

# The Markov Chain Imbedding Technique

Review by *Amărioarei Alexandru*

In this paper we will describe a method for computing exact distribution of runs and patterns in a sequence of discrete trial outcomes generated by an i.i.d or Markov source. The method, described for the first time by J. Fu and M.V. Koutras in [4], is entitled *Markov Chain Imbedding Technique* (for short MCIT) and shows a different approach than the traditional one which is based on combinatorial arguments. The main advantage of this method is that it can be applied not only in the i.i.d. cases, as most of the combinatorial approaches does, but also to Markov dependent multi-state trials (not matter what counting procedure is adopted:overlap vs. non-overlap). The paper is organized as follows: in the first section we will describe the general method and we will find the exact distribution of the number of occurrences of a pattern in  $n$  trials; in the second section we will describe the waiting time distributions of simple and compound patterns; the third section studies generating functions, mean, variance and large deviation approximations and in the forth section we will give worked examples. We mention that most of the results form are taken from [9].

## 1. DESCRIPTION OF THE METHOD

We begin this section by setting the framework and defining the necessary notions used throughout the paper. Let us consider that we have a sequence  $(X_t)_{t=\overline{1,n}}$  of  $n$  multi-state trials, each of which has  $m \geq 2$  states or symbols, labeled by  $\mathcal{S} = \{b_1, \dots, b_m\}$ .

**Definition 1.1.** We say that  $\Lambda$  is a *simple pattern* if  $\Lambda = b_{i_1} b_{i_2} \dots b_{i_k}$  where  $b_{i_j}$  is a symbol from  $\mathcal{S}$  for all  $j = \overline{1,k}$ .

Notice that the length of the pattern is fixed ( $k$ ) and the symbols can be repeated, for example the pattern  $\Lambda = b_1 b_1 b_1 b_2 b_2$  is a simple pattern.

**Definition 1.2.** Let  $\Lambda_1$  and  $\Lambda_2$  be two simple patterns of lengths  $k_1 \neq k_2$ . We say that the patterns are *distinct* if neither  $\Lambda_1 \subset \Lambda_2$  nor  $\Lambda_2 \subset \Lambda_1$ . More, we define  $\Lambda_1 \cup \Lambda_2$  to denote the occurrence of either  $\Lambda_1$  or  $\Lambda_2$ .

**Definition 1.3.** We say that a pattern  $\Lambda$  is a *compound pattern* if it can be written as the union of simple distinct patterns, *i.e.*  $\Lambda = \cup_{i=1}^l \Lambda_i$  where  $1 < l < \infty$  and  $\Lambda_i$  are simple distinct patterns.

An example of a compound pattern can be  $\Lambda = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$  where  $\Lambda_1 = aaccga$ ,  $\Lambda_2 = acggt$  and  $\Lambda_3 = ggactc$ . Notice that in this example we used the alphabet  $\mathcal{S} = \{a, c, g, t\}$ .

Now, given a simple or a compound pattern  $\Lambda$  we will denote by  $X_n(\Lambda)$  the random variable that represents the number of occurrences of the pattern  $\Lambda$  in a sequence of  $n$  multi-state trials, using both overlapping and non-overlapping counting scheme. We have to be careful about these counting procedures since  $X_n(\Lambda)$  differs from one another as the following example shows: lets suppose that we have a simple pattern  $\Lambda = ACA$  and a realization of a sequence of 20 four-state trials over the alphabet  $\mathcal{S} = \{A, C, G, T\}$

$$ATC\underline{ACACAT}AGAC\underline{ACAGT}AC$$

then we have  $X_{20}(\Lambda) = 2$  under non-overlapping scheme and  $X_{20}(\Lambda) = 4$  under overlapping scheme. One of the purposes of this section is to find the exact distribution of  $X_n(\Lambda)$  under both counting procedures (overlap and non-overlap), and when the source is in one of the following cases:

- (1)  $(X_t)_{t=\overline{1,n}}$  is a sequence of *i.i.d.* multi-state trials,
- (2)  $(X_t)_{t=\overline{1,n}}$  is a sequence of independent but not identically distributed multi-state trials,
- (3)  $(X_t)_{t=\overline{1,n}}$  is a sequence of homogeneous Markov dependent multi-state trials,
- (4)  $(X_t)_{t=\overline{1,n}}$  is a sequence of non-homogeneous Markov dependent multi-state trials.

We will define next the central concept of the paper, the *Markov chain imbeddable random variable*, as in [4] and [13]. For a given  $n$  let  $\Gamma_n = \{0, 1, \dots, n\}$  be a index set,  $\Omega = \{a_1, a_2, \dots, a_s\}$  be a finite state space and  $l_n = \max \{x | \mathbb{P}(X_n(\Lambda) = x) > 0\}$ .

**Definition 1.4.** We say that a non-negative integer valued random variable  $X_n(\Lambda)$  is *finite Markov chain imbeddable* if:

- a) there exists a finite Markov chain  $\{Y_t | t \in \Gamma_n\}$  defined on a finite state space  $\Omega$  with initial probability vector  $\boldsymbol{\xi}_0$ ,
- b) there exists a finite partition  $\{C_x | x = 0, 1, \dots, l_n\}$  on the state space, and
- c) for every  $x = 0, 1, \dots, l_n$  we have

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

For the next result we need to do a little work. Lets remember that given a non-homogeneous Markov chain,  $Z_t$ , on a finite state space  $\Omega = \{a_1, a_2, \dots, a_s\}$  the conditional probabilities

$$\mathbb{P}(Z_t = j | Z_{t-1} = i) = p_{ij}(t),$$

with  $i, j \in \Omega$  are called one-step transition probabilities at time  $t$ , and the corresponding one-step  $s \times s$  transition matrix at time  $t$  is given by:

$$\mathbf{M}_t = (p_{ij}(t))_{1 \leq i, j \leq s} = \begin{pmatrix} p_{11}(t) & p_{12}(t) & \dots & p_{1s}(t) \\ p_{21}(t) & \dots & \dots & \dots \\ \vdots & \ddots & \ddots & \vdots \\ p_{s1}(t) & p_{s2}(t) & \dots & p_{ss}(t) \end{pmatrix}_{s \times s}$$

If the Markov chain is homogeneous then  $\mathbf{M} = \mathbf{M}_t$  for all  $t = 1, 2, \dots$ . The  $n$  step transition probabilities,  $\mathbb{P}(Z_t = j | Z_{t-n} = i) = p_{ij}^{(n)}(t)$ , can be found using the Chapman-Kolmogorov equation:

$$p_{ij}^{(n)}(t) = \sum_{k \in \Omega} p_{ik}^{(r)}(t) p_{kj}^{(n-r)}(t+r)$$

for every intermediate values of  $r = \overline{1, n-1}$ . The  $n$  step transition matrix will be  $\mathbf{M}^{(n)} = \prod_{t=1}^n \mathbf{M}_t$  and in the homogeneous case  $\mathbf{M}^{(n)} = \mathbf{M}^n$ . Based on these equations, and given that the initial probability vector is  $\boldsymbol{\xi}_0 = (\mathbb{P}(Z_0 = a_1), \dots, \mathbb{P}(Z_0 = a_s))$ , we can find that the conditional probability of the chain state  $Z_n \in C$  for  $C \subset \Omega$  is given:

$$\mathbb{P}(Z_n \in C | \boldsymbol{\xi}_0) = \boldsymbol{\xi}_0 \left( \prod_{t=1}^n \mathbf{M}_t \right) \mathbf{U}^\top(C)$$

where  $\mathbf{U}(C) = \sum_{a_j \in C} e_j$  and  $e_j = (0, \dots, 1, \dots, 0)_{1 \times s}$  is a row vector with 1 on the  $j$  position and 0 otherwise. The main result in [4], and an important part of the MCIT approach, is a direct application of the Chapman-Kolmogorov equations:

**Theorem 1.5.** *If  $X_n(\Lambda)$  is finite Markov chain imbeddable, then*

$$\mathbb{P}(X_n(\Lambda) = x) = \boldsymbol{\xi}_0 \left( \prod_{t=1}^n \mathbf{M}_t \right) \mathbf{U}^\top(C_x),$$

where  $\boldsymbol{\xi}_0$  is the initial probability vector, and  $\mathbf{M}_t$ ,  $t = 1, \dots, n$  are the transition probability matrices of the imbedded Markov chain.

We will define next a new concept based on a particular form of the transition matrix corresponding to the imbedded chain, namely the notion of *Markov chain imbeddable*

*random variables of binomial type* (MVB) as it is named in [13]. Koutras and Alexandrou ([13]) showed that many run statistics ( $N_{n,k}$ ,  $M_{n,k}$ , etc. as we will see in examples section) can be viewed as MVB's with this particular type of matrix, a bi-diagonal matrix with non zero blocks only on the main diagonal and on the diagonal next to it. Let first assume that we are in the framework of the Definition 1.4 and suppose that every subset of the partition of  $\Omega$  has the same cardinality, more exactly  $r = |C_x|$  and  $C_x = \{c_{x0}, \dots, c_{xr}\}$  for  $x = \overline{0, l_n}$  (such a partition always exists since we can expand the cardinality of  $C_x$  by adding hypothetical states which not affect the chain behavior).

**Definition 1.6.** A random variable  $X_n(\Lambda)$  is MVB if:

- a)  $X_n(\Lambda)$  can be imbedded in a Markov chain like in Definition 1.4,
- b)  $\mathbb{P}(Y_t = c_{yj}|Y_{t-1} = c_{xi}) = 0$  for all  $y \neq x, x + 1$ .

