1

Martingale Methods for Patterns and Scan Statistics

Vladimir Pozdnyakov and J. Michael Steele

Department of Statistics, University of Connecticut, Storrs, USA
Department of Statistics, University of Pennsylvania, Philadelphia, USA

Abstract: We show how martingale techniques (both old and new) can be used to obtain otherwise hard-to-get information for the moments and distributions of waiting times for patterns in independent or Markov sequences. In particular, we show how these methods provide moments and distribution approximations for certain scan statistics, including about variable length scan statistics. Each general problem that is considered is also illustrated with a concrete example confirming the computational tractability of the method.

Keywords and phrases: Scan, pattern, martingale

1.1 Introduction

The martingale method for waiting times for patterns in an independent sequence was pioneered in Li (1980), and in the intervening time many variations on the original idea have been developed. Our first aim here is to survey these developments using the unifying language of gambling teams. We further show how the martingale method can be extended to cover a great variety of problems in applied probability, including the occurrence of patterns in Markov sequences. One of the key intermediate steps is the development of a clear understanding of the distribution of the first time of occurrence of a pattern from a finite set of patterns. It is this general problem that leads to methods that are applicable to the theory of scan statistics.

1.2 Patterns in an Independent Sequence

By $\{Z_n, n \geq 1\}$ we denote a sequence of independent identically distributed random variables with values from a finite set $\Omega = \{1, 2, ..., M\}$ which we call the process alphabet. To specify the distribution of Z_n, we then set

$$p_1 = \mathbf{P}(Z_n = 1) > 0,\ p_2 = \mathbf{P}(Z_n = 2) > 0, ...,\ p_M = \mathbf{P}(Z_n = M) > 0.$$

By a *pattern* A we mean a finite ordered sequence of letters $a_1 a_2 \cdots a_m$ over the alphabet Ω. The random variable that is of most interest here is τ_{A}, the first time that one observes the pattern A as a run in the sequence $\{Z_n, n \geq 1\}$. Our main goal is to provide methods — and often explicit formulas — for the expected value, the higher moments, and the probability generating function of τ_{A}.

1.2.1 A gambling approach to the expected value

We begin with a construction that originates with Li (1980) and which we frame as a gambling scheme. Consider a casino game that generates the sequence $\{Z_n,\ n \geq 1\}$, say as the out-put of a biased roulette wheel. Next consider a sequence of gamblers who arrive sequentially so that the n'th gambler arrives right before n'th round when Z_n is generated. We also assume that this casino pays fair odds, so that a dollar bet on an event that has probability p would pay $1/p$ dollars to a winner (and zero to a loser).

Now we consider the strategy that is followed by the n'th gambler, the one who arrives just before the n'th round of play. For specificity, we first consider the gambler who enters just before the first round. This gambler bets one dollar that $Z_1 = a_1$. If Z_1 is not a_1 the gambler stops betting after having lost one dollar. If Z_1 yields a_1, the gambler wins $1/\mathbf{P}(Z_1 = a_1)$. He then continues to play, now betting his entire capital on $Z_2 = a_2$. If he loses, he stops gambling; otherwise he increases his bet by the factor $1/\mathbf{P}(Z_2 = a_2)$. The gambler then continues in the same fashion until the entire pattern A is exhausted or until he has lost his original dollar, which ever comes first.

If the first gambler is very lucky and pattern A is observed after m rounds, the gambler stops and has total winnings of

$$\Big(\mathbf{P}(Z_1 = a_1)\mathbf{P}(Z_2 = a_2) \times \cdots \times \mathbf{P}(Z_m = a_m)\Big)^{-1}$$

dollars. Otherwise, the first gambler simply loses his initial bet of \$1.

In the meanwhile, additional gamblers enter the casino at successive times $2, 3, ...$ and each of these gamblers uses the same strategy that was used by the first gambler. That is, he bets successively on the letters of the pattern, each time "letting his stake ride." We then let X_n denote the total net gain of the casino at the end of the n'th round of play. The game was fair at each stage, so the stochastic process $\{X_n, \sigma(Z_1, ..., Z_n)\}$ is a martingale.

Now consider the random variable $X_{\tau_{\mathrm{A}}}$; this is the casino's net gain at the time when the pattern A is first observed. This random variable is well-defined since τ_{A} is finite with probability one. In fact, it is easy to show that τ_{A} is bounded by a geometrically distributed random variable, so by Wald's lemma (or the optional stopping theorem, Williams (1991, p. 100)), we have the basic relation

$$\mathbf{E}(X_{\tau_{\mathrm{A}}}) = 0.$$

Fortunately, we know more about $X_{\tau_{\mathrm{A}}}$. Specifically, we know that

$$X_{\tau_{\mathrm{A}}} = \tau_{\mathrm{A}} - W,$$

where W is the total amount of money that has been won by gamblers by time τ_{A}. The key observation is that W is not a random variable. The value of W is fully determined by the way in which the pattern A overlaps with itself.

Moreover, it is reasonably easy to calculate W. For a gambler to have any capital left when pattern A is first observed, that gambler needs to still be gambling, so in particular the gamblers who entered the game before $\tau_{\mathrm{A}} - m + 1$ must have all lost their dollar. The gambler who enters the game at time $\tau_{\mathrm{A}} - m + 1$ is the lucky guy who wins the most, but also some of those gamblers who entered after him may have some amount in their pockets.

The total amount of money that these few players have is represented by a certain measure of the overlapping of pattern A with itself. To describe this measure, we first consider $0 \leq i, j \leq m$ and set

$$\delta_{ij} = \begin{cases} 1/\mathbf{P}(Z_1 = a_i), & \text{if } a_i = a_j, \\ 0, & \text{otherwise.} \end{cases}$$

With this notation we then find the explicit formula

$$W = \delta_{11}\delta_{22} \cdots \delta_{mm} + \delta_{21}\delta_{32} \cdots \delta_{mm-1} + \cdots + \delta_{m1}, \tag{1.1}$$

so from our earlier observation that $\mathbf{E}(X_{\tau_{\mathrm{A}}}) = 0$, we find

$$\mathbf{E}(\tau_{\mathrm{A}}) = \delta_{11}\delta_{22} \cdots \delta_{mm} + \delta_{21}\delta_{32} \cdots \delta_{mm-1} + \cdots + \delta_{m1}. \tag{1.2}$$

The relation (1.2) really is quite explicit, and it provides an easily applied answer to our first question, say as one sees in the following:

Example 1.2.1 Let $\Omega = \{1, 2\}$ and consider the pattern 1121 of length 4. We then have

$$\mathbf{E}(\tau_{\mathrm{A}}) = W = (p_1 \times p_1 \times p_2 \times p_1)^{-1} + (p_1)^{-1}$$

1.2.2 Gambling on a generating function

By a natural modification of the preceding method, one can obtain a formula for the generating function of $\tau_{\rm A}$. The trick is to change the initial bet for each gambler. Now instead of \$1, the n'th gambler starts his betting by placing a bet of size α^n, where $0 < \alpha < 1$. Let $\alpha^{\tau_{\rm A}} W(\alpha)$ be the total winnings of all the gamblers by time $\tau_{\rm A}$. As before, we let X_n denote the casino's gain at the end of the n'th round, and as before $\{X_n\}$ is a martingale. For convenience, we denote the total accumulated winnings of the gamblers when the pattern A is first observed by $\alpha^{\tau_{\rm A}} W(\alpha)$. Again, the key is that we have a nice relation for the casino's net gain $X_{\tau_{\rm A}}$. Specifically, we have

$$\begin{aligned} X_{\tau_{\rm A}} &= \alpha^1 + \alpha^2 + \cdots + \alpha^{\tau_{\rm A}} - \alpha^{\tau_{\rm A}} W(\alpha) \\ &= \alpha \frac{\alpha^{\tau_{\rm A}} - 1}{\alpha - 1} - \alpha^{\tau_{\rm A}} W(\alpha) \\ &= \alpha^{\tau_{\rm A}} \left(\frac{\alpha}{\alpha - 1} - W(\alpha) \right) - \frac{\alpha}{\alpha - 1}. \end{aligned}$$

As in previous subsection, $W(\alpha)$ is not a random variable, and it has an explicit representation:

$$W(\alpha) = \delta_{11}\delta_{22} \cdots \delta_{mm-1}/\alpha^{m-1} + \delta_{21}\delta_{32} \cdots \delta_{mm-1}/\alpha^{m-2} + \cdots + \delta_{m1}/1.$$

The optional stopping theorem implies

$$0 = \mathbf{E}(X_{\tau_{\rm A}}) = \mathbf{E}(\alpha^{\tau_{\rm A}}) \left(\frac{\alpha}{\alpha - 1} - W(\alpha) \right) - \frac{\alpha}{\alpha - 1}.$$

When we solve this relation for $\mathbf{E}(\alpha^{\tau_{\rm A}})$, we obtain

$$\mathbf{E}(\alpha^{\tau_{\rm A}}) = \left(1 + \frac{1-\alpha}{\alpha} W(\alpha) \right)^{-1}.$$

Again, this is an explicit useable formula, as one sees in the following example.

