

COMPUTING EXPECTED TRANSITION EVENTS IN REDUCIBLE MARKOV CHAINS*

BRIAN D. EWALD[†], JEFFREY HUMPHERYS[‡], AND JEREMY M. WEST[‡]

Abstract. We present a closed-form, computable expression for the expected number of times any transition event occurs during the transient phase of a reducible Markov chain. Examples of events include time to absorption, number of visits to a state, traversals of a particular transition, loops from a state to itself, and arrivals to a state from a particular subset of states. We give an analogous expression for time-average events, which describe the steady-state behavior of reducible chains as well as the long-term behavior of irreducible chains.

Key words. Markov chains, reducible matrices, generalized inverses

AMS subject classifications. 60J22, 15A09

DOI. 10.1137/080726082

1. Introduction. In this paper we present a method for computing the expectation of transition events that occur during the transient phase of a reducible Markov chain using the Hadamard product and a generalized $(1, 2)$ -inverse; see [1]. Some examples of transition events are the time to absorption, the number of visits to a state, the number of traversals of a transition, the number of loops from a state to itself, and the arrivals to a state from a particular subset of states.

Meyer [12] showed among other things that the group generalized inverse, a special case of the Drazin inverse, can be used to determine (i) the expected number of visits to any 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 over all transitions entering a given state, one can also compute the expected number of occurrences of state events; therefore, counting transitions is not an alternative to counting states, but a generalization. Indeed, we provide examples of quantities that are easily determined using transition information but cumbersome, at best, to compute with state events.

We describe a transition event with a matrix, which we call a *mask*, where the entries of the mask are the weights assigned to each transition of the Markov chain. For example, the mask for the expected time to absorption assigns a unit weight to each transition that leaves a transient state and zero to all other transitions. The main result of this paper is a closed-form, computable expression for the expected value of a transition event; see Theorem 2.6.

In section 2, we define the random variables associated with a mask and give an expression for the expectation of these random variables on reducible Markov chains. Next we examine the time-average of a mask, which yields both the long-term behavior

*Received by the editors June 3, 2008; accepted for publication (in revised form) by I. C. F. Ipsen June 30, 2009; published electronically August 21, 2009.

<http://www.siam.org/journals/simax/31-3/72608.html>

[†]Department of Mathematics, Florida State University, Tallahassee, FL 32306 (ewald@math.fsu.edu).

[‡]Department of Mathematics, Brigham Young University, Provo, UT 84602 (jeffh@math.byu.edu, jwest@math.byu.edu). The second author’s work was partially supported by NSF grant DMS-6007721. The third author’s work was partially supported by BYU Graduate Research Fellowship Award.

of irreducible chains and the steady-state behavior of reducible chains. In section 3, we give examples of masks. In section 4, we address the numerical issues of conditioning, stability, and complexity. In section 5, we compare expectations computed using our results to a Monte Carlo simulation as a verification of our methods.

2. Main results. In this section we develop a closed-form expression for the expectation of a transition event. After dispensing with the preliminaries, we treat the absorbing chain case, that is, where the ergodic classes are single states. Next, we generalize to reducible chains. Finally, we show how masks can be used to determine the steady-state behavior of reducible chains and the time-average of irreducible chains.

2.1. 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, $(A \odot B)_{i,j} = A_{i,j}B_{i,j}$.

THEOREM 2.1 (see [6]). *Let $x \in \mathbb{R}^n$ and $A, B \in \mathbb{R}^{m \times n}$ be given, and let $D = \text{diag}(x)$. 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, [3]. Here, $\mathcal{S} = \{s_1, \dots, s_n\}$ 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 $\Omega = \mathcal{S} \times \mathcal{S} \times \dots$ satisfying $P_\mu(X_0 = s_i) = \mu_i$ 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

$$(2.1) \quad T_{i,j} = P(X_{k+1} = s_i \mid X_k = s_j)$$

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

$$(2.2) \quad P_\mu(X_k = s_i) = [T^k \mu]_i.$$

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

$$(2.3) \quad Y_M = \sum_{k=0}^{\infty} M_{X_{k+1}, X_k}.$$

LEMMA 2.2. *Given $M \in \mathbb{R}^{n \times n}$,*

$$(2.4) \quad E_\mu M_{X_{k+1}, X_k} = \sum_{i=1}^n [(M \odot T)T^k \mu]_i.$$

Proof. By total probability,

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

2.2. Cumulative events on absorbing chains. Let $\mathcal{A} \subset \mathcal{S}$ denote the absorbing states. That is, $s_j \in \mathcal{A}$ if $P(X_{k+1} = s_j \mid X_k = s_j) = 1$, or, equivalently, $T_{j,j} = 1$. A Markov chain X_k is absorbing if $\mathcal{A} \neq \emptyset$ and there exists $k \in \mathbb{N}$ such that

$$(2.5) \quad P(X_k \in \mathcal{A} \mid X_0 = s_j) > 0, \quad j = 1, \dots, n.$$

In other words, an absorbing chain is a reducible chain in which the ergodic classes are single states; see, for example, [1, 8, 9, 11]. Without loss of generality, the transition matrix of an absorbing chain assumes the form

$$(2.6) \quad T = \begin{bmatrix} A_T & 0 \\ B_T & I \end{bmatrix},$$

where $A_T \in \mathbb{R}^{t \times t}$ and $t = n - |\mathcal{A}|$ is the number of transient states. Thus, A_T and B_T are the transitions leaving the t transient states. In particular, the diagonal entries of A_T are strictly less than 1. Furthermore,

$$(2.7) \quad T^k = \begin{bmatrix} A_T^k & 0 \\ B_T \sum_{m=0}^{k-1} A_T^m & I \end{bmatrix}.$$

