ON OCCURRENCE OF PATTERNS IN MARKOV CHAINS: METHOD OF GAMBLING TEAMS

VLADIMIR POZDNYAKOV

DEPARTMENT OF STATISTICS, UNIVERSITY OF CONNECTICUT

215 GLENBROOK ROAD, U-4120, STORRS, CT 06269-4120

EMAIL: VLADIMIR.POZDNYAKOV@UCONN.EDU

ABSTRACT. A new martingale technique is developed to find formulas for the first two moments and generating function of the waiting time until one observes an element of a finite collection of patterns in a finite multi-state Markov chain. KEYWORDS: Gambling, waiting times, patterns, runs, Markov chains, martingales, stopping times, generating functions. MATHEMATICS SUBJECT CLASSIFICATION (2000): Primary 60J10, Secondary 60G42.

1. Introduction

The central object of interest of the paper is the waiting time till a member of a finite collection of patterns is observed in a stochastic sequence of letters from a finite alphabet. The distribution of the waiting time is a key to many real-world questions in various fields: quality control, hypothesis testing, molecular biology, DNA sequencing and others. Because of the practical importance the occurrence of patterns has been studied extensively by many different techniques. Here is a list of some recent works in the area: Antzoulakos (2001), Chang (2005), Fu (2001), Fu and Chang (2002), Fu and Lou (2006), Hirano and Aki (2003), Pozdnyakov et. al. (2005), Rukhin (2002) and Stefanov (2000, 2003). Useful reviews of different approaches can be found in two recent books: Balakrishnan and Koutras (2002) and Fu and Lou (2003).

We will demonstrate how the occurrence of patterns in Markov chains can be treated with help of the martingale methods which were introduced by Li (1980) and Li and Gerber (1981) in their investigation of independent sequences. Their key observation was that information on the waiting time can be obtained from values assumed by a specially constructed martingale at a relevant stopping time. The common perception is that the martingale technique is not suitable for the situation of Markov dependent trials. But as we have shown in our recent paper that is not exactly true. More specifically, in Glaz et al. (2006) the martingale technique was developed to treat occurrence of patterns in two-state Markov chains. Here we present another algorithm based on the method of gambling teams that works for multi-state Markov chains as well. Some ideas and devices in this article are similar to those in Glaz et al. (2006) but the overall technique is substantially different even when it is applied to two-state Markov chains.

This shows, contrary to common thinking, that the elegant martingale approach developed by Li (1980) and Gerber and Li (1981) for the occurrence of patterns in the case of independent trials can be, in fact, extended to treat the Markov chain case. The most important benefit of the new technique is that the higher moments computed *directly*—not via differentiation of the generating function.

2. Problem Statement and Assumptions

Let $\{Z_n, n \geq 1\}$ be a finite homogeneous Markov chain with a state space $\Omega = \{1, 2, ..., K\}$. We suppose the chain has the initial distribution $\mathbf{P}(Z_1 = k) = p_k$,

$1 \le k \le K$ and the transition matrix $P = \{p_{ij}\}_{1 \le i,j \le K}$ where

$$p_{ij} = \mathbf{P}(Z_{n+1} = j | Z_n = i).$$

Let us consider a finite collection $\mathcal{C}$ of finite ordered sequences (patterns)

$$A_i, \quad 1 \le i \le M,$$

over the alphabet Ω. Let τ_{A_i} denote the first time until the pattern A_i has been observed in the series $Z_1, Z_2, \ldots$. The random variable of main interest here is

$$\tau = \min\{\tau_{A_1}, \ldots, \tau_{A_M}\},$$

the first time when we observe a pattern from $\mathcal{C}$.

Let us list our assumptions.

(1) No pattern of $\mathcal{C}$ contains another pattern from $\mathcal{C}$ as a subpattern.

The longer pattern cannot occur before the shorter one, so the longer pattern can be deleted from our list without loss of generality.