From the above definition we observe that for any MVB we can define the following  $r \times r$  matrices:

$$\mathbf{A}_t(x) = (a_{ij}(t))_{i,j} = (\mathbb{P}(Y_t = c_{xj}|Y_{t-1} = c_{xi}))_{i,j},$$

and

$$\mathbf{B}_t(x) = (b_{ij}(t))_{i,j} = (\mathbb{P}(Y_t = c_{x+1j}|Y_{t-1} = c_{xi}))_{i,j}.$$

Using these matrices, the transition probability matrices of the imbedded chain will be a bi-diagonal block matrix given by:

$$\mathbf{M}_t = \begin{pmatrix} \mathbf{A}_t(0) & \mathbf{B}_t(0) & & & & & \\ & \mathbf{A}_t(1) & \mathbf{B}_t(1) & & & & \mathbf{O} \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \\ & \mathbf{O} & & \mathbf{A}_t(l_n - 1) & & \mathbf{A}_t(l_n - 1) & \\ & & & & & & \mathbf{A}_t(l_n) \end{pmatrix}$$

for  $t = \overline{1, n}$ . Looking at the matrix  $\mathbf{M}_t$  we observe that it can written as a sum of a diagonal matrix (with  $\mathbf{A}_t(x)$  components) and an upper diagonal matrix (with  $\mathbf{B}_t(x)$  components). Also lets notice that if for  $t = \overline{1, n}$  we denote the row vector by  $\boldsymbol{\alpha}(x) = (\mathbb{P}(Y_t = c_{x1}), \dots, \mathbb{P}(Y_t = c_{xr}))$  then

$$\mathbb{P}(X_n(\Lambda) = x|\boldsymbol{\xi}_0) = \boldsymbol{\alpha}_n(x)\mathbf{1}^\top, \quad \text{for all } x = 0, 1, \dots, l_n.$$

where  $\mathbf{1} = (1, \dots, 1)_{1 \times r}$ . These remarks leads to the main theorem in [13], which gives a recursive way to find the distribution of  $X_n(\Lambda)$ :

**Theorem 1.7.** *The following recursive equations holds:*

$$\begin{aligned}\alpha_t(0) &= \alpha_{t-1}(0)\mathbf{A}_t(0) \\ \alpha_t(x) &= \alpha_{t-1}(x-1)\mathbf{B}_t(x-1) + \alpha_{t-1}(x)\mathbf{A}_t(x), \quad x = 1, \dots, l_n.\end{aligned}$$

The *binomial type* nomenclature comes from the similarity between the above equations and the binomial distribution. In the last section we will see examples with this type of random variables.

Until now we assumed that the random variable is Markov chain imbeddable (MCI), both in general or MVB case. The main question is: given a random variable  $X_n(\Lambda)$  associated with the simple or compound pattern  $\Lambda$ , can we always imbed a Markov chain like in Definition 1.4 and if the answer is yes, how can we construct a such a chain? The answer to this question, as is showed in [5], is given by the *forward and backward principle*. We will describe this principle in detail for both non-overlap and overlap counting schemes.

Lets first assume that we are in the case of **non-overlap counting** and that  $(X_t)_{t=\overline{1,n}}$  is a sequence of  $n$  Markov dependent multi-state trials over the alphabet  $\mathcal{S} = \{b_1, \dots, b_m\}$ . As before  $X_n(\Lambda)$  denotes the number of (non-overlapping) occurrences of the pattern  $\Lambda$  (simple or compound) in  $n$  trials. For a better understanding of the principle we will consider the following example:

**Example 1.8.** Lets consider that the sequence  $(X_t)_{t=\overline{1,n}}$  is Markov dependent (homogeneous) three-state trials with transition probability matrix

$$\mathbf{A} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

over the alphabet  $\mathcal{S} = \{b_1, b_2, b_3\}$ , and take the simple pattern  $\Lambda = b_1b_1b_1b_2$ . We will give the method in four steps:

**i)** First we will decompose the pattern  $\Lambda = b_1b_1b_1b_2$  into the set of all sequential subpatterns denoted by  $\mathcal{S}(\Lambda) = \{b_1, b_1b_1, b_1b_1b_1, b_1b_1b_1b_2\}$  and define

$$\mathcal{E} = \mathcal{S} \cup \mathcal{S}(\Lambda) = \{b_1, b_2, b_3, b_1b_1, b_1b_1b_1, b_1b_1b_1b_2\}$$

**ii)** Second, if  $\omega = (x_1, \dots, x_n)$  is a realization of the sequence  $(X_t)_{t=\overline{1,n}}$ , we define the state space

$$\Omega = \{(u, v) | u = 0, 1, \dots, [n/4], v \in \mathcal{E}\} \cup \{\emptyset\} \setminus \{(0, b_1b_1b_1b_2)\}$$

and a Markov chain

$$Y_t = (X_t(\Lambda), E_t), \quad t = 1, 2, \dots, n$$

such that  $Y_t(\omega) = (u, v) \in \Omega$ , where

$u = X_t(\Lambda)(\omega)$  - the total number of non-overlapping occurrences of the pattern  $\Lambda$  in the first  $t$  trials, counting *forward* from the first to the  $t$ -th trial

$v = E_t(\omega)$  - the longest ending block in  $\mathcal{E}$ , counting *backward* from  $X_t$ .

Lets consider the following realization  $\omega = (b_2b_3b_1b_1b_2b_1b_1b_2b_1b_3b_1)$  and try to apply the described principle:

$$\begin{array}{lll} Y_1(\omega) = (0, b_2) & Y_5(\omega) = (0, b_2) & Y_9(\omega) = (1, b_1b_1b_1b_2) \\ Y_2(\omega) = (0, b_3) & Y_6(\omega) = (0, b_1) & Y_{10}(\omega) = (1, b_1) \\ Y_3(\omega) = (0, b_1) & Y_7(\omega) = (0, b_1b_1) & Y_{11}(\omega) = (0, b_3) \\ Y_4(\omega) = (0, b_1b_1) & Y_8(\omega) = (0, b_1b_1b_1) & Y_{12}(\omega) = (0, b_1) \end{array}$$

Notice that for every given  $\omega$ , the imbedded Markov chain is uniquely determined by the above procedure.

**iii)** The imbedded Markov chain  $(Y_t)_{t=\overline{1, n}}$  is homogeneous, and if its transition probability matrix is  $\mathbf{M}$  we will show how to determine it. Lets suppose that we are in  $Y_8(\omega) = (0, b_1b_1b_1)$ , we will have

$$(0, b_1b_1b_1) \rightarrow \begin{cases} (0, b_1b_1b_1), & X_9 = b_1 \text{ and the probability is } p_{11} \\ (1, b_1b_1b_1b_2), & X_9 = b_2 \text{ and the probability is } p_{12} \\ (0, b_3), & X_9 = b_3 \text{ and the probability is } p_{13} \end{cases}$$

and zero to any other state. Notice that in the definition of the state space  $\Omega$  we inserted the state  $\emptyset$  (dummy state) so that the chain will begin from that state  $\mathbb{P}(Y_0 = \emptyset) = 1$  and the transition probabilities are given by the initial distribution, that is  $\mathbb{P}(Y_1 = b_i | Y_0 = \emptyset) = p_i$  with  $i = 1, 2, 3$ . Also in the definition of the state space we deleted the state  $(0, b_1b_1b_1b_2)$  since if the ending block is equal with the pattern then the number of occurrences of  $\Lambda$  is  $\geq 1$ .

**iv)** The last step deals with the construction of a partition of the state space such that the conditions in Definition 1.4 to be fulfilled. We consider the following partition:

$$C_x = \begin{cases} C_\emptyset = \{\emptyset\} \\ C_0 = \{(0, b_1), (0, b_2), (0, b_3), (0, b_1b_1), (0, b_1b_1b_1)\} \\ C_z = \{(z, v) | v \in \mathcal{E}, z = 1, \dots, [n/4]\}. \end{cases}$$

In our case if we take  $n = 5$  then the state space is given by:

$$\Omega = \{\emptyset, (0, b_1), (0, b_2), (0, b_3), (0, b_1 b_1), (0, b_1 b_1 b_1), (1, b_1 b_1 b_1 b_2), \\ (1, b_1), (1, b_2), (1, b_3), (1, b_1 b_1), (1, b_1 b_1 b_1)\}$$

and the transition matrix:

$$\mathbf{M} = \begin{matrix} \emptyset \\ (0, b_1) \\ (0, b_2) \\ (0, b_3) \\ (0, b_1 b_1) \\ (0, b_1 b_1 b_1) \\ (1, \Lambda) \\ (1, b_1) \\ (1, b_2) \\ (1, b_3) \\ (0, b_1 b_1) \\ (0, b_1 b_1 b_1) \end{matrix} \begin{pmatrix} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{13} & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{13} & 0 & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{13} & 0 & p_{11} & p_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{12} & p_{13} & p_{11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} & p_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{12} & p_{13} & 0 & p_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In what follows, based on the Example 1.8, we will describe the general procedure. First, suppose that we are in the state  $Y_{t-1} = (x, z) \in \Omega$  at time  $t - 1$ ,  $X_t = b_j \in \mathcal{S}$  and denote by

$$(u, v) = \langle (x, z), b_j \rangle_{\Omega}$$

the state  $(u, v) \in \Omega$  resulted from *forward-backward procedure* (described in Example 1.8) knowing that at time  $t$  we have  $X_t = b_j$ . Also given a state  $(x, z) \in \Omega$ , lets denote by  $l(z)$  the last symbol in the ending block  $z$ .