LEMMA 2.3 (see [11]). *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*

$$(2.8) \quad (I - A_T)^{-1} = \sum_{k=0}^{\infty} A_T^k.$$

LEMMA 2.4. *Let $M, T \in \mathbb{R}^{n \times n}$ be given, where T is the transition matrix of an absorbing Markov chain. If $M_{j,j} = 0$ whenever $s_j \in \mathcal{A}$, then*

$$(2.9) \quad \sum_{k=0}^{\infty} (M \odot T) T^k = (M \odot T) \begin{bmatrix} (I - A_T)^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof. If $s_j \in \mathcal{A}$, then $(M \odot T)_{i,j} = 0$ for $i = 1, \dots, n$. Using the block form (2.6) for M ,

$$M \odot T = \begin{bmatrix} A_M \odot A_T & 0 \\ B_M \odot B_T & 0 \end{bmatrix}.$$

Combining this with (2.7),

$$(M \odot T) T^k = \begin{bmatrix} (A_M \odot A_T) A_T^k & 0 \\ (B_M \odot B_T) A_T^k & 0 \end{bmatrix} = (M \odot T) \begin{bmatrix} A_T^k & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence,

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

Throughout the paper, let

$$(2.10) \quad Q = \begin{bmatrix} I - A_T & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad Q^- = \begin{bmatrix} (I - A_T)^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that Q^- satisfies $(I - T)Q^-(I - T) = (I - T)$ and $Q^-(I - T)Q^- = Q^-$ so that Q^- is a (1,2)-inverse of $I - T$; see, for example, [1]. However, it is not always the case that $((I - T)Q^-)^* = (I - T)Q^-$ or that $(Q^-(I - T))^* = Q^-(I - T)$. Hence, Q^- is neither the Moore–Penrose inverse nor the Drazin inverse of $I - T$ since $I - T$ and Q^- do not necessarily commute. However, it is straightforward to show that Q^- is both the Moore–Penrose inverse and the Drazin inverse of Q .

THEOREM 2.5. *Let $M, T \in \mathbb{R}^{n \times n}$ and $\mu \in \mathbb{R}^n$ be given, where T is the transition matrix of an absorbing Markov chain and μ is stochastic. Set $D = \text{diag}(Q^- \mu)$. If $M_{j,j} = 0$ for all $s_j \in \mathcal{A}$, then the random variable (2.3) has expectation*

$$(2.11) \quad E_\mu Y_M = \text{tr}(MDT^*).$$

Proof. Suppose that $M_{i,j} \geq 0$ for all i, j so that Y_M is an increasing series. Then by the monotone convergence theorem (see [3]) we may exchange the order of summation and expectation,

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

Applying Lemma 2.2,

$$E_\mu Y_M = \sum_{k=0}^{\infty} \sum_{i=1}^n [(M \odot T)T^k \mu]_i = \sum_{i=1}^n \left[\sum_{k=0}^{\infty} (M \odot T)T^k \mu \right]_i.$$

By Lemma 2.4 and Theorem 2.1,

$$E_\mu Y_M = \sum_{i=1}^n [(M \odot T)Q^- \mu]_i = \text{tr}(MDT^*).$$

For general M , let Z be the random variable $Z = \sum_{k=0}^{\infty} |M_{X_{k+1}, X_k}|$. For all $m \in \mathbb{N}$,

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

The nonnegative case indicates that $E_\mu |Z| = E_\mu Z < \infty$ so that the dominated convergence theorem allows us to exchange the order of summation with expectation. The remainder of the argument is identical to the nonnegative case. $\square$

Remark. Theorem 2.5 indicates that the condition $M_{j,j} = 0$ for $s_j \in \mathcal{A}$ is sufficient to guarantee that $E_\mu |Y_M| < \infty$. This condition is practically necessary; if $s_j \in \mathcal{A}$ satisfies $P_\mu(X_k = s_j) > 0$ for some $k \in \mathbb{N}$, then $M_{j,j} \neq 0$ implies that $E_\mu |Y_M| = \infty$. Thus, $M_{j,j} = 0$ is required of all absorbing states that are “reachable.”

2.3. Cumulative events on reducible Markov chains. We now generalize to a reducible Markov chain; see, for example, [9, 11, 12]. We assume that the transition

matrix T is in canonical form

$$(2.12) \quad T = \left[\begin{array}{cccc|cccc} T_{11} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ T_{21} & T_{22} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{r1} & T_{r2} & \cdots & T_{rr} & 0 & 0 & \cdots & 0 \\ \hline T_{r+1,1} & T_{r+1,2} & \cdots & T_{r+1,r} & T_{r+1,r+1} & 0 & \cdots & 0 \\ T_{r+2,1} & T_{r+2,2} & \cdots & T_{r+2,r} & 0 & T_{r+2,r+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{m2} & \cdots & T_{mr} & 0 & 0 & \cdots & T_{mm} \end{array} \right].$$

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. Here, $\rho(T_{ii}) < 1$ for $i \leq r$. The ergodic classes of a reducible chain generalize the notion of an absorbing state to a collection of states. We generalize the block form (2.6) for T to

$$(2.13) \quad T = \begin{bmatrix} A_T & 0 \\ B_T & E_T \end{bmatrix},$$

where A_T and B_T correspond to the transient states and E_T is block diagonal containing the ergodic classes. Let $\mathcal{E}$ denote the set of ergodic states.

