Stochastic Processes and their Applications 11 (1981) 101–108
North-Holland Publishing Company

THE OCCURRENCE OF SEQUENCE PATTERNS IN REPEATED EXPERIMENTS AND HITTING TIMES IN A MARKOV CHAIN

Hans U. GERBER
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Shuo-Yen Robert LI
Department of Mathematics, University of Illinois at Chicago Circle, Chicago, IL 60680, U.S.A.

Received 6 August 1979
Revised 2 January 1980

A martingale argument is used to derive the generating function of the number of i.i.d. experiments it takes to observe a given string of outcomes for the first time. Then, a more general problem can be studied: How many trials does it take to observe a member of a finite set of strings for the first time? It is shown how the answer can be obtained within the framework of hitting times in a Markov chain. For these, a result of independent interest is derived.

Hitting times	runs
sequence patterns	martingale

1. Introduction

In recent years there has been an increasing interest in the following type of questions: If letters are determined randomly and lined up in a string, how long do we have to wait to observe a given word for the first time? Or, how long do we have to wait to observe a member of a certain class of words for the first time, and which will it be? For a more complete history, the reader might wish to consult [2], [4], and [6], and the references quoted therein.

In this note, answers are given in terms of the expected value and the moment generating function of the waiting times, for which explicit formulas are developed. In the existing literature, mostly combinatorial arguments have been used. In contrast, our note first presents some general results for hitting times in a Markov chain (Sections 2 and 3) which are useful if the expected values and the moment generating functions of the hitting times are known. This is indeed the case for the questions at hand. In Section 4 martingale arguments are used to derive an explicit formula for the moment generating functions, which then can be substituted in the general results of Sections 2 and 3.

2. Hitting times in a Markov chain

Let $X_0, X_1, X_2, \ldots$ be a Markov chain with stationary transition probabilities. We denote the states by $a, b, c, \ldots$ and, for simplicity, assume a finite number of states. Let

$$N_{ab} = \min\{t: X_t = b \mid X_0 = a\} \tag{1}$$

be the waiting time for state b when the Markov chain starts at state a. Let

$$e_{ab} = \mathbf{E}[N_{ab}] \tag{2}$$

and

$$g_{ab}(z) = \mathbf{E}[z^{N_{ab}}], \quad 0 \leqslant z < 1 \tag{3}$$

denote its expectation and generating function, respectively. Often we shall omit the argument z and write g_{ab} instead of $g_{ab}(z)$. We assume that all states communicate with each other. Together with the assumption of a finite state space, this implies that e_{ab} is finite for all a, b.

Let $b_1, b_2, \ldots, b_n$ be n different states. We are interested in the questions which of these states will be hit first (if we start at $X_0 = 0$, say) and when will this happen. For simplicity, we shall write j and N_j instead of b_j and N_{0j}, respectively. Let

$$N = \min\{N_1, \ldots, N_n\} \tag{4}$$

and let

$$p_{0j} = \mathbf{P}(N = N_j) \tag{5}$$

denote the probability that state j is the first to be hit.

By conditioning, we see that

$$e_{0i} = \mathbf{E}[N_i] = \mathbf{E}[N] + \mathbf{E}[N_i - N] = \mathbf{E}[N] + \sum_{j=1}^{n} p_{0j} e_{ji}, \tag{6}$$

where $e_{ii} = 0$. Since

$$p_{01} + \cdots + p_{0n} = 1, \tag{7}$$

we have a system of $n+1$ linear equations for the unknowns $\mathbf{E}[N], p_{01}, \ldots, p_{0n}$:

$$\begin{pmatrix} 0 & 1 & 1 \cdots 1 \\ 1 & & \\ 1 & & e_{ji} \\ \vdots & & \\ 1 & & \end{pmatrix} \begin{pmatrix} \mathbf{E}[N] \\ p_{01} \\ p_{02} \\ \vdots \\ p_{0n} \end{pmatrix} = \begin{pmatrix} 1 \\ e_{01} \\ e_{02} \\ \vdots \\ e_{0n} \end{pmatrix}. \tag{8}$$

The entries of the coefficient matrix in row 0 and column 0 are as shown; the (i, j)th entry is e_{ji} for $i, j = 1, 2, \ldots, n$.