Example 1.2.2 Let $\Omega = \{1, 2\}$ and again consider the pattern 1121. One then has

$$W(\alpha) = \frac{\alpha^{-3}}{p_1^3 p_2} + \frac{1}{p_1},$$

so by substitution one has

$$\begin{aligned} \mathbf{E}(\alpha^{\tau_{\rm A}}) &= \frac{p_1^3 p_2 \alpha^4}{1 - \alpha + \alpha^3 (1 - p_2 \alpha) p_1^2 p_2} \\ &= p_1^3 p_2 \alpha^4 + p_1^3 p_2 \alpha^5 + p_1^3 p_1 \alpha^6 + p_1^3 p_2 (1 - p_1^2 p_2) \alpha^7 + o(\alpha^7). \end{aligned}$$

As a check, one should note that this formula can be used to confirm the calculation of the mean from our first example:

$$\frac{\partial \mathbf{E}(\alpha^{\tau_{\rm A}})}{\partial \alpha}\Big|_{\alpha=1} = \frac{1}{p_1^3 p_2} + \frac{1}{p_1} = E\tau_{\rm A}.$$

1.2.3 Second and higher moments

In theory, the ability to compute the probability generating function also gives one the higher moments, but in practice it is often useful to have an alternative method. Here it also seems instructive to show how the method of sequential gamblers can be used to find $\mathbf{E}(\tau_{\mathrm{A}}^2)$.

This time the trick is that the gambler who joins the game in the n'th round will bet n dollars. If, as always, we let X_n denote the casino's net gain after n rounds, then X_n is again a martingale. Moreover, in this case one can check that at the stopping time τ_{A} we have

$$\begin{aligned} X_{\tau_{\mathrm{A}}} &= 1+2+\cdots+\tau_{\mathrm{A}} \\ &\quad -(\tau_{\mathrm{A}}-m+1)\delta_{11}\delta_{22}\cdots\delta_{mm} \\ &\quad -(\tau_{\mathrm{A}}-m+2)\delta_{21}\delta_{32}\cdots\delta_{mm-1} \\ &\quad \cdots \\ &\quad -(\tau_{\mathrm{A}}-m+m)\delta_{m1} \\ &= 1+2+\cdots+\tau_{\mathrm{A}}-\tau_{\mathrm{A}}W-N \\ &= \frac{\tau_{\mathrm{A}}^2+\tau_{\mathrm{A}}}{2}-\tau_{\mathrm{A}}W-N, \end{aligned}$$

where

$$N = -\delta_{11}\delta_{22}\cdots\delta_{mm}(m-1)-\delta_{21}\delta_{32}\cdots\delta_{mm-1}(m-2)-\cdots-\delta_{m1}0.$$

It is now time to apply the optional stopping theorem, but in this case the increments of X_n are no longer uniformly bounded, so a more refined version of Doob's optional stopping theorem is needed. Here we can use the stopping time theorem of Shiryaev (1995, p. 485) since we have $X_n = O(n^2)$ and since $\mathbf{P}(\tau_{\mathrm{A}} > n)$ decays at an exponential rate. The application of this optional stopping theorem leads us to

$$0 = \mathbf{E}(X_{\tau_{\mathrm{A}}}) = \mathbf{E}(\tau_{\mathrm{A}}^2)/2 + \mathbf{E}(\tau_{\mathrm{A}})/2 - W\mathbf{E}(\tau_{\mathrm{A}}) - N.$$

Solving this equation for $\mathbf{E}(\tau_{\mathrm{A}}^2)$ gives us

$$\mathbf{E}(\tau_{\mathrm{A}}^2) = (2W-1)\mathbf{E}(\tau_{\mathrm{A}}) + 2N = 2W^2 - W + 2N,$$

and as a corollary we have the nice formula

$$\mathbf{Var}(\tau_{\mathrm{A}}) = W^2 - W + 2N.$$

Naturally, variations of this technique can be applied to obtain formulas for any moment. For example, to find an expression for the third moment, the n'th gambler's bet should now be taken to be n^2.

Example 1.2.3 For the traditional sample space $\Omega = \{1, 2\}$ and the pattern 1121 we now find

$$N = -\frac{3}{p_1 \times p_1 \times p_2 \times p_1},$$

and

$$\mathbf{Var}(\tau_{\mathrm{A}}) = \left(\frac{1}{p_1} + \frac{1}{p_1^3 p_2}\right)^2 - \frac{1}{p_1} - \frac{7}{p_1^3 p_2}.$$

Here it is interesting to note that when either $p_1 \to 0$ or $p_2 \to 0$ one has the limit relation

$$\frac{\mathbf{E}(\tau_{\mathrm{A}})}{\mathbf{Var}(\tau_{\mathrm{A}})^{1/2}} \to 1.$$

Moreover, there is an intuitive explanation for this limit. When either $p_1 \to 0$ or $p_2 \to 0$ the occurrence of the pattern 1121 becomes a rare event. By the clumping heuristic [c.f. Aldous (1989)], one then expects the distribution of τ_{A} to be well approximated by an exponential distribution, and for an exponential X we have the equality $\mathbf{E}(X) = \sqrt{\mathbf{Var}(X)}$.

1.3 Compound Patterns and Gambling Teams

In many important applications — such as scans — one is concerned about the waiting time until the first occurrence of one out of *many* patterns from a *finite list of patterns.* Here we call a finite collection of K patterns $\{\mathrm{A}_1, \mathrm{A}_2, \cdots, \mathrm{A}_K\}$ a *compound pattern* and denote it simply by $\mathcal{A}$. Now, if τ_{A_i} denotes the first time until the pattern A_i has been observed as a completed run in the i.i.d. series $Z_1, Z_2, ...$ then the new random variable of interest is

$$\tau_{\mathcal{A}} = \min\{\tau_{\mathrm{A}_1}, ..., \tau_{\mathrm{A}_K}\}.$$

In words, $\tau_{\mathcal{A}}$ is the first time when we observe a pattern from $\mathcal{A}$, and one should note that without loss of generality we assume that in $\mathcal{A}$ no pattern is a subblock of another.

Gerber and Li (1981) studied compound patterns with help of an appropriate Markov chain imbedding. We use an alternative method that has several benefits. In particular, the new method gives us clear hints on how we should extend the martingale approach to the case of Markov dependent trials. It also guides us when we consider the case of highly regular patterns, such as those associated with scans or structured motifs.

1.3.1 Expected time

It seems natural in the case of compound pattern $\mathcal{A}$ to introduce K *gambling teams*. The gamblers from each gambling team will bet on a pattern from the list $\mathcal{A}$. But now the problem is that the total amount of winnings of all the gamblers at time $\tau_{\mathcal{A}}$ is a random variable. It depends on how the game is stopped.

However, if one knows which simple pattern from $\mathcal{A}$ triggered the stop, then the winnings of a gambling team are not random. This amount is fully determined by overlapping of two patterns: (1) the pattern associated with the gambling team, and (2) the pattern associated with the *ending scenario*. An explicit expression for this amount will be given a bit later.

As we will demonstrate in a moment it is beneficial to allow every gambling team to have their own size for an initial bet. More specifically, let y_j will be an amount with which the gambler from the j'th gambling team (the team that bets on A_j) starts his betting. Let $W_{ij}y_j$ be total winnings of the j'th gambling team in case when game was ended by the i'th scenario (i.e., the pattern A_i is observed at time $\tau_{\mathcal{A}}$). If X_n is as before the net casino gain, then it is clear that it forms a martingale, because a weighted sum of martingales is a martingale. The stopped martingale $X_{\tau_{\mathcal{A}}}$ is given by

$$X_{\tau_{\mathcal{A}}} = \sum_{j=1}^{K} y_j \tau_{\mathcal{A}} - \sum_{i=1}^{K} \sum_{j=1}^{K} W_{ij} y_j 1_{E_i},$$

where 1_{E_i} is the indicator that the game is ended by the i'th scenario.

There is an analogy between the way gambling teams are used here and the notion of hedging in finance. The trick, analogous to arbitrage constructions, is to choose weights y_j in such a way that the total winnings of all the teams $\sum_{j=1}^{K} W_{ij} y_j$ is equal to 1 regardless of an ending scenario. Now, if the vector $\{y_j\}_{1 \le j \le K}$ is a solution of the linear system

$$\sum_{j=1}^{K} W_{ij} y_j = 1, \quad 1 \le i \le K, \tag{1.3}$$

then the stopped martingale is given by

$$X_{\tau_{\mathcal{A}}} = \sum_{j=1}^{K} y_j \tau_{\mathcal{A}} - 1.$$

This puts us on familiar ground. By another application of the optional stopping theorem we obtain a computationally effective representation of $\mathbf{E}(\tau_{\mathcal{A}})$.