(2) $\mathbf{P}(\tau = \tau_{A_i}) > 0$ for all $1 \le i \le M$.

If $\mathbf{P}(\tau = \tau_{A_i}) = 0$ for some i then we do not need to have A_i in our list. Note that in the case of independent trials this assumption is a consequence of the first one. However, in the case of Markov dependent trials there are other possibilities. For example, if pattern A_i contains subpattern km and $p_{km} = 0$ then A_i can not happen as a run of $\{Z_n, n \ge 1\}$.

(3) $\mathbf{P}(\tau < \infty) = 1$.

This excludes, for instance, a situation when all patterns from $\mathcal{C}$ can only occur while the Markov chain is in the set of transient states. In such a case, it is possible that the system will go to a close irreducible set without hitting $\mathcal{C}$ first. (The question of expected value is trivial in this case, it is the infinity.) But there is another benefit of this assumption. Since we deal here with a finite Markov chain one can easily show that $\mathbf{P}(\tau < \infty) = 1$, in fact, implies $\mathbf{E}[\tau] < \infty$. This will be useful when later we will employ the optional stopping theorem for martingales.

Our goal is to find the first and second moments of τ and its generating function.

3. Expected Waiting Time

Here we derive a formula for the expected value of τ. Following Li (1980) we describe our construction as a gambling system. First we decompose the occurrence of a single pattern A_i into a list of $1 + K + K^2$ (recall $K = |\Omega|$) possible ending scenarios:

(1) A_i occurs as an initial segment of the sequence $\{Z_n, n \geq 1\}$,

(2) kA_i $(1 \leq k \leq K)$ occurs as an initial segment of the sequence $\{Z_n, n \geq 1\}$,

(3) kmA_i $(1 \leq k, m \leq K)$ occurs.

The first $1+K$ ending scenarios are called *initial scenarios*, and the last K^2 scenarios are called *later scenarios.* Thus we have $(1 + K + K^2)M$ scenarios to begin with.

For every later scenario associated with the pattern kmA_i we introduce a kmA_i-*gambling team* of gamblers. Imagine that a casino produces the Markov chain $\{Z_n, n \geq 1\}$. Gambler $n+1$ from the kmA_i-gambling team arrives before round $n+1$ to observe the result of nth trial, Z_n. Then he starts his betting. If $Z_n = k$ he bets a certain amount of money (the same for all gamblers from the kmA_i-gambling team) on pattern mA_i. If $Z_n \neq k$ he bets on A_i.

Let us explain what we mean by "betting \$1 on pattern $Q = q_1 q_2 \cdots q_{l'}$", when $Z_n = q_0$. After observing Z_n the gambler bets a dollar that the next trial yields q_1.

If $Z_{n+1} \neq q_1$ he loses his dollar and leaves the game. If $Z_{n+1} = q_1$, he gets $1/p_{q_0 q_1}$. Note that the expected return of one-dollar bet is \$1, i.e., the betting is fair. If he is lucky he continues his betting. Now he bets his entire capital that the $n+2$ round yields q_2. If it is q_2 he increases his capital by factor $1/p_{q_1 q_2}$, otherwise he leaves the game with nothing. He continues in the same fashion till either pattern Q is exhausted (and the happy gambler goes home with his winning) or the gambler is broke (and the casino is happy to have gambler's dollar).

Now, note that not every ending scenario can occur before the waiting time τ. Also some scenarios are impossible simply because some new patterns associated with some ending scenarios cannot be observed at all in the Markov chain.

Thus from our original list of ending scenarios we need to eliminate those that cannot occur at all and those that can occur only after the time τ. Let K' denote the number of initial scenarios, and let N' denote the number of later scenarios that are still in our list after the elimination. For every jth later scenario in the new list we introduce the corresponding gambling team, and we assume that the *initial amount* with which the gamblers of the jth teams start their betting is y_j. The values of y_j will be specified a bit later.