**Theorem 1.9.** *Assuming that  $(X_t)_{t=1, \infty}$  is a homogeneous Markov chain with transition matrix  $\mathbf{A} = (p_{ij})_{s \times s}$ , and  $\Lambda = \cup_{i=1}^l \Lambda_i$  is a compound pattern of length  $k$  (each simple pattern  $\Lambda_i$  has the same length  $k$ ), then the imbedded Markov chain  $\{Y_t = (X_t(\Lambda), E_t)\}_{t=1, n}$ , corresponding to the r.v.  $X_n(\Lambda)$  has the following components:*

i) the state space

$$\Omega = \{\emptyset\} \cup \{(x, z) | x = 0, 1, \dots, [n/k], z \in \mathcal{E}\} \\ \setminus \{(0, \Lambda_i) | i = 1, 2, \dots, l\} \setminus \{([n/k], z) | k[n/k] + z(k) > n\},$$

where  $\mathcal{E} = \mathcal{S} \cup_{i=1}^l \mathcal{S}(\Lambda_i)$  and  $z(k) \equiv (\text{length of } z) \bmod k$

ii) the transition probabilities are defined by

$$P_{(x,z)(u,v)} = \begin{cases} p_j, & \text{if } (x, z) = \emptyset, u = 0, v = b_j \text{ for all } b_j \in \mathcal{S} \\ p_{ij}, & \text{if } (u, v) = \langle (x, z), b_j \rangle_{>\Omega}, x \leq [n/k], b_j \in \mathcal{S}, \\ & L(z) = b_i, kx + z(k) < n, \\ 1, & \text{if } (u, v) = (x, z), x = [n/k], \\ & \text{and } k[n/k] + z(k) = n, \\ 0, & \text{otherwise} \end{cases}$$

iii) the partition of  $\Omega$  is given by

$$\begin{aligned} C_\emptyset &= \{\emptyset\} \\ C_0 &= \{(0, z) | z \in \mathcal{E}\} \setminus \{(0, \Lambda_i) | i = 1, 2, \dots, l\} \\ C_x &= \{(x, z) | z \in \mathcal{E}, x = 1, \dots, [n/k] - 1\} \\ C_{[n/k]} &= \{([n/k], z) | z \in \mathcal{E}, k[n/k] + z(k) \leq n\}. \end{aligned}$$

From the above theorem we can deduce that the distribution of  $X_n(\Lambda)$  can be computed by

$$\mathbb{P}(X_n(\Lambda) = x) = \mathbb{P}(Y_n \in C_x) = \boldsymbol{\xi}_0 \mathbf{M}^n \mathbf{U}^\top(C_x)$$

where  $\mathbf{M}$  is the  $d \times d$  transition matrix ( $|\Omega| = d$ ),  $\xi_0$  is the initial distribution of the chain given by the dummy state (that is  $\mathbb{P}(Y_0 = \emptyset) = 1$ ) and  $\mathbf{U}(C_x) = \sum_{a_r \in C_x} e_r$ . The *forward-backward principle* it is still applicable in a more general framework, where the compound pattern is not compose from simple patterns of the same length, but the Theorem 1.9 is very difficult to write. This problem can be solved using the duality relationship between  $X_n(\Lambda)$  and waiting time  $W(\Lambda)$  as we will see in the next section.

We will end this section with a little discussion about the case of **overlap counting**. In this counting scheme we have to do some small modifications to the above procedure, since the main difference between this type of counting and the non-overlap one is that when the pattern is formed, a part of it will be counted toward forming the next pattern (up to at last  $length(\Lambda) - 1$ ). So, for a pattern  $\Lambda$  we will define the ending block  $\hat{E}$  as the longest ending block ( $\hat{E} \neq \Lambda$ ) that, after each occurrence of  $\Lambda$  under the overlap counting, can be assigned as the initial block for the next occurrence of  $\Lambda$ . To get a better understanding lets give some examples:

- if  $\Lambda = b_1 b_2 b_1$  then  $\hat{E} = b_1$
- if  $\Lambda = \underbrace{b_1 b_1 \dots b_1}_k$  then  $\hat{E} = \underbrace{b_1 b_1 \dots b_1}_{k-1}$



We have to be careful that in overlap counting the maximum number of occurrences of a pattern of length  $k$  in a sequence of  $n$  trials is given by

$$\hat{l}_n = 1 + \left\lceil \frac{n - k}{k - |\hat{E}|} \right\rceil.$$

The second example (Example 4.2) in the last section shows the main differences between these two counting scheme.

**Remark 1.10.** In [3] we encounter another form of the *forward-backward principle* which is a slightly modification of the above one, but it is interesting to be mentioned. Given the simple pattern  $\Lambda = b_{i_1}b_{i_2}\dots b_{i_k}$  we decompose it into  $k - 1$  subpatterns labeled  $1 = b_{i_1}, 2 = b_{i_1}b_{i_2}, \dots, k - 1 = b_{i_1}b_{i_2}\dots b_{i_{k-1}}$  and denote by 0 none of the subpatterns  $1, \dots, k - 1$ . We can see that these subpaterns are the ending blocks mentioned before (inclusive 0). For a realization  $\omega = (x_1, \dots, x_n)$  of the sequence  $X_t$ , where  $x_i$  is the outcome of the  $i$ -th trial, we define the imbedded Markov chain  $(Y_t)_{t=\overline{1, n}}$  operating on  $\omega$  by  $Y_t = (u, v)$ , where  $u$  is as in the number of occurrences of the pattern  $\Lambda$  until time  $t$  and  $v$  is the ending block (the definition is the same as in the previous case). The state space is given by:

$$\Omega = \{(u, v) | u = 0, 1, \dots, [n/k]; v = 0, 1, \dots, k - 1\}$$

and the transition probabilities:

$$\mathbb{P}(Y_t = (u', v') | Y_{t-1} = (u, v)) = \begin{cases} \sum_{v \rightarrow v'} p_{i_{j_1} i_{j_2}}, & \text{if } u' = u \in \{0, 1, \dots, [n/k]\} \\ & \text{and } v, v' \in \{0, 1, \dots, k - 1\} \\ p_{i_{k-1} i_k}, & \text{if } u' = u + 1 \text{ for } u = \overline{0, [n/k] - 1}, \\ & v' = 0 \text{ and } v = k - 1, \\ 0, & \text{otherwise,} \end{cases}$$

where  $\sum_{v \rightarrow v'}$  denotes the sum over all states such that the ending block  $v$  transforms into  $v'$ , and  $p_{i_{j_1} i_{j_2}}$  are the corresponding probabilities knowing that the last symbol was  $b_{i_{j_1}}$  and the new one is  $b_{i_{j_2}}$ . Like in the previous approach we define the partition of the state space by:

$$C_x = \{(x, v) | (x, v) \in \Omega, v = 0, 1, \dots, k - 1\}, \text{ for } x = 1, 2, \dots, [n/k].$$

## 2. WAITING TIME DISTRIBUTIONS

The main purpose of this section is to give a general method for finding the exact distribution of the waiting time random variable associated with a compound pattern. The development of what follows is according to [2], [7], [9], [11].

We will begin by some definitions. Lets suppose that we are in the framework described in Section 1.

**Definition 2.1.** We have the following notions:

- the waiting time for a simple pattern  $\Lambda = b_{i_1} b_{i_2} \dots b_{i_k}$

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

- the waiting time of a compound pattern  $\Lambda = \cup_{i=1}^l \Lambda_i$

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

- the waiting time of the  $r$ -th occurrence of the pattern  $\Lambda$

$$\begin{aligned} W(r, \Lambda) &= \text{minimum number of trials required to observe the } r\text{-th occurrence} \\ &\quad \text{of the pattern } \Lambda \end{aligned}$$

We can observe now that, in the context of Definition 2.1, we have the following relationship ([5]) between the random variables  $X_n(\Lambda)$  and  $W(r, \Lambda)$  corresponding to the pattern  $\Lambda$ :

$$\mathbb{P}(X_n(\Lambda) < r) = \mathbb{P}(W(r, \Lambda) > n).$$

The above relation is known also as *dual property* and justifies in a sense the study of the waiting time distributions. Next, we will give some basic results about Markov chains that will characterize the probability of the first-entry of the chain in an absorbing state. These results are the basis of the main results regarding the waiting time distributions.

An absorbing state is a state in which once entered, the chain cannot escape. If we have a (homogeneous) Markov chain,  $(Y_t)_{t=\overline{1, \infty}}$ , defined on the state space  $\Omega$ , and if we denote the set of the absorbing states by  $A = \{\alpha_1, \dots, \alpha_k\}$ , then the probability transition matrix  $\mathbf{M}$  can be written (after proper arrangements):

$$\mathbf{M} = \begin{array}{c} \Omega \setminus A \\ A \end{array} \begin{pmatrix} \mathbf{N}_{(m-k) \times (m-k)} & \mathbf{C}_{(m-k) \times k} \\ \mathbf{O}_{k \times (m-k)} & \mathbf{I}_{k \times k} \end{pmatrix}$$

where  $m = |\Omega|$ . The matrix  $\mathbf{N}$  is called the essential matrix and plays an important role in the development of MCIT, as we will see. Lets denote the initial distribution (assuming that the chain doesn't start from an absorbing state) by  $\boldsymbol{\xi}_0 = (\boldsymbol{\xi} : \mathbf{0})_{1 \times m}$ , where  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{m-k})$  such that  $\sum_{i=1}^{m-k} \xi_i = 1$ . The following result is due to [3]:

**Theorem 2.2.** *We have the following results:*

i) *the probability that the chain enters first the set of absorbing states at time  $n$  is given*

$$\mathbb{P}(Y_n \in A, Y_{n-1} \notin A, \dots, Y_1 \notin A | \boldsymbol{\xi}_0) = \boldsymbol{\xi} \mathbf{N}^{n-1} (\mathbf{I} - \mathbf{N}) \mathbf{1}^\top$$

where  $\mathbf{1} = (1, \dots, 1)_{1 \times (m-k)}$ ,

ii) *for any state  $i \in \Omega \setminus A$*

$$\mathbb{P}(Y_n = i, Y_{n-1} \notin A, \dots, Y_1 \notin A | \boldsymbol{\xi}_0) = \boldsymbol{\xi} \mathbf{N}^n \mathbf{e}_i^\top$$

where  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)_{1 \times (m-k)}$  with 1 on the  $i$ -th position,

iii) *for any  $j \in A$  we have*

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

where  $\mathbf{C}_j$  is the  $j$ -th column of the matrix  $\mathbf{C}$ .

