event $\{W(2, \Lambda) = n, LO(\Lambda_1)\}$ where the second pattern Λ_1 occurs at time n. As seen in Fig. 1, Λ_1 is observed at time n ($X_{n-2} = S, X_{n-1} = F, X_n = S$) and the first pattern occurs before $n - 3 + 1$, say at time k. Under this

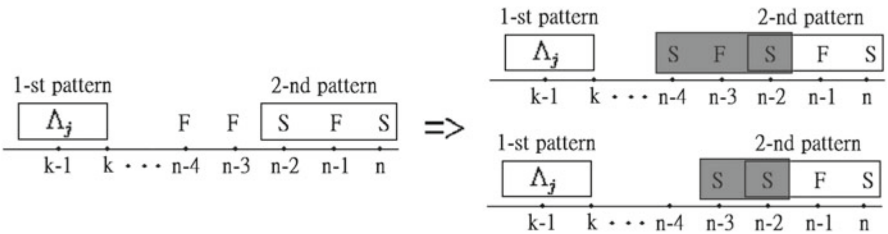

Fig. 1 An illustration of non-overlapping counting scheme

circumstance, if $X_{n-4} = F, X_{n-3} = F$, then the event $\{W(2, \Lambda) = n, LO(\Lambda_1)\}$ occurs; however, if $X_{n-4} = S, X_{n-3} = F$, then the second pattern Λ_1 actually occurs at time $n-2$. Similarly, if $X_{n-3} = S$, then the second pattern Λ_2 actually occurs at time $n-2$. Therefore, those possibilities should be excluded in the consideration of event $\{W(2, \Lambda) = n, LO(\Lambda_1)\}$, since the overlapping part will not be counted toward forming a new pattern under non-overlapping counting. The cases where the actual r-th pattern may occur earlier than time n can be determined by the indicator function $I_{\Lambda_k}(\Lambda_i \setminus j)$.

Throughout this section, the initial condition of the recursive equations for waiting time distributions is $P(W(r, \Lambda) > 0, X_0 = \emptyset) = 1$. To maintain consistent notation, $p_j = P(X_1 = a_j | X_0 = \emptyset)$ is sometimes written as $p_{\emptyset j}$. By convention, we let $P(W(r, \Lambda) = n) = 0$ and $P(W(r, \Lambda) > n, X_n = z) = 0$ if $n < 0$ and $r = 0$. In view of Example 1, by the concept of the forward and backward principle of the FMCI technique, we establish the following theorem.

Theorem 1 *Let $\{X_n\}$ be a sequence of Markov dependent m-state trials and $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$ be a compound pattern. Then the recursive equations for the distribution of $W(r, \Lambda)$ under non-overlapping counting are given by*

$$P(W(r, \Lambda) = n) = \sum_{i=1}^{L} P(W(r, \Lambda) = n, LO(\Lambda_i)),$$

where

$$\begin{aligned}
&P(W(r, \Lambda) = n, LO(\Lambda_i)) \\
&\quad = \sum_{z \in \Gamma} P(W(r, \Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b_1^i} \prod_{j=1}^{l_i - 1} p_{b_j^i; b_{j+1}^i} \\
&\qquad - \sum_{z \in \Gamma} P(W(r-1, \Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b_1^i} \prod_{j=1}^{l_i - 1} p_{b_j^i; b_{j+1}^i} \\
&\qquad - \sum_{j=1}^{l_i - 1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) P(W(r, \Lambda) = n - j, LO(\Lambda_k)) \prod_{t=1}^{j} p_{b_{l_i - t}^i; b_{l_i - t + 1}^i}, \\
&P(W(r, \Lambda) > n - l_i, X_{n-l_i} = z) \\
&\quad = \sum_{y \in \Gamma} P(W(r, \Lambda) > n - l_i - 1, X_{n-l_i-1} = y) p_{y;z} \\
&\qquad - \sum_{k=1}^{L} I_z(b_{l_k}^k) P(W(r, \Lambda) = n - l_i, LO(\Lambda_k)),
\end{aligned}$$

and $I_z(b_{l_k}^k)$ equals 1 if $z = b_{l_k}^k$, and 0 otherwise.