Let $y_j W_{ij}$, $1 \leq i \leq K' + N'$, $1 \leq j \leq N'$ be the amount of money that the jth team wins in the ith ending scenario. Let X_n denote the casino's net gain from all teams at the conclusion of the nth round. The main property of sequence $\{X_n\}$ is that it forms a martingale with respect to the filtration generated by the Markov chain $\{Z_n, n \geq 1\}$. Indeed, for every gambler in the game the bet size at a current round is fully determined by previous rounds, and odds—as we have seen—are fair. Now let us look at the stopped martingale X_τ. It is easy to see that we have the following expression for X_τ:

$$X_\tau = \sum_{j=1}^{N'} y_j(\tau - 1) - \sum_{i=1}^{K'} \sum_{j=1}^{N'} W_{ij} y_j 1_{E_i} - \sum_{i=K'+1}^{K'+N'} \sum_{j=1}^{N'} W_{ij} y_j 1_{E_i},$$

where E_i is the event that the ith scenario occurs.

Now, note that W_{ij} is *not* a random variable. It depends only on overlapping of the pattern associated with the ith scenario and the pattern associated with the jth gambling team. Indeed, let us consider kmA-gambling team, where

$$A = a_1 a_2 \cdots a_l.$$

Suppose that the game was ended by a scenario associated with pattern

$$Q = q_1 q_2 \cdots q_{l'}.$$

First, we introduce the following two measures of overlapping of a prefix of kmA with a suffix of Q:

$$\delta_t(Q, kmA) = \begin{cases} \dfrac{1}{p_{km} p_{ma_1} \cdots p_{a_{t-2} a_{t-1}}}, & \text{if } q_{l'-t} = k, q_{l'-t+1} = m, q_{l'-t+2} = a_1, \ldots \\ & \quad \ldots, q_{l'} = a_{t-1} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\delta'_t(Q, kmA) = \begin{cases} \dfrac{1}{p_{q_{l'-t} a_1} p_{a_1 a_2} \cdots p_{a_{t-1} a_t}}, & \text{if } q_{l'-t} \neq k, q_{l'-t+1} = a_1, q_{l'-t+2} = a_2, \ldots \\ & \quad \ldots, q_{l'} = a_t \\ 0, & \text{otherwise.} \end{cases}$$

Now, it is easy to see that if the game ends by the scenario associated with pattern Q, then the kmA-gambling team wins (if they bet \$1)

$$\sum_{t=1}^{\min(l'-1, l+1)} \delta_t(Q, kmA) + \delta'_t(Q, kmA).$$

The idea of gambling teams method is simple and somewhat similar to hedging in finance. In nutshell, we try to choose the free parameters $(y_1, y_2, ..., y_{N'})$ in such a way that the total winnings of the all the teams is \$1 regardless how the game ended. More specifically, assume that we can find y_j such that

$$\sum_{j=1}^{N'} W_{ij} y_j = 1, \text{ for all } K' + 1 \le i \le K' + N', \tag{1}$$

then

$$X_\tau = \sum_{j=1}^{N'} y_j(\tau - 1) - \sum_{i=1}^{K'} \sum_{j=1}^{N'} W_{ij} y_j 1_{E_i} - \sum_{i=K'+1}^{K'+N'} 1_{E_i}.$$

Since $\{X_n\}_{n\ge 1}$ has bounded increments and $\mathbf{E}[\tau] < \infty$, the optional stopping theorem (for instance, Williams (1991, p. 100)) tells us that $0 = \mathbf{E}[X_1] = \mathbf{E}[X_\tau]$. That gives us

$$0 = \mathbf{E}[X_\tau] = \sum_{j=1}^{N'} y_j(\mathbf{E}[\tau] - 1) - \sum_{i=1}^{K'} \sum_{j=1}^{N'} W_{ij} y_j \pi_i - (1 - \sum_{i=1}^{K'} \pi_i),$$

where π_i is the probability that the ith initial scenario occurs. Solving the equation with respect to $\mathbf{E}[\tau]$ we obtain the following result.

