The Analysis of Discrete Transient Events in Markov Games

Brian D. Ewald, Jeffrey Humpherys, and Jeremy M. West

Abstract— **The evolution of a system from the transient phase into a steady-state or asymptotic phase is an important area of study in engineering and the mathematical sciences. While analytic methods exist for determining the steady-state behavior of a system, the transient analysis is typically more difficult. Transient analysis is often approached in either an ad hoc, case-by-case manner or is performed by simulation. In this paper we explore the transient analysis of absorbing Markov chains by counting discrete-time events. We derive a closed-form expression for the expectation of these events and give some examples. We then show how several single-agent systems may be combined into a multi-agent system where the interactions between agents can be analyzed. This affords a model for analyzing competition. For example, we can determine advantages to specific players and determine the expected number of lead changes. After developing these ideas we present simulation results to verify our methods.**

I. Introduction

An important phenomenon in the mathematical sciences is the evolution of a system from a transient phase into a steady state. Damped harmonic motion is a classic example in mechanics. Problems in chemistry and biology often deal with the diffusion of a molecule across a membrane. In business and finance the convergence of interest rates and prices to market equilibrium exhibits this behavior. In electrical systems a voltage change causes oscillations for some time before establishing a steady state current. Structural oscillations, due to wind for example, factor into the design and construction of bridges and high-rise buildings.

Many of these examples have been analyzed mathematically and myriad results exist describing their asymptotic behavior. The transient analysis is often difficult; moreover, it is important for a viable solution. An excessively long period of below-equilibrium prices can ruin a company that cannot cover production costs or overhead. On the other hand, above-equilibrium prices can result in wasted inventory or the inability to sell enough product to cover fixed costs. Too frequent price changes can be costly and disconcerting to customers. Hence, approaching equilibrium price with the fewest changes is desirable. Similarly, spikes in current can damage electrical devices and large-amplitude or high-frequency oscillations in buildings can cause structural damage.

Transient analysis is particularly interesting in multi-agent, competitive systems. In environments such as business and athletics, competitive systems occur naturally. Political campaigns, both for ballot measures and for candidacy, can be modeled as a competition. Portfolio management can be seen as a competition between a portfolio of financial instruments and an index fund. If a fund manager is unable to consistently beat an index fund, he may have difficulty securing market share.

In this paper, we develop a method for analyzing the transient behavior of finite state, absorbing Markov chains. These models are familiar and provide a reasonable level of flexibility and sophistication. Our method allows us to determine the expected time to absorption and the frequency of specific states or transitions. We also develop a method for combining several single-agent systems into a multi-agent Markov chain. We use this composite system to analyze competitive systems. In particular, we measure the advantage of specific agents as well as the number of times that the lead changes.

Meyer [10] showed among other things that the group generalized inverse, a special case of the Drazin inverse, could be used to determine (i) the expected number of visits to any given transient state, and (ii) the probability of absorption into a particular state. Whereas Meyer's method determines the expected number of occurrences of state events, our method computes the expected number of occurrences of transition events. Nonetheless, by summing the expectation of all transitions that arrive into a given state, one can also compute the expected number of occurrences of state events. Therefore, this approach is a generalization of Meyer's method for determining expectations such as (i) and (ii). For example, with Meyer's method, the expected time to absorption is produced by summing the expected number of visits to each transient state, while our method sums the expected number of traversals of any transition leaving a transient state.

Although the emphasis is on transient analysis, our method can also be used to determine steady state behavior as well. The result of the paper is that computations which would otherwise be performed on a problem-by-problem basis or which would have been approximated using simulation can now be computed explicitly using a closed-form expression. This expression is a matrix computation that depends on a mask, the transition matrix, and the initial distribution. The expression can be easily computed using a matrix algebra platform such as MATLAB. Additionally, the paper enables theoretical results which may have previously been unattainable.

