MARTINGALE METHODS

VLADIMIR POZDNYAKOV AND J. MICHAEL STEELE

Abstract. We survey the ways that martingales and the method of gambling teams can be used to obtain otherwise hard-to-get information for the moments and distributions of waiting times for the occurrence of simple or compound patterns in an independent or a Markov sequence. We also survey how such methods can be used to provide moments and distribution approximations for a variety of scan statistics, including variable length scan statistics. Each of the general problems considered here is accompanied by one or more concrete examples that illustrate the computational tractability of the methods.

Keywords and phrases: Scan, pattern, martingale, waiting time, gambling teams, shifted exponential distribution

1. Introduction

To illustrate the notion of a scan statistic in the simplest possible context, one can begin with a sequence of independent identically distributed (i.i.d.) Bernoulli random variables $\{Z_i : 1 \le i \le T\}$. Now given a window size w with $1 \le w \le T$ and a set $\{i : 1 \le i \le T - w + 1\}$ of time indices one then considers any fixed function of the observations within the moving windows defined by w and a time index i. In the most classical case, one takes the function to be the sum of the observations in the window, and in this case one then has the window values

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

Finally, the scan statistic $S_{w,T}$ is defined to be the maximum of the window values. Thus, for the Bernoulli sequence and the sum function, the scan statistic is given by

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

A key feature of the scan statistic is that the sliding windows overlap, so one must deal with the maximum of a set of dependent random variables. As a consequence, the distribution theory of the scan statistic can be analytically demanding even in relatively simple contexts. Nevertheless, the scan statistic has a natural role in many problems and, despite its analytical challenges, it has seen a remarkably extensive development since its introduction in Naus (1965).

Vladimir Pozdnyakov: Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, Email: vladimir.pozdnyakov@uconn.edu.
J. M. Steele: The Wharton School, Department of Statistics, Huntsman Hall 447, University of Pennsylvania, Philadelphia, PA 19104. Email address: steele@wharton.upenn.edu.

In the context of the Bernoulli sequence and the sum function, the distribution of the scan statistic (1) has a dual interpretation that is often useful. Specifically, if we take $\tau_{k,w}$ to be the first time that one observes k or more occurrences of the value 1 in a window of length w, then one has the identity

$$\mathbf{P}(S_{w,T} \geq k) = \mathbf{P}(\tau_{k,w} \leq T). \tag{2}$$

The distribution theory of this scan statistic is therefore equivalent to the distribution theory of the waiting time $\tau_{\mathcal{A}}$ until the first occurrence of a pattern from a certain special *set* $\mathcal{A}$ of patterns. For example, if $k = 3$ and $w = 5$, then the associated set of patterns is given by $\mathcal{A} = \{111, 1101, 1011, 11001, 10101, 10011\}$. Such sets are also called *compound patterns*.

Analogous duality arguments apply much more broadly than this simple example might suggest. In particular, duality can be used to analyze variable length window scans (see e.g. Glaz and Zhang (2006) or Glaz et al. (2001)) or "double scans" (see e.g. Naus and Wartenberg (1997), Naus and Stefanov (2002), or Glaz et al. (2001)).

Finally, one should note that the validity of the duality relation (2) does not depend on the dependence structure of the underlying stochastic sequence. Still, in practice, one almost always assumes that the generating process $\{Z_n : n = 1, 2, \ldots\}$ is independent, or at least Markovian.

The main goal here is to give a unified review of the martingale techniques that have been developed for the analysis of the first occurrence time of simple and compound patterns. Li (1980) pioneered this development, and our first aim is to use the language and logic of gambling teams to give a treatment of the theory developed by Li (1980) and Gerber and Li (1981). We then focus on the devices that can be used to extend these martingale techniques so that they might deal with the occurrence times of patterns in Markov sequences. Finally, we consider how these moment calculations can be used to obtain effective approximations to the distributions of a wide range of scan statistics.

2. Occurrence of a Pattern in an Independent Sequence

To begin we let $\{Z_n : n = 1, 2, \ldots\}$ denote a sequence of i.i.d. random variables with values in the finite set $\Omega = \{1, 2, \ldots, M\}$, which we conventionally call the *alphabet*. Without loss of generality, we assume that each *letter* in this alphabet has non-zero probability of occurrence; moreover, we denote these probabilities by

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

By a *pattern* (or more explicitly, a *simple pattern*) we mean a finite ordered sequence $\mathrm{A} = a_1 a_2 \cdots a_m$ of letters from the alphabet Ω.

The random variable of central interest here is the waiting time until one observes the pattern A as a continuous run in the sequence $\{Z_n : n = 1, 2, \ldots\}$. This waiting time τ_{A} is clearly a stopping time and, with help of martingale theory, one can derive explicit formulas for its expected value, its higher moments, and its probability generating function.

2.1. **A gambling approach to the expected value.** To put the construction of Li (1980) and Gerber and Li (1981) in the language of gambling teams, we first consider a casino game that generates the sequence $\{Z_n : n \geq 1\}$. Next, we consider a *team of gamblers* that arrive one after another; in particular, the n'th gambler on the team arrives just before n'th round of play when the random letter Z_n is generated.

More remarkably, we assume that this unusual casino plays *fairly.* Thus, a dollar that is bet on an event that has probability p would pay $1/p$ dollars to a winner (and zero dollars to a loser). In other words, on each round of play the net expected return is zero for both the players and the casino.

We assume that each gambler on the team has an initial stake of one dollar. Moreover, each gambler follows a fixed strategy that is determined by the pattern $\mathrm{A} = a_1 a_2 \cdots a_m$. The n'th gambler on the team arrives just before the n'th round of play. At time n, when this gambler has his first opportunity to play, he bets his stake of one dollar on the event that $Z_n = a_1$. If Z_n is not a_1 this gambler loses his stake, and he stops playing. On the other hand, if $Z_n = a_1$, this gambler wins $1/\mathbf{P}(Z_n = a_1)$ dollars, and this pleasingly enlarged amount becomes the gamblers current stake. Having won, gambler number n continues to play in round $n+1$, and he bets his entire stake on the event $Z_{n+1} = a_2$. This gambler continues in this partially mad way until either he wins on all m bets in the pattern, or until he loses one of his bets — and goes broke.

At the time τ_{A} when the pattern $\mathrm{A} = a_1 a_2 \cdots a_m$ is first observed, the team decides to do a little bookkeeping. First, one should note that at time τ_{A} one has seen exactly τ_{A} gamblers have come into the casino. Moreover, each gambler has placed one or more bets.

At time τ_{A} the lucky gambler who came in at time $\tau_{\mathrm{A}} - m + 1$ has won all of his bet, and he has substantial winnings. On the other hand, many of the other gamblers are likely to have lost their stake, although a few modestly lucky gamblers may still be winners at time τ_{A}.

For example, suppose that one has $\Omega = \{1, 2, \ldots, 10\}$ and $P(Z_n = i) = 1/10$ for $i = 1, 2, \ldots, 10$. Further suppose that $A = 101$ and $\tau_{\mathrm{A}} = 20$. The lucky gambler who began to play time 18 will have won \$1000 while the gambler who began at time 20 will have won \$10. All of the other gamblers will have lost their stake.