Theorem 1.3.1 *If vector $\{y_j\}_{1\le j\le K}$ solves the linear system (1.3), then expected value of $\tau_{\mathcal{A}}$ is given by*

$$\mathbf{E}(\tau_{\mathcal{A}}) = \frac{1}{\sum_{j=1}^{K} y_j}.$$

Here we should make two technical comments. First, in the course of their Markov embedding method, Gerber and Li (1981) showed that the matrix W_{ij} is nonsingular if no pattern from $\mathcal{A}$ is a subpattern of another. Consequently, the solution $\{y_j\}_{1\le j\le K}$ always exists.

Second, there is an explicit formula for W_{ij}. For example, consider two patterns A $= a_1a_2\cdots a_m$ and B $= b_1b_2\cdots b_l$. Next, we consider the measure of t-overlap of a suffix of A with a prefix of B that is given by the formula

$$\delta_t(A,B) = \begin{cases} \dfrac{1}{\prod_{s=1}^{t}\mathbf{P}(Z_1=b_s)}, & \text{if } b_1=a_{m-t+1}, b_2=a_{m-t+2}, ..., b_t=a_m,\\ 0, & \text{otherwise.}\end{cases}$$

Now, if the j'th gambling team bets on A, and the i'th ending scenario is associated with pattern B then

$$W_{ij} = \sum_{t=1}^{\min(m,l)} \delta_t(A,B).$$

1.3.2 The generating function and the second moment

The method of gambling teams can be used to obtain a formula for the probability generating function $\mathbf{E}(\alpha^{\tau_{\mathcal{A}}}), 0<\alpha<1$, for any compound pattern $\mathcal{A}$. The solution is a little more complicated, but it mainly calls for the systematic elaboration of ideas that we have already seen. Here we consider the same number of gambling teams and ending scenarios that we used before, but now the gambler from the j'th team who joins the game in the n'th round will place an initial bet of size of $y_j\alpha^n$ where the weights $\{y_j\}_{1\le j\le K}$ will be chosen later.

Let $W_{ij}(\alpha)y_j\alpha^{\tau_{\mathcal{A}}}$ denote the winnings of the j'th gambling team when the game ends by the i'th ending scenario. If X_n denotes the martingale that gives us the casino's net gain at time n, then the stopped martingale $X_{\tau_{\mathcal{A}}}$ is given by

$$X_{\tau_{\mathcal{A}}} = \alpha\frac{\alpha^{\tau_{\mathcal{A}}}-1}{\alpha-1}\sum_{j=1}^{K} y_j - \sum_{i=1}^{K}\sum_{j=1}^{K} W_{ij}(\alpha)y_j\alpha^{\tau_{\mathcal{A}}}1_{E_i},$$

where as before 1_{E_i} is the indicator of the i'th ending scenario.

Again, the key fact is that $W_{ij}(\alpha)$ is not a random variable. If the j'th gambling team bets on pattern A, and the i'th ending scenario is linked with pattern B, then

$$W_{ij}(\alpha) = \sum_{t=1}^{\min(m,l)} \delta_t(A,B)\alpha^{1-t}, \tag{1.4}$$

where $\delta_t(A,B)$ is define as in the preceding section. If weights $\{y_j(\alpha)\}_{1\le j\le K}$ are chosen such that

$$\sum_{j=1}^{K} W_{ij}(\alpha)y_j(\alpha)=1, \quad 1\le i\le K, \tag{1.5}$$

then the stopped martingale $X_{\tau_{\mathcal{A}}}$ is given by

$$X_{\tau_{\mathcal{A}}}=\alpha\frac{\alpha^{\tau_{\mathcal{A}}}-1}{\alpha-1}\sum_{j=1}^{K} y_j(\alpha)-\alpha^{\tau_{\mathcal{A}}}.$$

After taking the expectation, a little algebra leads one to a strikingly simple formula for the generating function for $\tau_{\mathcal{A}}$.

Theorem 1.3.2 *If the vector $\{y_j(\alpha)\}_{1\le j\le K}$ solves the linear system (1.5), then*

$$\mathbf{E}(\alpha^{\tau_{\mathcal{A}}})=1-\frac{1}{1+\sum_{j=1}^{K} y_j(\alpha)\alpha/(1-\alpha)}.$$

We can use Theorem 1.3.2 to obtain the higher moments of $\tau_{\mathcal{A}}$, but it is also possible to use the method of gambling teams more directly. For example, to compute the second moment of $\tau_{\mathcal{A}}$ we ask the gambler from the j'th team that starts gambling in n'th round to place an initial bet of y_j+nz_j dollars on the first letter of A_j and to continue betting his fortune on the subsequent letters of A_j until he either loses or until some gambler observes a pattern from $\mathcal{A}$.

This time we write the winnings of the j'th team in the case of the i'th ending scenario by the sum

$$W_{ij}y_j+\tau_{\mathcal{A}}W_{ij}z_j+N_{ij}z_j$$

where W_{ij} is as before but where N_{ij} is a new quantity for which we will give an explicit formula shortly. The casino's net gain at time $X_{\tau_{\mathcal{A}}}$ then is given by

$$\begin{aligned} X_{\tau_{\mathcal{A}}} &= \sum_{j=1}^{K} y_j\frac{\tau_{\mathcal{A}}(\tau_{\mathcal{A}}+1)}{2}+\sum_{j=1}^{K} z_j\tau_{\mathcal{A}} \\ &\quad -\sum_{i=1}^{K}\left(\sum_{j=1}^{K} W_{ij}y_j\tau_{\mathcal{A}}+\sum_{j=1}^{K} N_{ij}y_j+\sum_{j=1}^{K} W_{ij}z_j\right)1_{E_i}. \end{aligned}$$

Now, if weights $\{y_j\}_{1\le j\le K}$ and $\{z_j\}_{1\le j\le K}$ are such that

$$\begin{aligned} \sum_{j=1}^{K} W_{ij}y_j &= 1, \quad 1\le i\le K, \\ \sum_{j=1}^{K}(N_{ij}y_j+W_{ij}z_j) &= 1, \quad 1\le i\le K, \end{aligned} \tag{1.6}$$

then the stopped martingale is equal to

$$\sum_{j=1}^{K} y_j \frac{\tau_{\mathcal{A}}(\tau_{\mathcal{A}}+1)}{2} + \sum_{j=1}^{K} z_j \tau_{\mathcal{A}} - \tau_{\mathcal{A}} - 1.$$

After the application of the optional stopping theorem we obtain a formula for the second moment.

Theorem 1.3.3 *If* $\{y_j\}_{1\le j\le K}$ *and* $\{z_j\}_{1\le j\le K}$ *solve the linear system (1.6), then*

$$\mathbf{E}(\tau_{\mathcal{A}}^2) = \frac{1 + (1 - \sum_{j=1}^{K} z_j - \sum_{j=1}^{K} y_j/2)\mathbf{E}(\tau_{\mathcal{A}})}{\sum_{j=1}^{K} y_j/2}.$$

As we mentioned above, N_{ij} is just another measure of the overlap of two patterns. Specifically, if the j'th gambling team bets on pattern A and the i'th ending scenario corresponds to pattern B, then we have the explicit recipe:

$$N_{ij} = \sum_{t=1}^{\min(m,l)} \delta_t(A,B)(1-t).$$

Here one should also note that from the representation (1.4) for $W_{ij}(\alpha)$ one also has the nice alternative formulas

$$W_{ij}(1) = W_{ij}, \qquad \frac{\partial W_{ij}(\alpha)}{\partial \alpha}\Big|_{\alpha=1} = N_{ij}.$$

As before, an example shows that these representations are all quite explicit.

Example 1.3.1 As usual we take $\Omega = \{1,2\}$, but now we consider the compound pattern $\mathcal{A} = \{11, 121\}$. If we further assume that

$$\mathbf{P}(Z_1 = 1) = \mathbf{P}(Z_1 = 2) = 1/2,$$

then we find

$$W_{ij} = \begin{bmatrix} 6 & 2 \\ 2 & 10 \end{bmatrix},$$

$$W_{ij}(\alpha) = \begin{bmatrix} 4\alpha^{-1}+2 & 2 \\ 2 & 8\alpha^{-2}+2 \end{bmatrix},$$

and

$$N_{ij} = \begin{bmatrix} -4 & 0 \\ 0 & -16 \end{bmatrix}.$$

The theorems of this section then give us the concrete answers:

$$\mathbf{E}(\tau_{\mathcal{A}}) = \frac{8}{3}, \quad \text{and} \quad \mathbf{Var}(\tau_{\mathcal{A}}) = 10,$$

and

$$\mathbf{E}(\alpha^{\tau_{\mathcal{A}}}) = \frac{\alpha^2(\alpha+2)}{8-4\alpha-\alpha^3} = \frac{\alpha^2}{4} + \frac{\alpha^3}{4} + \frac{\alpha^4}{8} + \frac{3\alpha^5}{32} + \frac{5\alpha^6}{64} + \frac{7\alpha^7}{128} + o(\alpha^7).$$

1.4 Patterns in Markov Dependent Trials

Gambling teams provide a handy way to deal with many questions about sequences of independent symbols, but, when the symbols are generated by a Markov chain, one finds that the method of gambling teams is especially powerful — even though some new subtleties are introduced. For example, in the Markov case one typically needs to introduce multiple teams of gamblers who gamble according to different rules. To illustrate the basic ideas in the simplest non-trivial case, we first apply the gambling team method to the calculation of the expected time until one observes a specified pattern in a sequence generated by a two-state Markov chain.