Theorem 1. *If $(y_1, y_2, ..., y_{N'})$ solves the linear system (1), then*

$$\mathbf{E}[\tau] = 1 + \frac{(1 - \sum_{i=1}^{K'} \pi_i) + \sum_{i=1}^{K'} \pi_i \sum_{j=1}^{N'} y_j W_{ij}}{\sum_{j=1}^{N'} y_j}. \tag{2}$$

Example 1. Let $\Omega = \{1, 2, 3\}$ and $\mathcal{C} = \{323, 313, 33\}$. Suppose now that

$$p_1 = 1/3, p_2 = 1/3, p_3 = 1/3, \tag{3}$$

and the transition matrix P is given by

$$P = \begin{bmatrix} 3/4 & 0 & 1/4 \\ 0 & 3/4 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}. \tag{4}$$

After the eliminating impossible scenarios we get 9 initial scenarios:

$$323\cdots, 313\cdots, 33\cdots, 1323\cdots, 2323\cdots, 1313\cdots, 2313\cdots, 133\cdots, 233\cdots$$

and only 6 later scenarios (because transitions $1 \to 2$ and $2 \to 1$ are impossible):

$$\cdots 11323, \cdots 22323, \cdots 11313, \cdots 22313, \cdots 1133, \cdots 2233.$$

Let us give some entries of matrix W. For instance, the 11323-gambling team in the initial scenario $323\cdots$ wins $1/p_{23} = 4$. The same team in the later scenario $\cdots 11323$ wins $1/(p_{11}p_{13}p_{32}p_{23}) + 1/p_{23} = 268/3$, and in the later scenario $\cdots 22323$ it wins $1/(p_{23}p_{32}p_{23}) + 1/p_{23} = 68$. The entries of matrix W that correspond to the later scenarios (those are needed for linear system (1)) are

$$\begin{bmatrix} 268/3 & 64 & 4 & 0 & 4 & 0 \\ 68 & 256/3 & 4 & 0 & 4 & 0 \\ 0 & 4 & 256/3 & 68 & 0 & 4 \\ 0 & 4 & 64 & 268/3 & 0 & 4 \\ 2 & 2 & 2 & 2 & 38/3 & 10 \\ 2 & 2 & 2 & 2 & 10 & 38/3 \end{bmatrix}.$$

Using formula (2) we obtain

$$\mathbf{E}[\tau] = 8\frac{7}{15}.$$

4. Second (and Higher) Moments

To derive a formula for the second moment of τ we will work with the same numbers of ending scenarios and gambling teams. But now the gambler from the jth team that places his bet the first time in the nth round will start his betting with $y_j + z_j n$ dollars. The weights y_j and z_j will be chosen later.

Now, let

$$W_{ij}y_j + \tau W_{ij}z_j + N_{ij}z_j$$

denote the winnings of jth team in the ith ending scenario. Here W_{ij} is as before, and N_{ij} is a new quantity. But the key observation, as earlier, is that N_{ij} is *not* random. More specifically, suppose that the ith scenario is associated with pattern $Q = q_1q_2\cdots q_{l'}$, and the jth gambling team—with pattern $kmA = kma_1a_2\cdots a_l$. Then one can find that

$$N_{ij} = \sum_{t=1}^{\min(l'-1,l+1)} (\delta_t(Q,kmA) + \delta_t'(Q,kmA))(1-t).$$

Next, we obtain the following expression for the casino net gain at moment τ.

