Finite Markov Chains

This note gives a brief introduction to the theory of finite Markov chains, omitting some proofs. This topic has many applications and will be revisited in several higher-level courses.

1 Introduction

Markov chains are models of random motion in a finite or countable set. These models are powerful because they capture a vast array of systems that we encounter in applications. Yet, the models are simple in that many of their properties can often be determined using elementary matrix algebra. In this course, we limit the discussion to the case of finite Markov chains, i.e., motions in a finite set.

Imagine the following scenario. You flip a fair coin until you get two consecutive 'heads'. How many times do you have to flip the coin, on average? You roll a balanced six-sided die until the sum of the last two rolls is 8. How many times do you have to roll the die, on average?

As another example, say that you play a game of 'heads or tails' using a biased coin that yields 'heads' with probability 0.48. You start with \$10. At each step, if the flip yields 'heads', you earn \$1. Otherwise, you lose \$1. What is the probability that you reach \$100 before \$0? How long does it take until you reach either \$100 or \$0?

You try to go up a ladder that has 20 rungs. At each time step, you succeed in going up by one rung with probability 0.9. Otherwise, you fall back to the ground. How many time steps does it take you to reach the top of the ladder, on average?

You look at a web page, then select randomly one of the links on that page, with equal probabilities. You then repeat on the next page you visit, and so on. As you keep browsing the web in this way, what fraction of the time do you open a given page? How long does it take until you reach a particular page? How likely is it that you visit a given page before another given page?

These questions can all be answered using the methods of Markov chains, as we explain in this note.

2 A First Example

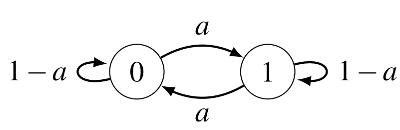

Figure 1: A simple Markov chain

Figure 1 illustrates a simple Markov chain. It describes a random motion in the set $\{0,1\}$. The position at time $n = 0,1,2,\ldots$ is $X_n \in \{0,1\}$. We call X_n the *state* of the Markov chain at step (or time) n. The set

$\{0,1\}$ is the *state space*, i.e., the set of possible values of the state. The motion, i.e., the time evolution, of X_n follows the following rules. One is given a number $a \in [0,1]$ and two nonnegative numbers $\pi_0(0)$ and $\pi_0(1)$ that add up to 1. Then,

$$\mathbb{P}[X_0 = 0] = \pi_0(0) \quad \text{and} \quad \mathbb{P}[X_0 = 1] = \pi_0(1). \tag{1}$$

That is, the *initial* state X_0 is equal to 0 with probability $\pi_0(0)$, otherwise it is 1. Then for $n \geq 0$,

$$\mathbb{P}[X_{n+1} = 0 \mid X_n = 0, X_{n-1}, \ldots, X_0] = 1 - a \tag{2}$$
$$\mathbb{P}[X_{n+1} = 1 \mid X_n = 0, X_{n-1}, \ldots, X_0] = a \tag{3}$$
$$\mathbb{P}[X_{n+1} = 0 \mid X_n = 1, X_{n-1}, \ldots, X_0] = a \tag{4}$$
$$\mathbb{P}[X_{n+1} = 1 \mid X_n = 1, X_{n-1}, \ldots, X_0] = 1 - a \tag{5}$$

Figure 1 summarizes rules (2) to (5). These rules specify the *transition probabilities* of the Markov chain. Rules (2) to (3) specify that if the Markov chain is in state 0 at step n, then at the next step it stays in state 0 with probability $1-a$ and it moves to state 1 with probability a, regardless of what happened in the previous steps. Rules (4) to (5) are similar. Figure 1 is called the *state transition diagram* of the Markov chain. It captures the transition probabilities in a graphical form.

The key property of a Markov chain is that it is *amnesic*: at step n, it forgets what it did before getting to the current state and its future steps depend only on that current state. Here is one way to think of the rules of motion. When the Markov chain gets to state 0, it flips a coin with heads probability a. If the outcome is H then it goes to state 1; otherwise, it stays in state 0 and flips the coin again. The situation is similar when the Markov chain gets to state 1.