1.4.1 Two-state Markov chains and a single pattern

In next two sections we take $\{Z_n, n \geq 1\}$ to be a Markov chain with state space $\Omega = \{1, 2\}$. We suppose the chain has the initial distribution $\mathbf{P}(Z_1 = 1) = p_1$, $\mathbf{P}(Z_1 = 2) = p_2$ and the transition matrix

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}.$$

Here, as usual, p_{ij} is shorthand for $\mathbf{P}(Z_{n+1} = j \,|\, Z_n = i)$. Given a pattern $\mathrm{A} = a_1 a_2 \cdots a_m$, we then let τ_{A} denote the first time that the pattern is observed in the sequence generated by the Markov chain.

Next, we need to make explicit the Markov version of a fair casino where a gambler who bets on the event $\{Z_{n+1} = a\}$ is assumed to have first observed Z_n. Here, if one first observes $Z_n = 1$, then the bettor of \$1 dollar on the event $\{Z_{n+1} = a\}$ receives p_{1a}^{-1} dollars if $Z_{n+1} = a$ occurs; otherwise the bettor receives 0. Similarly, if one first observes $Z_n = 2$ and then bets that $Z_{n+1} = a$ the payoffs are p_{2a}^{-1} and 0 respectively.

There are now three distinct scenarios under which the pattern A can be observed. Either

- the pattern A occurs at the beginning of the sequence $\{Z_n, n \geq 1\}$, or
- the pattern 1A occurs at the end of the sequence, or

- the pattern 2A occurs at the end of the sequence.

The probability of the first scenario is easy to find, but to determine the individual probabilities of the last two scenarios would be more subtle. Instead we will use another gambling team trick to avoid such calculation. The new trick is to consider *two* gambling teams and to allow by teams to bet *differently* on the pattern A. The added flexibility will permit us to set things up so that teams total winnings are known if we know how the game ended.

For each time n two new gamblers are ready to take action, one from each team. The gamblers now follow the two rules:

1. For each n a gambler from the first team arrives before round n and watches the result of the n'th trial. He then bets y_1 dollars on the first letter of the sequence A and continues to bet his accumulated winnings on the successive letters in the successive rounds until either he loses or until the patten A is observed, either by himself or by some other gambler from one of the two teams. We call the gamblers on this team *straightforward gamblers.*

2. Gamblers from the second team bet differently. If $Z_n \neq a_1$ then the n'th gambler from the second team bets y_2 dollars on the round $n+1$ on the first letter of the pattern A. This gambler then continues to "let his fortune roll" until either he loses or until A is observed, either by himself or by some other gambler. On the other hand, if $Z_n = a_1$ then this gambler (intelligently!) bets y_2 dollars on a_2 on round $n+1$ and then he continues to bet on the remaining letters of the pattern $a_3 \cdots a_m$ until he loses or until the pattern A is observed by himself or by some other gambler. We call the gamblers of the second team *smart gamblers.*

Now, we let $W_{ij}y_j$, $i = 1,2,3$, $j = 1,2$ be amount of money that the j'th team wins if the game ends in the i'th scenario. It is vital to note that the *deterministic* quantities W_{ij} are easy to compute. The stopped martingale $X_{\tau_{\mathrm{A}}}$ that represents the net casino gain at time τ_{A} is given by

$$X_{\tau_{\mathrm{A}}} = (y_1 + y_2)(\tau_{\mathrm{A}} - 1) - \sum_{i=1}^{3}\sum_{j=1}^{2} W_{ij} y_j 1_{E_i},$$

where 1_{E_i} is the indicator of i'th ending scenario. To see this note that no money was bet on the first round, and $y_1 + y_2$ was the amount bet by each of the first time bettors at each of the subsequent rounds.

Now, we assume that we can find $\{y_j\}_{1 \leq j \leq 2}$ such that

$$\sum_{j=1}^{2} W_{ij} y_j = 1, \quad 2 \leq i \leq 3.$$

The existence of y_1 and y_2 depends on the computed values $\{W_{ij}\}$, but they will exist except in isolated, degenerate cases. The stopped martingale is then given by the simpler formula

$$X_{\tau_{\mathrm{A}}} = (y_1 + y_2)(\tau_{\mathrm{A}} - 1) - (W_{11}y_1 + W_{12}y_2)1_{E_1} - 1_{E_1^c},$$

where $1_{E_1^c}$ is the indicator of the complement of the 1st ending scenario. Taking the expectation and employing the optional stopping theorem we obtain

$$0 = (y_1 + y_2)(\mathbf{E}(\tau_{\mathrm{A}}) - 1) - \pi_1(W_{11}y_1 + W_{12}y_2) - (1 - \pi_1),$$

where π_1 is the probability of the first scenario. As we noted earlier, it is always easy to compute π_1, so at the end of the day one just solves for $\mathbf{E}(\tau_{\mathrm{A}})$ to find

$$\mathbf{E}(\tau_{\mathrm{A}}) = 1 + \frac{\pi_1(W_{11}y_1 + W_{12}y_2) + (1 - \pi_1)}{y_1 + y_2}.$$

Example 1.4.1 To see that this is indeed an explicitly computable formula, consider the pattern A = 121. The straightforward gamblers start with a fortune of y_1 dollars and successively bet their accumulated fortunes on the successive values of 121. On the other hand, the smart gamblers start with y_2 dollars and bet their accumulated fortune one the successive values of 121 if they observed 2 before placing their first bet, but they bet their money on the successive values of 21 if they observed 1 before placing their first bet. The three scenarios are (1) the game ends with 121 at the beginning, or (2) the game ends with 2121 at the end of some indeterminate number of rounds, or (3) the game ends with 1121 at the end of some indeterminate number of rounds. The 3×2 (scenarios by teams) matrix $\{W_{ij}\}$ is then given by

$$\begin{bmatrix} \frac{1}{p_{21}} & \frac{1}{p_{12}p_{21}} + \frac{1}{p_{21}} \\ \frac{1}{p_{21}p_{12}p_{21}} + \frac{1}{p_{21}} & \frac{1}{p_{21}p_{12}p_{21}} + \frac{1}{p_{12}p_{21}} + \frac{1}{p_{21}} \\ \frac{1}{p_{11}p_{12}p_{21}} + \frac{1}{p_{21}} & \frac{1}{p_{12}p_{21}} + \frac{1}{p_{21}} \end{bmatrix}.$$

To determine the initial bet sizes y_1 and y_2 we then just solve the relations

$$\begin{aligned} y_1\Big(\frac{1}{p_{21}p_{12}p_{21}} + \frac{1}{p_{21}}\Big) + y_2\Big(\frac{1}{p_{21}p_{12}p_{21}} + \frac{1}{p_{12}p_{21}} + \frac{1}{p_{21}}\Big) &= 1, \\ y_1\Big(\frac{1}{p_{11}p_{12}p_{21}} + \frac{1}{p_{21}}\Big) + y_2\Big(\frac{1}{p_{12}p_{21}} + \frac{1}{p_{21}}\Big) &= 1, \end{aligned}$$

to find

$$y_1 = \frac{p_{11}p_{12}p_{21}}{p_{12} + p_{21} + p_{12}p_{21}} \quad \text{and} \quad y_2 = \frac{p_{12}p_{21}(p_{21} - p_{11})}{p_{12} + p_{21} + p_{12}p_{21}}.$$

The probability π_1 of the first scenario is just $p_1 p_{12} p_{21}$, so after substitution and simplification we obtain the pleasingly succinct formula

$$\mathbf{E}(\tau_{\mathrm{A}}) = 1 + \frac{p_2}{p_{21}} + \frac{1}{p_{21}^2} + \frac{1}{p_{12}p_{21}}.$$

1.4.2 Two-state Markov chains and compound patterns

The next natural challenge is to compute the expected value of $\tau_{\mathcal{A}}$ the fist time that one observes a pattern from the set $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \cdots, \mathrm{A}_K\}$. The gambling teams method again applies, but one more nuance emerges. In particular, it is useful to refine the split notion of ending scenarios into *initial-ending scenarios* and *later-ending scenarios.* Specifically, we consider K *initial-ending scenarios* where in the i'th initial-ending scenario the pattern $\mathrm{A}_i, 1 \leq i \leq K$ occurs in the beginning of the sequence $\{Z_n, n \geq 1\}$, and we consider $2K$ *later-ending scenarios* where either the pattern $1\mathrm{A}_i$ for some $1 \leq i \leq K$ occurs or else the pattern $2\mathrm{A}_i$ for some $1 \leq i \leq K$ occurs after some indeterminate number of rounds.

