SCAN STATISTICS: PATTERN RELATIONS AND 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 Bernoulli random variables \( \{Z_n : 1 \leq i \leq T\} \). Now given a window size \( w \) with \( 1 \leq w \leq T \) and a set \( \{i : 1 \leq i \leq T - w + 1\} \) 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 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

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

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).
A Dual Interpretation and a Related Stopping Time

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 at least \( k \) or more occurrences of the value 1 in a window of length \( w \), then one has the identity

\[
\Pr(S_{w,T} \geq k) = \Pr(\tau_{k,w} \leq T).
\]

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

Organization of the Survey

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

\[
p_1 = \Pr(Z_n = 1) > 0, \quad p_2 = \Pr(Z_n = 2) > 0, \ldots, \quad p_M = \Pr(Z_n = M) > 0.
\]

By a pattern (or more explicitly, a simple pattern) we mean a finite ordered sequence \( A = a_1a_2\cdots a_m \) of letters from the alphabet \( \Omega \).

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 \( \tau_A \) 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 \( A = a_1 a_2 \cdots a_m \). 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 \( 1/P(Z_n = a_1) \) 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 \( \tau_A \) when the pattern \( A = a_1 a_2 \cdots a_m \) is first observed, the team decides to do a little bookkeeping. First, one should note that at time \( \tau_A \) one has seen exactly \( \tau_A \) gamblers have come into the casino. Moreover, each gambler has placed one or more bets. At time \( \tau_A \) the lucky gambler who came in at time \( \tau_A - m + 1 \) 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 \( \tau_A \).

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 \( \tau_A = 20 \). 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 \( \sigma \)-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 \( X_{\tau_A} \). Since the stopping time \( \tau_A \) 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)), then tells us that

\[
\mathbb{E}(X_{\tau_A}) = 0.
\]

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

\[
X_{\tau_A} = \tau_A - W,
\]

where \( W \) is the amount of money in the pocket of the gambling team at time \( \tau_A \).

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 $\tau_A$ when the pattern $A$ is first observed, the gambler had to be gambling at time $\tau_A$ and he had to win at time $\tau_A$. In particular the gamblers who entered the game before $\tau_A - m + 1$ must have all lost their stake. At time $\tau_A$ the lucky gambler who entered at time $\tau_A - m + 1$ will have won all of his $m$ bets. Among the gamblers who arrived after time $\tau_A - m + 1$ 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 some notation; specifically, for $0 \leq i, j \leq m$ we set

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

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

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

From the earlier observation that $E(X_{\tau_A}) = 0$, we therefore find

$$E(\tau_A) = \delta_{11} \delta_{22} \cdots \delta_{mm} + \delta_{21} \delta_{32} \cdots \delta_{m,m-1} + \cdots + \delta_{m1}.$$  

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

$$E(\tau_A) = (p_1 \times p_1 \times p_2 \times p_1)^{-1} + (p_1)^{-1},$$

and, somewhat paradoxically, it also gives us

$$E(\tau_B) = (p_1 \times p_1 \times p_1 \times p_2)^{-1}.$$  

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

$$A = HHTHH, \quad B = HTHHT, \quad C = THHTH,$$

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 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 more precisely that

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

and

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

but also

$$P(C \text{ occurs before } A) = 0.62 \cdots > 0.50.$$
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 \( \tau_A \). 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 \( \alpha^n \) 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 \( \{X_n : n = 1, 2, \ldots \} \) 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 \( X_{\tau_A} \) can be written as the paid in amount minus the team’s current cash holdings,

\[
X_{\tau_A} = \alpha^1 + \alpha^2 + \cdots + \alpha^{\tau_A} - \alpha^{\tau_A} W(\alpha)
\]

\[
= \frac{\alpha^{\tau_A} - 1}{\alpha - 1} - \alpha^{\tau_A} W(\alpha)
\]

\[
= \alpha^{\tau_A} \left( \frac{\alpha}{\alpha - 1} - W(\alpha) \right) - \frac{\alpha}{\alpha - 1}.
\]

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

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

Here the factors \( \delta_{ij}, 1 \leq i, j \leq m \), 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 then find

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

from which we get the probability generating function,

\[
E(\alpha^{\tau_A}) = \left( 1 + \frac{1 - \alpha}{\alpha} W(\alpha) \right)^{-1}.
\]