Proof Let $B_i = \prod_{j=1}^{l_i-1} p_{b^i_j;b^i_{j+1}}$, and define the subsets

$$C_\emptyset = \{\emptyset\},$$
$$C_\ell(z) = \{(\ell, w) : w \in \Gamma \cup \Gamma^+(\Lambda) \text{ and } w \text{ ends with element } z\},$$
$$\ell = 0, \dots, r-1,\ z \in \Gamma.$$

It follows from the FMCI technique that

$$\begin{aligned}
&P(W(r,\Lambda) > n - l_i, X_{n-l_i+1} = b^i_1, \dots, X_n = b^i_{l_i}) \\
&= \sum_{z\in\Gamma} P(W(r,\Lambda) > n - l_i, X_{n-l_i} = z) P(X_{n-l_i+1} = b^i_1, \dots, X_n = b^i_{l_i} | X_{n-l_i} = z) \\
&= \sum_{z\in\Gamma} \xi_0 N^{n-l_i} \left[U^{'}(C_0(z)) + \cdots + U^{'}(C_{r-2}(z))\right] p_{z;b^i_1} B_i \\
&\quad + \xi_0 N^{n-l_i} U^{'}(C_{r-1}(z)) p_{z;b^i_1} B_i, \quad i = 1, \dots, L,
\end{aligned}$$

where $U^{'}(C_\ell(z))$ is a column vector $(0, \dots, 0, 1, \dots, 1, 0, \dots, 0)^{'}$ with ones at the locations associated with the states in $C_\ell(z)$. Along with $\{X_{n-l_i+1} = b^i_1, \dots, X_n = b^i_{l_i}\}$, some states in $C_{r-1}(z)$ may comprise a new pattern before time n and some may not (r-th pattern occurs exactly at time n) as illustrated in Fig. 1. Thus, for each Λ_i and $z \in \Gamma$, $C_{r-1}(z)$ can be partitioned into subsets $C^+_{r-1}(z)$ and $C^*_{r-1}(z)$, where $C^+_{r-1}(z)$ stands for the set of states which will comprise a new pattern when combined with subpatterns of $\Lambda_i = b^i_1 \cdots b^i_{l_i}$, and states in $C^*_{r-1}(z)$ will not. Note that $C^+_{r-1}(z)$ and $C^*_{r-1}(z)$ depend on pattern Λ_i but we suppress the index i for simplicity. Then we have

$$\begin{aligned}
&P(W(r,\Lambda) > n - l_i, X_{n-l_i+1} = b^i_1, \dots, X_n = b^i_{l_i}) \\
&\qquad = \sum_{z\in\Gamma} P(W(r-1,\Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b^i_1} B_i \\
&\qquad\quad + \xi_0 N^{n-l_i} \left[U^{'}(C^+_{r-1}(z)) + U^{'}(C^*_{r-1}(z))\right] p_{z;b^i_1} B_i \\
&\qquad = \sum_{z\in\Gamma} P(W(r-1,\Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b^i_1} B_i \\
&\qquad\qquad + P(W(r,\Lambda) = n, LO(\Lambda_i)) \\
&\qquad\quad + \sum_{j=1}^{l_i-1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) P(W(r,\Lambda) = n - j, LO(\Lambda_k)) \prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}}.
\end{aligned}$$

The last term follows from the definition of the indicator function given in Eq. (5). For any $z \in \Gamma$,

$$
\begin{aligned}
P(W(r,\Lambda) > n-l_i, X_{n-l_i}=z) \\
&= \sum_{y\in\Gamma} P(W(r,\Lambda) > n-l_i-1, X_{n-l_i-1}=y)p_{y;z} \\
&\quad - P(W(r,\Lambda) = n-l_i, X_{n-l_i}=z) \\
&= \sum_{y\in\Gamma} P(W(r,\Lambda) > n-l_i-1, X_{n-l_i-1}=y)p_{y;z} \\
&\quad - \sum_{k=1}^{L} I_z(b_{l_k}^k) P(W(r,\Lambda) = n-l_i, LO(\Lambda_k)).
\end{aligned}
$$