$$\begin{aligned}
X_\tau &= (\tau-1)\sum_{j=1}^{N'} y_j + (2+3+...+\tau)\sum_{j=1}^{N'} z_j \\
&\quad -\sum_{i=1}^{K'} 1_{E_i}\sum_{j=1}^{N'}[W_{ij}y_j + \tau W_{ij}z_j + N_{ij}z_j] \\
&\quad -\sum_{i=K'+1}^{K'+N'} 1_{E_i}\sum_{j=1}^{N'}[W_{ij}y_j + \tau W_{ij}z_j + N_{ij}z_j] \\
&= (\tau-1)\sum_{j=1}^{N'} y_j + \frac{(\tau-1)(\tau+2)}{2}\sum_{j=1}^{N'} z_j \\
&\quad -\sum_{i=1}^{K'} 1_{E_i}\sum_{j=1}^{N'}[W_{ij}y_j + \tau W_{ij}z_j + N_{ij}z_j] + \sum_{i=1}^{K'} 1_{E_i}(1+\tau) \\
&\quad -\sum_{i=K'+1}^{K'+N'} 1_{E_i}\sum_{j=1}^{N'}[W_{ij}y_j + \tau W_{ij}z_j + N_{ij}z_j] - \sum_{i=1}^{K'} 1_{E_i}(1+\tau),
\end{aligned}$$

where 1_{E_i} is the indicator that the ith scenario occurs.

Suppose that we can find y_i and z_i such that for all $K'+1 \le i \le K'+N'$

$$\sum_{j=1}^{N'} W_{ij} z_j = 1 \text{ and } \sum_{j=1}^{N'} W_{ij} y_j + N_{ij} z_j = 1. \tag{5}$$

In this case the stopped martingale X_τ is given by

$$\begin{aligned} X_\tau &= (\tau-1)\sum_{j=1}^{N'} y_j + \frac{(\tau-1)(\tau+2)}{2}\sum_{j=1}^{N'} z_j \\ &\quad -\sum_{i=1}^{K'} 1_{E_i} \sum_{j=1}^{N'} [W_{ij} y_j + \tau W_{ij} z_j + N_{ij} z_j] + \sum_{i=1}^{K'} 1_{E_i}(1+\tau) \\ &\quad -(1+\tau). \end{aligned}$$

Finally, once again the optional stopping theorem gives us a formula for $\mathbf{E}[\tau^2]$. Note, however, that the increments of martingale X_n are no longer bounded almost sure, so we need a different version. For example, the optional stopping theorem from Shiryaev (1995, p. 485) will work here—just note that X_n is at most $O(n^2)$, but $\mathbf{P}(\tau > n)$ goes to zero at exponential rate. Let us now summarize our findings.

Theorem 2. *If $(y_1, y_2, ..., y_{N'})$ and $(z_1, z_2, ..., z_{N'})$ solves the linear system (5) then*

$$\mathbf{E}[\tau^2] = \frac{2}{\sum_{j=1}^{N'} z_j}\left[1 + \mathbf{E}[\tau] + (1-\mathbf{E}[\tau])\sum_{j=1}^{N'} y_j + A\right] + 2 - \mathbf{E}[\tau], \tag{6}$$

where

$$A = \sum_{i=1}^{K'} \pi_i \left[\sum_{j=1}^{N'} [W_{ij} y_j + \Lambda_i W_{ij} z_j + N_{ij} z_j] - 1 - \Lambda_i\right],$$

π_i is the probability that the ith initial scenario occurs, and Λ_i is the value of τ when the ith initial scenario occurs.

Example 2. Consider the same alphabet Ω, compound pattern $\mathcal{C}$ and Markov chain $\{Z_n, n \geq 1\}$ as in Example 1. The values of (new) matrix N_{ij} for the later scenarios are

$$\begin{bmatrix} -256 & -128 & 0 & 0 & 0 & 0 \\ -128 & -256 & 0 & 0 & 0 & 0 \\ 0 & 0 & -256 & -128 & 0 & 0 \\ 0 & 0 & -128 & -256 & 0 & 0 \\ 0 & 0 & 0 & 0 & -64/3 & -8 \\ 0 & 0 & 0 & 0 & -8 & -64/3 \end{bmatrix}.$$

Solving linear system (5) (in fact, we need to solve consecutively two linear systems of size 6) and applying formula (6) we obtain

$$\mathbf{E}[\tau^2] = 125\frac{4}{25}.$$

Of course, in this case a symbolic differentiation of the generating function is possible, and it leads to the same answer (see next section).