As before, the usefulness of this formula depends on our ability to evaluate the 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

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

By expansion one finds

\[
E(\alpha^{\tau_A}) = p_1^3 p_2 \alpha^4 + p_1^3 p_2 \alpha^5 + p_1^3 p_1 \alpha^6 + p_1^3 p_2 (1 - p_1^2 p_2) \alpha^7 + \cdots.
\]
The first term of this expansion is obviously correct, but, for a more serious check of our formula for $E(\alpha \tau_A)$, one can use it to compute the first moment
\[
\frac{\partial E(\alpha \tau_A)}{\partial \alpha} \bigg|_{\alpha = 1} = \frac{1}{p_1^2 p_2} + \frac{1}{p_1},
\]
and this does recover what we found earlier for $E(\tau_A)$.

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

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 $\{X_n : n = 1, 2, \ldots\}$ is again a martingale. Moreover, at the stopping time $\tau_A$ we have
\[
X_{\tau_A} = 1 + 2 + \cdots + \tau_A - (\tau_A - m + 1)\delta_{11}\delta_{22} \cdots \delta_{mm} - (\tau_A - m + 2)\delta_{21}\delta_{32} \cdots \delta_{m,m-1} - \cdots - (\tau_A - m + m)\delta_{m1} = 1 + 2 + \cdots + \tau_A - \tau_A W - N,
\]
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 $\{X_n : n = 1, 2, \ldots\}$ 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 $P(\tau_A > n)$ decay at an exponential rate, so the stopping time theorem of Shiryaev (1995, p. 485) gives us
\[
0 = E(X_{\tau_A}) = E\left(\frac{1}{2} \tau_A (\tau_A - 1)\right) - W E(\tau_A) - N.
\]
Now, when we solve for $E(\tau_A^2)$ and use our earlier formula $E(\tau_A) = W$, we get
\[
E(\tau_A^2) = (2W - 1)E(\tau_A) + 2N = 2W^2 - W + 2N,
\]
and, as an immediate corollary, we also have
\[
\text{Var}(\tau_A) = W^2 - W + 2N.
\]
At this point the formula for $\text{Var}(\tau_A)$ 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.

\[
\text{Var}(\tau_A) = \left( \frac{1}{p_1} + \frac{1}{p_1^2 p_2^2} \right)^2 - \frac{1}{p_1} - \frac{7}{p_1^2 p_2^2}.
\]

**Remarks.** First, one should note that when either \( p_1 \to 0 \) or \( p_2 \to 0 \) one has the limit relation

\[
\frac{\text{E}(\tau_A)}{\text{Var}(\tau_A)^{1/2}} \to 1.
\]

In fact, this relation could have been anticipated without explicit knowledge of the formulas for \( \text{E}(\tau_A) \) and \( \text{Var}(\tau_A) \). 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 \( \tau_A \) should be approximately exponential and the ratio \( \text{E}(\tau_A)/\text{Var}(\tau_A)^{1/2} \) should be approximately one.

Second, one should also note that method used to calculate \( \text{E}(\tau^2_A) \) can be modified in a natural way to calculate \( \text{E}(\tau^3_A) \). 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 \( \text{E}(\tau_A) \) and \( \text{E}(\tau^2_A) \).

### 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 \( \tau_A \) until one observes a pattern from a specified finite set \( \mathcal{A} = \{A_1, A_2, \ldots, A_K\} \) of patterns. If we write \( \tau_{A_i} \) for the first time that one observes the simple pattern \( A_i \in \mathcal{A} \), then we have

\[
\tau_A = \min\{\tau_{A_1}, \ldots, \tau_{A_K}\},
\]

and \( \tau_A \) 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 \( A = a_1 a_2 \cdots a_j \) is a subpattern of \( B \) if the letters \( a_1 a_2 \cdots a_j \) appear consecutively in the pattern \( B \). What makes this notion particularly useful here is that if \( A \) is a subpattern of \( B \) then

\[
\min(\tau_A, \tau_B) = \tau_A.
\]

Thus, in the specification of the waiting time \( \tau_A \) 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 $\tau_A$ for compound pattern $A = \{A_1, A_2, \cdots, A_K\}$ we use $K$ different gambling teams, and, for each $1 \leq j \leq K$, 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 $\tau_A$ the gambling is stopped, and we calculate the casino’s net winnings $X_{\tau_A}$. If $A_i$ is the first pattern in $A$ that is observed, then we have $\tau_A = \tau_{A_i}$. 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 $\{X_n : n = 1, 2, \cdots\}$ and at the stopping time $\tau_A$, the casino’s net win is $X_{\tau_A}$. Moreover, by our usual bookkeeping of paid in capital minus cash on hand, we have