We will turn now our attention to the waiting time problem. Given a sequence of Markov dependent  $m$ -state random variables  $(X_t)_{\overline{1, \infty}}$  over the alphabet  $\mathcal{S} = \{b_1, \dots, b_m\}$ , and a compound pattern  $\Lambda = \cup_{i=1}^l \Lambda_i$ , we want to show that the waiting time random variable corresponding to this pattern,  $W(\Lambda)$ , is Markov chain imbeddable in the sense of Definition 1.4. For that we will define the imbedded chain, the state space where it is defined and its transition probabilities. The state space is (see the resemblance with the *forward-backward principle*):

$$\Omega = \{\emptyset\} \cup \mathcal{S} \cup_{i=1}^l \mathcal{S}(\Lambda_i)$$

where  $\mathcal{S}(\Lambda_i)$  is the collection of all subpatterns of  $\Lambda_i$  (in the sense described in the first section). Let  $\alpha_i$  be the absorbing state with respect to the simple pattern  $\Lambda_i$ , for  $i = \overline{1, l}$ , and denote by  $A = \{\alpha_1, \dots, \alpha_l\}$  the set of all these states.

For  $Y_{t-1} = u \in \Omega \setminus A \setminus \{\emptyset\}$  and  $X_t = z \in \mathcal{S}$  we define

$$v = \langle u, z \rangle_\Omega$$

= the longest ending block in  $\Omega$  with respect to

the *forward and backward procedure*,

and the set

$$[u : \mathcal{S}] = \{v | v \in \Omega, v = \langle u, z \rangle_{\Omega}, z \in \mathcal{S}\}$$

Now we are ready to give the main result of the section (due to [9] and [3]):

**Theorem 2.3.** *Assuming that  $(X_t)_{t=1, \infty}$  is a sequence of Markov-dependent  $m$ -state trials with transition probabilities  $p_{ij}$ , and  $\Lambda = \cup_{i=1}^l \Lambda_i$  is a compound pattern, then*

*i) the imbedded Markov chain  $Y_t$ , defined on the state space  $\Omega$  has the transition probabilities given by*

$$p_{u,v} = \mathbb{P}(Y_t = v | Y_{t-1} = u) = \begin{cases} p_z, & \text{if } u = \emptyset, v = z, z \in \mathcal{S} \\ p_{xz}, & \text{if } u \in \Omega \setminus A \setminus \{\emptyset\} \\ & v \in [u : \mathcal{S}] \text{ and } X_t = z \\ 1, & \text{if } u \in A \text{ and } v = u \\ 0, & \text{otherwise} \end{cases}$$

where  $x$  is the last symbol of  $u$ ,  $p_z$  is the initial probability given  $Y_0 = \emptyset$  to  $z$ .

The transition matrix has the form:

$$\mathbf{M} = \begin{matrix} \Omega \setminus A \\ A \end{matrix} \begin{pmatrix} \mathbf{N}_{(d-l) \times (d-l)} & \mathbf{C} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}_{d \times d}$$

where  $|\Omega| = d$ .

*ii) given the initial distribution  $\boldsymbol{\xi}_0 = (\boldsymbol{\xi} : \mathbf{0})$ , the waiting time distribution of the compound pattern  $\Lambda$  is given by:*

$$\mathbb{P}(W(\Lambda) = n) = \boldsymbol{\xi} \mathbf{N}^{n-1} (\mathbf{I} - \mathbf{N}) \mathbf{1}^\top$$

*iii) for every  $j = 1, 2, \dots, l$*

$$\mathbb{P}(W(\Lambda) = n, W(\Lambda_j) = n) = \boldsymbol{\xi} \mathbf{N}^{n-1} \mathbf{C}_j$$

where  $\mathbf{C}_j$  is the  $j$ -th column of the matrix  $\mathbf{C}$ .

An important aspect is to observe that the imbedded chain highly depends on its initial distribution. Based on Theorem 2.3, Fu and Chang [7] developed an algorithm for computing the waiting time distribution. Also Chang developed three more algorithms in [3] for the waiting time distribution of the  $r$ -th occurrence of a pattern. Examples regarding the application of these results can be found in the last section.

**Remark 2.4.** In what follows we will speak about some extensions of the above results in the case of  $W(r, \Lambda)$ . Considering the approach developed by Chang in [2] and the above hypothesis, we construct the imbedded Markov chain  $\{Y_t | t = 0, 1, \dots\}$  on the state space  $\Omega^{(r)}$  having the form

$$\Omega^{(r)} = \{\emptyset\} \cup \Omega_1 \cup \Omega_2 \cup A$$

where  $A = \{\alpha_1, \dots, \alpha_l\}$  is the set of all the absorbing states  $\alpha_j$  corresponding to the  $r$ -th occurrence of the pattern  $\Lambda_j$ , and  $\Omega_1$  and  $\Omega_2$  are defined

$$\begin{aligned} \Omega_1 &= \{(u, v) | u = 0, \dots, r-1; v \in \mathcal{S} \cup \mathcal{S}(\Lambda_1) \cdots \cup \mathcal{S}(\Lambda_l) \setminus A\}, \\ \Omega_2 &= \{(u, v) | u = 1, \dots, r-1; v \in \mathcal{B}\} \end{aligned}$$

where  $\mathcal{B}$  is the collection of the last symbols of  $\Lambda_i$ ,  $i = 1, \dots, l$ . To distinguish between the elements of  $\mathcal{B}$  and  $\mathcal{S}$  we will add some marks (asterisk, hat) to the former ones. Now, it is easy to observe that we can relabel the elements of the state space and obtain  $\Omega^{(r)} = \{1, \dots, k, \alpha_1, \dots, \alpha_l\}$ , which ensures us that the probability transition matrix  $\mathbf{M}^{(r)}$  of the imbedded chain associated with  $W(r, \Lambda)$  has the form mentioned in Theorem 2.3. An example that will illustrate this procedure it is given in the last section (Example 4.4).

### 3. PROBABILITY GENERATING FUNCTIONS, MEAN, VARIANCE AND APPROXIMATIONS

Since probability generating functions (p.g.f.) play an important role in the study of the probability distributions we dedicate a whole section to their study and related aspects. We will deal mostly with the p.g.f. of waiting time distributions associated to a simple or compound pattern, and we will develop some recurrence relations for finding the mean and the variance (which will be very useful in simulations).

Considering the context of Theorem 1.5 we can easily write the p.g.f. of the random variable  $X_n(\Lambda)$  as:

$$\varphi_{X_n(\Lambda)}(s) = \mathbb{E}[s^{X_n(\Lambda)}] = \boldsymbol{\xi}_0 \left( \prod_{t=1}^n \mathbf{M}_t \right) \left( \sum_{x=0}^{l_n} s^x \mathbf{U}^\top(C_x) \right)$$

and the  $k$ -th moment

$$\mathbb{E}[X_n^k(\Lambda)] = \boldsymbol{\xi}_0 \left( \prod_{t=1}^n \mathbf{M}_t \right) \left( \sum_{x=0}^{l_n} x^k \mathbf{U}^\top(C_x) \right).$$

Referring us to the waiting time random variable associated to a pattern  $\Lambda$  we have the following result ([7] and [3]) which is a direct application of Theorem 2.2:

**Theorem 3.1.** *Suppose that  $\mathbf{M}$  is the probability transition matrix (with the form given in Theorem 2.3) of the imbedded Markov chain for the waiting time random variable  $W(\Lambda)$  of a compound pattern  $\Lambda$ , then*

*i) the p.g.f. is given by*

$$\varphi_W(s) = 1 + (s - 1)\boldsymbol{\xi}(\mathbf{I} - s\mathbf{N})^{-1}\mathbf{1}^\top,$$

*ii) and the mean*

$$\mathbb{E}[W(\Lambda)] = \boldsymbol{\xi}(\mathbf{I} - \mathbf{N})^{-1}\mathbf{1}^\top.$$

The above theorem gives a way of computing the mean and the p.g.f. of the random variable  $W(\Lambda)$ , but as we can see it is based on the inverse of the essential matrix, and in many cases this is very difficult to compute. The next theorem is due to Fu and Chang ([7] and [3]) and gives a recurrence relationship for the mean and the p.g.f.:

**Theorem 3.2.** *For the waiting time random variable  $W(\Lambda)$  we have that*

*i) the mean waiting time can be expressed as*

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

*where  $(S_1, S_2, \dots, S_k)$  is the solution of the system of equations*

$$S_i = \boldsymbol{\xi}e_i^\top + (S_1, \dots, S_k)\mathbf{N}(i), \text{ for } i = 1, 2, \dots, k,$$

*where  $\mathbf{N}(i)$  are the column vectors of the essential matrix  $\mathbf{N}$ ,*

*ii) the p.g.f. is given by*

$$\varphi_W(s) = 1 + \left(1 - \frac{1}{s}\right)\boldsymbol{\Phi}_W(s),$$

*where  $\boldsymbol{\Phi}_W(s)$  is the probability generating function of the sequence of cumulative probabilities  $\{\mathbb{P}(W(\Lambda) \geq n)\}_{n=1}^\infty$  which can be expressed as*

$$\boldsymbol{\Phi}_W(s) = \phi_1(s) + \phi_2(s) + \cdots + \phi_k(s)$$

*with  $(\phi_1, \dots, \phi_k)$  being the solution of the system*

$$\phi_i(s) = s\boldsymbol{\xi}e_i^\top + s(\phi_1(s), \dots, \phi_k(s))\mathbf{N}(i), \text{ for } i = 1, 2, \dots, k.$$

It is worth mentioning that due to Theorem 2.2 we have that the cumulative probability of the waiting time is given by:

$$\mathbb{P}(W(\Lambda) \geq n) = \boldsymbol{\xi}\mathbf{N}^{n-1}\mathbf{1}^\top,$$

and that the expression of  $S_i$  in Theorem 3.2 is

$$S_i = \sum_{n=1}^{\infty} \boldsymbol{\xi} \mathbf{N}^{n-1} \mathbf{e}_i^\top$$

and that of  $\phi_i(s)$  is

$$\phi_i(s) = \sum_{n=1}^{\infty} s^n \boldsymbol{\xi} \mathbf{N}^{n-1} \mathbf{e}_i^\top.$$

As we mentioned before, this theorem is very important in computer simulations and algorithms for symbolic computations was developed by Chang in [3]. For the computation of the variance of the waiting time distribution  $W(\Lambda)$  similar recursive relations were deduced in [17]. Before the main result of the paper [17], let's observe that due to Theorem 3.2, in order to deduce the variance of  $W(\Lambda)$ , we will need to compute only the second moment of the waiting time random variable.