This gives us complete coverage of how the one of the patterns from $\mathcal{A}$ can appear; in fact the coverage is over complete since it is possible that some of the later-ending scenarios need not be achievable as final blocks of the Markov sequence at time $\tau_{\mathcal{A}}$. For example, if $\mathcal{A} = \{212, 22\}$, then *doubling step* formally gives us four later-ending scenarios: $\{\cdots 1212, \cdots 2212, \cdots 122, \cdots 222\}$, but 221 and 222 cannot occur as a substring of the string $Z_1, Z_2, ..., Z_{\tau_{\mathcal{A}}}$. Similarly, if the initial collection is $\mathcal{A} = \{21, 111\}$, then the only observable later-scenarios are $\{\cdots 121, \cdots 221\}$.

Thus, one typically needs to do some cleaning of the initial list of later-ending scenarios, and, if a later-ending scenario cannot be observed in a sequence that ends at time $\tau_{\mathcal{A}}$, then the scenario is eliminated from the original list of $2K$ later-ending scenarios. The *final list* of ending scenarios is then the set of initial-ending scenarios and later-ending scenarios that have not been eliminated. We let N' denote the number of later-ending scenarios in the final list.

Now we introduce N' gambling teams, one for each of the later-ending scenarios. The rule is simple. If in the final list of scenarios there are *two* later-ending scenarios associated with the pattern A_i, then we introduce *two* gambling teams. One bets on A_i in a straightforward way, and one team bets on A_i in the smart way of the previous section. On the other hand, if in the final list we have only one later-ending scenario associated with the pattern A_i we will use only one gambling team of straightforward gamblers. Finally, if there are no later-ending scenarios in the final list associated with A_i, no gambling teams linked with A_i are needed.

We let X_n denote the casino's net gain at time n. We take $y_j, 1 \leq j \leq N'$ to be the initial bet with which a gambler from the j'th gambling team starts his betting, and we let $W_{ij}y_j$, $1 \le i \le K$ be total winnings of the j'th gambling team in the case of the i'th initial-ending scenario. Finally, we let $y_j W_{ij}$, $K+1 \le i \le K+N'$ be total winnings of the j'th gambling team in the case when the game is ended by the i'th later-ending scenario. Then the stopped martingale $X_{\tau_{\mathcal{A}}}$ is given by

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

where E_i is the event that the i'th scenario occurs. Again, the W_{ij} are not random, and, parallel to our earlier calculations, we assume that one can find $\{y_j\}_{1\le j\le N'}$ such that

$$\sum_{j=1}^{N'} W_{ij}y_j = 1, \text{ for all } K+1 \le i \le K+N'. \tag{1.7}$$

We then have the representation

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

so the optional stopping theorem tells us that

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

where π_i is the probability that the i'th initial-ending scenario occurs. Solving this equation, we obtain a slightly untidy but still completely computable formula for $\mathbf{E}(\tau_{\mathcal{A}})$.

Theorem 1.4.1 *If $\{y_j\}_{1\le j\le N'}$ solves the linear system (1.7), then*

$$\mathbf{E}(\tau_{\mathcal{A}}) = 1 + \frac{(1 - \sum_{i=1}^{K}\pi_i) + \sum_{i=1}^{K}\pi_i\sum_{j=1}^{N'} y_j W_{ij}}{\sum_{j=1}^{N'} y_j}. \tag{1.8}$$

Example 1.4.2 For the collection of patterns $\mathcal{A} = \{11, 212\}$ we find after the doubling and cleaning steps that the final list of later-ending scenarios is $\{211, 1212, 2212\}$. Together with our initial-ending scenarios, so we have a total of five ending scenarios which we order as

$$\{11, 212, 211, 1212, 2212\}.$$

The scenario-by-team win matrix $\{W_{ij}\}$ is then given by

$$\begin{bmatrix} \frac{1}{p_{11}} & 0 & 0 \\ 0 & \frac{1}{p_{12}} & \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}} \\ \frac{1}{p_{21}p_{11}} + \frac{1}{p_{11}} & 0 & 0 \\ 0 & \frac{1}{p_{12}p_{21}p_{12}} + \frac{1}{p_{12}} & \frac{1}{p_{12}p_{21}p_{12}} + \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}} \\ 0 & \frac{1}{p_{22}p_{21}p_{12}} + \frac{1}{p_{12}} & \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}} \end{bmatrix},$$

and, after solving the corresponding linear system, we find that the appropriate initial team bets are given by

$$y_1 = \frac{p_{21}p_{11}}{1+p_{21}}, \quad y_2 = \frac{p_{22}p_{21}p_{12}}{p_{21}+p_{12}+p_{21}p_{12}}, \quad y_3 = \frac{p_{21}p_{12}(p_{12}-p_{22})}{p_{21}+p_{12}+p_{21}p_{12}}.$$

The probabilities π_1 and π_2 that 11 and 212 are initial segments of the process $\{Z_n, n \geq 1\}$ are given by p_1p_{11} and $p_2p_{21}p_{12}$ respectively, so the formula (1.8) leads one to the following result

$$\mathbf{E}(\tau_{\mathcal{A}}) = 2 + p_1p_{12} + \frac{1-p_1p_{11}}{p_{21}},$$

which we see was not so complicated after all.

Finally, one should note that when a martingale method for the expected waiting time is developed, it is usually straightforward to extend the method to obtain formulas for higher moments or generating functions. We have already seen how this can be done in the independent model, and Glaz et. al. (2006) give a more detailed exposition that covers the case of the two-state Markov chains.

1.4.3 Finite state Markov chains

Now consider a temporally homogeneous Markov chain $\{Z_n, n \geq 1\}$ with a finite state space $\Omega = \{1, 2, ..., M\}$, initial distribution $\mathbf{P}(Z_1 = m) = p_m$, $1 \leq m \leq M$, and transition matrix $P = \{p_{ij}\}_{1 \leq i,j \leq M}$ where as always

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

We let $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_K\}$ denote a compound pattern, and let

$$\tau_{\mathcal{A}} = \min\{\tau_{\mathrm{A}_1}, ..., \tau_{\mathrm{A}_K}\}$$

denote the first time when we observe a pattern from $\mathcal{A}$ in the Markov sequence. We also assume that the Markov chain has the following normalization and regularization properties:

- We assume that no pattern of $\mathcal{A}$ contains another pattern of $\mathcal{A}$ as a subpattern. This property holds without loss of generality since if one pattern is a subpattern of another the longer one can be excluded from our list.
- We assume that $\mathbf{P}(\tau_{\mathcal{A}} = \tau_{\mathrm{A}_i}) > 0$ for all $1 \leq i \leq K$. If to the contrary one were to have $\mathbf{P}(\tau = \tau_{\mathrm{A}_i}) = 0$ for some i then A_i could simply be excluded from the list. This possibility is excluded by the first assumption for independent sequences, but for Markov sequences it often needs our attention. For example, if the pattern A_i contains subpattern km and $p_{km} = 0$, then A_i can not happen as a run of $\{Z_n, n \geq 1\}$.
- We assume that $\mathbf{P}(\tau_{\mathcal{A}} < \infty) = 1$. If the patterns of $\mathcal{A}$ all contain transient states this condition can easily fail even for a finite Markov chain. Here we should note that for finite Markov chains the basic finiteness condition $\mathbf{P}(\tau_{\mathcal{A}} < \infty) = 1$ already implies the formally stronger condition $\mathbf{E}[\tau_{\mathcal{A}}] < \infty$.

The Multi-state Chain Martingale Construction

When $M = |\Omega| > 2$ the critical martingales require a more elaborate description. We begin by decomposing the possible occurrence of a single pattern A_i into an *initial list* of $1 + M + M^2$ ending scenarios:

- Either the sequence A_i occurs as an initial segment of $\{Z_n, n \geq 1\}$, or
- for some $1 \leq k \leq M$, the pattern kA_i occurs as an initial segment of the sequence $\{Z_n, n \geq 1\}$, or
- for some pair (k, m), $1 \leq k, m \leq M$, the pattern kmA_i occurs after some indeterminant number of rounds.

The first $1 + M$ ending scenarios are called *initial* scenarios. The last M^2 scenarios are called *later* scenarios. Since we have K patterns, we have an initial list of $(1 + M + M^2)K$ scenarios.

For every later scenario associated with the pattern $km\mathrm{A}_i$ we introduce a team of gamblers that we call the *$km\mathrm{A}_i$-gambling team.* Gambler $n + 1$ from the $km\mathrm{A}_i$-gambling team arrives before round $n + 1$ to observe the result of n'th trial, Z_n.