If we now let X_n denote the total net gain of the casino at the end of the n'th round of play then, since the game was fair at each round of play, the stochastic process $\{X_n : n = 1, 2, \ldots\}$ is a martingale with respect to the sequence of σ-fields $\mathcal{F}_n = \sigma\{Z_1, \ldots, Z_n\}$ that are determined by sequence of observed letters. The casino's net gain at the time when the pattern A is first observed is $X_{\tau_{\mathrm{A}}}$. Since the stopping time τ_{A} is bounded by a geometrically distributed random variable and since the martingale $\{X_n : n = 1, 2, \ldots\}$ has bounded increments, Doob's stopping theorem (e.g. Williams (1991, p. 100)) tells us that

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

Moreover, since each gambler enters with a stake of one dollar, we have

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

where W is the amount of money in the pocket of the gambling team at time τ_{A}. The key observation here is that W is not random; its value is fully determined by the structure of pattern A. Moreover, it is easy to calculate W.

For a gambler to have any capital left time τ_{A} when the pattern A is first observed, the gambler had to be gambling at time τ_{A} and he had to win at time τ_{A}. In particular, the gamblers who entered the game before $\tau_{\mathrm{A}} - m + 1$ have all lost their stake. At time τ_{A} the lucky gambler who entered at time $\tau_{\mathrm{A}} - m + 1$ will have won all of his m bets. Among the gamblers who arrived after time $\tau_{\rm A} - m + 1$ some may have lost everything and some may still have money in their pocket.

To describe more precisely the money W in the team's hands, we need to introduce some notation. Specifically, for $0 \le i, j \le m$ we set

$$\delta_{ij} = \begin{cases} 1/\mathbf{P}(Z_1 = a_i), & \text{if } a_i = a_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

With this shorthand, one can check just by parsing the notation that we have the explicit representation

$$W = \delta_{11}\delta_{22}\cdots\delta_{mm} + \delta_{21}\delta_{32}\cdots\delta_{m,m-1} + \cdots + \delta_{m1}. \tag{4}$$

From the earlier observation that $\mathbf{E}(X_{\tau_{\rm A}}) = 0$, we therefore find

$$\mathbf{E}(\tau_{\rm A}) = \delta_{11}\delta_{22}\cdots\delta_{mm} + \delta_{21}\delta_{32}\cdots\delta_{m,m-1} + \cdots + \delta_{m1}. \tag{5}$$

The general formula (5) may not look so attractive at first, but it comes to life in the context of concrete examples.

Example 1. Let $\Omega = \{1, 2\}$ and consider the two patterns A = 1121 and B = 1112 of length four. Formula (5) then gives us

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

and, somewhat paradoxically, it also gives us

$$\mathbf{E}(\tau_{\rm B}) = (p_1 \times p_1 \times p_1 \times p_2)^{-1}.$$

Thus, for all choices of $0 < p_1 < 1$, the expected waiting time to see A is strictly greater than the expected waiting time to see B, despite the fact that the probability of observing A in any given 4-block is the exactly the same as observing B in that 4-block. Phenomena of this kind seem to have been first observed by Solov'ev (1966).

In fact, there are many non-intuitive phenomena in the occurrence times of patterns. One of the most curious of these arises in *Penney's Game*, cf. Guibas and Odlyzko (1981) or Graham et. al. (1994, pp. 401–410). In this game, each of two players, Alice and Bob, picks one of the patterns

$$\mathrm{A} = HHTHH, \quad \mathrm{B} = HTHHT, \quad \mathrm{C} = THHTH,$$

and the player whose pattern comes up first in a sequence of fair coin tosses is the winner of the game.

It turns out that if Alice chooses first and Bob chooses second, then Bob always has a choice that will give him a probability of winning that is strictly larger than 50%. In fact, the general formulas of Pozdnyakov and Kulldorff (2006) tell us the following:

$$\mathbf{P}(\text{A occurs before B}) = 0.58\cdots > 0.50$$

and

$$\mathbf{P}(\text{B occurs before C}) = 0.59\cdots > 0.50,$$

but

$$\mathbf{P}(\text{C occurs before A}) = 0.62\cdots > 0.50.$$

On might have expected such "non-transitive" triples to be hard to find, but instances of this phenomenon are rather common.

2.2. **Gambling on a generating function.** A simple modification of the gambling team method also leads to a useful formula for the probability generating function of τ_A. The trick is to *change the initial bet for each gambler*. Specifically, we fix an $\alpha \in (0,1)$ and we instruct the n'th gambler to start his betting by placing a bet of size α^n on the first letter of the pattern A. He would then continue gambling as before — betting his entire stake on the successive letters of A until he either has seen A or has lost one of his bets.

If we let X_n denote the casino's net gain at the end of the n'th round, then as before, the process $\{X_n : n = 1, 2, \dots\}$ is a martingale with bounded increments. Now, if we let $\alpha^{\tau_A} W(\alpha)$ denote the money in the hands of the team when the pattern A is first observed, then the casino's net gain X_{τ_A} can be written as the amount paid minus the team's current cash holdings,

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

The key here is that $W(\alpha)$ is again deterministic, even though $\alpha^{\tau_A} W(\alpha)$ *does* depend on chance. Fortunately, $\alpha^{\tau_A} W(\alpha)$ depends on chance in a controlled way that leads to pleasant cancelations.

Bookkeeping like the one used before gives a formula for $\alpha^{\tau_A} W(\alpha)$, and, after cancelation of α^{τ_A} from each side of that formula, one finds

$$W(\alpha) = \delta_{11}\delta_{22}\cdots\delta_{m,m-1}/\alpha^{m-1} + \delta_{21}\delta_{32}\cdots\delta_{m,m-1}/\alpha^{m-2} + \cdots + \delta_{m1}/1. \tag{6}$$

Here the factors δ_{ij}, $1 \le i, j \le m$, are given by the deterministic combinatorial formula (3), so (6) confirms that $W(\alpha)$ does not depend on chance.

Finally, by Doob's stopping time theorem, we get

$$0 = \mathbf{E}(X_{\tau_A}) = \mathbf{E}(\alpha^{\tau_A}) \left(\frac{\alpha}{\alpha - 1} - W(\alpha) \right) - \frac{\alpha}{\alpha - 1},$$

therefore, the probability generating function is given by

$$\mathbf{E}(\alpha^{\tau_A}) = \left(1 + \frac{1 - \alpha}{\alpha} W(\alpha) \right)^{-1}. \tag{7}$$

As before, the usefulness of this formula depends on our ability to evaluate formula (6) for $W(\alpha)$. This is often an easy calculation.

Example 2. If we again take $\Omega = \{1, 2\}$ and consider the pattern 1121, then we have

$$W(\alpha) = \frac{\alpha^{-3}}{p_1^3 p_2} + \frac{1}{p_1} \quad \text{so we have} \quad \mathbf{E}(\alpha^{\tau_A}) = \frac{p_1^3 p_2 \alpha^4}{1 - \alpha + \alpha^3(1 - p_2\alpha) p_1^2 p_2}.$$

By expansion one finds

$$\mathbf{E}(\alpha^{\tau_A}) = p_1^3 p_2 \alpha^4 + p_1^3 p_2 \alpha^5 + p_1^3 p_1 \alpha^6 + p_1^3 p_2 (1 - p_1^2 p_2)\alpha^7 + \cdots.$$

The first term of this expansion is obviously correct. For a more serious check of our formula for $\mathbf{E}(\alpha^{\tau_A})$, one can use it to compute the first moment

$$\frac{\partial \mathbf{E}(\alpha^{\tau_A})}{\partial \alpha}\Big|_{\alpha=1} = \frac{1}{p_1^3 p_2} + \frac{1}{p_1},$$

and this does recover what we found earlier for $\mathbf{E}(\tau_{\rm A})$.