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

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 $y_j$, $1 \leq j \leq K$ in 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 $\{y_j : 1 \leq j \leq K\}$ that solve the linear system

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

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,

$$X_{\tau_A} = K \sum_{j=1}^{K} y_j \tau_A - 1.$$

Doob’s stopping theorem gives us $E X_{\tau_A} = 0$, so, in the end, have a formula for $E(\tau_A)$ that is simple and computationally effective.

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

$$E(\tau_A) = \frac{1}{\sum_{j=1}^{K} y_j}.$$  

**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 $A$ is a subpattern of another pattern in $A$. Moreover, as noted earlier in this section, one can always assume that $A$ satisfies this condition. The solvability of (12) and related systems has also been further investigated by Zajkowski (2014).
More Explicit Calculations

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

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

we set

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

In terms of this shorthand, one then finds without difficulty that

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

3.2. The generating function and compound patterns. The method of gambling teams can also be used to get the probability generating function $E(\alpha^{\tau_A})$ for compound pattern $A = \{A_1, A_2, \ldots, A_K\}$. 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 $\{y_j : 1 \leq j \leq K\}$ will be determined later.

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

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

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

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

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

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

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

then our formula (16) for $X_{\tau_A}$ simplifies to

$$X_{\tau_A} = \alpha^{\tau_A} - 1 \sum_{j=1}^{K} y_j(\alpha) - \alpha^{\tau_A},$$

and from $E X_{\tau_A} = 0$ one quickly gets the probability generating function for $\tau_A$. 

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

\[
\mathbb{E}(\alpha^{\tau_A}) = 1 - \frac{1}{1 + \sum_{j=1}^{K} y_j(\alpha)\alpha/(1 - \alpha)}.
\]

3.3. Second moments and compound patterns. One can use the probability generating function (19) to compute the second moment of \( \tau_A \), but this may not always be so pleasant. Alternatively, one can appeal more directly the method of gambling teams. In this instance, there is even an instructive twist. If we use affinely sized bets 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 \( ny_j + z_j \) on the first letter of \( \Lambda_j \). Team gambling now continues as before, and we just need to understand the amount of cash held by the teams at time \( \tau_A \) when we first observe a pattern from \( \Lambda = \{\Lambda_1, \Lambda_2, \ldots, \Lambda_K\} \).

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

\[
W_{ij}(\tau_A y_j + z_j) + N_{ij} y_j
\]

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

\[
N_{ij} = \min(m, l) \sum_{t=1}^{K} \delta_t(\Lambda_i, \Lambda_j)(1 - t).
\]

After another, more sustained, calculation one also finds that the casino’s net gain \( X_{\tau_A} \) at time \( \tau_A \) is given by

\[
\sum_{j=1}^{K} y_j \frac{\tau_A(\tau_A + 1)}{2} + \sum_{j=1}^{K} z_j \tau_A - \sum_{i=1}^{K} \left( \sum_{j=1}^{K} W_{ij} y_j \tau_A + \sum_{j=1}^{K} N_{ij} y_j + \sum_{j=1}^{K} W_{ij} z_j \right) 1_{E_i}.
\]

Now, if we now choose real weights \( \{y_j\}_{1 \leq j \leq K} \) and \( \{z_j\}_{1 \leq j \leq K} \) such that

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

for all \( 1 \leq i \leq K \), then we get a much tidier formula for \( X_{\tau_A} \). Specifically we have

\[
X_{\tau_A} = \sum_{j=1}^{K} y_j \frac{\tau_A(\tau_A + 1)}{2} + \sum_{j=1}^{K} z_j \tau_A - \tau_A - 1,
\]

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 \( \{y_j\}_{1 \leq j \leq K} \) and \( \{z_j\}_{1 \leq j \leq K} \) solve the linear system (21), then

\[
\mathbb{E}(\tau_A^2) = \frac{1 + (1 - \sum_{j=1}^{K} z_j - \sum_{j=1}^{K} y_j/2)\mathbb{E}(\tau_A)}{\sum_{j=1}^{K} y_j/2}.
\]