This gambler then starts his betting. If $Z_n = k$ he bets a certain amount of money (which is the same for all gamblers from the $km\mathrm{A}_i$-gambling team) on the pattern $m\mathrm{A}_i$. If $Z_n \neq k$ he bets on A_i. Here, of course, by "betting \$1 on the pattern $\mathrm{A} = a_1a_2 \cdots a_m$, when $Z_n = a_0$" we mean the following:

- After observing Z_n the gambler bets a dollar that the next trial yields a_1. If $Z_{n+1} \neq a_1$ he loses his dollar and leaves the game. If $Z_{n+1} = a_1$, he gets $1/p_{a_0 a_1}$. Note that the odds are fair. If he wins he continues his betting.

- Now he bets his entire capital that the $n+2$ round yields a_2. If it is a_2 he increases his capital by factor $1/p_{a_1 a_2}$, otherwise he leaves the game with nothing. He continues to bet his full fortune on the successive letters of the pattern A until either the pattern A is observed, or until some other gambler has succeeded.

Now recall that it can happen that some of the scenarios on our initial list simply cannot occur before the waiting time $\tau_{\mathcal{A}}$. Moreover, some ending scenarios are impossible simply because some new patterns associated with some ending scenarios cannot be observed at all in the Markov chain. Thus we need to clean the initial list of ending scenarios.

Those scenarios that cannot occur at all and those that can occur only after the time τ_{A} must be eliminated. Let K' denote the number of initial scenarios, and let N' denote the number of later scenarios that we have in our list after cleaning. For each j'th later scenario in the new list, we introduce the corresponding gambling team, and we assume that the inial amount with which the gamblers of the j'th team start their betting is y_j. The values $\{y_j\}$ will be chosen later.

Let $y_j W_{ij}$, $1 \leq i \leq K' + N'$, $1 \leq j \leq N'$ be the amount of money that the j'th team wins in the i'th ending scenario. Let X_n denote the casino's net gain from all teams at time n. The sequence $\{X_n\}$ 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. By bookkeeping, one finds for the stopped martingale $X_{\tau_{\mathcal{A}}}$ that

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

where E_i is the event that the i'th scenario occurs. Here, again W_{ij} is not a random variable; it depends only on overlap properties of the pattern associated with the i'th scenario and the pattern associated with the j'th gambling team.

If we now assume that we can find $\{y_j\}_{1 \leq j \leq N'}$ such that

$$\sum_{j=1}^{N'} W_{ij} y_j = 1, \text{ for all } K' + 1 \leq i \leq K' + N', \tag{1.9}$$

then $X_{\tau_{\mathcal{A}}}$ has the more tractable representation

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

Since $\{X_n\}_{n\geq 1}$ has bounded increments and $\mathbf{E}[\tau_{\mathcal{A}}] < \infty$, the Doob's optional stopping theorem gives us

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

where π_i is the probability that the i'th initial scenario occurs. Solving the equation with respect to $\mathbf{E}(\tau_{\mathcal{A}})$ we obtain the main result of this section.

Theorem 1.4.2 *If $\{y_j\}_{1\leq j\leq N'}$ solves the linear system (1.9), then*

$$\mathbf{E}(\tau_{\mathcal{A}}) = 1 + \frac{(1 - \sum_{i=1}^{K'} \pi_i) + \sum_{i=1}^{K'} \pi_i \sum_{j=1}^{N'} y_j W_{ij}}{\sum_{j=1}^{N'} y_j}. \qquad (1.10)$$

Example 1.4.3 Let $\Omega = \{1, 2, 3\}$ and $\mathcal{A} = \{323, 313, 33\}$. Let the initial distribution be given by

$$p_1 = 1/3, \quad p_2 = 1/3, \quad p_3 = 1/3,$$

and let the transition matrix P be given by

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

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 because transitions 1 → 2 and 2 → 1 are impossible we get just six later scenarios:

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

Now we need to calculate the matrix W and we first consider some sample entries. 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 $1/(p_{23}p_{32}p_{23}) + 1/p_{23} = 68$. Finally, the entries of matrix W that correspond to the later scenarios — the ones that are needed for linear system (1.9) — are given by

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

Finally, from formula (1.10) we have the bottom line:

$$\mathbf{E}(\tau_{\mathcal{A}}) = 8\frac{7}{15}.$$

Higher Moments, the Generating Functions, and Efficiency

In parallel with our earlier examples, one can now take initial bets of size $y_j + z_j n$ to obtain a formula for the second moment, or take initial bets of size $y_j \alpha^n$ to obtain the corresponding generating function, c.f. Pozdnyakov (2008).

Here we should note that while the method of this subsection is also applicable to two-state Markov chains, it is certainly less efficient than the one given in the Subsection 1.4.2. Here, in the case of two-state Markov chains we would have $4K$ ending scenarios but the method of Subsection 1.4.2 needs only $2K$.

Finally, one should remark some of the computational differences between the martingale technique to the Markov chain imbedding method. To find the expected time $\mathbf{E}(\tau_{\mathcal{A}})$ via an appropriate Markov chain imbedding one needs to solve a linear system associated with the transition matrix of imbedded Markov, c.f. Fu and Chang (2002, p. 73). The size of the matrix depends on the cardinality K of the compound pattern $\mathcal{A}$ and lengths of single patterns in $\mathcal{A}$. Our matrix depends on K and cardinality M of the alphabet. Thus, there are situations when the martingale approach is computationally more effective. For a very simple example, one can take $\mathcal{A}$ to consist of just one very long pattern.

1.5 Applications to Scans

In its simplest form [c.f. Naus (1965)], the scan statistic is the largest number of “events that occur” in a window of a given fixed length when we scan the window over a realization of a temporally homogeneous process up to a specified terminal time. For a concrete example, consider a sequence of independent Bernoulli trials $\{Z_n, n \geq 1\}$ with

$$\mathbf{P}(Z_i = 1) = p = 1 - \mathbf{P}(Z_i = 0).$$

Now given $1 \leq w \leq T$ and $1 \leq i \leq T - w + 1$, we consider the sums

$$Y_{i,w} = \sum_{j=i}^{i+w-1} Z_j,$$

and we define the *scan statistic* $S_{w,T}$ to be the maximum of $Y_{i,w}$; that is,

$$S_{w,T} = \max_{1 \leq i \leq T-w+1} Y_{i,w}.$$

If $\tau_{k,w}$ denotes the first time when one first observes at least k occurrences of the value 1 in a window of length w, then $\tau_{k,w}$ is related to the scan statistic by

$$\mathbf{P}(S_{w,T} \geq k) = \mathbf{P}(\tau_{k,w} \leq T).$$

For us, the key observation is that the waiting time $\tau_{k,w}$ can as be viewed as the waiting time $\tau_{\mathcal{A}}$ for an appropriate compound pattern $\mathcal{A}$. For example, for $k = 3$ and $w = 5$ the compound pattern $\mathcal{A}$ is given by

$$\{111, 1101, 1011, 11001, 10101, 10011\}.$$

The bottom line message is that knowledge of the distribution of $\tau_{\mathcal{A}}$ gives us distribution of associated scan statistics. Moreover, this method of association goes well beyond the simple scan of this example. Analogous transformations permit one to treat the variable window scans of Glaz and Zhang (2006) or the double scans considered by Naus and Stefanov (2002) and Naus and Wartenberg (1997).

1.5.1 Second moments and distribution approximations

Since martingale methods yield effective computations of the moments of the waiting time $\tau_{\mathcal{A}}$, it is natural to ask if martingales methods also suggest *approximations* of the distribution of $\tau_{\mathcal{A}}$ that use the first two (or perhaps more) moments of the waiting time.

It is reasonable from the clumping heuristic that the stopping time $\tau_{\mathcal{A}}$ that one associates with a scan statistic should have tail probabilities $\mathbf{P}(\tau_{\mathcal{A}} \leq n)$ that are close to those of the exponential distribution. Still, when one considers the whole distribution, there are natural competitors to the exponential such as the Gamma, the Weibull and the shifted exponentials. The main finding in Pozdnyakov et al. (2005) was that in many natural situations it is in fact the class of shifted exponential distribution provides the most accurate approximation to the distribution of $\tau_{\mathcal{A}}$.

To make this approximation explicit, we first recall that X' called a shifted exponential provided that $X' = X + c$ where X has an exponential distribution. We take X' as our moment matching approximation to $\tau_{\mathcal{A}}$ provided c is chosen so that

$$\mathbf{E}(X + c) = \mathbf{E}(\tau_{\mathcal{A}}), \quad \mathbf{Var}(X + c) = \mathbf{Var}(\tau_{\mathcal{A}}).$$

For the tail probabilities this approximation gives us the relation

$$\mathbf{P}(\tau_{\mathcal{A}} \leq n) \approx 1 - \exp(-(n + 0.5 + \sigma - \mu))/\sigma), \tag{1.11}$$

where $\mu = \mathbf{E}(\tau_{\mathcal{A}})$, $\sigma = \mathbf{Var}(\tau_{\mathcal{A}})$, and the 0.5 term provides a continuity correction. As following examples demonstrate, this approximation works remarkably well for a wide variety of scan statistics.