2.3. **Second and higher moments.** In theory, the probability generating function $\mathbf{E}(\alpha^{\tau_{\rm A}})$ can always be used to calculate the second and higher moments of $\tau_{\rm A}$. Still, such calculations can be awkward, and it is useful to know that the higher moments of $\tau_{\rm A}$ can also be found by a direct modification of the method of gambling teams. This is most nicely illustrated by the calculation of $\mathbf{E}(\tau_{\rm A}^2)$.

This time the gambler who joins the game just before the n'th round will bet n dollars on the first letter of the pattern A. If X_n denotes the casino's net gain after n rounds of play, then $\{X_n : n = 1, 2, \dots\}$ is again a martingale. Moreover, at the stopping time $\tau_{\rm A}$ we have

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

where, as in (4), we have the deterministic sum

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

and now we also have a second deterministic term

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

One wants to apply an optional stopping theorem here, but in this case the increments of $\{X_n : n = 1, 2, \dots\}$ are no longer uniformly bounded, so one needs a more refined version of Doob's stopping time theorem. Nevertheless, the martingale differences satisfy $X_{n+1} - X_n = O(n)$ and the probabilities $\mathbf{P}(\tau_{\rm A} > n)$ decay at an exponential rate, so the stopping time theorem of Shiryaev (1995, p. 485) gives us

$$0 = \mathbf{E}(X_{\tau_{\rm A}}) = \mathbf{E}\left(\frac{1}{2}\tau_{\rm A}(\tau_{\rm A} - 1)\right) - W\mathbf{E}(\tau_{\rm A}) - N.$$

Now, when we solve for $\mathbf{E}(\tau_{\rm A}^2)$ and use our earlier formula $\mathbf{E}(\tau_{\rm A}) = W$, we get

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

and, as an immediate corollary, we also have

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

At this point the formula for $\mathbf{Var}(\tau_{\rm A})$ may seem abstract, but in many cases the deterministic terms W and N are easy to compute.

Example 3. For the traditional sample space $\Omega = \{1, 2\}$ and the pattern 1121, we now find from our formula for N that

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

Thus, in the end, we have a reasonably simple formula for the variance.

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

REMARKS. First, one should note that when either $p_1 \to 0$ or $p_2 \to 0$ one has the limit relation

$$\frac{\mathbf{E}(\tau_{\mathrm{A}})}{\mathbf{Var}(\tau_{\mathrm{A}})^{1/2}} \to 1. \tag{8}$$

In fact, this relation could have been anticipated without explicit knowledge of the formulas for $\mathbf{E}(\tau_{\mathrm{A}})$ and $\mathbf{Var}(\tau_{\mathrm{A}})$. When either $p_1 \to 0$ or $p_2 \to 0$, the occurrence of the pattern 1121 becomes a rare event. As a consequence, the clumping heuristic of Aldous (1989), suggests that the distribution of the random variable τ_{A} should be approximately exponential and the ratio $\mathbf{E}(\tau_{\mathrm{A}})/\mathbf{Var}(\tau_{\mathrm{A}})^{1/2}$ should be approximately one.

Second, one should also note that method used to calculate $\mathbf{E}(\tau_{\mathrm{A}}^2)$ can be modified in a natural way to calculate $\mathbf{E}(\tau_{\mathrm{A}}^3)$. The obvious idea is to have n'th gambler's begin his bets with a bet of size n^2 on the first letter of A. Everything then proceeds just as before, although the calculations do become more demanding and one must first calculate the values of both $\mathbf{E}(\tau_{\mathrm{A}})$ and $\mathbf{E}(\tau_{\mathrm{A}}^2)$.

3. Compound Patterns and Gambling Teams

By the duality relation (2), the analytical understanding the scan statistic (1) depends in a direct way on the waiting time $\tau_{\mathcal{A}}$ until one observes a pattern from a specified finite set $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \cdots, \mathrm{A}_K\}$ of patterns. If we write τ_{A_i} for the first time that one observes the simple pattern $\mathrm{A}_i \in \mathcal{A}$, then we have

$$\tau_{\mathcal{A}} = \min\{\tau_{\mathrm{A}_1}, \ldots, \tau_{\mathrm{A}_K}\}, \tag{9}$$

and $\tau_{\mathcal{A}}$ is called the waiting time until the occurrence of the *compound pattern* $\mathcal{A}$.

Now, given any simple patterns A and B, we say that $A = a_1a_2\cdots a_j$ is a subpattern of B if the letters $a_1a_2\cdots a_j$ appear consecutively in the pattern B. What makes this notion particularly useful here is that if A is a subpattern of B then

$$\min(\tau_{\mathrm{A}}, \tau_{\mathrm{B}}) = \tau_{\mathrm{A}}. \tag{10}$$

Thus, in the specification of the waiting time $\tau_{\mathcal{A}}$ for the compound pattern $\mathcal{A}$, one can always assume without loss of generality that no pattern in $\mathcal{A}$ is a subpattern of any other pattern in $\mathcal{A}$.

Gerber and Li (1981) used Markov chain embedding in their seminal study of the waiting time for a compound patterns, but here we will use martingale methods. This approach has several benefits. In particular, martingale methods suggest more clearly how one should proceed when the i.i.d. process $\{Z_j : j = 1, 2, \ldots\}$ is replaced by a Markov chain. Moreover, the martingale method also deals more efficiently the waiting times for certain highly regular patterns, such as those associated with scan statistics.

3.1. **Expected time.** To calculate the expected value of the waiting time $\tau_{\mathcal{A}}$ for compound pattern $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \cdots, \mathrm{A}_K\}$ we use K different gambling teams, and, for each $1 \le j \le K$, the gamblers from the j'th team place their bets according to the successive letters of the pattern A_j. For the moment, the initial bet for a player from the j'th team will just be denoted by y_j, and later we will find that a wise choice of these bet sizes can greatly aid our calculations.

At time $\tau_{\mathcal{A}}$ the gambling is stopped, and we calculate the casino's net winnings $X_{\tau_{\mathcal{A}}}$. If A_i is the first pattern in $\mathcal{A}$ that is observed, then we have $\tau_{\mathcal{A}} = \tau_{A_i}$. In this case, we say that the game has ended in the i'th scenario, and we denote this event by E_i. We also let $W_{ij}y_j$ denote the cash in hand of the j'th team when the game ends by the i'th scenario.

As usual, the casino's net win determines a martingale $\{X_n : n = 1, 2, \dots\}$ and at the stopping time $\tau_{\mathcal{A}}$, the casino's net win is $X_{\tau_{\mathcal{A}}}$. Moreover, by our usual bookkeeping of paid in capital minus cash on hand, we have

$$X_{\tau_{\mathcal{A}}} = \sum_{j=1}^{K} y_j \tau_{\mathcal{A}} - \sum_{i=1}^{K}\sum_{j=1}^{K} W_{ij} y_j 1_{E_i}, \tag{11}$$

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

The crucial trick here is that one can choose the initial bet sizes y_j, $1 \le j \le K$, in a way that greatly simplifies this sum. One first needs to note that each of the factors W_{ij} is purely deterministic; in fact we will shortly give a simple formula for W_{ij}. Next, we consider bet sizes $\{y_j : 1 \le j \le K\}$ that solve the linear system

$$\sum_{j=1}^{K} W_{ij} y_j = 1, \quad 1 \le i \le K. \tag{12}$$

Here one should note that we do not insist upon a non-negative solution for (12), and bets of negative amounts are given the natural interpretation, *viz.* if in a given game the winning of a bet of one dollar would return w dollars, then for any real value y a corresponding winning bet of y dollars would return wy dollars.