THEOREM 2.6. *Set $D = \text{diag}(Q^- \mu)$. For a reducible Markov chain T , if $M_{i,j} = 0$ whenever $s_i, s_j \in \mathcal{E}$, then the random variable (2.3) has expectation*

$$(2.14) \quad E_\mu Y_M = \text{tr}(MDT^*).$$

Proof. Since $\rho(T_{ii}) < 1$ for all the transient classes, it follows that $\rho(A_T) < 1$ as in Lemma 2.3. The condition $M_{i,j} = 0$ for $s_i, s_j \in \mathcal{E}$ guarantees the result of Lemma 2.4. With these results, the remainder of the proof is identical to the proof of Theorem 2.5. $\square$

2.4. Time-average events. Masks may be used for a general Markov chain, although the random variable in (2.3) may not converge. However, the limit

$$(2.15) \quad G = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^N T^k$$

does exist. Set $S = I - T$, and let $G = I - SS^\#$, where $S^\#$ is the group generalized inverse, or Drazin inverse, of S ; see, for example, [1, 9, 12]. Then G is the projector onto the null space of S along the range of S .

THEOREM 2.7. *Set $D = \text{diag}(G\mu)$. Then for any stochastic T , the random variable*

$$(2.16) \quad Y_M = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^N M_{X_{k+1}, X_k}$$

has expectation

$$(2.17) \quad E_\mu Y_M = \text{tr}(MDT^*).$$

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

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

so that we may apply the dominated convergence theorem. This and the linearity of expectation give

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

Remark. If T is reducible, the value of M on the transitions leaving transient states is irrelevant; the value of Y_M on any realization depends only on the ergodic class that is entered. Thus, Y_M represents the steady-state behavior of T in this case. For example, in the case of an absorbing chain,

$$(2.18) \quad E_\mu Y_M = \sum_{s_j \in \mathcal{A}} P_\mu(X_k \rightarrow s_j) M_{j,j}.$$

If we fix $s_j \in \mathcal{A}$ and set $M_{j,j} = 1$ with all other entries zero, then $E_\mu Y_M$ is the probability of absorption into s_j given the initial distribution μ .

3. Examples. In this section we present examples of masks for determining some of the canonical quantities for reducible chains given by Meyer [12]. Then we give some novel examples that show the flexibility of transition events.

3.1. Canonical examples. Meyer showed that $S^\#$ and $I - SS^\#$ contain the following values for absorbing chains:

(i) For $s_i \in \mathcal{A}$, $(I - SS^\#)_{i,j}$ is the probability of being absorbed into state s_i when initially in state s_j .

(ii) If $s_i, s_j \notin \mathcal{A}$, then $S^\#_{i,j}$ is the expected number of times the chain will be in state s_i when initially in state s_j .

(iii) The expected number of steps until absorption when initially in state $s_j \notin \mathcal{A}$ is $\sum_{s_i \notin \mathcal{A}} S^\#_{i,j}$.

For general reducible chains, Meyer suggests representing the ergodic class by a single state and using the above results to determine the same quantities. We can express these quantities in terms of transition events. Furthermore, we may do so on any reducible chain without having to convert to an absorbing representation. For any ergodic class $\mathcal{E}_m$, let

$$(3.1) \quad M_{i,j} = \begin{cases} 1, & s_i \in \mathcal{E}_m, s_j \notin \mathcal{E}_m, \\ 0 & \text{otherwise.} \end{cases}$$

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 absorption into $\mathcal{E}_m$ which gives (i) for any reducible chain.

For (ii), given $s_h \notin \mathcal{E}$, let

$$(3.2) \quad M_{i,j} = \begin{cases} 1, & i = h, s_j \notin \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

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

To find (iii) let

$$(3.3) \quad M_{i,j} = \begin{cases} 1, & s_j \notin \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

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

In the next two examples, the desired quantities correspond directly with transitions and not states. Therefore, using transition events is more natural than attempting to reproduce the results using state information.

3.2. Expected path lengths. Consider an object that moves between n states with transition probabilities $T_{i,j}$, and suppose that s_n is absorbing. Let $d(s_j, s_i)$ be the distance between s_j and s_i , and set

$$(3.4) \quad M_{i,j} = \begin{cases} 0, & j = n, \\ d(s_j, s_i) & \text{otherwise.} \end{cases}$$

The random variable (2.3) describes the distance traveled on any realization. If the initial position of the object has distribution μ , then Theorem 2.5 indicates that the expected distance traveled is given by (2.11).

3.3. Application from physics. Consider a hydrogen atom that is excited by an external energy source so that the atom's electron is perpetually changing energy states. Let $\{s_1, \dots, s_n\}$ be the various allowable energy levels and $T_{i,j}$ be the probability that the atom's electron moves from s_j to s_i . Also, let μ be the distribution on the electron's initial position. To determine the portion of light emitted by the hydrogen atom that is in a particular range, say, the visible light range, we set $M_{i,j} = 1$ for any transition that emits visible light and $M_{i,j} = 0$ otherwise. Then the portion of light that is visible in any realization is the time-average random variable given by (2.16). Applying Theorem 2.7, the expected portion of visible light is given by (2.17).

3.4. 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, 6, 10]. 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 and is equal to $T_{(i_1, i_2), (j_1, j_2)} = [T_1]_{i_1, j_1} [T_2]_{i_2, j_2}$. It is straightforward to check that T is also stochastic. Indeed, if X_k is the Markov chain of T_1 with states $\{s_1, \dots, s_{n_1}\}$ and Y_k is the Markov chain of T_2 with states $\{t_1, \dots, t_{n_2}\}$, then

$$(3.5) \quad T_{(i_1, i_2), (j_1, j_2)} = P(X_{k+1} = s_{i_1}, Y_{k+1} = t_{i_2} \mid X_k = s_{j_1}, Y_k = t_{j_2}).$$

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:

$$(3.6) \quad P_\mu(X_0 = s_{i_1}, Y_0 = t_{i_2}) = \mu_{(i_1, i_2)}.$$

This generalizes in the obvious way for any finite number of transition matrices.

Suppose T_0 is a competitive system and the states are ordered such that higher indices represent being closer to winning. Then the p -wise Kronecker product $T = T_0 \otimes \cdots \otimes T_0$ represents competition between p players taking turns. It is natural to ask what the expected number of lead changes is.