In the special case $n = 2$, this matrix equation can be solved readily; we find that

$$\mathbf{E}[N] = \frac{e_{01}e_{12} + e_{02}e_{21} - e_{12}e_{21}}{e_{12} + e_{21}} \tag{9}$$

and a formula that appears as Corollary 2, p. 65, in [1]:

$$p_{01} = \frac{e_{02} + e_{21} - e_{01}}{e_{12} + e_{21}}, \qquad p_{02} = \frac{e_{01} + e_{12} - e_{02}}{e_{12} + e_{21}}. \tag{10}$$

Eq. (8) has a unique solution in the general case:

Theorem 2.1. *The coefficient matrix of* (8) *is nonsingular.*

Proof. Let M denote the coefficient matrix, and let

$$A = \begin{pmatrix} 1 & 1 & 1 \cdots 1 \\ e_{01} & & \\ e_{02} & & e_{ji} \\ \vdots & & \\ e_{0n} & & \end{pmatrix}. \tag{11}$$

We apply Cramer's rule in (8) to isolate the unknown $\mathbf{E}[N]$, and find that

$$\mathbf{E}[N] \cdot \det M = \det A. \tag{12}$$

We shall use this and the fact that $\mathbf{E}[N]$ is positive to show by induction with respect to n that

$$(-1)^n \det M > 0. \tag{13}$$

First, for $n = 1$, $\det M = -1$. For the induction step, we assume that (13) holds for any set of n different states and denote by M_{n+1} a matrix corresponding to M for $n+1$. We shall show that $\text{sign}(\det M_{n+1}) = -\text{sign}(\det M)$.

We expand $\det M_{n+1}$ by minors using column 0 to see that

$$\det M_{n+1} = \sum_{k=1}^{n+1} (-1)^k \det B_k, \tag{14}$$

where the matrix B_k is obtained by deleting column 0 and row k in M_{n+1}. Let A_k denote the matrix that results when the kth column of matrix B_k is moved to the front. Hence $\det B_k = (-1)^{k-1} \det A_k$, and (14) reduces to

$$\det M_{n+1} = -\sum_{k=1}^{n+1} \det A_k. \tag{15}$$

Each of the matrices A_k is of the same type as the matrix A. Thus, because of (12) and the induction assumption, $\text{sign}(\det A_k) = \text{sign}(\det M)$. But from this and (15) it follows that $\text{sign}(\det M_{n+1}) = -\text{sign}(\det M)$.

3. Generating functions

From (8) we can compute the probabilities for each state among $b_1, \ldots, b_n$ to be hit first and also the expected waiting time until the first hit. This computational method is useful for only those Markov chains for which e_{ij} can be easily calculated. One such example is the process of repeated experiments with sequence patterns of experimental outcomes as the states. For this Markov chain, a martingale method for calculating e_{ij} is given in [6]. In the next section we shall extend this method to the calculation of g_{ij}. In this section we derive a system of equations that allow us to compute $\mathbf{P}(N = t)$ and $\mathbf{P}(N = N_j = t)$ for any state b_j and positive integer t in terms of g_{ij}.

Let I_j denote the indicator function of the event that $N = N_j$. By writing $N_i = N + (N_i - N)$ and distinguishing according to which of the states is hit first, we see that

$$g_{0i} = \mathbf{E}[z^{N_i}] = \sum_{j=1}^{n} g_{ji}\mathbf{E}[z^N I_j]. \tag{16}$$