Now, since the sum of the indicators 1_{E_i} is equal to one, the representation (11) reduces to the much simpler sum,

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

Doob's stopping theorem gives us $\mathbf{E}X_{\tau_{\mathcal{A}}} = 0$. Therefore, in the end, we have a formula for $\mathbf{E}(\tau_{\mathcal{A}})$ that is simple and computationally effective.

Theorem 1. *If the real values $\{y_j : 1 \le j \le K\}$ solve the linear system (12), then expected value of $\tau_{\mathcal{A}}$ is given by*

$$\mathbf{E}(\tau_{\mathcal{A}}) = \frac{1}{\sum_{j=1}^{K} y_j}. \tag{13}$$

Remark. The system (12) may have a solution even when the matrix $\{W_{ij}\}$ is singular. Still, it is useful to note that Gerber and Li (1981) proved that the matrix $\{W_{ij}\}$ is always nonsingular provided that no pattern from $\mathcal{A}$ is a subpattern of another pattern in $\mathcal{A}$. Moreover, as noted earlier in this section, one can always assume that $\mathcal{A}$ satisfies this condition. The solvability of (12) and related systems has also been further investigated by Zajkowski (2014).

To make good use of the formula (13) one needs a tractable formula for W_{ij}, and such a formula is easier to give than one might guess. First, given any two patterns

$$\mathrm{A} = a_1 a_2 \cdots a_m \quad \text{and} \quad \mathrm{B} = b_1 b_2 \cdots b_l,$$

we set

$$\delta_t(A, B) = \begin{cases} \dfrac{1}{\prod_{s=1}^{t} \mathbf{P}(Z_1 = b_s)} & \text{if } b_1 = a_{m-t+1},\, b_2 = a_{m-t+2}, \dots, b_t = a_m, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Then one can find without difficulty that

$$W_{ij} = \sum_{t=1}^{\min(m,l)} \delta_t(A_i, A_j). \tag{15}$$

3.2. **The generating function and compound patterns.** The method of gambling teams can also be used to get the probability generating function $\mathbf{E}(\alpha^{\tau_{\mathcal{A}}})$ for compound pattern $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \dots, \mathrm{A}_K\}$. A small but familiar modification of the preceding method will do the trick.

We again use K gambling teams, and the bets of the members of the j'th team are determined by the successive letters of A_j. The only novelty is that members of the j'th team begin their gambling with a bet of size $y_j \alpha^n$, where $0 < \alpha < 1$ is fixed and the real values $\{y_j : 1 \le j \le K\}$ will be determined later.

By analogy with our previous calculation, we let $W_{ij}(\alpha) y_j \alpha^{\tau_{\mathcal{A}}}$ denote the cash in hand of the j'th gambling team at time $\tau_{\mathcal{A}}$ if the play stops in the i'th ending scenario. If $\{X_n : n = 1, 2, \dots\}$ denotes the martingale that gives us the casino's net gain at times $n = 1, 2, \dots$, then at the stopping time $\tau_{\mathcal{A}}$ we have

$$X_{\tau_{\mathcal{A}}} = \alpha \frac{\alpha^{\tau_{\mathcal{A}}} - 1}{\alpha - 1} \sum_{j=1}^{K} y_j - \sum_{i=1}^{K} \sum_{j=1}^{K} W_{ij}(\alpha) y_j \alpha^{\tau_{\mathcal{A}}} 1_{E_i}, \tag{16}$$

where, as before, 1_{E_i} is the indicator of the event $E_i = \{\omega : \tau_{\mathcal{A}} = \tau_{A_i}\}$.

Again, $W_{ij}(\alpha)$ is a deterministic quantity. Moreover, if we take $\delta_t(A, B)$ as defined by (14), then our now familiar calculations give us that

$$W_{ij}(\alpha) = \sum_{t=1}^{\min(m,l)} \delta_t(A_i, A_j)\, \alpha^{1-t}. \tag{17}$$

Now, if we choose real weights $\{y_j(\alpha) : 1 \le j \le K\}$ so that

$$\sum_{j=1}^{K} W_{ij}(\alpha) y_j(\alpha) = 1, \quad \text{for } 1 \le i \le K, \tag{18}$$

then our formula (16) for $X_{\tau_{\mathcal{A}}}$ simplifies to

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

and from $\mathbf{E} X_{\tau_{\mathcal{A}}} = 0$ one quickly gets the probability generating function for $\tau_{\mathcal{A}}$.

Theorem 2. *If the real values $\{y_j(\alpha) : 1 \le j \le K\}$ solve the linear system (18), then one has*

$$\mathbf{E}(\alpha^{\tau_{\mathcal{A}}}) = 1 - \frac{1}{1 + \sum_{j=1}^{K} y_j(\alpha)\alpha/(1-\alpha)}. \tag{19}$$

3.3. **Second moments and compound patterns.** One can use the probability generating function (19) to compute the second moment of $\tau_{\mathcal{A}}$, but this may not always be so pleasant. Alternatively, one can appeal more directly to the method of gambling teams. In this instance, there is even an instructive twist. If we use affinely sized bets as initial bets, then we put two weight vectors at our disposal.

More specifically, we ask the n'th gambler from the j'th team to place an initial bet of size $n y_j + z_j$ on the first letter of A_j. Team gambling now continues as before, and we just need to understand the amount of cash held by the teams at time $\tau_{\mathcal{A}}$, when we first observe a pattern from $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_K\}$.

After some calculation Pozdnyakov et al. (2005), showed that at time $\tau_{\mathcal{A}}$ the cash held by the j'th team in the i'th ending scenario is given by

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

where W_{ij} is again defined as in (15) and where now N_{ij} is given by

$$N_{ij} = \sum_{t=1}^{\min(m,l)} \delta_t(A_i, A_j)(1-t). \tag{20}$$

After another, more sustained, calculation one also finds that the casino's net gain $X_{\tau_{\mathcal{A}}}$ at time $\tau_{\mathcal{A}}$ is given by

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

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

$$\sum_{j=1}^{K} W_{ij} y_j = 1, \quad \text{and} \quad \sum_{j=1}^{K} (N_{ij} y_j + W_{ij} z_j) = 1, \tag{21}$$

for all $1 \le i \le K$, then we get a much tidier formula for $X_{\tau_{\mathcal{A}}}$. Specifically, we have

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

so the usual invocation of Doob's stopping theorem gives us a formula for the second moment in terms of the first.

Theorem 3. *If $\{y_j\}_{1\le j\le K}$ and $\{z_j\}_{1\le j\le K}$ solve the linear system (21), then*

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

Example 4. As before, we take $\Omega = \{1, 2\}$, but this time 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

$$W_{ij} = \begin{pmatrix} 6 & 2 \\ 2 & 10 \end{pmatrix} \text{ and } N_{ij} = \begin{pmatrix} -4 & 0 \\ 0 & -16 \end{pmatrix}.$$

Theorems 2 and 3 then give us

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

together with an explicit formula for formula for the probability generating function,

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

4. Occurrence of Patterns in Markov Dependent Trials

The method of gambling teams extends in a natural way to problems where the driving sequence $\{Z_n : n = 1, 2, \dots\}$ is a Markov chain. In fact *teams* are even more critical in the theory of waiting times for patterns in a Markov chain. In particular, one needs teams even for the analysis of the waiting time until a simple pattern.