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

$$\mathbb{E}[W^2(\Lambda)] = \sum_{j=1}^l (V_1, \dots, V_k) \mathbf{C}_j$$

where  $(V_1, \dots, V_k)$  is the solution of the system of equations

$$V_i = (V_1, \dots, V_k) \mathbf{N}(i) + 2T_i - S_i, \text{ for } i = 1, \dots, k,$$

and where  $S_i$  are given in Theorem 3.2, and  $(T_1, \dots, T_k)$  is the solution of the system of equations

$$T_i = (T_1, \dots, T_k) \mathbf{N}(i) + S_i, \text{ for } i = 1, \dots, k.$$

In the case of the waiting time until the  $r$ -th occurrence, we notice that we can express the random variable  $W(r, \Lambda)$  as the sum of  $r$  inter-waiting times

$$W(r, \Lambda) = W_1(\Lambda) + \dots + W_r(\Lambda).$$

If we consider the case in which  $X_t$  are i.i.d.  $m$ -state trials under non-overlapping counting, then  $\{W_i(\Lambda) | i = \overline{1, r}\}$  are i.i.d. random variables and the p.g.f. of  $W(r, \Lambda)$  is given by

$$\varphi_{W(r, \Lambda)}(s) = (\phi_{W_1(\Lambda)}(s))^r.$$

In the overlapping counting scheme, then  $W_2(\Lambda), \dots, W_r(\Lambda)$  are i.i.d. random variables and the p.g.f. of  $W(r, \Lambda)$  is given by

$$\varphi_{W(r, \Lambda)}(s) = \phi_{W_1(\Lambda)}(s) (\phi_{W_2(\Lambda)}(s))^{r-1}.$$

Detailed explanations about how to derive the p.g.f. in the case of the waiting time until the  $r$ -th occurrence are given by Chang in [2] where he use the generating functions of waiting time to the first occurrence of  $\Lambda_j$ , and  $\Lambda_j$  occurs first among all simple patterns  $\Lambda_1, \dots, \Lambda_l$  (denoted by  $W(\Lambda_j|\Lambda_1, \dots, \Lambda_l)$ ) as is described in the following theorem:

**Theorem 3.4.** *The probability generating function for the waiting time random variable  $W(r, \Lambda)$  under the non-overlapping counting is given by*

$$\varphi_{W(r, \Lambda)}(s) = \sum_{i=1}^r \prod_{j=1}^r \psi_{W_j(\Lambda_{i_j}|\Lambda_1, \dots, \Lambda_l)}(s | \boldsymbol{\xi}_{\emptyset_{L(\Lambda_{i_{j-1}})}}(\Lambda_{i_{j-1}}))$$

where  $i_j \in \{1, 2, \dots, l\}$ ,  $L(\Lambda_{i_{j-1}})$  is the last symbol of  $\Lambda_{i_{j-1}}$  for  $j = 2, \dots, l$ , and  $\boldsymbol{\xi}_{\emptyset_{L(\Lambda_{i_0})}} = \boldsymbol{\xi}$  by convention.

Next, based on the work of Fu, Wang and Lou in [8] and Fu and Johnson in [12], we will give some approximation results using the methods of spectrum analysis and the large deviation approach. We will begin with some preliminary notations. Let  $\lambda_1, \dots, \lambda_w$  be the ordered eigenvalues, in the sense  $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_w|$ , of the essential matrix  $\mathbf{N}$  as described in Theorem 2.3. Due to Perron-Frobenius Theorem, and the fact that  $\mathbf{N}$  is a sub-stochastic matrix, we know that the largest eigenvalue is unique, real and between zero and one ( $0 < |\lambda_1| < 1$ ). Since the vector  $\mathbf{1}^\top$  can always be decomposed

$$\mathbf{1}^\top = \sum_{i=1}^w a_i \boldsymbol{\eta}_i$$

where  $\boldsymbol{\eta}_i$  is the column eigenvector associated to the eigenvalue  $\lambda_i$  ( $\mathbf{N}\boldsymbol{\eta}_i = \lambda_i \boldsymbol{\eta}_i$ ) we have the following

**Theorem 3.5.** *If the transition matrix  $\mathbf{M}$  corresponding to the waiting time  $W(\Lambda)$  has the form given by Theorem 2.2, then*

i)

$$\mathbb{P}(W(\Lambda) \geq n) = \sum_{i=1}^w C_i \lambda_i^{n-1},$$

where  $C_i = a_i \boldsymbol{\xi} \boldsymbol{\eta}_i$ , for  $i = 1, 2, \dots, w$ ,

ii)

$$\Phi_W(s) = \sum_{n=1}^{\infty} s^n \mathbb{P}(W(\Lambda) \geq n) = \sum_{i=1}^w \frac{C_i s}{1 - s \lambda_i},$$

and  $\Phi_W(s)$  exists for  $|s| < \frac{1}{\lambda_1}$ ,



iii)

$$\varphi_W(s) = 1 + \left(1 - \frac{1}{s}\right) \sum_{i=1}^w \frac{C_i s}{1 - s\lambda_i}, \quad |s| < \frac{1}{\lambda_1}.$$

In [8] the authors develop a large deviation approximation for the longest run in a sequence of two state Markov-dependent trials. Their main result is based on the following theorem, which shows that the tail probability of the waiting time random variable  $W(\Lambda)$ , converge to zero with an exponential rate:

**Theorem 3.6.** *We have the following relation:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(W(\Lambda) \geq n) = -\beta(\Lambda)$$

where  $\beta(\Lambda) = -\log \lambda_1$ . More,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}(W(\Lambda) \geq n)}{C_1 \lambda_1^{n-1}} = 1.$$

Fu and Johnson gives in [12] an approximation for the number of non-overlapping occurrences of a simple pattern  $(\Lambda = b_{i_1} \dots b_{i_k})$  in  $n$ -trials generated by a Markov source.

**Theorem 3.7.** *For any fixed  $r \geq 0$ ,*

$$\mathbb{P}(X_n(\Lambda) = r) \sim ab^r \binom{n - r(k-1)}{r} (1 - \lambda_1)^k \lambda_1^{n-k},$$

where

$$a = \sum_{j=1}^d a_j \boldsymbol{\xi} \boldsymbol{\eta}_j \quad \text{and} \quad b = \sum_{j=1}^d a_j \boldsymbol{\xi}_\Lambda \boldsymbol{\eta}_j$$

and where  $d$  is the algebraic multiplicity of  $\lambda_1$ , and  $\boldsymbol{\xi}_\Lambda$  is the row vector corresponding to the inter-arrival times  $W_2(\Lambda), \dots, W_k(\Lambda)$  (due to the duality relation between  $X_n(\Lambda)$  and  $W(r, \Lambda)$ ) such that  $\mathbb{P}(W_j(\Lambda) = n) = \boldsymbol{\xi}_\Lambda \mathbf{N}^{n-1} (\mathbf{I} - \mathbf{N}) \mathbf{1}^\top$  for  $j \geq 2$ .

In the ending of this section we should mention that in [14] the distribution of  $X_n(\Lambda)$  is studied with the help of double generating functions  $G(s, t)$ , which is shown that can be written in terms of p.g.f. of the waiting time  $W(\Lambda)$ . A more general context for the study of the waiting time distributions of simple and compound patterns is presented by Fu and Wendy Lou in [11], where the framework consists in a sequence of  $r$ -th order Markov-dependent multi-state trials. They take into consideration the ergodic probability of a irreducible, aperiodic  $r$ -th order homogeneous Markov chain and express the distribution of the waiting time  $W(\Lambda)$  with their help giving a similar result with Theorem 2.3.

## 4. WORKED EXAMPLES

In this section we will give some examples for a better understanding of the MCIT.

**Example 4.1.** In this example we will show that the number of non-overlapping consecutive  $k$  successes,  $N_{n,k}$ , in a sequence of  $n$  two-state trials generated by a Markov source, is an imbeddable random variable in the sense of Definition 1.6. Lets suppose that  $(X_i)_{i=\overline{1,n}}$  is a sequence of homogeneous Markov dependent two-state trials, with outcomes  $S$  and  $F$ , and with the probability matrix

$$P = \begin{pmatrix} p_{SS} & p_{SF} \\ p_{FS} & p_{FF} \end{pmatrix}$$

It is easy to see that for the random variable  $N_{n,k}$ , the associated pattern  $(X_n(\Lambda))$  is given by  $\Lambda = \underbrace{SS \dots S}_k$ . In the view of Theorem 1.9, let us define the state space:

$$\Omega = \{(x, i) | x = 0, 1, \dots, l_n = [n/k] \text{ and } i = 0, 1, \dots, k - 1\}$$

and the Markov chain  $\{Y_t | t = 0, 1, \dots, n\}$  on  $\Omega$  by

$$Y_t = (N_{t,k}, E_t)$$

where  $N_{t,k}$  is the number of non-overlapping consecutive  $k$  successes that occurs in the first  $t$  trials. To make things a little easier we will define the ending block  $E_t$  to be equal with the number of trailing successes modulo  $k$ , with an ending block of zero only in the case where the last outcome is  $F$  and for the cases in which the modulo is zero to be denoted by  $\gamma$ . In this way the *forward-backward procedure* is simplified a bit. So, we will have

$$\begin{aligned} Y_t &= (x, \gamma) \\ &= \text{in the first } t \text{ trials there are } x \text{ runs of } k \text{ consecutive successes} \\ &\quad \text{and the last } m > 0 \text{ trailing successes satisfy } m \equiv 0(\text{mod } k) \end{aligned}$$

and

$$\begin{aligned} Y_t &= (x, 0) \\ &= \text{in the first } t \text{ trials there are } x \text{ runs of } k \text{ consecutive successes} \\ &\quad \text{and the last trial is } F \text{ (} X_t = F \text{)} \end{aligned}$$

