J. Appl. Prob. **51,** 287–292 (2014)
Printed in England

STOPPING PROBABILITIES FOR PATTERNS IN MARKOV CHAINS

RENATO JACOB GAVA,* *Universidade de São Paulo*
DANILO SALOTTI,** *Fundação Educacional Inaciana Padre Sabóia de Medeiros*

Abstract

Consider a sequence of Markov-dependent trials where each trial produces a letter of a finite alphabet. Given a collection of patterns, we look at this sequence until one of these patterns appears as a run. We show how the method of gambling teams can be employed to compute the probability that a given pattern is the first pattern to occur.

Keywords: Stopping time; waiting time; gambling team; martingale; Markov chain; pattern; stopping probability

2010 Mathematics Subject Classification: Primary 60J10
Secondary 60G42

1. Introduction

Let $\{Z_n\}_{n\geq 1}$ be a homogeneous Markov chain on a finite set Ω, which we call an alphabet. Let us call a pattern a finite sequence of elements of Ω. We then consider a finite collection $\mathcal{C} = \{A_1, A_2, \dots, A_K\}$ of patterns, possibly with different lengths. Let τ_{A_q} be the waiting time until A_q occurs as a run in the series $Z_1, Z_2, \dots$. Define the stopping time

$$\tau = \min\{\tau_{A_1}, \dots, \tau_{A_K}\}. \tag{1.1}$$

Many authors have studied waiting time problems for specific and general choices of $\mathcal{C}$ and their probability generating functions. Several distinct techniques have been used to solve these problems for both independent or Markov-dependent trials; see [1], [2], [4], [5], [6], [7], [8], and the references therein. We are interested in computing the stopping probabilities

$$\mathbb{P}(\tau = \tau_{A_q}), \qquad q = 1, \dots, K. \tag{1.2}$$

In order to well define (1.1) and avoid ties, we assume that no pattern from $\mathcal{C}$ contains another pattern as a subpattern. We shall show that the martingale methods introduced in [5] and [8], and further developed in [6], [9], and [10] may also be applied to compute stopping probabilities of a sequence of patterns for finite-state Markov chains.

Feller [3] studied the occurrence of patterns in independent Bernoulli trials using recurrent event theory. In a more general setting, when the trials are independent and identically distributed discrete random variables, Li [8] elegantly proposed the martingale approach to waiting time problems and provided a means of computing (1.2) and $\mathbb{E}(\tau)$ for any collection

Received 12 December 2012; revision received 26 March 2013.
* Postal address: Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, CEP 05508-090, São Paulo, SP, Brazil. Email address: gava@ime.usp.br
Research supported by FAPESP fellowship 2012/01432-9.
** Postal address: Centro Universitário da FEI, Av. Humberto de Alencar Castelo Branco 3972, CEP 09850-901, São Bernardo do Campo, SP, Brazil. Email address: dsalotti@fei.edu.br

$\mathcal{C}$ of patterns. In [1] the mean waiting time and stopping probabilities were obtained for a $\mathcal{C}$ whose patterns have the same length and trials are independent, identically distributed, uniform, N-state random variables. In [11] run probabilities were computed as a function of the number n of trials for a collection of runs, but not as a function of the variable in (1.1) or (1.2). In this paper we bring together the methods developed for gambling teams in [6] and [9] to obtain the probabilities in (1.2). First, in Section 2 we use the procedure developed in [6] to get (1.2) for a two-state Markov chain. In Section 3 we apply the algorithm stated in [9] to get (1.2) for a multistate Markov chain. Both methods and their results are similar, but even when we consider only two-state Markov chains, these methods show us that the way we calculate the probabilities in (1.2) are different.