4.1. **Two-state Markov chains and a single pattern.** Before we deal with more general Markov chains, it is useful to consider first the special case of two state chains. For specificity, we take the state space $\Omega = \{1, 2\}$, initial distribution $\mathbf{P}(Z_1 = 1) = p_1$, $\mathbf{P}(Z_1 = 2) = p_2$, and transition matrix $\{p_{ij}\}_{\Omega\times\Omega}$ where, as usual, $p_{ij} = \mathbf{P}(Z_{n+1} = j \,|\, Z_n = i)$.

Naturally we want our gambles on the Markov chain to be fair. Thus, a dollar bet on $j \in \Omega$ in the first round must pay $1/p_j$ to a winning gambler, and a dollar bet on $j \in \Omega$ on the $n + 1$'st round must pay a winner $1/p_{ij}$ if $Z_n = i$.

A pattern A is just a sequence $a_1a_2 \cdots a_m$ with $a_i \in \Omega$ for each $1 \le i \le m$, and there are *three scenarios* under which the pattern A can be observed. In the first scenario the pattern A is observed at the very beginning of the process $\{Z_n : n = 1, 2, \dots\}$. In the second scenario the pattern 1A first occurs at time τ_{A}, and in the third scenario the pattern 2A first occurs at time τ_{A}.

We now consider *two* gambling teams, one team of *straightforward gamblers* and one team of *smart gamblers*. Just before the n'th round a new gambler from each team enters the game. Both gamblers observe the n'th round but they only start betting on the $n+1$'th round. The straightforward gamblers and the smart gamblers use different strategies, though it turns out that neither team bets any money on the first round. For $n = 1, 2, \dots$ we have the following situation for the newly arriving gamblers:

- The straightforward gambler observes the n'th round, but he does not use this information. On the $n+1$'th round he bets y_1 dollars on the first letter of the sequence A, and he continues to bet his accumulated winnings on the successive letters of A until he either loses or until he observes the full pattern.
- The smart gambler observes the n'th round, and he uses this information. If $Z_n \neq a_1$ he bets y_2 dollars on the round $n + 1$ on the first letter of the pattern A, and he then continues to "let his fortune roll" until he either loses or until he observes A. On the other hand, if $Z_n = a_1$, then on round $n + 1$ the smart gambler bets y_2 dollars on a_2. He then continues to bet on the successive letters of the pattern $a_3 \cdots a_m$ until he loses or until the pattern A is observed.

Now, we let $W_{ij}y_j$, $i = 1, 2, 3$, $j = 1, 2$ be cash on hand of the j'th team wins when the game ends in the i'th scenario. Just as before, the quantities W_{ij} are deterministic, and they are easy to compute. For the casino's net gain $X_{\tau_{\mathrm{A}}}$ at time τ_{A} is given by

$$X_{\tau_{\mathrm{A}}} = (y_1 + y_2)(\tau_{\mathrm{A}} - 1) - \sum_{i=1}^{3}\sum_{j=1}^{2} W_{ij}y_j 1_{E_i}, \tag{22}$$

where 1_{E_i} is the indicator of i'th ending scenario. To see why we have the factor $(\tau_{\mathrm{A}} - 1)$ in (22) one should note that no money was bet on the first round and exactly $y_1 + y_2$ was bet by each of the first time bettors at each of the subsequent rounds.

If we now assume that we can choose (y_1, y_2) such that

$$\sum_{j=1}^{2} W_{ij} y_j = 1, \quad \text{for } i = 2 \text{ and } 3, \tag{23}$$

then (22) reduces to the much 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. If π_1 denotes the probability of the first scenario, Doob's stopping theorem gives us

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

so solving for the expected value we find

$$\mathbf{E}(\tau_{\mathrm{A}}) = 1 + \frac{\pi_1 (W_{11} y_1 + W_{12} y_2) + (1 - \pi_1)}{y_1 + y_2}. \tag{24}$$

Here one should further note that the computation of π_1 is always trivial.

Example 5. Consider the pattern A = 121. In this case, we have three possible scenarios to consider: (1) we observe 121 right at the beginning, i.e. we have $\tau_{\mathrm{A}} = 3$ or (2) the game stops with 2121 at the end of some random number of rounds, or (3) the game stops with 1121 at the end of some random number of rounds.

A straightforward calculation gives us our 3×2 (scenario by team) matrix $\{W_{ij}\}$,

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

The linear system (23) for initial bet sizes y_1 and y_2 is given by

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

from which we 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 obviously $p_1 p_{12} p_{21}$, so, after substitutions are made into (24), simplification gives us

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

4.2. **Two-state chains and compound patterns.** If $\mathcal{A}$ is a compound pattern with the alphabet $\{1,2\}$, then one can compute the expected value of $\tau_{\mathcal{A}}$ with the method of gambling teams, but some substantial modifications are required. In particular, one needs to split the ending scenarios into two classes that we call *initial-ending scenarios* and *later-ending scenarios*.

More precisely, if $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_K\}$ we need to consider a set of K different *initial-ending scenarios* where in the i'th initial-ending scenario the pattern A_i occurs at the beginning of the sequence $\{Z_n : n \geq 1\}$. We also need to consider a candidate set of $2K$ *later-ending scenarios*; in K of these there is an $1 \leq i \leq K$ such that pattern $1\mathrm{A}_i$ occurs at time $\tau_{\mathcal{A}}$ and in K more of these there is an $1 \leq i \leq K$ such that the pattern $2\mathrm{A}_i$ occurs at time $\tau_{\mathcal{A}}$.

The process that passes from $\mathcal{A}$ to the set of $2K$ later-ending scenarios is called the *doubling step*, and it is important to note that some of these potential ending scenarios may not really be possible. For example, if $\mathcal{A} = \{212, 22\}$, then the doubling step suggest the possibility of four later-ending scenarios which we can write as $*1212$, $*2212$, $*122$ and $*222$ where the wild-card symbol $*$ stands for an arbitrary (possibly empty) string of 1's and 2's. By the definition of $\tau_{\mathcal{A}}$ as the first occurrence time of a pattern from $\mathcal{A} = \{212, 22\}$, we see that neither of the strings 221 and 222 can occur as a substring of the string $Z_1, Z_2, \ldots, Z_{\tau_{\mathcal{A}}}$; thus, the only later-ending scenarios for $\mathcal{A} = \{212, 22\}$ that are actually observable are $*1212$ and $*122$. Similarly, one can check that if $\mathcal{A} = \{21, 111\}$, then the only observable later-ending scenarios are $*121$ and $*221$.

These observations permit us to do some useful cleaning of our initial list of $2K$ candidates for later-ending scenarios. If some candidate scenario cannot be observed in a sequence that ends at time $\tau_{\mathcal{A}}$, then we simply delete that scenario from the candidate list. The remaining list of later-ending scenarios is called the *final list*, and we denote its cardinality by of N. If $A_i \in \mathcal{A}$ then the final list could contain just one of the candidates $*1A_i$ and $*2A_i$ or it could contain both of these candidates. In the first case we say A_i has *type one* and in the second case we say A_i has *type two*. It is even possible that A_i is neither of type one nor type two, but this case will not figure into our calculations.

We now consider the instructions for a set of N gambling teams, one for each of the later-ending scenarios on the final list. If A_i is of type two, then we associate two gambling teams with A_i. One team bets on A_i in the straightforward way of subsection 4.1, and one team bets on A_i in the smart way of that subsection. Finally, if A_i is of type one, then we just associate a single gambling team with A_i. This team places then places its bets on A_i in the straightforward way.