For clarity, let $p = 2$. We count a lead change if a player comes from behind and ends in the lead. If 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

$$(3.7) \quad M_{(i_1, i_2), (j_1, j_2)} = \begin{cases} 0, & s_{j_1} \in \mathcal{A} \text{ or } s_{j_2} \in \mathcal{A}, \\ 1, & j_2 < j_1 \text{ and } i_2 > i_1, \\ 1, & j_2 > j_1 \text{ and } i_2 < i_1, \\ 1/2, & j_2 = j_1 \text{ and } i_2 \neq i_1, \\ 1/2, & j_2 \neq j_1 \text{ and } i_2 = i_1, \\ 0 & \text{otherwise.} \end{cases}$$

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

$$(3.8) \quad M_{(i_1, \cdot), (j_1, \cdot)} = \begin{bmatrix} 0 & \dots & 0 & 1/2 & 1 & \dots & 1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1/2 & 1 & \dots & 1 \\ 1/2 & \dots & 1/2 & 0 & 1/2 & \dots & 1/2 \\ 1 & \dots & 1 & 1/2 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & \dots & 1 & 1/2 & 0 & \dots & 0 \end{bmatrix}.$$

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. 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 complete lead change gets a weight of $M_{(i_1, \dots, i_p), (j_1, \dots, j_p)} = 1 + \cdots + p = p(p+1)/2$.

4. Computation. In this section we discuss the conditioning of (2.14). We then provide an algorithm and analyze its complexity and stability.

4.1. Conditioning. We take our definition of conditioning from Trefethen and Bau [13, p. 90]. For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, denote

$$(4.1) \quad \delta f(x) = f(x + \delta x) - f(x),$$

where $x, \delta x \in \mathbb{R}^n$. The relative condition number of f at $x \in \mathbb{R}^n$ is

$$(4.2) \quad \kappa(x) = \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f(x)\|}{\|f(x)\|} \bigg/ \frac{\|\delta x\|}{\|x\|} = \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f(x)\|}{\|\delta x\|} \frac{\|x\|}{\|f(x)\|}.$$

In this section, we give bounds on $\kappa = \sup_x \kappa(x)$ for the function $f(T, M, \mu) = \text{tr}(MDT^*)$ defined by (2.14). We treat this as three separate conditioning problems by analyzing the conditioning of f with respect to each input M, T , and μ individually;

see, for example, [13, Lecture 18]. This affords an understanding of the sensitivity of (2.14) to perturbations in each input.

The factor $\|x\|/\|f(x)\|$ in (4.2) is independent of the perturbation δx and may be pulled outside the limit. Our general approach in bounding the condition number is to evaluate, for each input separately, the absolute condition number

$$(4.3) \quad \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f(x)\|}{\|\delta x\|}$$

and then multiply by $\|x\|/\|f(x)\|$ to obtain (4.2).

Although the 1-norm is a natural choice for column-stochastic matrices, M and D are not stochastic, and the trace in (2.14) corresponds more naturally to the Frobenius inner product on the space of matrices. Therefore, we give bounds on the condition number κ in terms of the Frobenius norm $\|A\|_F = \sqrt{\text{tr}(A^*A)}$. By Cauchy–Schwarz, $|\text{tr}(A^*B)| \leq \|A\|_F \|B\|_F$. Furthermore, the Frobenius norm satisfies the submultiplicative property, that is, $\|AB\|_F \leq \|A\|_F \|B\|_F$. Therefore,

$$(4.4) \quad |\text{tr}(MDT^*)| = |\text{tr}(T^*MD)| \leq \|T\|_F \|M\|_F \|D\|_F.$$

THEOREM 4.1. *Set*

$$(4.5) \quad \kappa = \frac{\|M\|_F \|T\|_F \|(I - A_T)^{-1}\|_2}{|\text{tr}(MDT^*)|}.$$

The relative condition numbers for the expectation of transition events have the following bounds:

$$(4.6a) \quad \kappa_M \leq \kappa,$$

$$(4.6b) \quad \kappa_T \leq \kappa(1 + \|T\|_F \|(I - A_T)^{-1}\|_2),$$

$$(4.6c) \quad \kappa_\mu \leq \kappa.$$

Proof. Recall from Theorem 2.6 and (2.10) that $D = \text{diag}(\nu)$, where $\nu = Q^-\mu$. Since μ is stochastic and $\|\cdot\|_2 \leq \|\cdot\|_1$, we obtain the bound $\|\mu\|_2 \leq \|\mu\|_1 = 1$. Therefore,

$$(4.7) \quad \|D\|_F = \left(\sum_{i=1}^n d_{ii}^2 \right)^{1/2} = \|\nu\|_2 = \|Q^-\mu\|_2 \leq \|(I - A_T)^{-1}\|_2.$$

For κ_M , fix T and μ and treat $f(M) = \text{tr}(MDT^*)$ as a function of M only. We remark that D is independent of M . Therefore, for a perturbation δM of M we obtain

$$\begin{aligned} \|\delta f(M)\|_F &= |\text{tr}((M + \delta M)DT^*) - \text{tr}(MDT^*)| \\ &= |\text{tr}(\delta MDT^*)| \leq \|\delta M\|_F \|T\|_F \|(I - A_T)^{-1}\|_2 \end{aligned}$$

by (4.4) and (4.7). Hence,

$$\lim_{\delta \rightarrow 0} \sup_{\|\delta M\|_F \leq \delta} \frac{\|\delta f(M)\|_F}{\|\delta M\|_F} \leq \|T\|_F \|(I - A_T)^{-1}\|_2.$$

Multiplying by $\|M\|_F/\|f(M)\|_F = \|M\|_F/|\text{tr}(MDT^*)|$, we obtain (4.6a).