This completes the proof. □

Next we study three special cases of our general Theorem 1 for $r = 1$, including compound Makov dependent, compound i.i.d. and simple i.i.d.. The distribution of waiting time until the first occurrence $W(1,\Lambda) \equiv W(\Lambda)$ can be deduced from Theorem 1. For simplicity, let $P(W^*(\Lambda_i) = n) = P(W(\Lambda) = n, LO(\Lambda_i))$. Let $\Phi_{zW(\Lambda)}(s) = \sum_{n=2}^{\infty} s^n P(W(\Lambda) > n-1, X_{n-1} = z)$, and $\varphi_{W(\Lambda)}(s)$ and $\phi_{W^*(\Lambda_i)}(s)$ be the probability generating functions of the sequences $\{P(W(\Lambda) = n)\}_{n=1}^{\infty}$ and $\{P(W^*(\Lambda_i) = n)\}_{n=1}^{\infty}$, respectively. Then we have the following corollary for compound Markov dependent case.

Corollary 1 *Let $\{X_n\}$ be a sequence of Marakov dependent m-state trials and $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$ be a compound pattern, then*

(i) *the recursive equations for the distribution of $W(\Lambda)$ are given by*

$$
P(W(\Lambda) = n) = \sum_{i=1}^{L} P(W^*(\Lambda_i) = n),
$$

where for $i = 1, 2, \ldots, L$,

$$
\begin{aligned}
P(W^*(\Lambda_i) = n) &= \sum_{z\in\Gamma} P(W(\Lambda) > n-l_i, X_{n-l_i}=z)p_{z;b_1^i} B_i \\
&\quad - \sum_{j=1}^{l_i-1}\sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) P(W^*(\Lambda_k) = n-j) \prod_{t=1}^{j} p_{b^i_{l_i-t}; b^i_{l_i-t+1}},
\end{aligned}
$$

and

$$
\begin{aligned}
P(W(\Lambda) > n-l_i, X_{n-l_i}=z) &= \sum_{y\in\Gamma} P(W(\Lambda) > n-l_i-1, X_{n-l_i-1}=y)p_{y;z} \\
&\quad - \sum_{k=1}^{L} I_z(b_{l_k}^k) P(W^*(\Lambda_k) = n-l_i),
\end{aligned}
$$

and

(ii) *the probability generating function for the distribution of* $W(\Lambda)$ *is given by*

$$\varphi_{W(\Lambda)}(s) = \phi_{W^*(\Lambda_1)}(s) + \cdots + \phi_{W^*(\Lambda_L)}(s),$$

where $(\phi_{W^*(\Lambda_1)}(s), \ldots, \phi_{W^*(\Lambda_L)}(s))$ *satisfies the following simultaneous equations, for* $i = 1, \ldots, L$,

$$\begin{aligned}\phi_{W^*(\Lambda_i)}(s) &= s^{l_i} p_{b_1^i} B_i + s^{l_i-1} \sum_{z\in\Gamma} p_{z;b_1^i} B_i \Phi_{zW(\Lambda)}(s) \\ &\quad - \sum_{j=1}^{l_i-1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) s^j \prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}} \phi_{W^*(\Lambda_k)}(s),\end{aligned}$$

where $\Phi_{zW(\Lambda)}(s),\ z \in \Gamma$, *are solutions, in terms of* $\phi_{W^*(\Lambda_k)}(s)$, *of the following simultaneous equations*

$$\Phi_{zW(\Lambda)}(s) = s^2 p_z + \sum_{y\in\Gamma} s p_{y;z} \Phi_{yW(\Lambda)}(s) - \sum_{k=1}^{L} s I_z(b^k_{l_k}) \phi_{W^*(\Lambda_k)}(s).$$

Proof For part (i), it is a direct consequence of Theorem 1. Now we prove part (ii). We have

$$\begin{aligned}\varphi_{W(\Lambda)}(s) &= \sum_{n=1}^{\infty} s^n P(W(\Lambda) = n) = \sum_{n=1}^{\infty} s^n \sum_{i=1}^{L} P(W^*(\Lambda_i) = n) \\ &= \phi_{W^*(\Lambda_1)}(s) + \cdots + \phi_{W^*(\Lambda_L)}(s).\end{aligned}$$