We now order union of the set of K initial-ending scenarios and the final set of N later-ending scenarios. We also let the initial-ending scenarios lead this list, and for $i \in \{1, 2, \ldots, K+N\}$ we let E_i denote the event that at time $\tau_{\mathcal{A}}$ we are in the i'th ending scenario of this ordered list. Finally, for $j \in \{1, 2, \ldots, N\}$ we let y_j denote the initial bet size of the j'the team. These real values will be determined by the solution of a linear system.

For $i \in \{1, 2, \ldots, K+N\}$ we let $W_{ij} y_j$ be the cash in hand of the j'th gambling team at time $\tau_{\mathcal{A}}$ in the case of the i'th ending scenario. For $i \in \{1, 2, \ldots, K\}$ this is an initial-ending scenario, and for $i \in \{K+1, K+2, \ldots, K+N\}$ this is one of the later-ending scenarios on the final list.

If X_n denotes the casino's net gain at time n, then at time $\tau_{\mathcal{A}}$ the usual cash flow calculation tells us

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

where E_i is the event that the i'th scenario occurs. Again, the W_{ij} terms are not random, and we assume that there are real values $\{y_j\}_{1\le j\le N}$ that solve the $N \times N$ system

$$\sum_{j=1}^{N} W_{ij} y_j = 1, \quad \text{for all } i \in \{K+1, K+2, \dots, K+N\}. \tag{25}$$

For such initial bet sizes, we then have the representation

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

Next, for $1 \le i \le K$ we let $\pi_i = \mathbf{P}(E_i)$ denote the probability that the i'th initial-ending scenario occurs, so by Doob's stopping theorem we have

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

which one again solves for $\mathbf{E}(\tau_{\mathcal{A}})$.

Theorem 4. *If the real values $\{y_j\}_{1\le j\le N}$ solve the linear system (25), then*

$$\mathbf{E}(\tau_{\mathcal{A}}) = 1 + \frac{(1 - \sum_{i=1}^{K} \pi_i) + \sum_{i=1}^{K} \pi_i \sum_{j=1}^{N} y_j W_{ij}}{\sum_{j=1}^{N} y_j}. \tag{26}$$

Example 6. For the compound pattern $\mathcal{A} = \{11, 212\}$ 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 11 and 212, we then have a total of five ending scenarios that order and write simply as

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

The 5×3 scenario-by-team matrix $\{W_{ij}\}$ is then given by

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

When we solve the corresponding linear system (25), we find that the initial team bets are given by

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

Here $\pi_1 = p_1 p_{11}$ and $\pi_2 = p_2 p_{21} p_{12}$ are the respective probabilities that 11 and 212 are the initial segments of the process $\{Z_n : n \geq 1\}$, so the general formula (26) simply gives one

$$\mathbf{E}(\tau_{\mathcal{A}}) = 2 + p_1 p_{12} + \frac{1 - p_1 p_{11}}{p_{21}}.$$

Remark. We should note here that whenever one finds a martingale method that gives the expected value of a waiting time, it is reasonable to expect that the method can be extended to obtain formulas for higher moments or the generating function. We have already reviewed how such extensions can be done for the independent model, and Glaz et al. (2006) give a more detailed explanation of how for the two-state Markov chains one can also obtain the higher moments and generating function for the waiting time until one observes a compound pattern $\mathcal{A}$. In a similar vein, Gava and Salotti (2014) also show how one can use the method of gambling teams to calculate the scenario probabilities $\mathbf{P}(\tau_{\mathcal{A}} = \tau_{\mathrm{A}_i})$.

4.3. **General finite state Markov chains.** Now we consider a homogeneous Markov chain $\{Z_n : n \geq 1\}$ with a finite state space $S = \{1, 2, \ldots, M\}$. We take the initial distribution to be $\mathbf{P}(Z_1 = m) = p_m$, $1 \leq m \leq M$, and we take the transition matrix to be $P = \{p_{ij}\}_{1 \leq i,j \leq M}$, where, as always, $p_{ij} = \mathbf{P}(Z_{n+1} = j | Z_n = i)$.

In this situation, the analysis of the waiting time $\tau_{\mathcal{A}}$ until the occurrence of compound pattern $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_K\}$ requires us to places some restrictions on $\mathcal{A}$ that were not needed either for independent sequences or for two-state chains. Specifically, we confine our attention to compound patterns that satisfy three assumptions:

- We assume that no pattern $B \in \mathcal{A}$ contains any $A \in \mathcal{A}$ as a subpattern. As we noted before, we can make this assumption without any loss of generality since by the reasoning of (10) one can drop B from the compound pattern $\mathcal{A}$ and not change the distribution of $\tau_{\mathcal{A}}$.
- We assume that $\mathbf{P}(\tau_{\mathcal{A}} = \tau_{\mathrm{A}_i}) > 0$ for all $1 \leq i \leq K$. If to the contrary we were to have $\mathbf{P}(\tau = \tau_{\mathrm{A}_i}) = 0$ for some i, then we could just drop A_i from $\mathcal{A}$ without changing the distribution of $\tau_{\mathcal{A}}$. Here we should note that for independent sequences one can never have $\mathbf{P}(\tau = \tau_{\mathrm{A}_i}) = 0$, but for Markov sequences one must explicitly exclude this possibility. For example, if A_i contains the subpattern km and one has $p_{km} = 0$, then A_i can not occur in $\{Z_n : n \geq 1\}$.
- We assume that $\mathbf{P}(\tau_{\mathcal{A}} < \infty) = 1$. For example, if each simple patterns of $\mathcal{A}$ contains a transient state one can have $\mathbf{P}(\tau_{\mathcal{A}} = \infty) > 0$ even for a finite Markov chain, and we must exclude this possibility. Here one should also recall that for a finite Markov chain the condition of pointwise finiteness $\mathbf{P}(\tau_{\mathcal{A}} < \infty) = 1$ already implies the formally stronger condition $\mathbf{E}[\tau_{\mathcal{A}}] < \infty$.

For chains with more than two states, one also needs to take a more refined view of the ending scenarios for the occurrence of a simple pattern. Specifically, one decomposes the occurrence of the simple pattern A_i into an initial list of $1 + M + M^2$ *feasible ending scenarios*:

- either the sequence A_i occurs as an initial segment of $\{Z_n : n \geq 1\}$, or
- for some state $k \in S$, the pattern kA_i occurs as an initial segment of the sequence $\{Z_n : n \geq 1\}$, or

- for some ordered pair of states (k, m), $k \in S, m \in S$, the pattern kmA_i occurs after some indeterminant number of rounds.

The first $1 + M$ ending scenarios are called *initial-ending* scenarios. The last M^2 scenarios are called *later-ending* scenarios. Since $\mathcal{A} = \{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_K\}$ contains K simple patterns, we have an *a priori* candidate list of $(1+M+M^2)K$ feasible ending scenarios — $(1+M)K$ of the *initial-ending* kind and M^2K of the *later-ending* kind.

As before, some of these ending scenarios simply cannot occur in the sequence $Z_1, Z_2, \ldots, Z_{\tau_\mathrm{A}}$. Some are impossible because of the definition of τ_A and some are impossible because of the structure of the transition matrix $\{p_{ij}\}$. We need to clean up this candidate list.

Let $\mathcal{I}$ be the set of feasible initial-ending scenarios that can actually occur, and let $\mathcal{N}$ be the set of feasible later-ending scenarios that can actually occur. We then take I and N to be the respective cardinalities. Next, we form an ordered list of the $I + N$ elements of $\mathcal{I} \cup \mathcal{N}$. For specificity, we assume that the elements of $\mathcal{I}$ lead this list of observable ending scenarios.