The matrix D depends on both T and μ . We denote by $D_{T+\delta T}$ and $D_{\mu+\delta\mu}$ the diagonal matrix obtained from $T + \delta T$ and $\mu + \delta\mu$, respectively, and use a similar

notation for ν . A perturbation δT of T causes a perturbation in Q^- . If δA_T is the submatrix of δT corresponding to A_T , then

$$(4.8) \quad \begin{bmatrix} (I - A_T - \delta A_T)^{-1} & 0 \\ 0 & 0 \end{bmatrix} = (Q - \delta Q)^-,$$

where δQ is δA_T padded with zeros. In the limit as $\delta \rightarrow 0$, $\|\delta A_T\|_F \leq \|\delta T\|_F \leq \delta$ implies that the inverse $(I - A_T - \delta A_T)^{-1}$ exists. Note that

$$(4.9) \quad \nu = Q^- \mu = \begin{bmatrix} (I - A_T)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mu = \begin{bmatrix} \tilde{\nu} \\ 0 \end{bmatrix},$$

where $\tilde{\nu} \in \mathbb{R}^t$ is the transient portion of ν . If $\tilde{\mu}$ and $\delta \tilde{\nu}_T$ also represent the transient portions of μ and $\delta \nu_T$, respectively, then

$$(4.10) \quad (I - A_T - \delta A_T)(\tilde{\nu} + \delta \tilde{\nu}_T) = \tilde{\mu}.$$

Since $(I - A_T)\tilde{\nu} = \tilde{\mu}$, it follows that

$$(4.11) \quad \delta \tilde{\nu}_T = (I - A_T - \delta A_T)^{-1} \delta A_T \tilde{\nu}.$$

Consider $f(T) = \text{tr}(MDT^*)$ as a function of T only, where M and μ are fixed. Then

$$(4.12) \quad \begin{aligned} \|\delta f(T)\|_F &= |\text{tr}(MD_{T+\delta T}(T + \delta T)^*) - \text{tr}(MDT^*)| \\ &\leq |\text{tr}(M(D_{T+\delta T} - D)T^*)| + |\text{tr}(MD_{T+\delta T}\delta T^*)| \\ &\leq \|M\|_F \|D_{T+\delta T} - D\|_F \|T\|_F + \|M\|_F \|D_{T+\delta T}\|_F \|\delta T\|_F \end{aligned}$$

by (4.4). We require bounds on $\|D_{T+\delta T} - D\|_F$ and $\|D_{T+\delta T}\|_F$. In terms of ν we have $\|D_{T+\delta T} - D\|_F = \|\nu + \delta \nu_T - \nu\|_2 = \|\delta \nu_T\|_2$. Since $\|\nu\|_2 = \|\tilde{\nu}\|_2$ and $\|\delta \nu_T\|_2 = \|\delta \tilde{\nu}_T\|_2$, applying (4.11) yields

$$(4.13) \quad \|\delta \nu_T\|_2 \leq \|(I - A_T - \delta A_T)^{-1}\|_2 \|\delta A_T\|_2 \|\nu\|_2.$$

Clearly, $\|\delta A_T\|_2 \leq \|\delta T\|_2 \leq \|\delta T\|_F$. Combining this fact with (4.7), we obtain

$$(4.14) \quad \|D_{T+\delta T} - D\|_F = \|\delta \nu_T\|_2 \leq \|(I - A_T - \delta A_T)^{-1}\|_2 \|(I - A_T)^{-1}\|_2 \|\delta T\|_F.$$

We now turn our attention to $\|D_{T+\delta T}\|_F = \|\nu + \delta \nu_T\|_2 \leq \|\nu\|_2 + \|\delta \nu_T\|_2$. Using (4.7) and (4.14), we obtain

$$(4.15) \quad \|D_{T+\delta T}\|_F \leq \|(I - A_T)^{-1}\|_2 + \|(I - A_T - \delta A_T)^{-1}\|_2 \|(I - A_T)^{-1}\|_2 \|\delta T\|_F.$$

Putting (4.12), (4.14), and (4.15) together, we have

$$(4.16a) \quad \frac{\|\delta f(T)\|_F}{\|\delta T\|_F} \leq \|M\|_F \|(I - A_T)^{-1}\|_2$$

$$(4.16b) \quad + \|M\|_F \|T\|_F \|(I - A_T - \delta A_T)^{-1}\|_2 \|(I - A_T)^{-1}\|_2$$

$$(4.16c) \quad + \|M\|_F \|(I - A_T - \delta A_T)^{-1}\|_2 \|(I - A_T)^{-1}\|_2 \|\delta T\|_F.$$

In the limit as $\delta \rightarrow 0$, (4.16c) is zero and $(I - A_T - \delta A_T)^{-1} = (I - A_T)^{-1}$ in (4.16b); hence

$$\lim_{\delta \rightarrow 0} \sup_{\|\delta T\|_F \leq \delta} \frac{\|\delta f(T)\|_F}{\|\delta T\|_F} \leq \|M\|_F \|(I - A_T)^{-1}\|_2 (1 + \|T\|_F \|(I - A_T)^{-1}\|_2).$$

Multiplying by $\|T\|_F/\|f(T)\|_F = \|T\|_F/|\text{tr}(MDT^*)|$, we obtain (4.6b).