2. Stopping probabilities in a two-state Markov chain

Let $\Omega = \{S, F\}$ be our two-letter alphabet. Let $\{Z_n\}_{n\geq 1}$ be a time-homogeneous, two-state Markov chain in Ω with initial distributions $\mathbb{P}(Z_1 = S) = p_S$ and $\mathbb{P}(Z_1 = F) = p_F$, and transition matrix

$$\begin{bmatrix} p_{SS} & p_{SF} \\ p_{FS} & p_{FF} \end{bmatrix},$$

where $p_{SF} = \mathbb{P}(Z_{n+1} = F \mid Z_n = S)$. We also assume that

$$0 < p_{SS} < 1 \quad \text{and} \quad 0 < p_{FF} < 1. \tag{2.1}$$

The assumptions in (2.1) imply that $\mathbb{P}(\tau_A < \infty) = 1$ for any pattern A, which in turn implies that $\mathbb{E}(\tau) < \infty$.

We apply here the same procedure and notation as given in [6]. We invite the reader to look at Sections 3.1 and 3.2 of that article to understand the construction that follows.

Let X_n be the amount of money that the casino saves at the end of round n. Let $y_j W_{ij}$ be the amount of money that the jth team earns when the ith ending scenario occurs. By the rules of betting, it is clear that $X_1 = 0$. Moreover, it is not hard to see that $\{X_n\}_{n\geq 1}$ is a martingale with respect to the filtration $\{Z_n\}_{n\geq 1}$ and that at the moment τ we have

$$X_\tau = \sum_{j=1}^{L+2M} y_j(\tau - 1) - S(y_1, \dots, y_{L+2M}) \tag{2.2}$$

with

$$S(y_1, \dots, y_{L+2M}) = \sum_{i=1}^{K+L+2M} \mathbf{1}_{E_i} \sum_{j=1}^{L+2M} y_j W_{ij},$$

where E_i stands for the event that the ith scenario occurs and $\mathbf{1}_{E_i}$ is its indicator function. A general method to compute the profit matrix $W = \{W_{ij}\}$ is given in Section 3.3 of [6].

For $i = 1, \dots, K + L + 2M$, let $\mu_i = \mathbb{P}(E_i)$ be the probability of occurrence of the ith ending scenario. Suppose that $(y_1^*, \dots, y_{L+2M}^*)$ is a solution of the linear system

$$y_1^* W_{i,1} + \cdots + y_{L+2M}^* W_{i,L+2M} = 1 \quad \text{for } i \in \{K+1, \dots, K+L+2M\}. \tag{2.3}$$

Theorem 2.1. ([6, Theorem 1].) *If $(y_1^*, \dots, y_{L+2M}^*)$ solves the linear system (2.3) then*

$$\mathbb{E}(\tau) = 1 + \frac{\sum_{i=1}^{K} \mu_i \sum_{j=1}^{L+2M} y_j^* W_{ij} + (1 - \sum_{i=1}^{K} \mu_i)}{\sum_{j=1}^{L+2M} y_j^*}. \tag{2.4}$$

Observe that the expression in (2.4) depends on the μ_is which are easily calculated and that

$$\mathbb{P}(\tau = \tau_{A_q}) = \mu_q + \mathbb{P}(B \in \mathcal{C}'' \text{ such that the game ends with } B \text{ and } B \text{ is associated to } A_q).$$

Let $B_l \in \mathcal{C}''$ be a pattern associated to A_q. It corresponds to some ending scenario E_r, where $K < r \le K + L + 2M$. In addition, let us write $\mathbb{P}(E_r) = \mu_r^{(q)}$ just to emphasize the link between E_r and A_q. Next assume that $(z_1, \dots, z_{L+2M})$ is a solution of the linear system

$$\begin{aligned} z_1 W_{1,1} + \cdots + z_{L+2M} W_{1,L+2M} &= 1, \\ z_1 W_{i,1} + \cdots + z_{L+2M} W_{i,L+2M} &= 1 \quad \text{for } i \in \{K+1, \dots, K+L+2M\} \setminus \{r\}. \end{aligned} \tag{2.5}$$