Now, for each of the N elements in the sublist $\mathcal{N}$ we now introduce a gambling team with team members who gamble in a way that is reminiscent of the smart gamblers of subsection 4.1. More specifically, for each simple pattern $km\mathrm{A}_i \in \mathcal{N}$ we introduce a team of gamblers that we call the $km\mathrm{A}_i$*-gambling team*. For each such team we also introduce an initial stake which we call the *team stake*.

The $n+1$'st gambler from the $km\mathrm{A}_i$-gambling team arrives before round $n+1$, and he observes the value Z_n of the n'th round. If $Z_n = k$ he bets an amount equal to the team stake on the pattern $m\mathrm{A}_i$. On the other hand, if $Z_n \neq k$ he bets the team stake on A_i. Here by "betting the team stake on the pattern $\mathrm{A} = a_1 a_2 \cdots a_m$, when $Z_n = a_0$" we mean that one follows two rules:

- After observing Z_n, the gambler bets the team stake that the next trial yields a_1. If $Z_{n+1} \neq a_1$, he loses his money and leaves the game. If $Z_{n+1} = a_1$, he wins $1/p_{a_0 a_1}$ times his bet since the odds are fair. If he wins he continues his betting.
- He now 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 total capital on the successive letters of the pattern A until either the pattern A is observed, or until he loses an amount equal to his team stake.

Having fixed the ordering of the set $\mathcal{N}$ of later-ending scenarios that can occur, we let y_j denote the team stake size. Here again y_j is a real number and if y_j is negative then ones winnings are given the natural interpretation. The values $\{y_j\}$ will be determined later by the solution of a linear system. Now for $1 \leq i \leq I+N$ and $1 \leq j \leq N$ we let $y_j W_{ij}$ be the amount of money that the j'th team wins in the i'th ending scenario. As before, the values W_{ij} are non-random. We also take X_n to be the casino's net gain at time n, By the usual bookkeeping, one finds that at time $\tau_\mathcal{A}$ we have

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

where E_i is the event that the i'th scenario occurs.

If the real values $\{y_j\}_{1\le j\le N}$ satisfy the system

$$\sum_{j=1}^{N} W_{ij} y_j = 1, \text{ for all } I+1 \le i \le I+N, \tag{27}$$

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

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

Again, $\{X_n\}_{n\ge1}$ has bounded increments and $\mathbf{E}[\tau_{\mathcal{A}}] < \infty$, so if we write π_i for the probability $\mathbf{P}(E_i)$ that the i'th initial-ending scenario occurs, then Doob's stopping theorem gives us

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

One then solves for $\mathbf{E}(\tau_{\mathcal{A}})$ to obtain the main result of this section.

Theorem 5. *If the real values $\{y_j\}_{1\le j\le N}$ solve the linear system (27), then*

$$\mathbf{E}(\tau_{\mathcal{A}}) = 1 + \frac{(1 - \sum_{i=1}^{I} \pi_i) + \sum_{i=1}^{I} \pi_i \sum_{j=1}^{N} y_j W_{ij}}{\sum_{j=1}^{N} y_j}. \tag{28}$$

Example 7. Take $S = \{1, 2, 3\}$ 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 the scenarios that cannot occur, our set $\mathcal{I}$ observable initial-ending scenarios is given by

$$\mathcal{I} = \{323, 313, 33, 1323, 2323, 1313, 2313, 133, 233\}.$$

Also, because the transitions $1 \to 2$ and $2 \to 1$ are impossible, only six of the later-ending scenarios are observable. Again using the symbol $*$ for a wild-card string (which may be the empty string) we order the observable later-ending scenarios as

$$\mathcal{N} = \{*11323, *22323, *11313, *22313, *1133, *2233\}.$$

We now need the entries of the matrix $\{W_{ij}\}$, and for this example it is easy to compute these directly. For instance, the 11323-gambling team in the initial-ending scenario 323 would win $1/p_{23} = 4$. On the other hand, in the later-ending scenario $*11323$ this team would win

$$1/(p_{11}p_{13}p_{32}p_{23}) + 1/p_{23} = 268/3,$$

and in the later-ending scenario $*22323$ it would win $1/(p_{23}p_{32}p_{23}) + 1/p_{23} = 68$. One can continue with such direct computations, or one can appeal to the general formula for W_{ij} given by Pozdnyakov (2008). In either case, one finds that the payoffs for all the later-ending scenarios — the ones that are needed for linear system (27) — are given by the matrix

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

When we use these values in the general formula (28) we find at last that

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

Remarks. In parallel with our earlier examples, one can use the initial bets sizes $ny_j + z_j$ to get the second moment of $\tau_{\mathcal{A}}$, or one can use the initial bets sizes $y_j\alpha^n$ to get the corresponding generating function of $\tau_{\mathcal{A}}$, cf. Pozdnyakov (2008). Moreover, Fisher and Cui (2010) combined the martingale method with the occupation measure method of Benevento (1984) to get corresponding formulas for higher-order Markov chains.

One should note that the method of this subsection is also applicable to two-state Markov chains, but it would inefficient compared to the method of subsection 4.2. If one applies the method of this subsection to a two-state Markov chains, then one needs $4K$ ending scenarios but the method of subsection 4.2 only needs $2K$ ending scenarios.

There are also computational differences between the martingale method and the Markov chain embedding method. To compute $\mathbf{E}(\tau_{\mathcal{A}})$ by the Markov chain embedding of Fu and Chang (2002, p. 73) one needs to solve a linear system that depends on the cardinality K of the compound pattern $\mathcal{A}$ and lengths of the single patterns in $\mathcal{A}$. Here one solves a system that just depends on K and cardinality M of the alphabet, and in some situations this is a much smaller system. For a simple but important example, one can take $\mathcal{A} = \{A_1\}$ where the pattern A_1 is very long.

5. Applications to Scans

5.1. **Second moments and distribution approximations.** Martingales give us highly effective methods for the computation of the moments of the waiting time $\tau_{\mathcal{A}}$, and it is natural to ask if these computations might also lead to effective *approximations* of the distribution of $\tau_{\mathcal{A}}$. This is true, but the path to good approximations is not as direct as one might expect.

Since the clumping heuristic (see Aldous (1989)) typically applies to the stopping time $\tau_{\mathcal{A}}$ associated with a scan statistic, one certainly expects the tail probabilities $\mathbf{P}(\tau_{\mathcal{A}} > n)$ to be close to those of the exponential distribution. Still, the exponential approximation faces natural competition from several other families of distributions, including the gamma, the Weibull, and even the shifted exponentials. The main finding of Pozdnyakov et al. (2005) is that in a wide range of situations the shifted exponential family provides the most appropriate approximation to the distribution of $\tau_{\mathcal{A}}$.

To make this assertion explicit, we first recall that X' has the shifted exponential distribution if one can write $X' = X + c$ where X has an exponential distribution and c is a constant. To approximate $\tau_{\mathcal{A}}$, we choose c and X so that two moments match:

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

If we set $\mu = \mathbf{E}(\tau_{\mathcal{A}})$ and $\sigma^2 = \mathbf{Var}(\tau_{\mathcal{A}})$, we obtain the shifted exponential approximation for the tail probabilities of $\tau_{\mathcal{A}}$:

$$\mathbf{P}(\tau_{\mathcal{A}} > n) \approx \exp(-(n + 1/2 + \sigma - \mu))/\sigma), \tag{29}$$

where the 1/2 term may be viewed as a kind of "continuity correction."