It follows from part (i) that, for $i = 1, \ldots, L$,

$$\begin{aligned}\phi_{W^*(\Lambda_i)}(s) &= \sum_{n=1}^{\infty} s^n P(W^*(\Lambda_i) = n) \\ &= s^{l_i} p_{b_1^i} B_i + \sum_{n=l_i+1}^{\infty} s^n \sum_{z\in\Gamma} P(W(\Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b_1^i} B_i \\ &\quad - \sum_{n=l_i}^{\infty} s^n \sum_{j=1}^{l_i-1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) P(W^*(\Lambda_k) = n - j) \prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}} \\ &= s^{l_i} p_{b_1^i} B_i + s^{l_i-1} \sum_{z\in\Gamma} p_{z;b_1^i} B_i \Phi_{zW(\Lambda)}(s) \\ &\quad - \sum_{j=1}^{l_i-1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) s^j \prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}} \phi_{W^*(\Lambda_k)}(s),\end{aligned}$$

and for each $z \in \Gamma$,

$$\begin{aligned}\Phi_{zW(\Lambda)}(s) &= \sum_{n=2}^{\infty} s^n P(W(\Lambda) > n-1, X_{n-1} = z) \\ &= s^2 p_z + \sum_{n=3}^{\infty} s^n \sum_{y\in\Gamma} P(W(\Lambda) > n-2, X_{n-2} = y) p_{y;z} \\ &\quad - \sum_{k=1}^{L} s I_z(b^k_{l_k}) \sum_{n=1}^{\infty} s^n P(W^*(\Lambda_k) = n) \\ &= s^2 p_z + \sum_{y\in\Gamma} s p_{y;z} \Phi_{yW(\Lambda)}(s) - \sum_{k=1}^{L} s I_z(b^k_{l_k}) \phi_{W^*(\Lambda_K)}(s).\end{aligned}$$

The proof is completed. □

The second special case is when $\{X_n\}$ is a sequence of i.i.d. trials and Λ is a compound pattern. Let $\Phi_{W(\Lambda)}(s) = \sum_{n=1}^{\infty} s^n P(W(\Lambda) > n-1)$ and the relationship between $\varphi_{W(\Lambda)}(s)$ and $\Phi_{W(\Lambda)}(s)$ is given by

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

The following corollary can be derived immediately from Theorem 1 or Corollary 1.

Corollary 2 *Let $\{X_n\}$ be a sequence of i.i.d. m-state trials and $\Lambda = \bigcup_{i=1}^{L} \Lambda_i$ be a compound pattern, then*

(i) *the recursive equations for the distribution of $W(\Lambda)$ are given by*

$$P(W(\Lambda) = n) = \sum_{i=1}^{L} P(W^*(\Lambda_i) = n),$$

where for each $i = 1, 2, \ldots, L$,

$$\begin{aligned}P(W^*(\Lambda_i) = n) &= P(W(\Lambda) > n - l_i) \prod_{t=1}^{l_i} p_{b^i_t} \\ &\quad - \sum_{j=1}^{l_i - 1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) P(W^*(\Lambda_k) = n-j) \prod_{t=1}^{j} p_{b^i_{l_i - t+1}},\end{aligned}$$

and

$$P(W(\Lambda) > n - l_i) = P(W(\Lambda) > n - l_i - 1) - P(W(\Lambda) = n - l_i),$$

and

(ii) *the probability generating function for the distribution of* $W(\Lambda)$ *is given by*

$$\varphi_{W(\Lambda)}(s) = \phi_{W^*(\Lambda_1)}(s) + \cdots + \phi_{W^*(\Lambda_L)}(s),$$

where $(\phi_{W^*(\Lambda_1)}(s), \ldots, \phi_{W^*(\Lambda_L)}(s))$ *satisfies the following simultaneous equations*

$$\begin{aligned}\phi_{W^*(\Lambda_i)}(s) &= s^{l_i-1}\prod_{t=1}^{l_i} p_{b_t^i}\Phi_{W(\Lambda)}(s)\\ &\quad -\sum_{j=1}^{l_i-1}\sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j)s^j \prod_{t=1}^{j} p_{b^i_{l_i-t+1}}\phi_{W^*(\Lambda_k)}(s).\end{aligned}$$

The third and simplest case, where $\{X_i\}$ is a sequence of i.i.d. trials and Λ is a simple pattern, is given in Corollary 3.