Then

$$S(z_1, \dots, z_{L+2M}) = 1\,\mathbf{1}_{E_1} + \sum_{i=2}^{K} \mathbf{1}_{E_i} \sum_{j=1}^{L+2M} z_j W_{ij} + \sum_{i>K,\, i \ne r} \mathbf{1}_{E_i} + \mathbf{1}_{E_r} \sum_{j=1}^{L+2M} z_j W_{rj}.$$

As $\mathbb{E}(\tau) < \infty$ and the sequence $\{X_n\}_{n\ge1}$ has bounded increments, we can apply the optional stopping theorem (see [12, p. 100]) and take expectations of both sides of (2.2) for $S(z_1, \dots, z_{L+2M})$. Since $0 = \mathbb{E}(X_1) = \mathbb{E}(X_\tau)$,

$$0 = \sum_{j=1}^{L+2M} z_j(\mathbb{E}(\tau) - 1) - \sum_{i=2}^{K} \mu_i \sum_{j=1}^{L+2M} z_j W_{ij} - \left(1 - \sum_{i=2}^{K} \mu_i\right) + \mu_r^{(q)}\left(1 - \sum_{j=1}^{L+2M} z_j W_{rj}\right),$$

which in turn implies that

$$\mu_r^{(q)} = \frac{\sum_{j=1}^{L+2M} z_j(1 - \mathbb{E}(\tau)) - \sum_{i=2}^{K} \mu_i(1 - \sum_{j=1}^{L+2M} z_j W_{ij}) + 1}{1 - \sum_{j=1}^{L+2M} z_j W_{rj}}. \tag{2.6}$$

Theorem 2.2. *Let $\{Z_n\}_{n\ge1}$ be a homogeneous Markov chain on $\{S, F\}$. Consider a pattern A_q.*

(i) *If A_q is associated to an unmatched pattern $B_l \in \mathcal{C}''$ corresponding to the rth ending scenario of the game, $r > K$, then*

$$\mathbb{P}(\tau = \tau_{A_q}) = \mu_q + \frac{\sum_{j=1}^{L+2M} z_j(1 - \mathbb{E}(\tau)) - \sum_{i=2}^{K} \mu_i(1 - \sum_{j=1}^{L+2M} z_j W_{ij}) + 1}{1 - \sum_{j=1}^{L+2M} z_j W_{rj}},$$

where $(z_1, \dots, z_{L+2M})$ is a solution of (2.5).

(ii) *If A_q is associated to a pair of matched patterns $B_m, B_p \in \mathcal{C}''$ respectively corresponding to the sth and tth ending scenarios of the game, $s, t > K$, then*

$$\mathbb{P}(\tau = \tau_{A_q}) = \mu_q + \mu_s^{(i)} + \mu_t^{(i)},$$

where $(x_1, \dots, x_{L+2M})$ satisfies (2.5) and (2.6) with s rather than r and $(w_1, \dots, w_{L+2M})$ satisfies (2.5) and (2.6) with t rather than r.

Remark 2.1. We could readily apply Theorems 2.1 and 2.2 for instance to the well-known problem of Feller [3] in which A_1 is a run of α consecutive successes and A_2 is a run of β failures. The formulae for $\mathbb{P}(\tau = \tau_{A_1})$ and $\mathbb{E}(\tau)$ are too long and so we do not present them here. However, when $\{Z_n\}_{n\geq 1}$ is a sequence of independent Bernoulli random variables with $\mathbb{P}(Z_n = S) = p$ and $\mathbb{P}(Z_n = F) = q$, the omitted formulae reduce to the expressions found in Chapters VIII.1 and XIII.8 of [3]:

$$\mathbb{P}(\tau = \tau_{A_1}) = \frac{p^{\alpha-1}(1-q^{\beta})}{p^{\alpha-1}+q^{\beta-1}-p^{\alpha-1}q^{\beta-1}} \quad \text{and} \quad \mathbb{E}(\tau) = \frac{(1-p^{\alpha})(1-q^{\beta})}{p^{\alpha}q+pq^{\beta}-p^{\alpha}q^{\beta}}.$$