This approximation seems to work remarkably well, and Fu and Lou (2006, p. 307) suggest one explanation for its efficacy. There is also further discussion in Pozdnyakov and Steele (2009, p. 311), but neither of these explanations captures the full force of the numerical examples.

Example 8. (*Fixed Window Scans*). Here we take $\{Z_n : n \geq 1\}$ to be a sequence of Bernoulli trials, and we consider two kinds of scans: Table 1 reports on the at-least-3-out-of-10 scan and Table 2 reports on the at-least-4-out-of-20 scan.

For the fixed window scan statistics, Glaz and Naus (1991) developed tight lower and upper bounds which are provided in Tables 1 and 2 along with the approximations based on the exponential, shifted exponential, and gamma distributions. From these tables one sees that 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 substantially from μ, then the approximations for the distribution that are based on the exponential and gamma distributions do not perform nearly as well as those based on the shifted exponential approximation. In these tables (and the ones that follow) we omit the approximations based on the Weibull distribution because these approximations are so much worse than those given by the other methods.

Example 9. (*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. Here in order for τ and the waiting time $\tau_{\mathcal{A}}$ to have the same distribution we need a compound pattern $\mathcal{A}$ that contains 224 simple patterns.

The corresponding approximations are summarized in Table 3. Here analytical bounds are not available, so the approximations are judged by comparison with estimated probabilities based on $100,000$ replications. Again, we find superior performance of the approximation that is determined by a shifted exponential and calibration by two moments.

Example 10. (*Double Scans*). Here we take $\{Z_n : n \geq 1\}$ to be an i.i.d. sequence of random variables with the three-valued distribution given by

$$\mathbf{P}(Z_n = 1) = .04, \quad \mathbf{P}(Z_n = 2) = .01, \quad \text{and} \quad \mathbf{P}(Z_n = 0) = 0.95.$$

Table 1. Approximate values of $\mathbf{P}(\tau_{\mathcal{A}} \leq n)$ for the fixed window scan for least 3 failures out of 10 consecutive trials. Here we have $\mathbf{P}(Z_n = \text{Failure}) = .01$, $\mu = 30822$, and $\sigma = 30815$.

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 2. Approximate values of $\mathbf{P}(\tau_{\mathcal{A}} \leq n)$ for the fixed window scan for at least 4 failures out of 20 consecutive trials. Here we have $\mathbf{P}(Z_n = \text{Failure}) = .05$, $\mu = 481.59$, and $\sigma = 469.35$.

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

We then consider two types of "failures": a *type I* failure corresponds to observing a 1 and a *type II* failure corresponds to observing a 2. Next, we take a scan window with length 10. Finally, we let τ denote the first time until we either observe at least 2 failures of type II within the window, or we observe a total of least 3 failures of any combination either of the two types of failures within the window. Table 4 shows that the shifted exponential approximation again works well even in the challenging case when μ and σ are both relatively small and substantially different.

6. Summary and Concluding Observations

The duality relation $\mathbf{P}(S_{w,T} \geq k) = \mathbf{P}(\tau_{k,w} \leq T)$ creates a fundamental link between the scan statistic $S_{w,T}$ and the pattern-based stopping time $\tau_{k,w}$. By Doob's theorem, this stopping time, and more general stopping times, are tightly bound with the theory of martingales. Still, to exploit these connections, one needs a rich class of pattern-based martingales. Over time it has been found that the search for such martingales is powerfully served by the metaphor of gambling teams.

TABLE 3. Approximate values of $\mathbf{P}(\tau_{\mathcal{A}} \leq n)$ for the variable window scan for at least 2 failures out of 10 trials or at least 3 failures out of 50 trials. Here $\mathbf{P}(Z_n = \text{Failure}) = .01$, $\mu = 795.33$, and $\sigma = 785.85$.

n	exponential	shifted exponential	gamma	simulated 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 4. Approximate values of $\mathbf{P}(\tau_{\mathcal{A}} \leq n)$ for "double" scans for at least 2 type II failures out of 10 trials or at least 3 failures of any kind out of 10 trials. Here $\mathbf{P}(Z_n = 1) = .04$, $\mathbf{P}(Z_n = 2) = .01$, $\mu = 324.09$, and $\sigma = 318.34$.

n	exponential	shifted exponential	gamma	simulated 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

When one considers the waiting time τ_A until occurrence of a single pattern A, there is a striking observation that contributes in an essential way to the solution of the problem. Namely, we observe that at time τ_A the cash on hand of the gambling team is a simple *deterministic* function of just the pattern A and the probability distribution of the independent observations $\{Z_n : n = 1, 2, \ldots\}$. Much of the subsequent theory is then guided by the desire to find more refined refections of this *invariance* property.

When we looked at the time $\tau_{\mathcal{A}}$ until the first occurrence of a pattern from a list $\mathcal{A} = \{A_1, A_2, \ldots, A_K\}$ we needed to consider K gambling teams. Moreover, we distinguished these teams by having the members of the j'th teams begin betting with a stake of size y_j with $-\infty < y_j < \infty$. In this case, the invariance property was reflected by the fact that if we define $y_j W_{ij}$ to be the cash on hand of the j'th team "in the i'th winning scenario," then W_{ij} is deterministic. Some craft was needed to define these ending scenarios, and, in more complicated problems, the need can become substantial.

Finally, once methods for computing $\mathbf{E}(\tau_{\mathcal{A}})$ and $\mathbf{Var}(\tau_{\mathcal{A}})$ are in hand for both independent and Markov sequences, they help materially with the approximation of the distribution of $\tau_{\mathcal{A}}$. In particular, moment matching within the family of shifted exponential distributions turns out to give surprisingly good approximations to the distribution of $\tau_{\mathcal{A}}$, though here much of the evidence is numerical.

References

[1] Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*, Springer Publishing, New York.

[2] Benevento, R.V. (1984). The occurrence of sequence patterns in ergodic Markov chains, *Stochastic Processes and their Applications*, **17**, 369-373.

[3] Fisher, E., and Cui, S. (2010) Patterns generated by m-order Markov chains, *Statistics & Probability Letters*, **80**, 1157-1166.

[4] 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.

[5] 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.

[6] Gava, R.J. and Salotti, D. (2014). Stopping probabilities for patterns in Markov chains, *Journal of Applied Probability*, **51**, 287-292.

[7] 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.

[8] Graham, R.L. Knuth, D.E., and Patashnik, O. (1994). *Concrete mathematics. A foundation for computer science.* Second edition. Addison-Wesley Publishing Company, Reading, MA.

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

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

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

[13] Guibas, L.J. and Odlyzko, A.M. (1981). String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory. Series A* , **30**, 183-208.

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

[15] 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.

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

[17] 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.

[18] Pozdnyakov, V. (2008). On Occurrence of Patterns in Markov Chains: Method of Gambling Teams, *Statistics and Probability Letters*, **78**, 2762-2767.

[19] 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.

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

[21] Pozdnyakov V. and Steele J.M., (2009). Martingale Methods for Patterns and Scan Statistics, *Scan Statistics: Methods and Applications, J. Glaz, V. Pozdnyakov, and S. Wallenstein, eds.*, Birkhauser, Boston, 2009, 289-318.

[22] Shiryaev, A.N. (1995). *Probability*, Springer, New York.

[23] Solov'ev, A.D. (1966). A combinatorial identity and its application to the problem on the first occurrence of a rare event, *Teorija Verojatnostei i ee Primenenija*, **11**, 313-320.

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

[25] Zajkowski, K. (2014) A note on the gambling team method, *Statistics & Probability Letters*, **85**, 45-50.