Example 4. As usual we take \( \Omega = \{1, 2\} \), but this time we consider the compound pattern \( \Lambda = \{11, 121\} \). If we further assume that

\[
P(Z_1 = 1) = P(Z_1 = 2) = 1/2,
\]

then we find
\[ W_{ij} = \begin{pmatrix} 6 & 2 \\ 2 & 10 \end{pmatrix} \quad \text{and} \quad N_{ij} = \begin{pmatrix} -4 & 0 \\ 0 & -16 \end{pmatrix}. \]

Theorems 2 and 3 then give us
\[ \mathbb{E}(\tau_A) = \frac{8}{3}, \quad \text{and} \quad \text{Var}(\tau_A) = 10, \]

together with an explicit formula for the probability generating function,
\[ \mathbb{E}(\alpha^{\tau_A}) = \frac{\alpha^2(\alpha + 2)}{8 - 4\alpha - \alpha^3} = \frac{\alpha^2}{4} + \frac{\alpha^3}{8} + \frac{3\alpha^5}{32} + \frac{5\alpha^6}{64} + \frac{7\alpha^7}{128} + \cdots. \]

4. Occurrence of Patterns in Markov Dependent Trials

The method of gambling teams extends in a natural way to problems where the driving sequence \( \{Z_n : n = 1, 2, \ldots\} \) 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, but it is useful to first consider the special case of two state chains. For specificity, we take the state space \( \Omega = \{1, 2\} \), initial distribution \( P(Z_1 = 1) = p_1, \quad P(Z_1 = 2) = p_2 \), and transition matrix \( \{p_{ij}\}_{\Omega \times \Omega} \) where, as usual, \( p_{ij} = P(Z_{n+1} = j \mid Z_n = i) \).

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 \( a_1 a_2 \cdots a_m \) with \( a_i \in \Omega \) for each \( 1 \leq i \leq m \), 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 \( \{Z_n : n = 1, 2, \ldots\} \). In the second scenario the pattern \( 1A \) first occurs at time \( \tau_A \), and in the third scenario the pattern \( 2A \) first occurs at time \( \tau_A \).

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 \( n = 1, 2, \ldots \) 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 $X_{\tau_A}$ at time
$\tau_A$ is given by

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

where $1_{E_i}$ is the indicator of $i$'th ending scenario. To see why we have the factor
$(\tau_A - 1)$ 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

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

then (22) reduces to the much simpler formula,

$$X_{\tau_A} = (y_1 + y_2)(\tau_A - 1) - (W_{11}y_1 + W_{12}y_2)1_{E_1} - 1_{E_{\bar{1}}},$$

where $1_{E_{\bar{1}}}$ is the indicator of the complement of the 1st ending scenario. If $\pi_1$
denotes the probability of the first scenario, Doob’s stopping theorem gives us

$$0 = (y_1 + y_2)(E(\tau_A) - 1) - \pi_1(W_{11}y_1 + W_{12}y_2) - (1 - \pi_1),$$

so solving for the expected value we find

$$E(\tau_A) = 1 + \frac{\pi_1(W_{11}y_1 + W_{12}y_2) + (1 - \pi_1)}{y_1 + y_2}.$$}

Here one should further note that the computation of $\pi_1$ is always trivial.

**Example 5.** Consider the pattern $A = 121$. In this case, we have the three possible
scenarios to consider: (1) we observe 121 right at the beginning, i.e. we have $\tau_A = 3$
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 then gives us our $3 \times 2$ (scenario by team) matrix
$\{W_{ij}\}$,

$$
\begin{pmatrix}
\frac{1}{p_2} + \frac{1}{p_21} & \frac{1}{p_212} + \frac{1}{p_21} \\
\frac{1}{p_112} + \frac{1}{p_21} & \frac{1}{p_122} + \frac{1}{p_21} + \frac{1}{p_21} \\
\frac{1}{p_112} + \frac{1}{p_21} & \frac{1}{p_112} + \frac{1}{p_21}
\end{pmatrix}.
$$