If  $X_t$  were i.i.d. we could define the ending block without the intermediary state  $\gamma$ . Now it is easy to see that the transition probabilities are given by

$$\mathbb{P}(Y_t = (y, j) | Y_{t-1} = (x, i)) = \begin{cases} p_{FF}, & \text{if } y = x, j = 0, i = 0 \\ p_{FS}, & \text{if } y = x, j = 1, i = 0 \\ p_{SF}, & \text{if } y = x, j = 0, i = 1, \dots, k-1 \\ p_{SS}, & \text{if } y = x, j = i+1, i = 1, \dots, k-2 \\ p_{SS}, & \text{if } y = x+1, j = \gamma, i = k-1 \\ 1, & \text{if } y = x = l_n, j, i = k-1 \\ 0, & \text{otherwise} \end{cases}$$

The probability transition matrix will be

$$\mathbf{M}(N_{n,k}) = \begin{pmatrix} \mathbf{A}^* & \mathbf{B}^* & & & \\ & \mathbf{A} & \mathbf{B} & & \mathbf{O} \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ \mathbf{O} & & \mathbf{A} & & \mathbf{A} \\ & & & & \mathbf{A}^{**} \end{pmatrix}$$

where the matrix  $\mathbf{A}^*$  is

$$\mathbf{A}^* = \begin{matrix} (0,0) \\ (0,1) \\ \vdots \\ (0,k-2) \\ (0,k-1) \end{matrix} \begin{pmatrix} p_{FF} & p_{FS} & \dots & \dots & 0 \\ p_{SF} & 0 & p_{SS} & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ p_{SF} & 0 & \dots & \dots & p_{SS} \\ p_{SF} & 0 & \dots & \dots & 1 \end{pmatrix}_{k \times k},$$

the matrix and  $\mathbf{B}^*$  is

$$\mathbf{B}^* = \begin{matrix} (1,\gamma) & (1,0) & \dots & (1,k-2) & (1,k-1) \end{matrix} \begin{pmatrix} 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 0 \\ p_{SS} & 0 & \dots & \dots & 0 \end{pmatrix}_{k \times k+1},$$

and the matrix  $\mathbf{A}$  is

$$\mathbf{A} = \begin{matrix} & (i, \gamma) & (i, 0) & (i, 1) & \dots & (i, k-2) & (i, k-1) \\ \left( \begin{array}{cccccc} 0 & p_{FF} & p_{FS} & \dots & \dots & 0 \\ 0 & p_{SF} & 0 & p_{SS} & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ 0 & p_{SF} & 0 & \dots & \dots & p_{SS} \\ 0 & p_{SF} & 0 & \dots & \dots & 0 \end{array} \right) & & & & & & \end{matrix} \Bigg)_{k+1 \times k+1}.$$

The matrix  $\mathbf{B}$  has the same form as  $\mathbf{B}^*$  only its  $k+1 \times k+1$ , and the matrix  $\mathbf{A}^{**}$  is the same as  $\mathbf{A}$  with the exception that the last line in  $(0, \dots, 0, 1)$ . As we can see, the imbedded chain is MVB as is described in Definition 2.1. We need to pay attention on the initial distribution, but as we described in the *forward-backward principle* we can consider that the chain starts in the dummy state  $\emptyset$  and the initial probabilities are  $(p_S, p_F)$ . This run statistic was studied by Fu and Koutras in [4] and by Koutras and Alexandrou in [13]. They also showed that others related run statistics, like  $M_{n,k}$ -the number of overlapping consecutive  $k$  successes and  $L_n(S)$ -the longest success run, are Markov chain imbeddable.

The next example shows how to apply the *forward-backward procedure* in both non-overlap and overlap counting.

**Example 4.2.** Lets suppose that we are in the framework of Example 1.8, that is the sequence  $(X_t)_{t=\overline{1,n}}$  is Markov dependent (homogeneous) three-state trials with transition probability matrix

$$\mathbf{A} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

over the alphabet  $\mathcal{S} = \{b_1, b_2, b_3\}$ , and lets consider the pattern  $\Lambda = b_1 b_2 b_1$ . We will give the state space and the transition probability for both non-overlap and overlap counting in the case of  $n = 5$  trials. The state space for non-overlap is:

$$\Omega = \{\emptyset, (0, b_1), (0, b_2), (0, b_3), (0, b_1 b_2), (1, b_1 b_1 b_2 b_1), \\ (1, b_1), (1, b_2), (1, b_3), (1, b_1 b_2)\}$$

and for overlap

$$\hat{\Omega} = \{\emptyset, (0, b_1), (0, b_2), (0, b_3), (0, b_1 b_2), (1, b_1 b_2 b_1), \\ (1, b_1), (1, b_2), (1, b_3), (1, b_1 b_2), (2, b_1 b_2 b_1)\}$$

Notice that in the case of non-overlap counting the state  $(1, b_1 b_2 b_1)$  goes in  $(1, b_2)$  for  $X_t = b_2$  while for overlap counting it goes to  $(1, b_1 b_2)$ , and the additional state  $(2, b_1 b_2 b_1)$  appears ( $\hat{l}_n = 2$ ). Now we give the transition probability matrices:

- for non-overlap counting

$$M = \begin{matrix} \emptyset \\ (0, b_1) \\ (0, b_2) \\ (0, b_3) \\ (0, b_1 b_2) \\ (1, \Lambda) \\ (1, b_1) \\ (1, b_2) \\ (1, b_3) \\ (1, b_1 b_2) \end{matrix} \begin{pmatrix} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & 0 & p_{13} & p_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{13} & 0 & p_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & p_{12} & p_{13} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & 0 & p_{13} & p_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} & p_{23} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} & p_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- for overlap counting

$$\hat{M} = \begin{matrix} \emptyset \\ (0, b_1) \\ (0, b_2) \\ (0, b_3) \\ (0, b_1 b_2) \\ (1, \Lambda) \\ (1, b_1) \\ (1, b_2) \\ (1, b_3) \\ (1, b_1 b_2) \\ (2, \Lambda) \end{matrix} \begin{pmatrix} 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & 0 & p_{13} & p_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{31} & p_{32} & p_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{13} & 0 & p_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & 0 & p_{13} & p_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & 0 & p_{13} & p_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} & p_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{22} & p_{23} & 0 & p_{21} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The vector  $(p_1, p_2, p_3)$  corresponds to the initial probability vector given that at time 0 we are in the dummy state, that is  $\mathbb{P}(Y_1 = b_j | Y_0 = \emptyset) = p_j$  for  $j = 1, 2, 3$ . We notice that in the case of overlap scheme the transition matrix row corresponding to the state  $(1, \Lambda)$  is different from the one in the case of non-overlap transition matrix.

The following example shows how to compute the distribution, mean and generating function of a waiting time random variable associated with a compound pattern.



Next table gives some selected values for the waiting time probabilities:

$n$	$\mathbb{P}(W(\Lambda) = n)$	$\mathbb{P}(W(\Lambda) \geq n)$
3	0.009550000000000	1.000000000000000
5	0.019575760000000	0.972953250000000
10	0.017692272475628	0.878888187133464
20	0.014436759278876	0.717160459635556
25	0.013041004221151	0.647824930843360
30	0.011780191649146	0.585192805018191
50	0.007843639732009	0.389640650426870
100	0.002837467150353	0.140954019282224

and the mean is  $\mathbb{E}[W(\Lambda)] = 52.327942042091962$  (using the relation described in Theorem 3.1). It is worth mentioning that in the context of Theorem 3.6, we have  $\lambda_1 \approx 0.9798$  and taking  $n = 1000$  the error is:

$$\epsilon = \mathbb{P}(W(\Lambda) \geq 1000) - e^{1000 \log(\lambda_1)} \approx 1.1356 \times 10^{-10}.$$

The probability generating function is computed using the relation from Theorem 3.1 and is given by:

$$\varphi_W(s) = 1 + (s - 1)\boldsymbol{\xi}(\mathbf{I} - s\mathbf{N})^{-1}\mathbf{1}^\top = \frac{\Delta_1}{\Delta_2}$$

where

$$\begin{aligned} \Delta_1 = & -s^2(2519621181s^8 + 21622505850s^7 - 109947108150s^6 + 697873893750s^5 + \\ & + 21153645 \times 10^5s^4 - 3066425 \times 10^6s^3 + 48085 \times 10^9s^2 + 61 \times 10^{12}s - \\ & - 125 \times 10^{11}) \end{aligned}$$

and

$$\begin{aligned} \Delta_2 = & -16871893731s^9 + 134865799200s^8 - 579258098100s^7 + 337339728 \times 10^4s^6 \\ & - 95999415 \times 10^5s^5 + 704918 \times 10^8s^4 + 5345 \times 10^{10}s^3 - 5135 \times 10^{11}s^2 + \\ & + 53 \times 10^{14}s - 5 \times 10^{15} \end{aligned}$$

The above formulas were obtained in Matlab and if we compute the derivative of the generating function in  $s = 1$ , then we get the result for the mean. The variance can be obtained by:

$$\text{Var}[W(\Lambda)] = \mathbb{E}[W^2(\Lambda)] - \mathbb{E}^2[W(\Lambda)]$$

using

$$\mathbb{E}[W^2(\Lambda)] = \left. \frac{d^2\varphi_W(s)}{ds^2} \right|_{s=0}$$

The result is  $Var[W(\Lambda)] = 2365.67691827216$ . Another method to compute the variance is given by Theorem 3.3.