3. Stopping probabilities in a multistate Markov chain

In this section the state space is $\Omega = \{1, \dots, N\}$. Let $\{Z_n\}_{n\geq 1}$ be a homogeneous Markov chain on Ω. The initial distribution is $\mathbb{P}(Z_1 = i) = p_i$ and the transition matrix is $\{p_{ij}\}$, where $p_{ij} = \mathbb{P}(Z_{n+1} = j \mid Z_n = i)$. Since we do not impose assumptions on the transition probabilities as in (2.1), we need to make the following assumptions on τ.

1. For $q = 1, \dots, K$, $\mathbb{P}(\tau = \tau_{A_q}) > 0$.
2. $\tau < \infty$ almost surely.

Remarks on these assumptions can be found in Section 2 of [9]. We just add the remark that assumptions 1 and 2 together ensure that $\sum_q \mathbb{P}(\tau = \tau_{A_q}) = 1$.

We now use the method of gambling teams given in [9]. Before continuing, we suggest that the reader view the details of Section 3 of [9] to understand what follows. Once more we work with $\mathcal{C} = \{A_q\}_{q=1}^{K}$, but we need to introduce notation not used in [9]. Define the sets

$$\mathcal{D}' = \{lA_q;\, l = 1, \dots, N\}_{q=1}^{K} \quad \text{and} \quad \mathcal{C}' = \{lmA_q;\, l, m = 1, \dots, N\}_{q=1}^{K}.$$

Denote by $\mathcal{D}''$ and $\mathcal{C}''$ the list of patterns after excluding from $\mathcal{D}'$ and from $\mathcal{C}'$, respectively, the patterns which can occur only after the waiting time τ.

Let $K' := K + |\mathcal{D}''|$ and $M' := |\mathcal{C}''|$. Note that here M' plays the role of N' in [9]. Let X_n be the amount of money that the casino saves at the end of round n. Let $y_j W_{ij}$ be the amount of money that the jth team earns when the ith ending scenario occurs. As before, we have $X_1 = 0$, and $\{X_n\}_{n\geq 1}$ is a martingale with respect to the filtration $\{Z_n\}_{n\geq 1}$. At the moment τ we have

$$X_{\tau} = \sum_{j=1}^{M'} y_j(\tau - 1) - S(y_1, \dots, y_{M'}) \tag{3.1}$$

with

$$S(y_1, \dots, y_{M'}) = \sum_{i=1}^{K'+M'} \mathbf{1}_{E_i} \sum_{j=1}^{M'} y_j W_{ij},$$

where E_i stands for the event that the ith scenario occurs and $\mathbf{1}_{E_i}$ is its indicator function. See Section 3 of [9] for a general explanation of how to calculate the profit matrix $W = \{W_{ij}\}$.

For $i = 1, \dots, K' + M'$, let $\mu_i = \mathbb{P}(E_i)$ be the probability of occurrence of the ith ending scenario. Suppose that $(y_1^*, \dots, y_{M'}^*)$ is a solution of the linear system

$$y_1^* W_{i,1} + \cdots + y_{M'}^* W_{i,M'} = 1 \quad \text{for } i \in \{K'+1, \dots, K'+M'\}. \tag{3.2}$$

https://doi.org/10.1239/jap/1395771430 Published online by Cambridge University Press

Theorem 3.1. ([9, Theorem 1].) *If* $(y_1^*, \dots, y_{M'}^*)$ *solves the linear system (3.2) then*

$$\mathbb{E}(\tau) = 1 + \frac{\sum_{i=1}^{K'} \mu_i \sum_{j=1}^{M'} y_j^* W_{ij} + (1 - \sum_{i=1}^{K'} \mu_i)}{\sum_{j=1}^{M'} y_j^*}.$$