Note that this generalizes (6), which can be retrieved by taking the derivative at $z = 1$. Combining (16) with the obvious identity

$$\mathbf{E}[z^N I_1] + \cdots + \mathbf{E}[z^N I_n] = \mathbf{E}[z^N] \tag{17}$$

we get the following matrix equation:

$$\begin{pmatrix} -1 & 1 & 1 \cdots 1 \\ 0 & & \\ 0 & & g_{ji} \\ \vdots & & \\ 0 & & \end{pmatrix} \begin{pmatrix} \mathbf{E}[z^N] \\ \mathbf{E}[z^N I_1] \\ \mathbf{E}[z^N I_2] \\ \vdots \\ \mathbf{E}[z^N I_n] \end{pmatrix} = \begin{pmatrix} 0 \\ g_{01} \\ g_{02} \\ \vdots \\ g_{0n} \end{pmatrix}. \tag{18}$$

Then, the coefficient of z^t in the power series of $\mathbf{E}[z^N I_j]$ is the probability $\mathbf{P}(N = N_j = t)$.

There is a connection between the coefficient matrix M of (8) and the coefficient matrix, say $G(z)$, of (18). Using

$$e_{ij} = \frac{\mathrm{d}}{\mathrm{d}z} g_{ij}(z)|_{z=1} \tag{19}$$

and some well-known rules for the calculation of determinants, we find that

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k} \det G(z)|_{z=1} = 0 \quad \text{for } k = 0, \ldots, n-2, \tag{20}$$

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}} \det G(z)|_{z=1} = (n-1)! \det M. \tag{21}$$

In other words, if $\det G(z)$ is expanded by powers of $(z-1)$, the first nonzero term is $\det M \cdot (z-1)^{n-1}$. In particular, this shows that the function $\det G(z)$ is not identical

to zero. Therefore, (18) can be solved - at least in principle. In the special case $n = 2$ we find that

$$\mathbf{E}[z^N] = \frac{g_{01} + g_{02} - g_{02}g_{21} - g_{01}g_{12}}{1 - g_{12}g_{21}} \tag{22}$$

and

$$\mathbf{E}[z^N I_1] = \frac{g_{01} - g_{02}g_{21}}{1 - g_{12}g_{21}}, \qquad \mathbf{E}[z^N I_2] = \frac{g_{02} - g_{01}g_{12}}{1 - g_{12}g_{21}}. \tag{23}$$

Remark. As pointed out and elegantly shown by a referee, (18) has a unique solution for any Markov chain with countable state space, since $G_n = (g_{ij})$, $i, j = 1, \ldots, n$, is nonsingular for all n. This is obvious for $n = 1$. By the induction assumption G_n is nonsingular. Then the system

$$x_1 g_{1j} + \cdots + x_n g_{nj} = g_{n+1\,j}, \tag{24}$$

$j = 1, \ldots, n$, has a unique solution which (because of (18) with 0 replaced by $n + 1$) must be $x_i = E_{n+1}[z^N I_i]$. The matrix G_{n+1} is singular if and only if this solution satisfies (24) for $j = n + 1$, i.e., if

$$E_{n+1}[z^N I_1] g_{1\,n+1} + \cdots + E_{n+1}[z^N I_n] g_{n\,n+1} = g_{n+1\,n+1}. \tag{25}$$

The left-hand side is the generating function of the time needed to reach b_{n+1} starting in b_{n+1} if one first has to visit the set $\{b_1, \ldots, b_n\}$. For $0 \leq z < 1$ its value is strictly less than one, which is the value of the right-hand side.

4. Sequence patterns in repeated experiments

We shall use the following notation: If $B = (b_1, \ldots, b_k)$ and $C = (c_1, \ldots, c_i)$ are two ordered sequences, the symbol $(B/C)_j$ stands for the condition that

$$b_1 = c_{i+1-j}, \ldots, b_j = c_i, \tag{26}$$

i.e., that the first j members of B are the last j members of C. Note that this condition cannot be satisfied unless $j \leq \min(k, i)$.