Another way of finding the p.g.f. is by using the Theorem 3.2, which leads to the following system:

$$\left\{ \begin{array}{l} \phi_1(s) = s \\ \phi_2(s) = \phi_2(s)p_{AA} + \phi_3(s)p_{CA} + \phi_4(s)p_{GA} + \phi_5(s)p_{TA} + \phi_7(s)p_{GA} + \phi_8(s)p_{CA} + \\ \quad + \phi_9(s)p_{GA} + \phi_{10}(s)p_{TA} + p_{AS} \\ \phi_3(s) = \phi_3(s)p_{CC} + \phi_4(s)p_{GC} + \phi_5(s)p_{TC} + \phi_6(s)p_{CC} + \phi_8(s)p_{CC} + \phi_9(s)p_{GC} + \\ \quad + \phi_{10}(s)p_{TC} + p_{CS} \\ \phi_4(s) = \phi_2(s)p_{AG} + \phi_5(s)p_{TG} + \phi_{10}(s)p_{TG} + p_{GS} \\ \phi_5(s) = \phi_2(s)p_{AT} + \phi_3(s)p_{CT} + \phi_4(s)p_{GT} + \phi_5(s)p_{TT} + \phi_6(s)p_{CT} + \phi_7(s)p_{GT} + \\ \quad + \phi_8(s)p_{CT} + p_{TS} \\ \phi_6(s) = \phi_2(s)p_{AC} \\ \phi_7(s) = \phi_4(s)p_{GG} + \phi_7(s)p_{GG} + \phi_9(s)p_{GG} \\ \phi_8(s) = \phi_7(s)p_{GC} \\ \phi_9(s) = \phi_3(s)p_{CG} + \phi_6(s)p_{CG} \\ \phi_{10}(s) = \phi_9(s)p_{GT} \end{array} \right.$$

The system was deduced using MuPAD (symbolic notebook from Matlab).

Next example will show how to construct the state space and the transition matrix in the case of the  $r$ -th waiting time of a compound patter.

**Example 4.4.** For simplicity, lets take  $\{X_t|t = 1, 2, \dots\}$  to be a sequence of Markov-dependent two-state trials over the alphabet  $\mathcal{S} = \{1, 2\}$ , with initial probability vector  $(p_1, p_2, p_3)$ , and transition matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

Lets take the compound pattern  $\Lambda = \Lambda_1 \cup \Lambda_2$ , where  $\Lambda_1 = 121$  and  $\Lambda_2 = 22$ . We will define the state space of the imbedded Markov chain associated with the waiting time  $W(2, \Lambda)$  as described in Remark 2.4:

$$\Omega^{(2)} = \{\emptyset\} \cup \Omega_1 \cup \Omega_2 \cup \{\alpha_1, \alpha_2\}$$

where  $\alpha_1, \alpha_2$  correspond to the second occurrence of the patterns  $\Lambda_1, \Lambda_2$ , and

$$\begin{aligned} \Omega_1 &= \{(0, 1), (0, 2), (0, 12), (1, 1), (1, 2), (1, 12)\} \\ \Omega_2 &= \{(0, 1^*), (1, 2^*)\} \end{aligned}$$



Notice that  $\mathcal{B} = \{1^*, 2^*\}$  since the last symbols in the patterns  $\Lambda_1$  and  $\Lambda_2$  are  $\{1, 2\}$ , and  $(1, 1^*)$  means that we have one occurrence of the pattern  $\Lambda$  with the ending symbol 1 (just occurred for the first time). The corresponding transition matrix (under non-overlapping counting),  $\mathbf{M}^{(2)}$  is given by:

$$\mathbf{M}^{(2)} = \begin{pmatrix} \emptyset & (0, 1) & (0, 2) & (0, 12) & (1, 1^*) & (1, 2^*) & (1, 1) & (1, 2) & (1, 12) & \alpha_1 & \alpha_1 \\ 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & 0 & p_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{21} & 0 & 0 & 0 & p_{22} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{21} & p_{22} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & p_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{11} & 0 & p_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & 0 & 0 & 0 & p_{22} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{22} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We will end this section with an example dedicated to the scan statistic in a sequence of two-state trials, and is based primarily on [6].

**Example 4.5.** In this example we will show how to compute the exact distribution of the scan statistic for a sequence of two-state trials generated by a Markov source, using the MCIT. Lets suppose that we have a sequence of homogeneous Markov-dependent two-state trails (0 and 1) with probability transition matrix:

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

and consider the scan statistic of window size  $r$  defined as:

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

In [13] it is showed that the scan statistic is an imbeddable random variable in the sense Definition 1.4, but the transition matrix corresponding to the associated chain is rather big (has order  $2^r$ ). The next approach is based on [6]. The idea is to express  $S_n(r)$  in terms of the waiting time distribution of a special compound pattern. For



where  $(q, p)$  is the initial probability vector. Finally we get

$$\mathbb{P}(S_n(r) < k) = \boldsymbol{\xi} \mathbf{N}^n \mathbf{1}^\top,$$

where  $\boldsymbol{\xi} = (1, 0, \dots, 0)$  and  $\mathbf{N}$  is the essential matrix.

In the following table, some selected values for  $n, r, k$  and  $\mathbb{P}(S_n(r) < k)$  are given, both in i.i.d. and Markov case:

$n$	$r$	$k$	$p = q = 0.5$	$p_{11} = 0.35, p_{21} = 0.45$
100	5	2	$2.0002 \times 10^{-18}$	$1.6621 \times 10^{-14}$
		3	$1.1864 \times 10^{-9}$	$2.1371 \times 10^{-6}$
	7	4	$1.8407 \times 10^{-7}$	$1.2684 \times 10^{-4}$
		5	$8.7094 \times 10^{-4}$	0.0522
	10	5	$3.4100 \times 10^{-7}$	$1.9939 \times 10^{-4}$
		6	$2.3130 \times 10^{-4}$	0.0231
300	5	2	$3.3094 \times 10^{-54}$	$2.2786 \times 10^{-42}$
		3	$7.7147 \times 10^{-28}$	$5.8484 \times 10^{-18}$
	7	4	$2.5286 \times 10^{-21}$	$1.1610 \times 10^{-12}$
		5	$3.9156 \times 10^{-10}$	$1.1228 \times 10^{-4}$
	10	5	$1.1746 \times 10^{-20}$	$3.6584 \times 10^{-12}$
		6	$5.4061 \times 10^{-12}$	$8.0995 \times 10^{-6}$
500	5	2	$5.4756 \times 10^{-90}$	$3.1237 \times 10^{-70}$
		3	$5.0167 \times 10^{-46}$	$1.6005 \times 10^{-29}$
	7	4	$3.4735 \times 10^{-35}$	$1.0627 \times 10^{-20}$
		5	$1.7604 \times 10^{-16}$	$2.4150 \times 10^{-7}$
	10	5	$4.0461 \times 10^{-34}$	$6.7122 \times 10^{-20}$
		6	$1.2635 \times 10^{-19}$	$2.8431 \times 10^{-9}$

All the values from the above table, were obtained using MATLAB<sup>®</sup>.

## 5. APPENDIX

In this section we will give the matlab code used for computations in Example 4.3 and Example 4.5.

```

1 %Waiting time distribution for compound pattern
2 %=====
3 %number of letters in the alphabet
4 m=input('Give the number of symbols in the alphabet:');

```

```

5 l=input('Give the number of patterns:');
6 LL='';
7 for i=1:l
8     L=input(['Give the ',num2str(i),' pattern:'],'s');
9     LL=char(LL,L);
10 end
11 %the collection of given patterns in column format
12 LL=LL(2:end,:);
13 sz=size(LL);
14 sz=sz(2);
15 for i=1:l
16     for j=1:sz
17         if str2double(LL(i,j))>m
18             disp(['The pattern in not correct since the symbol ',...
19                 LL(i,j),' is not in the alphabet']);
20             break;
21         end
22     end
23 end
24 %=====
25 %defining the state space
26 %=====
27 S=[];
28 for k=1:m
29     S=char(S,num2str(k));
30 end
31 S=S(2:end,:); %the alphabet
32 omega=[];
33 for i=1:l
34     L=strtrim(LL(i,:));
35     SL=[];
36     if length(L)<=2
37         SL=[];
38     else
39         for j=2:length(L)-1
40             SL=char(SL,L(1:j));
41         end
42         SL=SL(2:end,:);
43     end
44     if isempty(SL)==0
45         omega=char(omega,SL);
46     end
47 end
48 omega=omega(2:end,:);
49 A=cellstr(LL);
50 c=[];
51 for i=1:length(omega)-1
52     if isempty(strmatch(omega(i,:),omega((i+1):end,:),...
53         'exact'))==0