The paper is organized as follows. In Section II, we show how to represent transition events using masks and define

B. Ewald is with the Department of Mathematics at Florida State University, Tallahassee, FL 32306, USA. `ewald@math.byu.edu`

J. Humpherys and J. West are with Department of Mathematics, Brigham Young University, Provo, UT 84602, USA. `jeffh@math.byu.edu`, `jwest@math.byu.edu`

the associated random variables. We then give an expression for the expectation of these random variables on reducible Markov chains. Next we examine the time-average case, which yields the steady-state behavior of reducible chains. In Section III, we give examples of masks. In Section V, we present a series of simulations and compare them to the expectations directly computed from Section II.

II. Main Results

In this section we develop the main results of the paper. After dispensing with the preliminaries, we give an expression for the expected number of occurrences of a transition event on an absorbing Markov chain. Following this we generalize to arbitrary reducible chains. Finally, we show how masks may be used to determine the steady-state behavior of reducible chains.

A. Preliminaries

To avoid confusion with the transition matrix T we denote the transpose of a matrix by A^*. Let $A \odot B$ denote the Hadamard product, that is $(A \odot B)_{i,j} = A_{i,j}B_{i,j}$. The following theorem, found in [4, pg. 305], relates the Hadamard product to matrix multiplication.

Theorem 2.1: Let $x \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{m\times n}$ be given and let $D = \operatorname{diag}(x)$. Then $(ADB^*)_{i,i} = [(A \odot B)x]_i$.

In this paper we consider finite, stationary (temporally homogeneous) Markov chains, denoted X_k; see for example [2]. Here, $\mathcal{S} = \{1, \dots, n\}$ is the state space. If $\mu \in \mathbb{R}^n$ is stochastic, that is $\mu_i \geq 0$ and $\|\mu\|_1 = 1$, then P_μ is the unique probability measure on $\Omega = \mathcal{S} \times \mathcal{S} \times \cdots$ satisfying

$$P_\mu(X_0 = i) = \mu_i$$

and having transition probabilities associated with the Markov chain X_k. Furthermore, E_μ is expectation with respect to P_μ. The (column)-stochastic matrix $T \in \mathbb{R}^{n\times n}$ with entries

$$T_{i,j} = P(X_{k+1} = i \mid X_k = j)$$

is the transition matrix. The k-step transition probabilities are found in T^k. To summarize,

$$P_\mu(X_k = i) = \left[T^k\mu\right]_i .$$

A mask is a matrix $M \in \mathbb{R}^{n\times n}$ that describes the weight assigned to the transitions of a Markov chain. Here $M_{i,j}$ is the weight assigned to the transition from the j^{th} state to the i^{th} state. The transition event for M is the random variable whose value is the sum of the mask entries on any realization,

$$Y_M = \sum_{k=0}^{\infty} M_{X_{k+1},X_k}.$$

Example 2.2: Consider a three-state Markov chain. The number of times the chain transitions from the second state to the first state is a transition event which can be described by the mask

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

For this mask, the random variable $Y_M(\omega)$ defined as above gives the number of times the chain transitions from the second state to the first during the instance ω.

Lemma 2.3: Let $M, T \in \mathbb{R}^{n\times n}$ be given with T stochastic. For any stochastic $\mu \in \mathbb{R}^n$,

$$E_\mu M_{X_{k+1},X_k} = \sum_{i=1}^{n} \left[(M \odot T)T^k\mu\right]_i .$$

Proof: By application of the law of total probability,

$$\begin{aligned} E_\mu M_{X_{k+1},X_k} &= \sum_{i,j=1}^{n} M_{i,j} P_\mu(X_{k+1} = i, X_k = j) \\ &= \sum_{i,j=1}^{n} M_{i,j} T_{i,j} \left[T^k\mu\right]_j \\ &= \sum_{i=1}^{n} \left[(M \odot T)T^k\mu\right]_i \end{aligned}$$

■

B. Cumulative Events on Absorbing Chains

We now consider transition events on absorbing chains. In the next section we generalize to reducible chains. For details on terminology see [1], [6], [7], [9]. An absorbing state $a \in \mathcal{S}$ is any state satisfying

$$P(X_{k+1} = a \mid X_k = a) = 1.$$

Denote the set of absorbing states by $\mathcal{A}$. A Markov chain X_k is absorbing if $\mathcal{A} \neq \emptyset$ and for every $s \in \mathcal{S}$ there exists $k \in \mathbb{N}$ such that

$$P(X_k \in \mathcal{A} | X_0 = s) > 0.$$

In other words, an absorbing chain is a reducible chain in which all the ergodic classes are single states.