Denote by $\delta\nu_\mu$ the change in ν due to a perturbation $\delta\mu$ of μ . This satisfies

$$(4.17) \quad \nu + \delta\nu_\mu = Q^-(\mu + \delta\mu).$$

By multiplying and canceling equal terms, we obtain

$$(4.18) \quad \delta\nu_\mu = Q^-\delta\mu.$$

We now consider $f(\mu) = \text{tr}(MDT^*)$ as a function of μ , where M and T are fixed. Applying (4.4), we obtain

$$\|\delta f(\mu)\|_F = |\text{tr}(MD_{\mu+\delta\mu}T^*) - \text{tr}(MDT^*)| \leq \|M\|_F\|T\|_F\|\delta\nu_\mu\|_2.$$

Using (4.18), we have $\|\delta\nu_\mu\|_2 \leq \|(I - A_T)^{-1}\|_2\|\delta\mu\|_2$. Thus,

$$(4.19) \quad \lim_{\delta \rightarrow 0} \sup_{\|\delta\mu\|_2 \leq \delta} \frac{\|\delta f(\mu)\|_F}{\|\delta\mu\|_2} \leq \|M\|_F\|T\|_F\|(I - A_T)^{-1}\|_2.$$

Since μ is stochastic, $\|\mu\|_F = \|\mu\|_2 \leq \|\mu\|_1 = 1$. Hence, multiplying (4.19) by $\|\mu\|_F/\|f(\mu)\|_F \leq 1/|\text{tr}(MDT^*)|$, we obtain (4.6c). $\square$

Since T is stochastic, $\|T\|_F \leq \sqrt{n}$. In all the examples in section 3, $\|M\|_F$ is no more than order n . Therefore, the magnitude of κ depends primarily on two factors: $\|(I - A_T)^{-1}\|_2$ and $|\text{tr}(MDT^*)|$. As $I - A_T$ becomes singular, $\|(I - A_T)^{-1}\|_2$ is unbounded. In this case, the conditioning may be poor, which is to be expected since the conditioning of the linear system $(I - A_T)\nu = \mu$ is also poor.

The conditioning may also be poor if $\text{tr}(MDT^*)$ is close to zero, particularly when $\|M\|_F\|T\|_F\|(I - A_T)^{-1}\|_2$ is relatively large. As the trace is a summation, cancellation of large magnitude terms with opposite signs results in poor conditioning. However, in all the examples in section 3, M is nonnegative. Since D and T are always nonnegative, cancellation is not a problem in this case, although the order of summation may affect roundoff errors; see [5, p. 63].

Even when M is nonnegative, $\text{tr}(MDT^*)$ may be small due to orthogonality. Recall that $\text{tr}(A^*B)$ is the Frobenius inner product on $\mathbb{R}^{m \times n}$. Therefore, $\text{tr}(MDT^*) = \text{tr}(DT^*M) = \langle TD, M \rangle_F = \|TD\|_F\|M\|_F \cos \theta$, where θ is the angle between TD and M . If these matrices are nearly orthogonal, the condition number may be large. This orthogonality often results from measuring events that are very unlikely to occur.

The quadratic dependence on $\|T\|_F\|(I - A_T)^{-1}\|_2$ in the upper bound for κ_T is to be expected since $E_\mu Y_M = \text{tr}(MDT^*)$ depends on T in two places—the product DT^* and the computation of ν .

4.2. Implementation. In this section we provide an algorithm for computing (2.14). As before, $\tilde{\mu}$ and $\tilde{\nu}$ are the first t entries of μ and ν , respectively, where t is the number of transient states. Let $M_j \cdot T_j$ denote the standard inner product of the j th columns of M and T . Then (2.14) may be expressed as

$$(4.20) \quad \text{tr}(MDT^*) = \sum_{i=1}^n [(M \odot T)\nu]_i = \sum_{j=1}^t \nu_j M_j \cdot T_j.$$

Algorithm 4.2. The following computes (4.20) for the inputs T, M , and μ , where T is in canonical form (2.12).

1. Solve $(I - A_T)\tilde{\nu} = \tilde{\mu}$ by forming the QR factorization of $(I - A_T)$ using Householder reflections; see, for example, [5, 13].
2. Compute the first t columns of $R = TD$, where $D = \text{diag}(\nu)$ by scaling the j th column of T by ν_j .
3. Compute $\psi = \sum_{j=1}^t M_j \cdot R_j$.

We refer to Steps 1–2 as the *setup*. This portion of the algorithm depends only on T and μ . Furthermore, Step 3 depends only on M . If several transition events are to be determined for the same chain and initial distribution, the setup need only be computed once.

Remark. The matrix $I - A_T$ is invertible and diagonally dominant by columns. Gaussian elimination on such a system requires no pivots and is stable [5]. However, the theoretical bounds for Gaussian elimination are insufficient to provide satisfactory bounds for Algorithm 4.2 beyond $n \approx 2300$. It is well known that Gaussian elimination generally performs much better in practice than numerical analysis suggests [5]. This does not change the asymptotic complexity of Algorithm 4.2 but does improve the constants.

4.3. Complexity. Recall that $I - A_T \in \mathbb{R}^{t \times t}$, where t is the number of transient states. It is well known that the temporal complexity of Step 1 is $O(t^3)$ and the spatial complexity is $O(t^2)$; see, for example, [2, 5, 13]. Steps 2 and 3 both have temporal and spatial complexity $O(nt)$. Therefore, the setup requires $O(t^3 + nt)$ time and $O(nt)$ space. Once the setup is completed, (4.20) may be computed in $O(nt)$ time and space for each mask representing a transition event.

4.4. Stability. In this section we give bounds on the backward errors introduced in the computation of Algorithm 4.2. We rely on the notation of Higham [5]. In particular, let u denote the unit roundoff, and let

$$(4.21) \quad \gamma_k = \frac{ku}{1 - ku} \quad \text{and} \quad \tilde{\gamma}_k = \frac{cku}{1 - cku},$$

where c is a small integer constant independent of k . The following result is useful in manipulating bounds involving γ_k .