Corollary 3 *Let* $\{X_n\}$ *be a sequence of i.i.d.* m*-state trials and* $\Lambda = b_1 \cdots b_l$ *be a simple pattern, then*

(i) *the recursive equation for the distribution of* $W(\Lambda)$ *is given by*

$$\begin{aligned}P(W(\Lambda) = n) &= P(W(\Lambda) > n - l)\prod_{t=1}^{l} p_{b_t}\\ &\quad -\sum_{j=1}^{l-1} I(\Lambda \setminus j)P(W(\Lambda) = n - j)\prod_{t=1}^{j} p_{b_{l-t+1}},\end{aligned}$$

where $I(\Lambda \setminus j) = I_\Lambda(\Lambda \setminus j)$, *and*

(ii) *the probability generating function for the distribution of* $W(\Lambda)$ *has a closed form given by*

$$\varphi_{W(\Lambda)}(s) = \frac{\prod_{t=1}^{l} p_{b_t}s^l}{1 - s + \prod_{t=1}^{l} p_{b_t}s^l - (s-1)\sum_{j=1}^{l-1} I(\Lambda \setminus j)s^j \prod_{t=1}^{j} p_{b_{l-t+1}}}. \quad (6)$$

Remark 1 Our Theorem 2 covers the general case, including r-th occurrence, compound pattern and Markov dependent trails. Han and Hirano (2003) mainly focused on the waiting time distributions of two simple patterns and gave the probability generating function for a compound pattern. Their results can be considered as special cases of our general Theorem 2. In particular, for $r = 1$, i.i.d. and simple pattern case, the probability generating function given in Eq. (6) coincides with the result in Remark 3.1 of Han and Hirano (2003).

3.2 Overlapping counting scheme

Under non-overlapping counting scheme, from the classification of states of the imbedded Markov chain, only states in $\cup_{z\in\Gamma} C^+_{r-1}(z)$ would cause the r-th pattern to occur earlier than time n and only states in $\cup_{z\in\Gamma} C^*_{r-1}(z)$ would cause the r-th pattern to occur at time n, when combined with pattern Λ_i located at time n. However, under overlapping counting scheme, certain states in $\cup_{z\in\Gamma} C_\ell(z)$, possibly for all $\ell = 0, 1, \ldots, r-2$, may cause the r-th pattern to occur earlier, while some states may cause the r-th pattern to occur at time n, when combined with pattern Λ_i located at time n. Thus, for each Λ_i, $C_\ell(z)$, $\ell = 0, \ldots, r-2$ and $C_{r-1}(z)$ can be partitioned as follows:

$$\begin{aligned} C_\ell(z) &= C^-_\ell(z,r) \cup C^*_\ell(z,r) \cup C^+_\ell(z,r),\ \ell = 0, \ldots, r-2,\\ C_{r-1}(z) &= C^*_{r-1}(z,r) \cup C^+_{r-1}(z,r),\ z \in \Gamma, \end{aligned}$$

where $C^+_\ell(z,r)$ consists of states that would lead to the occurrence of r-th pattern before time n when combined with some subpatterns of Λ_i, $C^*_\ell(z,r)$ consists of states that would lead to occurrence of r-th pattern at time n when combined with Λ_i, and $C^-_\ell(z,r)$ is the remaining subset.

We give an example for illustration. Consider a simple pattern $\Lambda = 11111$, $n = 100$ and $r = 5$. Let Λ be fixed and located at time 100. It is easy to see that state $(2, 111) \in C^+_2(1,5)$, at time $t = 95$, combined with the subpattern 1111 cause the fifth pattern to occur earlier at time $t = 99$, and the fifth pattern would occur at $t = 100$ for state (3,1)$\in C^*_3(1,5)$, while these will not happen under non-overlapping counting scheme. In a similar fashion to non-overlapping counting scheme, by classifying the states of the imbedded Markov chain, we establish the following theorem.