An equivalent, more compact way of describing a Markov chain is its *transition probability matrix* P, which for the above chain is given by

$$\begin{aligned} P(0,0) &= 1-a; \qquad & P(0,1) &= a; \\ P(1,0) &= a; \qquad & P(1,1) &= 1-a. \end{aligned}$$

That is,

$$P = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}.$$

Hence,

$$\mathbb{P}[X_{n+1} = j \mid X_n = i,\ X_{n-1}, \ldots,\ X_0] = P(i,j), \quad \text{for } n \geq 0 \text{ and } i,j \in \{0,1\}.$$

Figure 2 shows some simulations of the Markov chain with different values of a. When $a = 0.1$, it is unlikely that the state of the Markov chain changes in one step. As the figure shows, the Markov chain spends many steps in one state before switching. For larger values of a, the state of the Markov chain changes more frequently. Note that, by symmetry, over the long term the Markov chain spends half of the time in each state. (More about this phenomenon later.)

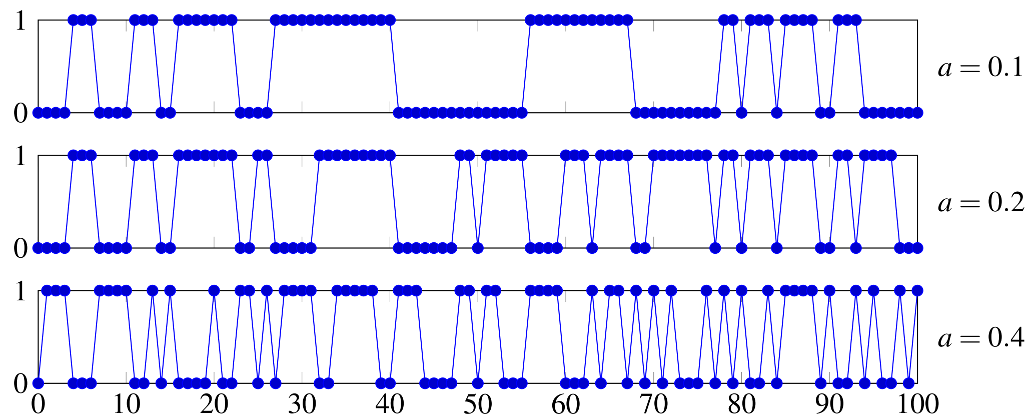

Figure 2: Simulations of the two-state Markov chain

3 A Second Example

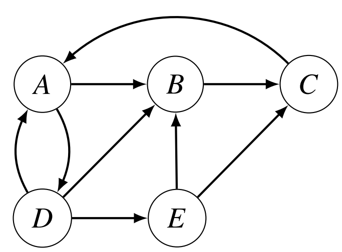

Figure 3: A five-state Markov chain. The outgoing arrows are equally likely.

Figure 3 shows the state transition diagram of a small web browsing experiment. Each state in the figure represents a web page. The arrows out of a state correspond to links on the page that point to other pages. The transition probabilities are not indicated on the figure, but the model is that each outgoing link from a particular page is equally likely. The figure corresponds to the following probability transition matrix, with rows/columns indexed by states A, B, C, D, E, respectively:

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

Figure 4 shows a simulation of the five-state Markov chain.

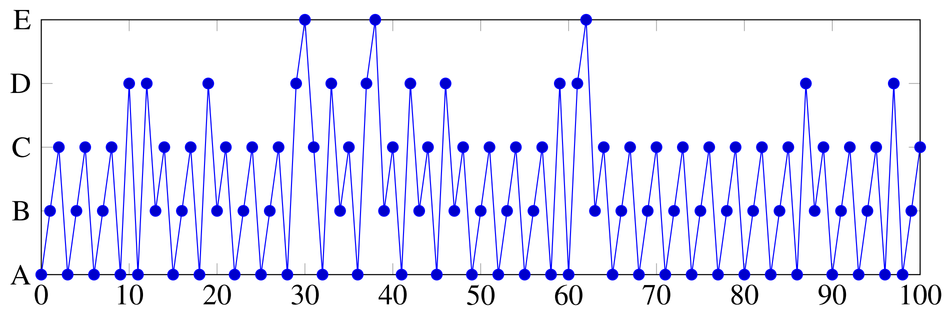

Figure 4: Simulation of the five-state Markov chain.