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

$$y_1\left(\frac{1}{p_212} + \frac{1}{p_21}\right) + y_2\left(\frac{1}{p_122} + \frac{1}{p_21} + \frac{1}{p_21}\right) = 1,$$

$$y_1\left(\frac{1}{p_112} + \frac{1}{p_21}\right) + y_2\left(\frac{1}{p_122} + \frac{1}{p_21}\right) = 1,$$
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 \( \tau_1 \) of the first scenario is obviously \( p_1p_{12}p_{21} \), so, after substitutions are made into (24), simplification gives us
\[ E(\tau_A) = 1 + \frac{p_2}{p_{21}} + \frac{1}{p_{21}} + \frac{1}{p_{12}p_{21}}. \]

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 \( \tau_A \) 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 \( \mathcal{A} = \{A_1, A_2, \ldots, A_K\} \) 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 \( 1A_i \) occurs at time \( \tau_A \) and in \( K \) more of these there is an \( 1 \leq i \leq K \) such that the pattern \( 2A_i \) occurs at time \( \tau_A \).

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 \(*1212, *2212, *122, *222\) where the wild-card symbol \(*\) stands for an arbitrary (possibly empty) string of 1’s and 2’s. By the definition of \( \tau_A \) as the first occurrence time of a pattern from \( \mathcal{A} = \{212, 22\} \), we see that neither of the strings \( 2121 \) and \( 2222 \) can occur as a substring of the string \( Z_1, Z_2, \ldots, Z_{\tau_A} \); 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 \( \tau_A \), 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 \( N \). 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 consider the instructions for a 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 \( \tau_A \) 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 \( \tau_A \) 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 \( \tau_A \) the usual cash flow calculation tells us

\[
X_{\tau_A} = \sum_{j=1}^{N} y_j (\tau_A - 1) - \sum_{i=1}^{K} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i} - \sum_{i=K+1}^{K+N} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i},
\]

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 \( \{y_j\}_{1 \leq j \leq N} \) that solve the \( N \times N \) system

\[
\sum_{j=1}^{N} W_{ij} y_j = 1, \quad \text{for all } i \in \{K + 1, K + 2, \ldots, K + N\}.
\]

For such initial bet sizes, we then have the representation

\[
X_{\tau_A} = \sum_{j=1}^{N} y_j (\tau_A - 1) - \sum_{i=1}^{K} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i} - \sum_{i=K+1}^{K+N} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i},
\]

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

\[
0 = \mathbb{E}(X_{\tau_A}) = \sum_{j=1}^{N} y_j (\mathbb{E}(\tau_A) - 1) - \sum_{i=1}^{K} \sum_{j=1}^{N} W_{ij} y_j \pi_i - \sum_{i=1}^{K} \pi_i - \sum_{i=K+1}^{K+N} \sum_{j=1}^{N} y_j W_{ij},
\]

which one again solves for \( \mathbb{E}(\tau_A) \).

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

\[
\mathbb{E}(\tau_A) = 1 + \frac{(1 - \sum_{i=1}^{K} \pi_i) + \sum_{i=1}^{K} \pi_i \sum_{j=1}^{N} y_j W_{ij}}{\sum_{j=1}^{N} y_j}.
\]

**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 \( \{\ast 211, \ast 1212, \ast 2212\} \). 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 \( 5 \times 3 \) scenario-by-team matrix \( \{W_{ij}\} \) is then given by
Here initial distribution to be  

\[
\begin{pmatrix}
\frac{1}{p_{11}} & 0 & 0 \\
0 & \frac{1}{p_{12}} & \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}} \\
\frac{1}{p_{21}p_{12}} + \frac{1}{p_{11}} & 0 & 0 \\
0 & \frac{1}{p_{12}p_{21}p_{12}} + \frac{1}{p_{12}} & \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}} \\
0 & \frac{1}{p_{22}p_{21}p_{12}} + \frac{1}{p_{12}} & \frac{1}{p_{21}p_{12}} + \frac{1}{p_{12}}
\end{pmatrix}
\]

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 \(\pi_1 = p_{11}p_{11}\) and \(\pi_2 = p_{22}p_{21}p_{12}\) 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

\[E(\tau_A) = 2 + p_{12}p_{11} + \frac{1 - p_{11}p_{11}}{p_{12}}.\]

**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 \(A\). In a similar vein, Gava and Salotti (2014) also show how one can use the method of gambling teams to calculate the scenario probabilities \(P(\tau_A = \tau_{A_i})\).

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 \(P(Z_1 = m) = p_m, 1 \leq m \leq M\), and we take the transition matrix to be \(P = \{p_{ij}\}_{1 \leq i, j \leq M}\), where, as always, \(p_{ij} = P(Z_{n+1} = j | Z_n = i)\).