We next note that

$$\mathbb{P}(\tau = \tau_{A_q}) = \mu_q + \mathbb{P}(B \in \mathcal{D}'' \cup \mathcal{C}'' \text{ such that game ends with } B \text{ and } B \text{ is associated to } A_q),$$

where $\mathcal{D}'' \cap \mathcal{C}'' = \varnothing$. For $r \in \{K+1, \dots, K', \dots, K'+M'\}$, write $\mathbb{P}(E_r) = \mu_r^{(q)}$ to emphasize the link between the rth ending scenario E_r and its associated pattern A_q, $q = 1, \dots, K$. Observe also that if a pattern $B_k \in \mathcal{D}''$ is generated by A_q and corresponds to the ending scenario u, with $K < u \le K'$, then $\mu_u^{(q)}$ is readily computable. In turn, let $B_l \in \mathcal{C}''$ be a pattern generated by A_q. It corresponds to some ending scenario E_r, where $K' < r \le K' + M'$. Next assume that $(z_1, \dots, z_{M'})$ is a solution of the linear system

$$\begin{gathered} z_1 W_{1,1} + \cdots + z_{M'} W_{1,M'} = 1, \\ z_1 W_{i,1} + \cdots + z_{M'} W_{i,M'} = 1 \quad \text{for } i \in \{K'+1, \dots, K'+M'\} \setminus \{r\}. \end{gathered} \tag{3.3}$$

Then

$$S(z_1, \dots, z_{M'}) = 1\,\mathbf{1}_{E_1} + \sum_{i=2}^{K'} \mathbf{1}_{E_i} \sum_{j=1}^{M'} z_j W_{ij} + \sum_{i > K'; i \neq r} \mathbf{1}_{E_i} + \mathbf{1}_{E_r} \sum_{j=1}^{M'} z_j W_{rj}.$$

Again, we apply the optional stopping theorem (see [12, p. 100]), and take the expectations of both sides of (3.1) for $S(z_1, \dots, z_{M'})$ to obtain

$$0 = \mathbb{E}(X_\tau) = \sum_{j=1}^{M'} z_j (\mathbb{E}(\tau) - 1) - \sum_{i=2}^{K'} \mu_i \sum_{j=1}^{M'} z_j W_{ij} - \biggl(1 - \sum_{i=2}^{K'} \mu_i\biggr) + \mu_r^{(q)} \biggl(1 - \sum_{j=1}^{M'} z_j W_{rj}\biggr),$$

which in turn implies that

$$\mu_r^{(q)} = \frac{\sum_{j=1}^{M'} z_j (1 - \mathbb{E}(\tau)) - \sum_{i=2}^{K'} \mu_i (1 - \sum_{j=1}^{M'} z_j W_{ij}) + 1}{1 - \sum_{j=1}^{M'} z_j W_{rj}}. \tag{3.4}$$

Theorem 3.2. *Let* $\{Z_n\}_{n \ge 1}$ *be a homogeneous Markov chain on* $\{1, \dots, N\}$. *For* A_q, $q = 1, \dots, K$, *suppose that the patterns* $B_k, \dots, B_{k+v} \in \mathcal{D}''$ *are associated to* A_q *and that they correspond to the ending scenarios* $u, u+1, \dots, u+v$. *Also, suppose that* $B_l, B_{l+1}, \dots, B_{l+s} \in \mathcal{C}''$ *are associated to* A_q *and that they correspond to the ending scenarios* $r, r+1, \dots, r+s$. *Then*

$$\mathbb{P}(\tau = \tau_{A_q}) = \mu_q + \mu_u^{(q)} + \cdots + \mu_{u+v}^{(q)} + \mu_r^{(q)} + \cdots + \mu_{r+s}^{(q)},$$

where each $\mu_l^{(q)}$, $l = r, \dots, r+s$, *demands a solution* $(z_1^l, \dots, z_{M'}^l)$ *for (3.3) with* j *rather than* r *and has the form given by (3.4).*

We now apply Theorem 3.2 to Example 1 of [9].