Theorem 2 *Let $\{X_n\}$ be a sequence of Marakov dependent m-state trials and $\Lambda = \bigcup_{i=1}^L \Lambda_i$ be a compound pattern. Then the recursive equations for the distribution of $W(r,\Lambda)$ under overlapping counting are given by*

$$P(W(r,\Lambda) = n) = \sum_{i=1}^{L} P(W(r,\Lambda) = n, LO(\Lambda_i)),$$

where

$$\begin{aligned} &P(W(r,\Lambda) = n, LO(\Lambda_i))\\ &\quad = \sum_{z\in\Gamma} P(W(r,\Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b^i_1} B_i\\ &\qquad - \sum_{z\in\Gamma} P(W(r-1,\Lambda) > n - l_i, X_{n-l_i} = z) p_{z;b^i_1} B_i\\ &\qquad - \sum_{j=1}^{l_i-1} \sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j) \big[P(W(r,\Lambda) = n - j, LO(\Lambda_k))\\ &\qquad - \ P(W(r-1,\Lambda) = n - j, LO(\Lambda_k)) \big] \prod_{t=1}^{j} p_{b^i_{l_i-t}; b^i_{l_i-t+1}}, \end{aligned}$$

and

$$
\begin{aligned}
&P(W(r,\Lambda) > n-l_i, X_{n-l_i} = z)\\
&\quad = \sum_{y\in\Gamma} P(W(r,\Lambda) > n-l_i-1, X_{n-l_i-1} = y)p_{y;z}\\
&\qquad - \sum_{k=1}^{L} I_z(b^k_{l_k})P(W(r,\Lambda) = n-l_i, LO(\Lambda_k)).
\end{aligned}
$$

Proof

$$
\begin{aligned}
&P(W(r,\Lambda) > n-l_i, X_{n-l_i+1} = b^i_1, \ldots, X_n = b^i_{l_i})\\
&\quad = \sum_{z\in\Gamma} \xi_0 N^{n-l_i}\left[U^{'}(C_0(z)) + \cdots + U^{'}(C_{r-1}(z))\right] p_{z;b^i_1} B_i\\
&\quad = \sum_{z\in\Gamma} \xi_0 N^{n-l_i}\left[U^{'}(C^-_0(z,r)) + \cdots + U^{'}(C^-_{r-2}(z,r))\right] p_{z;b^i_1} B_i\\
&\qquad + P(W(r,\Lambda) = n, LO(\Lambda_i))\\
&\qquad + \sum_{j=1}^{l_i-1}\sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j)P(W(r,\Lambda) = n-j, LO(\Lambda_k))\prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}}. \qquad (7)
\end{aligned}
$$

Note that we know

$$
C^+_\ell(z,r-1) = C^+_\ell(z,r) \cup C^*_\ell(z,r), \quad z \in \Gamma,
$$

then the first term on the right hand side of the last equality in Eq. (7) is

$$
\begin{aligned}
&\sum_{z\in\Gamma} \xi_0 N^{n-l_i}\left[U^{'}(C^-_0(z,r)) + \cdots + U^{'}(C^-_{r-2}(z,r))\right] p_{z;b^i_1} B_i\\
&\quad = \sum_{z\in\Gamma} \xi_0 N^{n-l_i}\left[U^{'}(C_0(z)) + \cdots + U^{'}(C_{r-2}(z))\right] p_{z;b^i_1} B_i\\
&\qquad - \sum_{z\in\Gamma} \xi_0 N^{n-l_i}\left[U^{'}(C^+_0(z,r-1)) + \cdots + U^{'}(C^+_{r-2}(z,r-1))\right] p_{z;b^i_1} B_i\\
&\quad = \sum_{z\in\Gamma} P(W(r-1,\Lambda) > n-l_i, X_{n-l_i} = z)p_{z;b^i_1} B_i\\
&\qquad - \sum_{j=1}^{l_i-1}\sum_{k=1}^{L} I_{\Lambda_k}(\Lambda_i \setminus j)P(W(r-1,\Lambda) = n-j, LO(\Lambda_k))\prod_{t=1}^{j} p_{b^i_{l_i-t};b^i_{l_i-t+1}}.
\end{aligned}
$$

Substituting the above result back to Eq. (7) completes the first part of the proof. The second part of the proof is the same as the proof of Theorem 1. □

Remark 2 The recursive equations for $P(W(r, \Lambda) > n)$ can be used to obtain the distribution of $X_n(\Lambda)$ via the relationship in Eq. (3). Nevertheless, the recursive equations for $P(X_n(\Lambda) = r)$ can also be derived using Theorem 1 and Theorem 2 along with Eq. (3) and

$$P(W(r, \Lambda) = n, LO(\Lambda_i)) = P(X_{n-1}(\Lambda) = r - 1, LO_n(\Lambda_i)),$$

where $\{X_{n-1}(\Lambda) = r - 1, LO_n(\Lambda_i)\}$ represents that $r - 1$ patterns occur until time $n - 1$ and also the r-th pattern occurs at time n. The details are left to the reader.