In this situation, the analysis of the waiting time \(\tau_A\) until the occurrence of compound pattern \(A = \{A_1, A_2, \ldots, A_K\}\) requires us to place some restrictions on \(A\) 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 \(B \in A\) contains any \(A \in A\) 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 \(A\) and not change the distribution of \(\tau_A\).
- We assume that \(P(\tau_A = \tau_{A_i}) > 0\) for all \(1 \leq i \leq K\). If to the contrary we were to have \(P(\tau = \tau_{A_i}) = 0\) for some \(i\), then we could just drop \(A_i\) from \(A\) without changing the distribution of \(\tau_A\). Here we should note that for independent sequences one can never have \(P(\tau = \tau_{A_i}) = 0\), but for Markov sequences one must explicitly exclude this possibility. For example, if \(A_i\)
The Multi-state Chain Martingale Construction

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 \( A \) into an initial list of \( 1 + M + M^2 \) feasible ending scenarios:

- either the sequence \( A \) occurs as an initial segment of \( \{Z_n : n \geq 1\} \), or
- for some state \( k \in S \), the pattern \( kA \) 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 \( kmA \) 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 \( \mathcal{A} = \{A_1, A_2, \ldots, A_K\} \) 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 \( Z_1, Z_2, \ldots, Z_{n+\tau} \). Some are impossible because of the definition of \( \tau_A \) and some are impossible because of the structure of the transition matrix \( \{p_{ij}\} \). We need to clean up this candidate list.

We let \( \mathcal{I} \) be the set of feasible initial-ending scenarios that can actually occur, and we 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 \( kmA_i \in \mathcal{N} \) we introduce a team of gamblers that we call the \( kmA_i \)-gambling team. For each such team we also introduce an initial stake which we call the team stake.

The \( n + 1 \)st gamble from the \( kmA_i \)-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 \( mA_i \). 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 \( A = a_1a_2 \cdots a_m \), 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 \( 1/p_{a_0a_1} \) 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 \( 1/p_{a_1a_2} \), otherwise he leaves the game with nothing. He continues to bet his total capital on the successive letters of \( a_3, a_4, \ldots, a_m \).
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 \( N \) 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 \( \tau_A \) we have

\[
X_{\tau_A} = \sum_{j=1}^{N} y_j (\tau_A - 1) - \sum_{i=1}^{I} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i} - \sum_{i=I+1}^{I+N} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i},
\]

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

If the real values \( \{y_j\}_{1 \leq j \leq N} \) satisfy the system

\[
\sum_{j=1}^{N} W_{ij} y_j = 1, \text{ for all } I + 1 \leq i \leq I + N,
\]

then \( X_{\tau_A} \) has the more tractable representation

\[
X_{\tau_A} = \sum_{j=1}^{N} y_j (\tau_A - 1) - \sum_{i=1}^{I} \sum_{j=1}^{N} W_{ij} y_j 1_{E_i} - \sum_{i=I+1}^{I+N} 1_{E_i}.
\]

Again, \( \{X_n\}_{n \geq 1} \) has bounded increments and \( E[\tau_A] < \infty \), so if we write \( \pi_i \) for the probability \( P(E_i) \) that the \( i \)'th initial-ending scenario occurs, then Doob’s stopping theorem gives us

\[
0 = E(X_{\tau_A}) = \sum_{j=1}^{N} y_j (E(\tau_A) - 1) - \sum_{i=1}^{I} \sum_{j=1}^{N} W_{ij} y_j \pi_i - (1 - \sum_{i=1}^{I} \pi_i).
\]

One then solves for \( E(\tau_A) \) to obtain the main result of this section.