```

```

54         v=strmatch(omega(i,:),omega((i+1):end,:),'exact');
55         v=i+v;
56         c=[c,v'];
57     end
58 end
59 omega(c,:)=[];
60 omega=char(S,omega,LL);
61 %=====
62 %Initial conditions
63 %=====
64 %the initial probability vector
65 ini=zeros(1,m);
66 disp('Give the initial probability vector:');
67 for i=1:m
68     ini(i)=input(['ini(',num2str(i),')=']);
69 end
70 if sum(ini)~=1
71     disp('The initial vector is wrong: the sum is not equal with 1.');
```

```

72     break;
73 end
74 %the transition matrix corresponding to the initial process
75 P=zeros(m);
76 disp('Give the transition matrix P:');
77 for i=1:m
78     for j=1:m
79         P(i,j)=input(['P(',num2str(i),',',num2str(j),')=']);
80     end
81 end
82 for s=1:m
83     if sum(P(s,:))~=1
84         disp(['The transition matrix is wrong since the sum on the ',...
85             num2str(s),' line is not equal with 1.']);
86         break;
87     end
88 end
89 k=length(omega)+1;
90 M=zeros(k,k);
91 %=====
92 %creating the transition matrix M
93 %for non-overlapping case
94 %=====
95 tic
96 for i=1:k
97     for j=1:m
98         if i==k
99             M(1,j+1)=ini(j);
100             continue;
101         end
102         if strcmp(strtrim(omega(i,:)),A)==0
```

```

103         state=strcat(strtrim(omega(i,:)),num2str(j));
104         if isempty(strmatch(state,omega,'exact')==1
105             g=str2mat(state);
106             ll=length(g);
107             h=zeros(1,ll-1);
108             for t=ll-1:-1:1
109                 if isempty(strmatch(g(t:1:end),omega,'exact')==1
110                     state=num2str(j);
111                 else
112                     h(ll-t)=length(g(t:1:end));
113                 end
114             end
115             for t=1:ll-1
116                 if h(t)==max(h)
117                     u=t;
118                 end
119             end
120             if isempty(strmatch(g(ll-u:1:end),omega,'exact')==0
121                 state=g(ll-u:1:end);
122             end
123         end
124     else
125         state=num2str(j);
126     end
127     v=strmatch(state,omega,'exact');
128     u=strtrim(omega(i,:));
129     M(i+1,v+1)=P(str2double(u(end)),str2double(state(end)));
130 end
131 end
132 toc
133 %=====
134 %computing the waiting time distribution
135 %=====
136 N=M(1:(end-1),1:(end-1));
137 N=sparse(N);
138 T=zeros(k);
139 T(1:(end-1),1:(end-1))=N;
140 T((end-1+1):end,(end-1+1):end)=eye(1); %full transition matrix
141 xsi=zeros(1,k-1);
142 xsi(1)=1;
143 xsi=sparse(xsi);
144 I=speye(k-1,k-1);
145 V=I-N;
146 un=ones(k-1,1);
147 PWn=zeros(length(1:50),1);
148 MM=0;
149 tic
150 %exact distribution of the waiting time
151 for i=1:50

```

```

152     Pwn(i,1)=xsi*(N^(i-1))*un-xsi*(N^i)*un;
153     MM=MM+i*Pwn(i,1);
154 end
155 Mn=xsi*(V\un); %mean
156 toc

```

```

1 %Computing scan statistic in a 2 state trails
2 %=====
3 %n=input('Give the number of trials:');
4 r=input('Give the size of the window:');
5 k=input('Give k:');
6 tic
7 if k>r
8     disp('It is not good!');
9     break;
10 else
11     LL=[];
12     for j=k:r
13         L=ones(1,j);
14         for s=2:j-1
15             if sum(L(2:j-1))≠k+2*(j-k)-2
16                 if L(s)==1
17                     L(s)=2;
18                 end
19             end
20         end
21         v=perms(L(2:j-1));
22         v=unique(v,'rows');
23         sv=size(v);
24         un=ones(sv(1),1);
25         d=[un,v,un];
26         sd=size(d);
27         for i=1:sd(1)
28             J=[];
29             for t=1:sd(2)
30                 J=strcat(J,num2str(d(i,t)));
31             end
32             LL=char(LL,J);
33         end
34     end
35 end
36 toc
37 LL=LL(2:end,:);
38 %all corresponding patterns associated with (r,k)
39 LL=unique(LL,'rows');
40 sLL=size(LL);

```

```

41 l=sLL(1);
42 %=====
43 %defining the state space
44 %=====
45 S=[];
46 for k=1:2
47     S=char(S,num2str(k));
48 end
49 S=S(2:end,:); %the alphabet
50 omega=[];
51 for i=1:l
52     L=strtrim(LL(i,:));
53     SL=[];
54     if length(L)≤2
55         SL=[];
56     else
57         for j=2:length(L)-1
58             SL=char(SL,L(1:j));
59         end
60         SL=SL(2:end,:);
61     end
62     if isempty(SL)==0
63         omega=char(omega,SL);
64     end
65 end
66 omega=omega(2:end,:);
67 A=cellstr(LL);
68 c=[];
69 for i=1:length(omega)-1
70     if isempty(strmatch(omega(i,:),omega((i+1):end,),'exact'))==0
71         v=strmatch(omega(i,:),omega((i+1):end,),'exact');
72         v=i+v;
73         c=[c,v'];
74     end
75 end
76 omega(c,:)=[];
77 omega=char(S,omega,LL); %the state space
78 %=====
79 %Initial conditions
80 %=====
81 %the initial probability vector
82 ini=zeros(1,2);
83 disp('Give the initial probability vector:');
84 for i=1:2
85     ini(i)=input(['ini(',num2str(i),')=']);
86 end
87 if sum(ini)≠1
88     disp('The initial vector is wrong: the sum is not equal with 1.');
```

```
89     break;
```



```

90 end
91 %the transition matrix corresponding to the initial process
92 P=zeros(2);
93 disp('Give the transition matrix P:')
94 for i=1:2
95     for j=1:2
96         P(i,j)=input(['P(',num2str(i),',',num2str(j),')=']);
97     end
98 end
99 for s=1:2
100     if sum(P(s,:))≠1
101         disp(['The transition matrix is wrong since the sum on the ',...
102             num2str(s), ' line is not equal with 1.']);
103         break;
104     end
105 end
106 lo=length(omega)+1;
107 M=zeros(lo,lo);
108 %=====
109 %creating the transition matrix M
110 %for non-overlapping case
111 %=====
112 tic
113 for i=1:lo
114     for j=1:2
115         if i==lo
116             M(1,j+1)=ini(j);
117             continue;
118         end
119         if strcmp(strtrim(omega(i,:)),A)==0
120             state=strcat(strtrim(omega(i,:)),num2str(j));
121             if isempty(strmatch(state,omega,'exact'))==1
122                 g=str2mat(state);
123                 ll=length(g);
124                 h=zeros(1,ll-1);
125                 for t=ll-1:-1:1
126                     if isempty(strmatch(g(t:1:end),omega,'exact'))==1
127                         state=num2str(j);
128                     else
129                         h(ll-t)=length(g(t:1:end));
130                     end
131                 end
132                 for t=1:ll-1
133                     if h(t)==max(h)
134                         u=t;
135                     end
136                 end
137                 if isempty(strmatch(g(ll-u:1:end),omega,'exact'))==0
138                     state=g(ll-u:1:end);

```

```
139         end
140     end
141     else
142         state=num2str(j);
143     end
144     v=strmatch(state,omega,'exact');
145     u=strtrim(omega(i,:));
146     M(i+1,v+1)=P(str2double(u(end)),str2double(state(end)));
147     end
148 end
149 toc
150 %=====
151 %computing the scan statistic distribution
152 %=====
153 N=M(1:(end-1),1:(end-1));
154 N=sparse(N);
155 clear M A J L LL SL c d g h i j k ll omega sLL state sv t u v
156 %T=zeros(lo);
157 %T(1:(end-1),1:(end-1))=N;
158 %T((end-1+1):end,(end-1+1):end)=eye(1); %full transition matrix
159 xsi=zeros(1,lo-1);
160 xsi(1)=1;
161 xsi=sparse(xsi);
162 un=ones(lo-1,1);
163 tic
164 for j=[100 300 500]
165     Sn=xsi*(N^j)*un; % an attempt for computin the scan statistic
166     disp(Sn);
167 end
168 toc
```

## REFERENCES

- [1] Boutsikas, M.V., Koutras, M.V.: Reliability approximation for Markov chain imbeddable systems. *Methodology and Computing in Applied Probability* **2** (2000), 393–411.
- [2] Chang, Y.M.: Distribution of waiting time until the  $r^{\text{th}}$  occurrence of a compound pattern. *Statistics & Probability Letters* **75** (2005), 29–38.
- [3] Chang, Y.M.: Waiting times distributions of runs and patterns. *Phd. Thesis* (2002).
- [4] Fu, J., Koutras, M.V.: Distribution theory of runs: a Markov chain approach. *J. American Statistical Association* **89** (1994), 1050–1056.
- [5] Fu, J.: Distribution theory of runs and patterns associated with a sequence of multistate trials. *Statistica Sinica* **6** (1996), 957–974.
- [6] Fu, J.: Distribution of the scan statistic for a sequence of bistate trials. *J. Appl. Prob.* **38** (2001), 908–916.
- [7] Fu, J., Chang, Y.M.: On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials. *J. Appl. Prob.* **39** (2002), 70–80.
- [8] Fu, J., Wang, L., Wendy Lou, W.Y.: On exact and large deviation approximation for the distribution of the longest run in a sequence of two state Markov dependent trials. *J. Appl. Prob.* **40** (2003), 346–360.
- [9] Fu, J., Wendy Lou, W.Y.: Distribution theory of runs and patterns and its applications: A Markov chain imbedding approach. *World Scientific-Singapore* (2003).
- [10] Fu, J., Chang, Y.M.: On ordered series and later waiting time distributions in a sequence of Markov dependent multistate trials. *J. Appl. Prob.* **40** (2003), 623–642.
- [11] Fu, J., Wendy Lou, W.Y.: Waiting time distributions of simple and compound patterns in a sequence of  $r^{\text{th}}$  order Markov dependent multistate trials *The Institute of Statistical Mathematics, Tokyo* (2006), 291–310.
- [12] Fu, J., Johnson, B.: Approximate probabilities for runs and patterns in i.i.d. and Markov dependent multistate trials. *Advances in Applied Probability* **41**(1) (2009).
- [13] Koutras, M.V., Alexandrou, V.A.: Runs, scans and urn model distributions: a unified Markov chain approach. *Ann. Inst. Statist. Math.* **47** (1995), 743–766.
- [14] Lin, H.Y., Chang, Y.M.: Duality of probabilities of runs and patterns. *Master Thesis* (2005).
- [15] Nuel, G.: On the first k moments of the random count of a pattern in a multi-scale sequence generated by a Markov source *J. Appl. Prob* **47** (2010), 1105–1123.
- [16] Te, L.T., Wu, C.T.: The waiting time for simple pattern with overlapping and nonoverlapping schemes. *Master Thesis* (2006).
- [17] Waiting time problems for a compound pattern. *Master Thesis* (2004).
- [18] Wendy Lou, W.Y.: On runs and longest run tests: a method of finite Markov chain imbedding. *J. American Statistical Association* **91** (1996), 1595–1601.