Example 1.5.1 (*Fixed window scans*). Here $\{Z_n, n \geq 1\}$ is a sequence of Bernoulli trials. We consider two scans: at-least-3-out-of-10 (Table 1.1) and at-least-4-out-of-20 (Table 1.2).

For the fixed window scan statistics, Glaz and Naus (1991) developed tight lower and upper bounds which are provided in Tables 1.1 and 1.2 along with the approximations based on the exponential, shifted exponential, and gamma distributions. The Weibull distribution based approximation is omitted, because the performance of Weibull approximations are significantly worse than those of the exponential and the gamma. As it can be seen, the shifted exponential approximation does consistently well. In the easy case when μ is large and σ is close to μ, the differences between the various approximations are marginal, and all of the estimates are close to the true probability. On the other hand, if μ is relatively small and σ differs from μ, then the approximations based on the exponential and gamma distributions do not perform nearly as well as the shifted exponential approximations.

Example 1.5.2 (*Variable window scans*). Again we let $\{Z_n, n \geq 1\}$ be a sequence of Bernoulli trials, but this time we scan for the occurrence of either of two situations: either we observe at least 2 failures in 10 consecutive trials, or we observe at least 3 failures in 50 consecutive trials. Here are interested in the approximation for the distribution of the waiting time τ until one of these two situations occurs. In this case we need a compound pattern $\mathcal{A}$ with 224 patterns in order for τ and $\tau_{\mathcal{A}}$ to have the same distribution.

The numerical results are given in Table 1.3. Since analytical bounds for this type of scans are not available, the performance of the approximation is judged by comparison with estimated probabilities based on $100,000$ replications. Here, again, we see that the shifted exponential distribution approximation that is calibrated by two moments performs quite well.

Table 1.1: Fixed window scans: at least 3 failures out of 10 consecutive trials, $\mathbf{P}(Z_n = 1) = .01$, $\mu = 30822$, $\sigma = 30815$

n	exponential	shifted exponential	gamma	upper bound	lower bound
500	0.01600	0.01589	0.01597	0.01588	0.01589
1000	0.03183	0.03173	0.03179	0.03171	0.03174
1500	0.04741	0.04731	0.04736	0.04729	0.04733
2000	0.06274	0.06265	0.06267	0.06262	0.06267
2500	0.07782	0.07773	0.07775	0.07770	0.07776
3000	0.09266	0.09258	0.09258	0.09254	0.09261
4000	0.12162	0.12155	0.12154	0.12150	0.12169
5000	0.14966	0.14960	0.14957	0.14954	0.14965

Table 1.2: Fixed window scans: at least 4 failures out of 20 consecutive trials, $\mathbf{P}(Z_n = 1) = .05$, $\mu = 481.59$, $\sigma = 469.35$

n	exponential	shifted exponential	gamma	upper bound	lower bound
50	0.09110	0.07827	0.08268	0.07713	0.07940
60	0.10977	0.09770	0.10059	0.09543	0.09989
70	0.12807	0.11672	0.11828	0.11337	0.11991
80	0.14599	0.13534	0.13573	0.13095	0.13949
90	0.16354	0.15357	0.15292	0.14819	0.15864
100	0.18073	0.17141	0.16985	0.16508	0.17736

Example 1.5.3 (*Double scans*). Let $\{Z_n, n \geq 1\}$ be an i.i.d. sequence of random variables with the three valued distribution specified by

$$\mathbf{P}(Z_n = 1) = .04, \quad \mathbf{P}(Z_n = 2) = .01, \quad \text{and} \quad \mathbf{P}(Z_n = 0).$$

Now we consider two types of "failures"; a type 1 failure corresponds to observing a 1 and a type 2 failure corresponds to observing a 2. Further, we assume that we scan with a window of length 10 for until we observe at least 2 failures of type 2 or observe at least 3 failures (of any combination of kinds). Table 1.4 shows that the shifted exponential approximation works well even when μ and σ are relatively small and significantly different.

The initial arguments of Pozdnyakov et al. (2005) in favor of the shifted exponential approximation were predominantly empirical, but subsequently a more theoretical motivation has emerged from work of Fu and Lou (2006, p. 307) which shows that for large n one has

$$\mathbf{P}(\tau_{\mathcal{A}} \geq n) \sim C^* \exp(-n\beta),$$

where the constants C^* and β are defined in terms of the largest eigenvalue (and corresponding eigenvector) of what Fu and Lou (2006) call the *essential transition probability matrix* of the imbedded finite Markov chain associated with compound pattern $\mathcal{A}$. One should note that this matrix is not a proper transition matrix; rather it is a restriction of a transition matrix.

Now, if omit the continuity factor correction in our shifted exponential approximation (1.11) we have an approximation of exactly the same form:

$$\mathbf{P}(\tau_{\mathcal{A}} \geq n) \approx \exp(-(n + \sigma - \mu)/\sigma) = \exp((\mu - \sigma)/\sigma) \exp(-n/\sigma).$$

These relations suggest that there is a strong connection between the largest eigenvalue of the essential transition matrix of the imbedded Markov chain and the first and second moments of $\tau_{\mathcal{A}}$. In particular, we conjecture that (in the typical case at least) the largest eigenvalue $\lambda_{[1]}$ of the essential transition probability matrix of the imbedded finite Markov chain associated with compound pattern $\mathcal{A}$ will satisfy the approximation

$$\lambda_{[1]} \approx \exp(-1/\sigma). \tag{1.12}$$

1.5.2 Scan for clusters of a certain word

Let $\{Z_n, n \geq 1\}$ be a sequence of independent identically distributed random variables that takes values over the alphabet $\Omega = \{1, 2, ..., M\}$ and let the distribution be given by

$$p_1 = \mathbf{P}(Z_n = 1) > 0,\, p_2 = \mathbf{P}(Z_n = 2) > 0,\, \ldots, p_M = \mathbf{P}(Z_n = M) > 0.$$

Table 1.3: Variable window: at least 2 failures out of 10 trials or at least 3 failures out of 50 trials, $\mathbf{P}(Z_n = 1) = .01$, $\mu = 795.33$, $\sigma = 785.85$

		shifted		simulated
n	exponential	exponential	gamma	N=100000
50	0.05857	0.05085	0.05542	0.05029
60	0.07033	0.06285	0.06685	0.06187
70	0.08195	0.07470	0.07817	0.07404
80	0.09342	0.08640	0.08939	0.08623
90	0.10474	0.09796	0.10050	0.09718
100	0.11593	0.10936	0.11150	0.11058

Table 1.4: Double scans: at least 2 type II failures out of 10 trials or at least 3 failures of any kind out of 10 trials, $\mathbf{P}(Z_n = 1) = .04$, $\mathbf{P}(Z_n = 2) = .01$, $\mu = 324.09$, $\sigma = 318.34$

		shifted		simulated
n	exponential	exponential	gamma	N=100000
10	0.02438	0.01480	0.02175	0.01401
15	0.03932	0.03015	0.03568	0.03084
20	0.05403	0.04527	0.04959	0.04508
25	0.06851	0.06015	0.06342	0.06169
30	0.08277	0.07479	0.07714	0.07590
35	0.09681	0.08921	0.09074	0.09134
40	0.11064	0.10340	0.10419	0.10529
45	0.12425	0.11738	0.11749	0.11878
50	0.13766	0.13113	0.13063	0.13342

Given a pattern $\mathrm{A} = a_1 a_2 \cdots a_m$ over the alphabet Ω, we then take a window of length $w \geq m$ and scan the sequence until the time τ when in the window of width w we have k (possibly overlapping) occurrences of pattern A.

One can show that τ is equal to the waiting time until the occurrence of a certain compound pattern $\mathcal{A}$, so formally the moments of τ follow from our previous results. Unfortunately, this approach runs into computational problems since the cardinality of compound pattern $\mathcal{A}$ grows exponentially as window width w increases. There seems to be no way to circumvent this problem entirely, but given that $\mathcal{A}$ can be computed we can greatly cut down on much of the other work.

A New Betting Scheme

The basic idea is to bet only on the pattern A and to *pause* the betting between non-overlapping occurrences of A. To make this explicit, we first take $\mathcal{A}$ as given and consider a certain equivalence relation on the patterns from $\mathcal{A}$. Specifically, we say that elements A_i and A_j from $\mathcal{A}$ are *similar* provided that

- lengths of A_i and A_j are the same,
- A_i and A_j have the same number of overlapping occurrences of A and,
- the patterns A_i and A_j have copies of A's at the same positions.