LEMMA 4.3 (see [5, p. 67]). *If $|\delta| \leq \gamma_k$ and $|\varepsilon| \leq \gamma_j$, then $(1 + \delta)(1 + \varepsilon) = (1 + \xi)$, where $|\xi| \leq \gamma_{k+j}$.*

THEOREM 4.4. *Given T , M , and μ , the value $\hat{\psi}$ computed by Algorithm 4.2 is the exact solution for the inputs $T + \Delta T$, $M + \Delta M$, and μ , where ΔT and ΔM satisfy the columnwise bounds*

$$(4.22) \quad \|\Delta T_j\|_2 \leq 2\sqrt{n}\tilde{\gamma}_{n^2}\|T_j\|_2 \quad \text{and} \quad \|\Delta M_j\|_2 \leq \frac{(1 + 2\sqrt{n})\tilde{\gamma}_{n^2}}{\sqrt{1 - 4\sqrt{n}\tilde{\gamma}_{n^2}}}\|M_j\|_2,$$

provided $1 - 4\sqrt{n}\tilde{\gamma}_{n^2} > 0$.

Proof. The computed solution obtained in Step 1 satisfies the following columnwise backward error bounds [5, p. 361]:

$$(I - A_T - \Delta A_T)\hat{\nu} = \tilde{\mu} + \Delta\tilde{\mu}, \quad \text{where } \|\Delta A_{T_j}\|_2 \leq \tilde{\gamma}_{n^2}\|(I - A_T)_j\|_2, \quad 1 \leq j \leq t.$$

Since T_j is stochastic, $1 = \|T_j\|_1 \leq \sqrt{n}\|T_j\|_2$, and hence

$$\|(I - A_T)_j\|_2 \leq 1 + \|T_j\|_2 \leq (\sqrt{n} + 1)\|T_j\|_2 \leq 2\sqrt{n}\|T_j\|_2.$$

Setting

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

we obtain the bound

$$(4.23) \quad \|\Delta T_j\|_2 = \|\Delta A_{T_j}\|_2 \leq 2\sqrt{n}\tilde{\gamma}_{n^2}\|T_j\|_2.$$

Since D is diagonal, the computation in Step 2 to produce the matrix $R = TD$ involves only a single multiplication in each entry of R . Therefore, the computed result satisfies

$$\hat{R}_{i,j} = (1 + \delta_{i,j})\hat{\nu}_j T_{i,j}, \quad |\delta_{i,j}| \leq u, \quad 1 \leq i \leq n, 1 \leq j \leq t,$$

where $\delta_{i,j}$ is the relative error caused by roundoff in the multiplication $\hat{\nu}_j T_{i,j}$. Step 3 is the inner product of two $nt \times 1$ vectors: $\text{vec } M \cdot \text{vec } \hat{R}$. The computed result satisfies the following bound on backward errors:

$$\hat{\psi} = \sum_{j=1}^t \sum_{i=1}^n (1 + \varepsilon_{i,j}) M_{i,j} \hat{R}_{i,j} = \sum_{j=1}^t \hat{\nu}_j \sum_{i=1}^n (1 + \varepsilon_{i,j})(1 + \delta_{i,j}) M_{i,j} T_{i,j},$$

where $\varepsilon_{i,j}$ is the backward error of the (i, j) entry that results from the computation of the inner product and satisfies $|\varepsilon_{i,j}| \leq \gamma_{nt}$. This error bound is independent of the order of summation; the bounds may be improved by a careful ordering of the terms [5, p. 63]. Since $|\delta_{i,j}| \leq u \leq \gamma_1$, Lemma 4.3 guarantees that $(1 + \varepsilon_{i,j})(1 + \delta_{i,j}) = (1 + \xi_{i,j})$, where $|\xi_{i,j}| \leq \gamma_{nt+1}$. To obtain (4.22), we require a perturbation ΔM satisfying

$$\sum_{i=1}^n (1 + \xi_{i,j}) M_{i,j} T_{i,j} = \sum_{i=1}^n (M + \Delta M)_{i,j} (T + \Delta T)_{i,j}, \quad 1 \leq j \leq t.$$

Recall that ΔT was fixed above when solving the system $(I - A_T)\tilde{\nu} = \tilde{\mu}$. Canceling the term $M_{i,j}T_{i,j}$ from the summation and regrouping,

$$(4.24) \quad \sum_{i=1}^n (\xi_{i,j} M_{i,j} T_{i,j} - M_{i,j} \Delta T_{i,j}) = \sum_{i=1}^n \Delta M_{i,j} (T_{i,j} + \Delta T_{i,j}), \quad 1 \leq j \leq t.$$

Let ξ_j be the j th column of the matrix $\xi = (\xi_{i,j})$. For each j , the left-hand side of (4.24) is the scalar quantity

$$b_j = (\xi_j \odot M_j) \cdot T_j - \Delta T_j \cdot M_j,$$

where $\xi_j \odot M_j$ is the Hadamard, or entrywise, product. The system (4.24) is equivalent to $(T_j + \Delta T_j) \cdot \Delta M_j = b_j$, which, for nonzero $T_j + \Delta T_j$, has as a solution

$$(4.25) \quad \Delta M_j = \frac{b_j}{\|T_j + \Delta T_j\|_2^2} (T_j + \Delta T_j).$$