Finally, let us note that in order to get the third moment we need to change the size of initial bet for the nth gambler to $y_j + z_j n + x_j n^2$, and after a similar bookkeeping we will arrive to a formula for the third moment.

5. Generating Function

To find the generating function for the waiting time τ, $\mathbf{E}[\alpha^\tau], 0 \leq \alpha \leq 1$, we need to introduce the same scenarios and the same gambling teams, but we need to change the size of initial bets. More specifically, a gambler from the jth team that starts his betting in the nth round will bet initially $y_j\alpha^n$ dollars. Let $\alpha^\tau W_{ij}(\alpha)y_j$ denote the total winnings of jth team in the ith scenario. As before the most important observation is that $W_{ij}(\alpha)$ is *not* a random variable, and it is fully determined by the relationship between the jth gambling team and the ith ending scenario.

If X_n again denotes the net gain of the casino at time n, then the stopped martingale X_τ is given by

$$\begin{aligned} X_\tau &= (\alpha^2 + ... + \alpha^\tau)\sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} 1_{E_i} \sum_{j=1}^{N'} \alpha^\tau y_j W_{ij}(\alpha) \\ &\quad - \sum_{i=K'+1}^{N'} 1_{E_i} \sum_{j=1}^{N'} \alpha^\tau y_j W_{ij}(\alpha) \\ &= \frac{\alpha^2 - \alpha\alpha^\tau}{1-\alpha} \sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} 1_{E_i} \sum_{j=1}^{N'} \alpha^\tau y_j W_{ij}(\alpha) + \sum_{i=1}^{K'} 1_{E_i}\alpha^\tau \\ &\quad - \sum_{i=1}^{K'} 1_{E_i}\alpha^\tau - \sum_{i=K'+1}^{N'} 1_{E_i} \sum_{j=1}^{N'} \alpha^\tau y_j W_{ij}(\alpha), \end{aligned}$$

where $E_i, 1 \le i \le K' + N'$ is the event that the ith scenario occurs. Suppose that for every $0 < \alpha < 1$ we can find $(y_1, ..., y_{N'})$ such that

$$\sum_{j=1}^{N'} W_{ij}(\alpha) y_j = 1, \text{ for all } K' + 1 \le i \le K' + N'. \tag{7}$$

Then the stopped martingale X_τ is given by

$$\begin{aligned} X_\tau &= \frac{\alpha^2 - \alpha\alpha^\tau}{1-\alpha} \sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} \alpha^\tau 1_{E_i} \left(\sum_{j=1}^{N'} y_j W_{ij}(\alpha) - 1 \right) - \alpha^\tau \\ &= \frac{\alpha^2 - \alpha\alpha^\tau}{1-\alpha} \sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} \alpha^{\Lambda_i} 1_{E_i} \left(\sum_{j=1}^{N'} y_j W_{ij}(\alpha) - 1 \right) - \alpha^\tau, \end{aligned}$$

where $\Lambda_i, 1 \le i \le K'$ is the value of τ when the ith initial scenario occurs. Note that Λ_i is not a random variable. After a routine application of the optional stopping theorem we obtain

$$0 = \frac{\alpha^2 - \alpha\mathbf{E}[\alpha^\tau]}{1-\alpha} \sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} \alpha^{\Lambda_i}\pi_i \left(\sum_{j=1}^{N'} y_j W_{ij}(\alpha) - 1 \right) - \mathbf{E}[\alpha^\tau].$$

After some algebra we get the following result.