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

\[
E(\tau_A) = 1 + \frac{(1 - \sum_{i=1}^{I} \pi_i) + \sum_{i=1}^{I} \pi_i \sum_{j=1}^{N} y_j W_{ij}}{\sum_{j=1}^{N} y_j}.
\]

**Example 7.** Take \( S = \{1, 2, 3\} \) and \( A = \{323, 313, 33\} \). 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 \( I \) observable initial-ending scenarios is given by

\[
I = \{323, 313, 33, 1323, 2323, 1313, 313, 133, 233\}.
\]
Also, because the transitions $1 \rightarrow 2$ and $2 \rightarrow 1$ 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

$$\mathcal{N} = \{\ast11323, \ast22323, \ast11313, \ast22313, \ast1133, \ast2233\}.$$  

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 $\ast11323$ this team would win

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

and in the later-ending scenario $\ast22323$ 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

$$E(\tau_A) = 8 + \frac{7}{15}.$$  

REMARKS. In parallel with our earlier examples, one can use the initial bets sizes $ny_j + z_j$ to get the second moment of $\tau_A$, or one can use the initial bets sizes $y_j \alpha^n$ to get the corresponding generating function of $\tau_A$, 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 $E(\tau_A)$ 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 $A$ and lengths of the single patterns in $A$. 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 $A = \{A_1\}$ 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 $\tau_A$, and it is natural to ask if these computations might also lead to effective approximations of the distribution of $\tau_A$. This is true, but the path to good approximations is not as direct as one might expect.

Since the clumping heuristic typically applies to the stopping time $\tau_A$ associated with a scan statistic, one certainly expects the tail probabilities $P(\tau_A > n)$ 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 $\tau_A$.

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 $\tau_A$ we then choose $c$ and $X$ so that two moments match:

$$E(X + c) = E(\tau_A), \quad Var(X + c) = Var(X) = Var(\tau_A).$$

If we set $\mu = E(\tau_A)$ and $\sigma^2 = Var(\tau_A)$, we then the shifted exponential approximation for the tail probabilities of $\tau_A$ is given by

$$P(\tau_A > n) \approx \exp(-(n + 1/2 + \sigma - \mu)/\sigma),$$

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 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 $\mu$ is large and $\sigma$ is close to $\mu$, the differences between the various approximations are marginal, and all of the estimates are close to the true probability. On the other hand, if $\mu$ is relatively small and $\sigma$ differs substantially from $\mu$, 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
Table 1. Approximate values of $P(\tau_A \leq n)$ for the fixed window scan for least 3 failures out of 10 consecutive trials. Here we have $P(Z_n = \text{Failure}) = .01$, $\mu = 30822$, and $\sigma = 30815$.

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

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

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

distribution of the waiting time $\tau$ until one of these two situations occurs. Here in order for $\tau$ and the waiting time $\tau_A$ 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

$$P(Z_n = 1) = .04, \quad P(Z_n = 2) = .01, \quad \text{and} \quad P(Z_n = 0) = .95.$$  

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>exponential</th>
<th>shifted</th>
<th>gamma</th>
<th>simulated N=100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.05857</td>
<td>0.05085</td>
<td>0.05542</td>
<td>0.05029</td>
</tr>
<tr>
<td>60</td>
<td>0.07033</td>
<td>0.06285</td>
<td>0.06865</td>
<td>0.06187</td>
</tr>
<tr>
<td>70</td>
<td>0.08195</td>
<td>0.07470</td>
<td>0.07817</td>
<td>0.07404</td>
</tr>
<tr>
<td>80</td>
<td>0.09342</td>
<td>0.08640</td>
<td>0.08939</td>
<td>0.08623</td>
</tr>
<tr>
<td>90</td>
<td>0.10474</td>
<td>0.09796</td>
<td>0.10050</td>
<td>0.09718</td>
</tr>
<tr>
<td>100</td>
<td>0.11593</td>
<td>0.10936</td>
<td>0.11150</td>
<td>0.11058</td>
</tr>
</tbody>
</table>

that the shifted exponential approximation again works well even in the challenging case when $\mu$ and $\sigma$ are both relatively small and substantially different.

Table 4. Approximate values of $P(\tau_A \leq n)$ 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 $P(Z_n = 1) = .04$, $P(Z_n = 2) = .01$, $\mu = 324.09$, and $\sigma = 318.34$.

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

6. Summary and Concluding Observations

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

When one considers the waiting time $\tau_A$ 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 $\tau_A$ 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 reflections of this invariance property.

When we looked at the time $\tau_A$ until the first occurrence of a pattern from a list $A = \{A_1, A_2, \ldots, A_K\}$ 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 $E(\tau_A)$ and $\text{Var}(\tau_A)$ are in hand for both independent and Markov sequences, they help materially with the approximation of the distribution of $\tau_A$. In particular, moment matching within the family of shifted exponential distributions turns out to give surprisingly good approximations to the distribution of $\tau_A$, though here much of the evidence is numerical.

References