Remark 3 Fu (1996) derived the recursive equations for $P(X_n(\Lambda) = r)$, based on the backward multiplication of transition probability matrices, which directly involved the transition probability sub-matrices of the imbedded Markov chain. The sizes of the sub-matrices also depend on the size of the compound pattern Λ. Our recursive equations do not involve the transition probability matrix of the imbedded Markov chain, however the results indirectly relate to the FMCI technique as the idea and proofs originated from the FMCI technique.

4 Numerical examples

We provide an example to illustrate the theoretical results and show the performance of the proposed method.

Example 2 Consider a compound pattern

$$\Lambda = CAACCTGTTG \cup AGAGCGA \cup AGAGAG.$$

1. Let $\{X_n\}$ be a sequence of i.i.d. four-state trials with probabilities $P(X_1 = A) = 0.6$, $P(X_1 = C) = 0.2$, $P(X_1 = G) = 0.1$ and $P(X_1 = T) = 0.1$. With respect to i.i.d. with overlapping and non-overlapping counting schemes, we denote by I-O and I-N, respectively.
2. Let $\{X_n\}$ be a sequence of Markov dependent four-state trials with initial probabilities $P(X_1 = A) = 0.6, P(X_1 = C) = 0.2, P(X_1 = G) = 0.1$ and $P(X_1 = T) = 0.1$ and transition probability matrix

$$\mathbf{P} = \begin{array}{c} A \\ C \\ G \\ T \end{array} \begin{bmatrix} 0.2 & 0.3 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.2 \\ 0.1 & 0.3 & 0.2 & 0.4 \\ 0.4 & 0.1 & 0.2 & 0.3 \end{bmatrix}.$$

   With respect to Markov dependent with overlapping and non-overlapping counting schemes, we denote by M-O and M-N, respectively.

The distributions of $W(r, \Lambda)$, for $r = 1, 5, 10, 20$, under case M-O are given in Fig. 2. Also, the distributions of $X_n(\Lambda)$ for $n = 500,000$ under two cases I-O and M-O are given in Fig. 3. Figure 3 gives us an observation that, for the same pattern, the probabilities can be rather different using different models or, say, if choosing a wrong model.

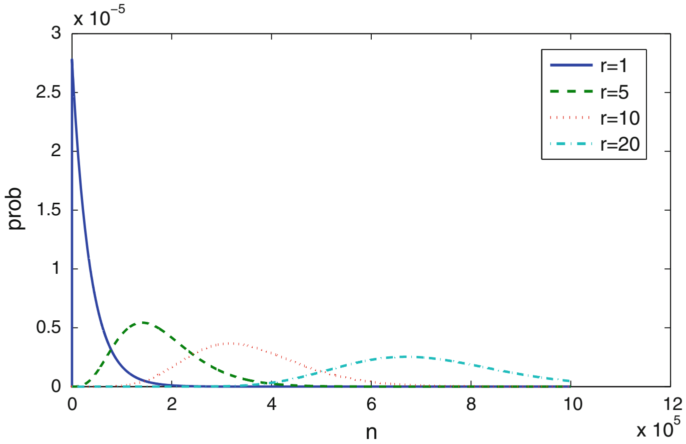

Fig. 2 The distributions of $W(r, \Lambda)$ for $r = 1, 5, 10, 20$ under case M-O

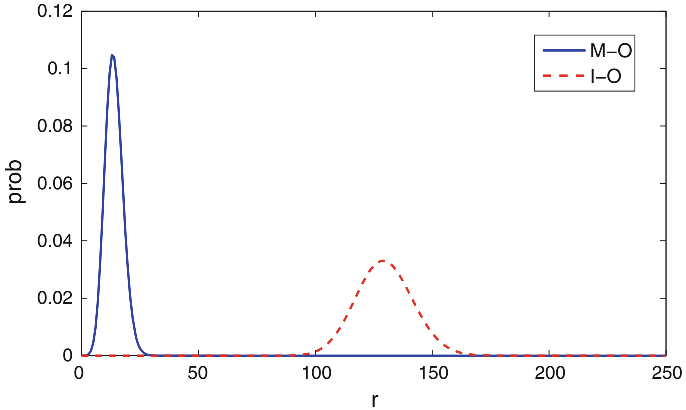

Fig. 3 The distributions of $X_n(\Lambda)$ under cases I-O and M-O with $n = 500{,}000$