We consider an experiment that has countably many possible outcomes. This experiment is performed repeatedly and independently. Let Z_t denote the tth observation. If $B = (b_1, \ldots, b_k)$ is an ordered sequence of possible outcomes, we are interested in the number of experiments it takes to observe B for the first time. We shall also study the more general (but perhaps less natural) case, where a sequence $A = (a_1, \ldots, a_m)$ is already given at the beginning. Then it is assumed that B is not a connected subsequence of A.

We associate to this problem a Markov chain as follows: The state space consists of the integers $0, 1, \ldots, k$, and the state at time t is

$$S_t = \max\{j: (B/W_t)_j\}, \tag{27}$$

where $W_t = (a_1, \ldots, a_m, Z_1, \ldots, Z_t)$. Note that a positive increment of the process $\{S_t\}$ is necessarily one. Then

$$N_{AB} = \min\{t: S_t = k\} = \min\{t: (B/W_t)_k\} \tag{28}$$

is the number of experiments it takes to observe B for the first time (if we are given A to start with).

Before we compute the expectation and the generating function of N_{AB}, we introduce the following notation. First let $P(x)$ denote the probability of outcome x in any particular experiment. If C is another sequence of possible outcomes, we define the function

$$C * B(z) = \sum \left\{ \frac{z^{-j}}{P(b_1) \cdots P(b_j)} : 1 \leq j \leq k \text{ and } (B/C)_j \right\}, \tag{29}$$

where $z \neq 0$.

It has been proved in [6] (and, for some special cases, in [2]) that

$$\mathbf{E}[N_{AB}] = B * B(1) - A * B(1). \tag{30}$$

We generalize this result as follows:

Theorem 4.1. *The generating function of N_{AB} is*

$$\mathbf{E}[z^{N_{AB}}] = \frac{1 + (1-z)A * B(z)}{1 + (1-z)B * B(z)}, \quad 0 < z < 1.$$

Note that this expression is particularly simple, if no initial sequence is given (in which case the numerator becomes one) and if the sequence B consists of identical outcomes, i.e., where B is a 'run' (of successes, for example). For this case, the generating function has been known for some time, see [3, Section XIII.7]. In [5, Section 7.3] and in [7, appendix] it is also shown in certain cases how the generating function can be found with a flow graph analysis combined with the method of collective marks (or additional event method).

Proof of Theorem 4.1. For $j \geq 1-m$ and $t \geq 0$ we define $M_t^{(j)}$ as follows. Set $M_t^{(j)} = 1$ for $t < j$,

$$M_t^{(j)} = \begin{cases} [P(b_1) \cdots P(b_{t-j+1})]^{-1}, & \text{if } (B/W_t)_{t-j+1}, \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

for $j \leq t < j+k$, and $M_t^{(j)} = M_{j+k-1}^{(j)}$ for $t \geq j+k$. Imagine a gambler whose initial fortune is 1, and who, starting with the jth experiment, wages his total fortune sequentially on the occurrence of the sequence B. Then, assuming fair bets, $M_t^{(j)}$ is the gambler's fortune at time t, and it follows that $\{M_t^{(j)}, t \geq 0\}$ is a martingale for every j. Thus

$$X_t = \sum_{j=1-m}^{\infty} z^{j-1} M_t^{(j)}, \quad t \geq 0 \tag{32}$$

is a martingale for $0<z<1$. Since $M_t^{(j)} \leq \kappa$, where $\kappa^{-1} = P(b_1) \cdots P(b_k)$, it follows that

$$X_t \leq \kappa \frac{z^m}{1-z}, \tag{33}$$

and from this that

$$X_0 = \mathbf{E}[X_\nu] \tag{34}$$

for any stopping time ν. In the following let $\nu = N_{AB}$. Then

$$\begin{aligned} X_\nu &= \left(\sum_{j=1-m}^{\nu-k} + \sum_{j=\nu-k+1}^{\nu} + \sum_{j=\nu+1}^{\infty} \right) z^{j-1} M_t^{(j)} \\ &= \sum_{j=\nu-k+1}^{\nu} z^{j-1} \{ [P(b_1) \cdots P(b_{\nu-j+1})]^{-1} : (B/W_\nu)_{\nu-j+1} \} + \frac{z^\nu}{1-z} \\ &= z^\nu \sum_{i=1}^{k} z^{-i} \{ [P(b_1) \cdots P(b_i)]^{-1} : (B/B)_i \} + \frac{z^\nu}{1-z} \\ &= z^\nu \left[B * B(z) + \frac{1}{1-z} \right]. \end{aligned} \tag{35}$$