Without loss of generality, the transition matrix of an absorbing chain assumes the form

$$T = \begin{bmatrix} A_T & 0 \\ B_T & I \end{bmatrix}, \tag{1}$$

where $A_T \in \mathbb{R}^{m\times m}$, $m = |\mathcal{S}| - |\mathcal{A}|$. Thus, A_T, B_T are the transitions leaving the m transient states and I represents the $n - m$ absorbing states. In particular, A_T does not have 1's on the diagonal. Furthermore,

$$T^k = \begin{bmatrix} A_T^k & 0 \\ B_T \sum_{m=0}^{k-1} A_T^m & I \end{bmatrix}, \tag{2}$$

which brings us to the following observation.

Lemma 2.4: If T is the transition matrix of an absorbing Markov chain then the spectral radius of A_T satisfies $\rho(A_T) < 1$. Moreover, $(I - A_T)^{-1}$ exists and

$$(I - A_T)^{-1} = \sum_{k=0}^{\infty} A_T^k.$$

Proof: Let k be chosen sufficiently large so that $P_{e_i}(X_k \in \mathcal{A}) > 0$ for all $1 \leq i \leq n$ where e_i is the i^{th} standard unit vector in $\mathbb{R}^n$. Consider the block form of T^k given by (1). Our choice of k implies that each column of

$B_{T^k} = B_T \sum_{m=0}^{n-1} A_T^m$ has a nonzero entry. Since T^k is column-stochastic each column of $A_{T^k} = A_T^k$ has a column sum strictly less than 1. Equivalently, $\|A_T^k\|_1 < 1$ and it follows that $\rho(A_T) \le \|A_T\|_1 < 1$. ■

Lemma 2.5: Let $M, T \in \mathbb{R}^{n\times n}$ where T is the transition matrix of an absorbing Markov chain. If $M_{i,i} = 0$ for each $i \in \mathcal{A}$ then

$$\sum_{k=0}^{\infty}(M \odot T)T^k = (M \odot T)\begin{bmatrix}(I - A_T)^{-1} & 0\\ 0 & 0\end{bmatrix}.$$

Proof: Since $T_{i,j} = 0$ for $j \in \mathcal{A}, i \ne j$, then $M_{i,i} = 0$ for $i \in \mathcal{A}$ guarantees that $(M \odot T)_{i,j} = 0$ for all $j \in \mathcal{A}$. Using the block form (1) for M and T, and (2) for T^k we have

$$\begin{aligned}(M \odot T)T^k &= \begin{bmatrix}A_M \odot A_T & 0\\ B_M \odot B_T & 0\end{bmatrix}\begin{bmatrix}A_T^k & 0\\ B_T\sum_{m=0}^{k-1} A_T^m & I\end{bmatrix}\\ &= \begin{bmatrix}(A_M \odot A_T)A_T^k & 0\\ (B_M \odot B_T)A_T^k & 0\end{bmatrix}\\ &= (M \odot T)\begin{bmatrix}A_T^k & 0\\ 0 & 0\end{bmatrix}.\end{aligned}$$

Summing over k yields

$$\begin{aligned}\sum_{k=0}^{\infty}(M \odot T)T^k &= (M \odot T)\sum_{k=0}^{\infty}\begin{bmatrix}A_T^k & 0\\ 0 & 0\end{bmatrix}\\ &= (M \odot T)\begin{bmatrix}(I - A_T)^{-1} & 0\\ 0 & 0\end{bmatrix}.\end{aligned}$$

■

Remark. Let $Q = I - T$. Notice that

$$Q^- = \begin{bmatrix}(I - A_T)^{-1} & 0\\ 0 & 0\end{bmatrix}$$

satisfies $QQ^-Q = Q$ and $Q^-QQ^- = Q^-$ so that Q^- is a (1,2)-inverse of Q; see for example [1]. However, it is not always the case that $(QQ^-)^* = QQ^-$ nor that $(Q^-Q)^* = Q^-Q$ so Q^- is not the Moore-Penrose inverse. Nor is it the Drazin inverse of Q since Q and Q^- do not necessarily commute.