Theorem 3. *If* $(y_1, y_2, ..., y_{N'})$ *solves the linear system (7), then*

$$\mathbf{E}[\alpha^\tau] = \frac{\alpha^2/(1-\alpha)\sum_{j=1}^{N'} y_j - \sum_{i=1}^{K'} \alpha^{\Lambda_i}\pi_i \left(\sum_{j=1}^{N'} y_j W_{ij}(\alpha) - 1\right)}{1 + \alpha/(1-\alpha)\sum_{j=1}^{N'} y_j}.$$

Example 3. Let $\Omega = \{1, 2, 3\}$ and $\mathcal{C} = \{323, 313, 33\}$. Suppose again that the initial probabilities are given by (3) and transition probability matrix is given by (4). We have 9 initial ending scenarios and 6 later ones (see Example 1). After solving linear system (7) and applying Theorem 3 we obtain

$$\mathbf{E}[\alpha^\tau] = \frac{\alpha^2(16-\alpha^2)}{96 - 9\alpha(8+\alpha^2)} = \frac{1}{6}\alpha^2 + \frac{1}{8}\alpha^3 + \frac{1}{12}\alpha^4 + \frac{5}{64}\alpha^5 + \frac{9}{128}\alpha^6 + \frac{31}{512}\alpha^7 + O(\alpha^8)$$

References

[1] Antzoulakos, D. (2001), Waiting times for patterns in a sequence of multistate trials, *Journal of Applied Probability*, **38**, 508–518.

[2] Balakrishnan, N. and Koutras, M. V. (2002), *Runs and Scans with Applications*, (John Wiley and Sons, Inc., New York).

[3] Chang, Y.-M. (2005), Distribution of waiting time until the rth occurrence of a compound pattern, *Statistics and Probability Letters*, **75**, 29–38.

[4] Fu, J. C. (2001), Distribution of the scan statistics for a sequence of bistate trials, *Journal of Applied Probability*, **38**, 908–916.

[5] Fu, J. C. and Chang, Y. (2002), On probability generating functions for waiting time distribution of compound patterns in a sequence of multistate trials, *Journal of Applied Probability*, **39**, 70–80.

[6] Fu, J. C. and Lou, W. Y. W. (2003), *Distribution Theory of Runs and Patterns*, (World Scientific Publishing, Singapore).

[7] Fu, J. C. and Lou, W. Y. W. (2006), Waiting time distributions of simple and compound patterns in a sequence of r-th order Markov dependent multi-state trials. *Annals of the Institute of Statistical Mathematics*, **58**, 291-310.

[8] Gerber, H. and Li, S. (1981), The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stochastic Processes and their Applications*, **11**, 101–108.

[9] Glaz, J., Kulldorff, M., Pozdnyakov, V., and Steele, J. M. (2006), Gambling teams and waiting times for patterns in two-state Markov chains. *Journal of Applied Probability*, **43**, 127–140.

[10] Hirano, K. and Aki, S. (2003), Number of occurrences of subpattern until the first appearance of a pattern and geometric distribution, *Statistics and Probability Letters*, **65**, 259–262.

[11] Li, S. (1980), A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *The Annals of Probability*, **8**, 1171-1176.

[12] Pozdnyakov, V., Glaz, J., Kulldorff, M., and Steele, J. M. (2005), A martingale approach to scan statistics. *Annals of the Institute of Statistical Mathematics*, **57**, 21–37.

[13] Rukhin, A. L. (2002), Distribution of the number of words with a prescribed frequency and tests of randomness. *Advances in Applied Probability*, **34**, 775–797.

[14] Shiryaev, A. N. (1995), *Probability*, (Springer, New York, 2nd Edition).

[15] Stefanov, V. T. (2000), On some waiting time problems. *Journal of Applied Probability*, **37**, 756–764.

[16] Stefanov, V. T. (2003), The intersite distances between pattern occurrences in strings generated by general discrete- and continuous-time models: an algorithmic approach. *Journal of Applied Probability*, **40**, 881–892.

[17] Williams, D. (1991), *Probability with martingales*, (Cambridge University Press, Cambridge).