Similarly,

$$\begin{aligned} X_0 &= \sum_{j=1-k}^{0} z^{j-1} \{ [P(b_i) \cdots P(b_{1-j})]^{-1} : (B/A)_{1-j} \} + \frac{1}{1-z} \\ &= A * B(z) + \frac{1}{1-z}. \end{aligned} \tag{36}$$

Finally, we substitute (35) and (36) in (34), and solve for $\mathbf{E}(z^\nu)$.

Now let n sequences $B_1, \ldots, B_n$ be given. Let N_i denote the number of experiments it takes to observe B_i for the first time, and let N be the minimum among $N_1, \ldots, N_n$. We define the generating functions

$$p_i(z) = \sum_{t=1}^{\infty} \mathbf{P}(N = N_i = t) z^t \tag{37}$$

and

$$p(z) = \sum_{t=1}^{\infty} \mathbf{P}(N = t) z^t. \tag{38}$$

We assume that none of these sequences is contained in any of the others, which excludes ties. In this context we consider the following Markov chain: The state at time t is an n-tuple, where, as in (27), the ith component is the length of the maximal overlap between the tail of the sequence W_t and the beginning of the sequence B_i.

Note that not all the n-tuples qualify as states. For example, if $n=2$, $B_1=(a,b,b,c)$, and $B_2=(b,a)$, the pairs $(0,1)$, $(1,0)$, $(1,2)$, $(2,1)$, $(3,1)$ and $(4,0)$ constitute the state space.

Applying (18) and Theorem 4.1 to this Markov chain, we obtain the following matrix equation:

$$\begin{pmatrix} -1 & 1 & 1\cdots 1 \\ \frac{1}{1-z} & & \\ \vdots & B_j * B_i(z) & \\ \frac{1}{1-z} & & \end{pmatrix} \begin{pmatrix} p(z) \\ p_1(z) \\ \vdots \\ p_n(z) \end{pmatrix} = \begin{pmatrix} 0 \\ A*B_1(z)+\frac{1}{1-z} \\ \vdots \\ A*B_n(z)+\frac{1}{1-z} \end{pmatrix}. \tag{39}$$

A system of equations comparable to this has been derived by combinatorial arguments, see [4, Theorem, 3.3].

If $n=2$, (39) can be solved explicitly. Alternatively, we can use Theorem 4.1 in (22) and (23) to obtain formulas for $p(z)$, $p_1(z)$, and $p_2(z)$.

Acknowledgment

This paper has benefited greatly from Professor Wendel's and a referee's suggestions. The work of the first author was partly done at the Forschungsinstitut für Mathematik, ETH Zürich. The research of the second author was supported in part by NSF Grant MCS77-03533.

References

[1] K.L. Chung, Markov Chains with Stationary Transition Probabilities (Springer, New York, 2nd ed., 1967).

[2] S. Collings, Improbable probabilities, to appear.

[3] W. Feller, An Introduction to Probability Theory and its Applications, Vol. I (Wiley, New York, 3rd ed. 1967).

[4] L.J. Guibas and A.M. Odlyzko, String overlaps, pattern matching, and nontransitive games, to appear.

[5] R.A. Howard, Dynamic Probabilistic Systems, Vol. I (Wiley, New York, 1971).

[6] S.-Y.R. Li, A martingale scheme for studying the occurrence of sequence patterns in repeated experiments, Ann. Probability, to appear.

[7] L. Råde, Thinning of renewal point processes, a flow graph study, Matematisk Statistik AB, Göteborg (1972).