Theorem 2.6: Let $M, T \in \mathbb{R}^{n\times n}$, $\mu \in \mathbb{R}^n$ be given where T is the transition matrix of an absorbing Markov chain and μ is stochastic. Set $D = \operatorname{diag}(Q^-\mu)$. If $M_{i,i} = 0$ for all $i \in \mathcal{A}$ then the random variable

$$Y_M = \sum_{k=0}^{\infty} M_{X_{k+1},X_k}$$

has finite expectation

$$E_\mu Y_M = \operatorname{tr}(MDT^*).$$

Proof: Suppose that $M_{i,j} \ge 0$ for all i, j so that Y_M is an increasing series. Then by the Monotone Convergence Theorem we may exchange the order of summation and expectation

$$\begin{aligned}E_\mu Y_M &= E_\mu \sum_{k=0}^{\infty} M_{X_{k+1},X_k}\\ &= \sum_{k=0}^{\infty} E_\mu M_{X_{k+1},X_k}\\ &= \sum_{k=0}^{\infty}\sum_{i=1}^{n}\left[(M \odot T)T^k\mu\right]_i\\ &= \sum_{i=1}^{n}\left[\left(\sum_{k=0}^{\infty}(M \odot T)T^k\right)\mu\right]_i\\ &= \sum_{i=1}^{n}\left[(M \odot T)Q^-\mu\right]_i\\ &= \operatorname{tr}(MDT^*).\end{aligned}$$

Notice that we have only used the assumption $M_{i,j} \ge 0$ in the first step where we exchanged the order of expectation and summation. For the general case let Z be the random variable given by

$$Z = \sum_{k=0}^{\infty} |M_{X_{k+1},X_k}|.$$

For all $m \in \mathbb{N}$ the triangle inequality indicates that

$$\left|\sum_{k=0}^{m} M_{X_{k+1},X_k}\right| \le \sum_{k=0}^{m} |M_{X_{k+1},X_k}| \le \sum_{k=0}^{\infty} |M_{X_{k+1},X_k}| = Z.$$

Our previous work indicates that $E_\mu|Z| = E_\mu Z < \infty$ so that the Dominated Convergence Theorem allows us to again exchange the order of summation with expectation in $E_\mu Y_M$. The remainder of the argument is identical to the nonnegative case. ■

In Theorem 2.6 we show that on an absorbing chain a sufficient condition to guarantee $E_\mu|Y_M| < \infty$ is that $M_{i,i} = 0$ for $i \in \mathcal{A}$. This condition is practically necessary in the sense that for $i \in \mathcal{A}$, if $P_\mu(X_k = i) > 0$ for some $k \in \mathbb{N}$ then $M_{i,i} \ne 0$ implies that $E_\mu|Y_M| = \infty$. Thus, $M_{i,i} = 0$ is required of all absorbing states that are "reachable."

C. Cumulative Events on Reducible Markov Chains

In the previous section we considered a special case of reducible Markov chains. We now generalize to any reducible chain using the canonical form for reducible matrices; see for example [10], [9], [7]. Every reducible Markov chain can be written in the *canonical form for reducible Markov chains*, given in Fig. 1.