Using our bound on ΔT , Cauchy–Schwarz guarantees

$$(4.26) \quad \begin{aligned} \|T_j + \Delta T_j\|_2^2 &= \|T_j\|_2^2 + 2T_j \cdot \Delta T_j + \|\Delta T_j\|_2^2 \geq \|T_j\|_2^2 - 2\|T_j\|_2 \|\Delta T_j\|_2 \\ &\geq \|T_j\|_2^2 - 4\sqrt{n}\tilde{\gamma}_{n^2}\|T_j\|_2^2 = (1 - 4\sqrt{n}\tilde{\gamma}_{n^2})\|T_j\|_2^2 > 0, \end{aligned}$$

under the assumption $1 - 4\sqrt{n}\tilde{\gamma}_{n^2} > 0$. Therefore, the computed $\hat{\psi}$ is the exact solution (4.20) for the inputs $M + \Delta M, T + \Delta T$, and μ . In (4.23) we gave bounds for ΔT . By Cauchy–Schwarz,

$$\begin{aligned}\|\Delta M_j\|_2 &= \frac{|b_j|\|T_j + \Delta T_j\|_2}{\|T_j + \Delta T_j\|_2^2} \leq \frac{|(\xi_j \odot M_j) \cdot T_j| + |M_j \cdot \Delta T_j|}{\|T_j + \Delta T_j\|_2} \\ &\leq \frac{\gamma_{nt+1}\|M_j\|_2\|T_j\|_2 + \|M_j\|_2\|\Delta T_j\|_2}{\|T_j + \Delta T_j\|_2} \leq \frac{\gamma_{nt+1} + 2\sqrt{n}\tilde{\gamma}_{n^2}}{\sqrt{1 - 4\sqrt{n}\tilde{\gamma}_{n^2}}} \|M_j\|_2\end{aligned}$$

by (4.23) and (4.26) and the observation $\|T_j\|_2 \leq \|T_j\|_1 = 1$, since T_j is stochastic. Finally, the bounds $t \leq n - 1$ and $n \geq 1$ imply that $nt + 1 \leq n^2$, so $\gamma_{nt+1} \leq \gamma_{n^2} \leq \tilde{\gamma}_{n^2}$, and we obtain (4.22). $\square$

Remark. For fixed n , (4.22) simplifies to

$$\frac{(1 + 2\sqrt{n})\tilde{\gamma}_{n^2}}{\sqrt{1 - 4\sqrt{n}\tilde{\gamma}_{n^2}}} \leq \frac{4\sqrt{n}\tilde{\gamma}_{n^2}}{1 - 4\sqrt{n}\tilde{\gamma}_{n^2}} = O(\sqrt{n}\tilde{\gamma}_{n^2}),$$

as $u \rightarrow 0$. The quantity ΔM_j obtained in (4.25) is the solution to the optimization problem

$$\begin{array}{ll}\text{minimize} & \|\Delta M_j\|_2 \\ \text{subject to} & (T_j + \Delta T_j) \cdot \Delta M_j = b_j.\end{array}$$

5. Simulations. We conducted a numerical study by computing expectations and comparing them to a Monte Carlo simulation. We used the game Chutes and Ladders (or Snakes and Ladders), which is characterized by a substantial number of states (82) and exhibits a gradual drift toward the absorbing state combined with occasional large jumps. Furthermore, this game is a good illustration of composite Markov chains as discussed in section 3.4. The MATLAB script used for computing expectations and the code for the simulations can be found in [7]. We simulated the following events in 100 million games and determined the sample mean for each. The results are summarized in Table 5.1.

- Second-to-Last Square: This is the number of times that a player gets “stuck” on the second-to-last square for spinning a number larger than 1.
- Large Ladder Traversal: The number of times a player traverses the largest ladder from square 28 to square 84.
- 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 as discussed in section 3.4.
- First-Player Advantage: This is the indicator event for the first player winning when both players finish on the same turn. In expectation, it is the probability that the first player wins by virtue of being the first player.
- First-Player Win Frequency: This is the indicator event for the first player winning. In expectation, this is the probability that the first player wins.

As can be seen in Table 5.1, the results agree up to at least three significant digits in every case. The execution time for computing expectations, shown in the last column of the table, indicates that even moderately large problems can feasibly be solved using this approach; the two-player Chutes and Ladders matrix has over 6500 rows. Parallelization would permit much larger problems; however, we expect that for large n , simulation will be faster, just as Monte Carlo integration is more efficient than quadrature for high-dimensional problems.

TABLE 5.1
Comparison of Monte Carlo simulations with computed expectations.

Event	Sample mean	Computed expectation	Computation time (sec)
Single-Player Events			
Setup			1.8(-3)
Second-to-Last Square	1.2954	1.2958	1.3(-4)
Large Ladder	0.5895	0.5896	1.0(-4)
Game Length	39.596	39.598	2.9(-4)
Two-Player Events			
Setup			2.5
Second-to-Last Square	1.1159	1.1166	8.1(-3)
Large Ladder	0.8181	0.8180	3.2(-2)
Game Length	26.513	26.513	3.1
Lead Changes	3.9679	3.9679	3.4
First-Player Advantage	0.0156	0.0156	6.2(-3)
First-Player Wins	0.5078	0.5078	1.4(-1)

Acknowledgments. Special thanks to C. Shane Reese for his useful conversations. We also thank our referees for their excellent comments and suggestions.

REFERENCES

- [1] S. L. CAMPBELL AND C. D. MEYER, JR., *Generalized Inverses of Linear Transformations*, Dover Publications, New York, 1991.
- [2] J. W. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [3] R. DURRETT, *Probability: Theory and Examples*, 2nd ed., Duxbury Press, Belmont, CA, 1996.
- [4] A. GRAHAM, *Kronecker Products and Matrix Calculus: With Applications*, Ellis Horwood Series in Mathematics and Its Applications, Ellis Horwood, Chichester, UK, 1981.
- [5] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.
- [6] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.
- [7] J. HUMPHERYS, *Markov Simulation Code*, <http://math.byu.edu/~jeffh/mathematics/games/chutes/index.htm>, 2008.
- [8] J. G. KEMENY AND J. L. SNELL, *Finite Markov Chains*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976.
- [9] A. N. LANGVILLE AND C. D. MEYER, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.
- [10] A. J. LAUB, *Matrix Analysis for Scientists & Engineers*, SIAM, Philadelphia, 2005.
- [11] C. MEYER, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
- [12] C. D. MEYER, JR., *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17 (1975), pp. 443–464.
- [13] L. N. TREFETHEN AND D. BAU, III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.