5 Summary and discussion

In this manuscript, based on the FMCI technique and the partition of the state space of the imbedded Markov chain, we have derived the recursive equations for distributions of waiting time of r-th occurrence of simple and compound patterns under overlapping and non-overlapping counting schemes when the trials are i.i.d. or Markov dependent. The probability generating functions for $r = 1$ are also derived. From the dual relationship between $W(r, \Lambda)$ and $X_n(\Lambda)$, the distributions of $X_n(\Lambda)$ can also be obtained.

The result for probability generating function is improved as a solution to two systems of simultaneous equations of fixed sizes L and m, respectively, where L is the number of simple patterns and m is the size of Γ, while the result for probability generating function in Chang (2005) is the solution to ω simultaneous equations where ω is the size of the state space of the imbedded Markov chain and it grows as the lengths of the simple patterns increase.

All the numerical results are calculated by the computing software MATLAB. From the figures in Sect. 4, our recursive equations are able to compute the probabilities for $r = 250$ and $n = 500,000$. As a byproduct, if we calculate the probability $P(X_{10,000}(\Lambda) = 10)$ for example, then we automatically obtain the probabilities $P(X_n(\Lambda) = r)$ for all $n \leq 10,000$ and $r \leq 10$.

Acknowledgments The authors would like to thank Professor James C. Fu for his invaluable suggestions. The authors also thank referees for their useful comments.

References

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

Chang, Y. M. (2005). Distribution of waiting time until the rth occurrence of acompound pattern. *Statistics and Probability Letters*, *75*, 29–38.

Chang, Y. M., Wu, T. L. (2011). On average run lengths of control charts for autocorrelated processes. *Methodology and Computing in Applied Probability*, *13*, 419–431.

Chang, Y. M., Fu, J. C., Lin, H. Y. (2012). Distribution and double generating function of number of patterns in a sequence of Markov dependent multistate trials. *Annals of the Institute of Statistical Mathematics*, *64*, 55–68.

Cui, L., Xu, Y., Zhao, X. (2010). Developments and applications of the finite Markov chain imbedding approach in reliability. *IEEE Transactions on Reliability*, *59*, 685–690.

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., Lou, W. Y. W. (2003). *Distribution theory of runs and patterns and its applications*. River Edge, NJ: World Scientific Publishing Co. Inc.

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

Fu, J. C., Shmueli, G., Chang, Y. M. (2003). A unified Markov chain approach for computing the run length distribution in control charts with simple or compound rules. *Statistics and Probability Letters*, *65*, 457–466.

Han, Q., Hirano, K. (2003). Sooner and later waiting time problems for patterns in Markov dependent trials. *Journal of Applied Probability*, *40*, 73–86.

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

Inoue, K., Aki, S. (2005). A generalized Pólya urn model and related multivariate distributions. *Annals of the Institute of Statistical Mathematics*, *57*, 49–59.

Inoue, K., Aki, S. (2007). On generating functions of waiting times and numbers of occurrences of compound patterns in a sequence of multistate trials. *Journal of Applied Probability*, *44*, 71–81.

Inoue, K., Aki, S. (2009). On waiting time distributions associated with compound patterns in a sequence of multi-state trials. *Annals of the Institute of Statistical Mathematics*, *61*, 499–516.

Koutras, M. V. (1997). Waiting times and number of appearances of events in a sequence of discrete random variables (pp. 363–384). In Advances in combinatorial methods and applications to probability and statistics. *Statistics for Industry and Technology*. Boston, MA: Birkhäuser Boston

Koutras, M. V., Milienos, F. S. (2012). Exact and asymptotic results for pattern waiting times. *Journal of Statistical Planning and Inference*, *142*, 1464–1479.

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.

Nuel, G. (2008). Pattern Markov chains: Optimal Markov chain embedding through deterministic finite automata. *Journal of Applied Probability*, *45*, 226–243.

Zhao, X., Cui, L. (2009). On the accelerated scan finite Markov chain imbedding approach. *IEEE Transactions on Reliability*, *58*, 383–388.