The blocks T_{11} through T_{rr} are the transient classes and the blocks $T_{r+1,r+1}$ through T_{mm} are the ergodic classes. It is well known that $\rho(T_{ii}) < 1$ for the transient classes, $i \le r$. The ergodic classes of a reducible chain generalize the notion of an absorbing state to a collection of states. We generalize the block form (1) for T to

$$T = \begin{bmatrix}A_T & 0\\ B_T & E_T\end{bmatrix},$$

Fig. 1. The canonical form for reducible Markov chain. The blocks $T_{11},\dots,T_{rr}$ represent the transient states and the block $T_{r+1,r+1},\dots,T_{mm}$ represent the ergodic classes.

$$T=\left[\begin{array}{cccc|cccc} T_{11} & 0 & \dots & 0 & 0 & 0 & \dots & 0\\ T_{21} & T_{22} & \dots & 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ T_{r1} & T_{r2} & \dots & T_{rr} & 0 & 0 & \dots & 0\\ \hline T_{r+1,1} & T_{r+1,2} & \dots & T_{r+1,r} & T_{r+1,r+1} & 0 & \dots & 0\\ T_{r+2,1} & T_{r+2,2} & \dots & T_{r+2,r} & 0 & T_{r+2,r+2} & & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ T_{m1} & T_{m2} & \dots & T_{mr} & 0 & 0 & \dots & T_{mm} \end{array}\right].$$

where A_T corresponds to the transient states, B_T is the transition into the ergodic classes, and E_T is block diagonal containing the ergodic classes. We denote the ergodic states by $\mathcal{E}$, the i^{th} ergodic class by $\mathcal{E}_i$, and the transient states by $\mathcal{T}$.

Theorem 2.7: Let $M, T \in \mathbb{R}^{n\times n}$ be given where T is a reducible stochastic matrix in canonical form and let $\mathcal{E}$ be the indices of the ergodic states. Let $\mu \in \mathbb{R}^n$ be stochastic and set $D = \operatorname{diag}(Q^-\mu)$. If $M_{i,j} = 0$ whenever either of $i, j \in \mathcal{E}$ then the random variable

$$Y_M = \sum_{k=0}^{\infty} M_{X_{k+1},X_k}$$

has expectation given by

$$E_\mu Y_M = \operatorname{tr}(MDT^*).$$

Proof: Since $\rho(T_{ii}) < 1$ for all the transient classes it follows that $\rho(A_T) < 1$ as in Lemma 2.4. The condition $M_{i,j} = 0$ for either of $i, j \in \mathcal{E}$ guarantees the result of Lemma 2.5. With these results, the remainder of the proof is identical to the proof of Theorem 2.6. ■

D. Time-Average Events

The previous sections address reducible Markov chains. Masks may be used for any general chain although the sum

$$\sum_{k=0}^{\infty} M_{X_{k+1},X_k}$$

does not converge in the general case. However, for any stochastic matrix T the limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} T^k = G \tag{3}$$

exists. If we let $Q = I - T$ as above, then G is the projector onto the null space $\mathcal{N}(Q)$ along the range $\mathcal{R}(Q)$. In terms of the group generalized inverse, or Drazin inverse $Q^{\#}$, we can write $G = I - QQ^{\#}$; see for example [1], [7], [10]. For a square matrix A, there exists a matrix P such that

$$A = P^{-1}\begin{bmatrix} B & 0\\ 0 & N\end{bmatrix} P$$

where N is nilpotent and B is invertible. Here N is the Jordan segment of A corresponding to the eigenvalue $\lambda = 0$. Then the group generalized inverse of A is given by

$$A^{\#} = P^{-1}\begin{bmatrix} B^{-1} & 0\\ 0 & 0\end{bmatrix} P.$$

Theorem 2.8: Let $M, T \in \mathbb{R}^{n\times n}$ be given with T stochastic. For any stochastic $\mu \in \mathbb{R}^n$, set $D = \operatorname{diag}(G\mu)$. Then the random variable

$$Y_M = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} M_{X_{k+1},X_k}$$

has expectation given by

$$E_\mu Y_M = \operatorname{tr}(MDT^*).$$

Proof: Let $\gamma = \max\{|M_{i,j}| \mid 1 \le i, j \le n\}$. Then for all $N \in \mathbb{N}$,

$$\frac{1}{N}\sum_{k=0}^{N} M_{X_{k+1},X_k} \le \frac{1}{N}\sum_{k=0}^{N}\gamma = \frac{N+1}{N}\gamma < 2\gamma$$