Example 3.1. Let $\Omega = \{1, 2, 3\}$, and let $\mathcal{C} = \{323, 313, 33\}$. Suppose that the initial distribution is $p_1 = p_2 = p_3 = \frac{1}{3}$ and that the transition matrix is

$$\begin{bmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

Then $\mathcal{D}'' = \{1323, 2323, 1313, 2313, 133, 233\}$ and $\mathcal{C}'' = \{11323, 22323, 11313, 22313, 1133, 2233\}$. Note that patterns such as $12323, 21323 \notin \mathcal{C}''$ are due to the absence of transitions $1 \to 2$ and $2 \to 1$.

Consider $A_1 = 323$, $A_2 = 313$, and $A_3 = 33$. In order to use Theorem 3.2, we need to compute the profit matrix W, the details of which we omit. After solving (3.2) and applying Theorem 3.1, we obtain $\mathbb{E}(\tau) = 8 + \frac{7}{15} = 8.466\,667$. With the profit matrix and $\mathbb{E}(\tau)$ determined, and having ordered the ending scenarios in the order that we have shown the sets $\mathcal{C}, \mathcal{D}'', \mathcal{C}''$, we use Theorem 3.2 to compute μ_i, $i = 10, \ldots, 15$, and obtain the stopping probabilities

$$\mathbb{P}(\tau = \tau_{A_1}) = \mu_1 + \mu_4^{(1)} + \mu_5^{(1)} + \mu_{10}^{(1)} + \mu_{11}^{(1)} = \frac{1}{48} + 2 \times \frac{1}{192} + 2 \times 0.034\,375 = \frac{1}{10},$$
$$\mathbb{P}(\tau = \tau_{A_2}) = \mu_2 + \mu_6^{(2)} + \mu_7^{(2)} + \mu_{12}^{(2)} + \mu_{13}^{(2)} = \frac{1}{48} + 2 \times \frac{1}{192} + 2 \times 0.034\,375 = \frac{1}{10},$$
$$\mathbb{P}(\tau = \tau_{A_3}) = \mu_3 + \mu_8^{(3)} + \mu_9^{(3)} + \mu_{14}^{(3)} + \mu_{15}^{(3)} = \frac{1}{6} + 2 \times \frac{1}{24} + 2 \times 0.275 = \frac{8}{10}.$$

References

[1] Blom, G. and Thorburn, D. (1982). How many random digits are required until given sequences are obtained? *J. Appl. Prob.* **19,** 518–531.
[2] Chrysaphinou, O. and Papastavridis, S. (1990). The occurrence of sequence patterns in repeated dependent experiments. *Theory Prob. Appl.* **35,** 145–152.
[3] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd edn. John Wiley, New York.
[4] Fu, J. S. and Chang, Y. M. (2002). On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials. *J. Appl. Prob.* **39,** 70–80.
[5] Gerber, H. U. and Li, S.-Y. R. (1981). The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stoch. Process. Appl.* **11,** 101–1086.
[6] Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Appl. Prob.* **43,** 127–140.
[7] Han, Q. and Aki, S. (2000). Waiting time problems in a two-state Markov chain. *Ann. Inst. Statist. Math.* **52,** 778–789.
[8] Li, S.-Y. R. (1980). A martigale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Prob.* **8,** 1171–1176.
[9] Pozdnyakov, V. (2008). On occurrence of patterns in Markov chain: method of gambling teams. *Statist. Prob. Lett.* **78,** 2762–2767.
[10] Pozdnyakov, V. and Kulldorff, M. (2006). Waiting times for patterns and a method of gambling teams. *Amer. Math. Monthly* **113,** 134–143.
[11] Schwager, S. J. (1983). Run probabilities in sequences of Markov-dependent trials. *J. Amer. Statist. Assoc.* **78,** 168–180.
[12] Williams, D. (1991). *Probability and Martigales*. Cambridge University Press.