Now, to each equivalence class under this relation, we can associate a unique pattern over the *extended* alphabet $\bar{\Omega} = \{1, 2, ..., M, *\}$ by a simple rule. If the simple pattern $\mathrm{A}_i \in \mathcal{A}$ is a representative of an equivalence class, then to construct what we will call the "star-pattern" for the class we replace each symbol of A_i that is not part of a block equal to A by the symbol $*$. This recipe is made clear with an example.

Example 1.5.4 Let $\Omega = \{1, 2, 3\}$ and let A = 121. Suppose we want to scan until we find the occurrence of at least two copies of A's in a window of 8 symbols. The compound pattern $\mathcal{A}$ associated with this scan consist of 11 simple patterns, none of which is subpattern of another):

1. exactly-2-in-5: 12121,
2. exactly-2-in-6: 121121,
3. exactly-2-in-7: 1211121 and 1213121,
4. exactly-2-in-8: 12111121, 12122121, 12113121, 12123121, 12131121, 12132121, and 12133121.

Now, although we have 11 simple patterns in $\mathcal{A}$ we have only 4 equivalence classes which we can enumerate with their star-patterns:

$$12121, 121121, 121*121, 121**121.$$

Here, it is important to note that $*$ does not mean just "any symbol", because, for example, 12121121 is not in $\mathcal{A}$, and, as a result, class $121**121$ does not include 12121121.

Given this reduction to equivalence classes, there are analogous reductions for the rest of our tools, such as the ending scenarios. Now, we introduce a list of ending scenarios associated with the list of equivalence classes (or star-patterns). As before, we associate a gambling team with each element of the final list of ending scenarios.

The real key is the new betting rule. Now, a gambler from a gambling team associated with a star-pattern bets on a symbol if it is a symbol from Ω but he simply passes when it is a star. For example, a gambler from the gambling team that corresponds to $121**121$ first bets on 121 in the sequential fashion that should now be quite familiar. If he is successful after those three bets, he then pause for two rounds. After the pause he bets then successively bets his entire capital on 121, the rest of the star-pattern.

Assume that we have N' ending scenarios and N' gambling teams. A gambler from the j'th gambling team that joins the game in n'th round will bet y_j dollars. Next let y_jW_{ij}, $1 \leq i, j \leq N'$ be the total winnings of the j'th gambling team in the case that game was ended by the i'th scenario. As before, W_{ij} is not random; it is fully determined by the pattern of overlap of the star-patterns associated with the given gambling team and ending scenario.

To make this explicit, we let $\mathrm{E} = e_1e_2\cdots e_m$ and $\mathrm{T} = t_1t_2\cdots t_l$ be two patterns over the extended alphabet $\bar{\Omega}$. We first define a measure of "two letters coincidence":

$$\delta(e_i, t_j) = \begin{cases} 1, & \text{if } t_j = * \\ 1/\mathbf{P}(Z_1 = t_j), & \text{if } t_j \neq *, e_i = t_j, \\ 0, & \text{if } t_j \neq *, t_j \neq e_i. \end{cases}$$

Next, we define a general measure of overlap for E and T:

$$W(\mathrm{E}, \mathrm{T}) = \sum_{i=1}^{\min(m,l)} \prod_{j=1}^{i} \delta(e_{m-i+j}, e_j).$$

Finally, if the i'th ending scenario is associated with the pattern E and the j'th gambling team bets on T, then we have the explicit (and deterministic) formula,

$$W_{ij} = W(\mathrm{E}, \mathrm{T}).$$

If X_n is the casino's total net gain at the end of the n'th round, then it is again a martingale, since taking a pause preserves a martingale property. At time $\tau_{\mathcal{A}}$ the stopped martingale is given by

$$X_{\tau_{\mathcal{A}}} = \sum_{j=1}^{N'} y_j \tau_{\mathcal{A}} - \sum_{i=1}^{N'} \sum_{j=1}^{N'} W_{ij} y_j 1_{E_i},$$

where 1_{E_i} is the indicator that the game is ended by the i'th scenario, so if the vector $\{y_j\}_{1 \le j \le N'}$ is a solution of the linear system

$$\sum_{j=1}^{N'} W_{ij} y_j = 1, \quad 1 \le i \le N', \tag{1.13}$$

then the stopped martingale has the tidy representation

$$X_{\tau_{\mathcal{A}}} = \sum_{j=1}^{N'} y_j \tau_{\mathcal{A}} - 1.$$

Since $\mathbf{E}(X_{\tau_{\mathcal{A}}})$, we come very quickly to our final formula.

Theorem 1.5.1 *If the vector $\{y_j\}_{1 \le j \le N'}$ solves the linear system (1.13), then expected value of $\tau_{\mathcal{A}}$ is given by*

$$\mathbf{E}(\tau_{\mathcal{A}}) = \frac{1}{\sum_{j=1}^{N'} y_j}.$$

Example 1.5.5 Let $\Omega = \{1, 2, 3\}$, and A $= 121$, and suppose we scan for at least two As in a window of 8 symbols. As we have seen there are only 4 equivalence classes:

$$12121, 121121, 121 * 121, 121 * *121.$$

The matrix W_{ij} in this case is

$$\begin{bmatrix} \frac{1}{p_1^3 p_2^2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^3 p_2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{2}{p_1^2 p_2} + \frac{1}{p_1} \\ \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^4 p_2^2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^3 p_2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} \\ \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^4 p_2^2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} \\ \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^2 p_2} + \frac{1}{p_1} & \frac{1}{p_1^4 p_2^2} + \frac{1}{p_1^2 p_2} + \frac{1}{p_1} \end{bmatrix},$$

Theorem 1.5.1 gives us the following formula for the expected value:

$$\mathbf{E}(\tau_{\mathcal{A}}) = \frac{1 + p_1 p_2 (1 + p_1 p_2)(1 + p_1(3 - p_2 - 2p_1 p_2))}{p_1^3 p_2^2 (1 + p_1(3 - p_2 - 2p_1 p_2))}$$

One obviously can extend this technique to the case of the higher moments and generating function.

1.6 Concluding Remarks

The martingale method for studying the waiting time until a compound pattern is now well developed — even the stubborn Markovian case. Still, from the examples given here, one can see that successful application of the method requires some detailed combinatorial information. Specifically, almost always needs to determine explicitly what we have called here the "final list of ending scenarios."

For problems, such as those that come from the theory of scan statistics, this final list can be large. Nevertheless, by the introduction of appropriate equivalence classes, one can still make steady progress. Explicit formulas for moments are possible more often than perhaps one might have guessed.

Going forward there are two problems that we believe deserve consideration: one general and one specific. The general problem is the identification of further problems like the one developed in subsection 1.5.2 for clusters of words. Generically, the challenge is to identify the problems in which one can find a substantial simplification of what would otherwise be the waiting time problem for a very large class of patterns. Correspondingly, it would be useful to identify as many problems as possible where one has a firm combinatorial understanding of the final list of ending scenarios.

The more specific problem is the conjecture given in equation (1.12). Historically, there has been considerable value to finding a good representation for the largest eigenvalue for even very special matrices. The class of matrices that are obtained as the essential (improper) transition probability matrix of imbedded finite Markov chain associated with compound pattern $\mathcal{A}$ is indeed special, yet it is still reasonably large. For this class the conjecture (1.12) provides an explicit — and novel — approach to the analysis of the largest eigenvalue.

References

1. Aldous, D. (1989) *Probability Approximations Via The Poisson Clumping Heuristic*, Springer Publishing, New York.

2. 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.

3. 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.

4. 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.

5. 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.

6. Glaz, J. and Naus, J.I. (1991). Tight bounds for scan statistics probabilities for discrete data, *Annals of Applied Probability*, **1**, 306-318.

7. Glaz, J., Naus, J.I. and Wallenstein, S. (2001). *Scan Statistics*, Springer, New-York.

8. Glaz, J. and Zhang, Z., (2006). Maximum scan score-type statistics, *Statistics and Probability Letters*, **76**, 1316-1322.

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

10. Naus, J.I. (1965). The distribution of the size of the maximum cluster of points on a line. *Journal of The American Statistical Association*, **60**, 532-538.

11. Naus, J.I. and Stefanov, V.T. (2002). Double-scan statistics. *Methodology and Computing in Applied Probability*, **4**, 163-180.

12. Naus, J.I. and Wartenberg, D. A. (1997). A double-scan statistic for clusters of two types of events. *Journal of The American Statistical Association*, **92**, 1105-1113.

13. Pozdnyakov, V. (2008). On Occurrence of Patterns in Markov Chains: Method of Gambling Teams, to appear in *Statistics and Probability Letters.*

14. 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.

15. Pozdnyakov, V., and Kulldorff, M. (2006). Waiting Times for Patterns and a Method of Gambling Teams, *The American Mathematical Monthly*, **113**, 134-143.

16. Shiryaev, A.N. (1995). *Probability*, Springer, New York.

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