so that we may apply the Dominated Convergence Theorem. This and the linearity of expectation give

$$\begin{aligned} E_\mu Y_M &= E_\mu \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} M_{X_{k+1},X_k}\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} E_\mu M_{X_{k+1},X_k}\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N}\sum_{i=1}^{n}\left[(M\odot T)T^k\mu\right]_i\\ &= \sum_{i=1}^{n}\left[(M\odot T)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} T^k\right)\mu\right]_i\\ &= \sum_{i=1}^{n}\left[(M\odot T)G\mu\right]_i\\ &= \operatorname{tr}(MDT^*). \end{aligned}$$

■

Time-average events may also be applied to reducible chains. In this case, the value of the mask M on the transitions leaving transient states is irrelevant since $Y_M(\omega)$ is determined by the time-average value on the ergodic class that ω enters. Thus, Y_M represents the steady-state behavior of T. For example, for an absorbing chain

$$E_\mu Y_M = \sum_{i\in\mathcal{A}} P_\mu(X_k \to i) M_{i,i}.$$

For example, if $M_{i,i} = 1$ for a given $i \in \mathcal{A}$ and $M_{i,j} = 0$ elsewhere then $E_\mu Y_M$ is the probability of absorption into i given the initial distribution μ.

III. Examples

In this section we present examples of masks for determining some of the canonical quantities for reducible chains, specifically, those presented by Meyer. We then give a novel example.

A. Canonical Examples

Meyer [10] showed that $Q^\#$ and $I - QQ^\#$ contain the following values for absorbing chains.

(a) For $j \in \mathcal{A}$, $(I - QQ^\#)_{i,j}$ is the probability of being absorbed into state j when initially in state i.
(b) If $i, j \notin \mathcal{A}$ then $(Q^\#)_{i,j}$ is the expected number of times the chain will be in state j when initially in state i.
(c) The expected number of steps until absorption when initially in state $i \notin \mathcal{A}$ is $\sum_{j\notin\mathcal{A}} (Q^\#)_{i,j}$.

For general reducible chains, Meyer suggests representing the ergodic class by a single state and using the above results to determine the same values. Notice that this method focuses on counting visits to states. By counting transitions we recover the same quantities and are not obligated to reduce a chain to its absorbing representation.

For any ergodic class $\mathcal{E}_m$, let

$$M_{i,j} = \begin{cases} 1 & j \in \mathcal{E}_m,\ i \in \mathcal{T}, \\ 0 & \text{otherwise.} \end{cases}$$

Then Y_M is 1 on any realization which enters $\mathcal{E}_m$ and zero elsewhere. Thus, $E_\mu Y_M$ is the probability of absorbtion into $\mathcal{E}$ which gives (a) for any reducible chain.

For (b), given any $s \in \mathcal{T}$, let

$$M_{i,j} = \begin{cases} 1 & i = s, \\ 0 & \text{otherwise.} \end{cases}$$

Then $E_\mu Y_M$ is the expected number of arrivals at state s given the initial distribution μ. Setting $M_{i,j} = 1$ when $j = s$ instead of $i = s$ gives the expected number of departures from state s. These quantities may differ depending on the initial distribution.

To find (c) let

$$M_{i,j} = \begin{cases} 1 & j \in \mathcal{T}, \\ 0 & \text{otherwise.} \end{cases}$$

$E_\mu Y_M$ is the expected number of steps until absorption into some ergodic class.

B. Composite Markov Chains

Suppose $T_1 \in \mathbb{R}^{n_1\times n_1}$ and $T_2 \in \mathbb{R}^{n_2\times n_2}$ are stochastic matrices. Let $T = T_1 \otimes T_2 \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$ be the Kronecker product of T_1 and T_2; see for example [4], [3], [8]. For simplicity, we label the entries of T by $T_{(i_1,i_2),(j_1,j_2)}$ which represents the i_2, j_2 entry of the i_1, j_1 block of T. It is straightforward to check that T is also column stochastic. Indeed, if X_k is the Markov chain of T_1 and Y_k is the Markov chain of T_2 then

$$\begin{aligned} &T_{(i_1,i_2),(j_1,j_2)} \\ &\quad = P(X_{k+1} = i_1, Y_{k+1} = i_2 \mid X_k = j_1, Y_k = j_2). \end{aligned}$$

Similarly, given stochastic $\mu_1 \in \mathbb{R}^{n_1}$ and $\mu_2 \in \mathbb{R}^{n_2}$, the vector $\mu = \mu_1 \otimes \mu_2 \in \mathbb{R}^{n_1 n_2}$ is stochastic and the same indexing scheme applies:

$$P_\mu(X_0 = i_1, Y_0 = i_2) = \mu_{(i_1,i_2)}.$$

Clearly, this generalizes to any $p \geq 2$.

Suppose T_1 represents a competitive system and the states are ordered such that higher indices are states closer to winning, that is, states that are closer to absorption. Then $T = T_1 \otimes \cdots \otimes T_p$ represents the competition between p players taking turns. It is natural to ask what the expected number of lead changes is, where a lead change is a permutation in the ordering of the players.

For the sake of clarity, let $p = 2$. We count a lead change if a player comes from behind and ends in the lead. In the case that a tie is either created or broken on a turn, we count a half a lead change. The mask for two-player lead changes is given by

$$M_{(i_1,i_2),(j_1,j_2)} = \begin{cases} 0 & j_1 \in \mathcal{A}_1 \text{ or } j_2 \in \mathcal{A}_2 \\ 1 & j_2 < j_1, i_2 > i_1 \\ 1 & j_2 > j_1, i_2 < i_1 \\ 1/2 & j_2 = j_1, i_2 \neq i_1 \\ 1/2 & j_2 \neq j_1, i_2 = i_1 \\ 0 & \text{otherwise.} \end{cases}$$

When $j_1 \in \mathcal{A}$ the (i_1, j_1) block is zero. For $j_1 \notin \mathcal{A}_1$ the (i_1, j_1) block is

$$M_{(i_1,j_1)} = \begin{bmatrix} 0 & \dots & 0 & 1/2 & 1 & \dots & 1 & \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \\ 0 & \dots & 0 & 1/2 & 1 & \dots & 1 & \vdots \\ 1/2 & \dots & 1/2 & 0 & 1/2 & \dots & 1/2 & 0 \\ 1 & \dots & 1 & 1/2 & 0 & \dots & 0 & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \\ 1 & \dots & 1 & 1/2 & 0 & \dots & 0 & \end{bmatrix}.$$

When $p > 2$, there are at least two natural ways to define a lead change. The first is to count a lead change whenever the player in the lead is passed by another. We count a half lead change for breaking or establishing a tie in the leading position. Define the lead set $L(\{x_1, \dots, x_p\})$ to be

$$L(\{x_1, \dots, x_p\}) = \{1 \leq m \leq p \mid x_m \geq x_l \text{ for } 1 \leq l \leq p\}.$$

Thus, $L(i_1, \ldots, i_p)$ is the set of indices of the players tied for the lead at the end of a turn and $L(j_1, \ldots, j_p)$ is the set of indices of the players tied for the lead at the beginning of a turn. The lead change mask is

$$M_{(i_1,\ldots,i_p),(j_1,\ldots,j_p)} = \begin{cases} 0 & j_m \in \mathcal{A}_m \text{ for some } 1 \le m \le p, \\ 1 & L(i_1,\ldots,i_p) \ne L(j_1,\ldots,j_p) \text{ and } \\ & \qquad |L(i_1,\ldots,i_p)| = |L(j_1,\ldots,j_p)| = 1, \\ 1/2 & |L(i_1,\ldots,i_p)| = 1 \text{ and } |L(j_1,\ldots,j_p)| > 1, \\ 1/2 & |L(i_1,\ldots,i_p)| > 1 \text{ and } |L(j_1,\ldots,j_p)| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The second way to extend lead changes for $p > 2$ is to count the permutations in the players positions. For example, if $j_1 > j_2 > \cdots > j_p$ and $i_1 < i_2 < \cdots < i_p$, then this is a complete reordering of position, which would count as $1 + \ldots + p$ lead changes. Of course, lead changes may have useful interpretations in contexts other than games or explicit competitions.

IV. Computation

In this section we discuss practical computation of Theorem 2.6. We treat only the common, absorbing case. The theorem assumes the states are ordered to obtain the submatrix A_T. We can perform the computation without any reordering.

(a) Set $Q = I - T$.
(b) Set $Q_{i,j} = 0$ for any i, j satisfying either $Q_{i,i} = 0$ or $Q_{j,j} = 0$. That is, zero the rows and columns of Q corresponding to the absorbing states of T.
(c) Solve the system $Q\nu = \mu$.
(d) Set $D = \operatorname{diag}(\nu)$ and compute $R = DT^*$.

The quantity R is independent of the mask. For each mask, we compute $E_\mu Y_M = \operatorname{tr}(MR)$. Of course, the trace requires the computation of only n scalar products of the rows of M and the columns of R, which correspond to the diagonal entries of MR.

V. Simulations

In this section we compute the expectations of several transition events on a specific Markov chain and compare the results to a Monte Carlo simulation as a verification of our results. We use the game Chutes and Ladders (or Snakes and Ladders), which is characterized by a substantial number of states (82) and exhibits a gradual drift towards the absorbing state combined with occasional large jumps. The MATLAB script used for computing expectations and the code for the simulations can be found in [5].

We simulated the following events in 100 million games.

- Second-To-Last Square: The number of times a player gets stuck on the second-to-last square.
- Large Ladder Traversal: The number of times the ladder from $28 \to 84$ is traversed.
- Game Length: The number of turns in the game.

In addition to the above events the following were simulated for a two-player game.

- Lead Changes: The number of lead changes in the game.
- First-player Advantage: The probability that both players finish in the same number of turns (player 1 is the winner in this case).
- First-player Win Frequency: The probability that player 1 wins.

Table I compares the sample mean obtained from simulation with the expectation computed using Theorem 2.6. The results agree up to at least three significant digits in every case.

TABLE I
Results of Event Simulations

Event	Sample Mean	Expectation Computed Using Theorem 2.6
Single-Player Events		
Second-To-Last Square	1.2954	1.2958
Large Ladder	0.5895	0.5896
Game Length	39.596	39.598
Two-Player Events		
Second-To-Last Square	1.1159	1.1166
Large Ladder	0.8181	0.8180
Game Length	26.513	26.513
Lead Changes	3.9679	3.9679
First-Player Advantage	0.0156	0.0156
First-Player Wins	0.5078	0.5078

VI. Acknowledgements

This work was supported in part by the National Science Foundation, Grant No. DMS-6007721. We thank C. Shane Reese for useful conversations.

References

[1] S. L. Campbell and C. D. Meyer, Jr. *Generalized inverses of linear transformations*. Dover Publications Inc., New York, 1991. Corrected reprint of the 1979 original.
[2] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
[3] Alexander Graham. *Kronecker products and matrix calculus: with applications*. Ellis Horwood Ltd., Chichester, 1981. Ellis Horwood Series in Mathematics and its Applications.
[4] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.
[5] Jeffrey Humpherys. http://math.byu.edu/˜jeffh/mathematics/games/chutes/index.htm. Markov simulation code, 2008.
[6] John G. Kemeny and J. Laurie Snell. *Finite Markov chains*. Springer-Verlag, New York, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics.
[7] Amy N. Langville and Carl D. Meyer. *Google's PageRank and beyond: the science of search engine rankings*. Princeton University Press, Princeton, NJ, 2006.
[8] Alan J. Laub. *Matrix analysis for scientists & engineers*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
[9] Carl Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. With 1 CD-ROM (Windows, Macintosh and UNIX) and a solutions manual (iv+171 pp.).
[10] Carl D. Meyer, Jr. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Rev.*, 17:443–464, 1975.