4 Finite Markov Chains

We define a general finite Markov chain as follows. The *state space* is $\mathscr{X} = \{1, 2, \ldots, K\}$ for some finite K. The *transition probability matrix* P is a $K \times K$ matrix such that

$$P(i, j) \geq 0, \quad \forall i, j \in \mathscr{X}$$

and

$$\sum_{j=1}^{K} P(i, j) = 1, \quad \forall i \in \mathscr{X}.$$

The *initial distribution* is a vector $\pi_0 = \{\pi_0(i) \mid i \in \mathscr{X}\}$ where $\pi_0(i) \geq 0$ for all $i \in \mathscr{X}$ and $\sum_{i \in \mathscr{X}} \pi_0(i) = 1$. (In many applications, π_0 is concentrated on a single state, i.e., $\pi_0(i) = 1$ for some i, which corresponds to starting deterministically in state i.)

One then defines the random sequence $\{X_n \mid n = 0, 1, 2, \ldots\}$ by

$$\begin{aligned} \mathbb{P}[X_0 = i] &= \pi_0(i), \quad \forall i \in \mathscr{X}; \\ \mathbb{P}[X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0] &= P(i, j), \quad \forall n \geq 0, \forall i, j \in \mathscr{X}. \end{aligned}$$

Note that

$$\begin{aligned} &\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\ &= \mathbb{P}[X_0 = i_0]\,\mathbb{P}[X_1 = i_1 \mid X_0 = i_0]\,\mathbb{P}[X_2 = i_2 \mid X_0 = i_0, X_1 = i_1] \cdots \mathbb{P}[X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}] \\ &= \pi_0(i_0) P(i_0, i_1) \cdots P(i_{n-1}, i_n). \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}[X_n = i_n] &= \sum_{i_0, \ldots i_{n-1}} \mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\ &= \sum_{i_0, \ldots i_{n-1}} \pi_0(i_0) P(i_0, i_1) \cdots P(i_{n-1}, i_n) \\ &= [\pi_0 P^n](i_n), \end{aligned}$$

where the last expression denotes the i_n component of the product of the row vector π_0 times the nth power of the matrix P.

Thus, if we designate by π_n the distribution of X_n, so that $\mathbb{P}[X_n = i] = \pi_n(i)$, then the last derivation proves the following result.

Theorem 25.1. *For all $n \geq 0$, one has*

$$\pi_n = \pi_0 P^n.$$

In particular, if $\pi_0(i) = 1$ for some i, then $\pi_n(j) = P^n(i,j) = \mathbb{P}[X_n = j \mid X_0 = i]$.

For the two-state Markov chain in Figure 1, one can verify that (see Appendix for details)

$$P^n = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}^n = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(1-2a)^n & \frac{1}{2} - \frac{1}{2}(1-2a)^n \\ \frac{1}{2} - \frac{1}{2}(1-2a)^n & \frac{1}{2} + \frac{1}{2}(1-2a)^n \end{bmatrix}. \tag{6}$$

Note that if $0 < a < 1$,

$$P^n \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad \text{as } n \to \infty.$$

Consequently, for $0 < a < 1$, one has $\pi_n = \pi_0 P^n \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ as $n \to \infty$, which means that after a long time the chain is equally likely to be in either of the two states.

5 Invariant Distribution

The following definition introduces the important notion of invariant distribution.

Definition 25.1. *A distribution π is* invariant *for the transition probability matrix P if it satisfies the following* balance equations*:*

$$\pi = \pi P. \tag{7}$$

The relevance of this definition is stated in the next result.

Theorem 25.2. *One has $\pi_n = \pi_0$ for all $n \geq 0$ if and only if π_0 is invariant.*

Proof. If $\pi_n = \pi_0$ for all $n \geq 0$, then $\pi_0 = \pi_1 = \pi_0 P$, so that π_0 satisfies (7) and is thus invariant.

If $\pi_0 P = \pi_0$, then $\pi_1 = \pi_0 P = \pi_0$. And similarly, by induction, we get that $\pi_n = \pi_{n-1} P = \pi_0 P = \pi_0$. □

For instance, in the case of the two-state Markov chain, the balance equations are

$$\pi(0) = \pi(0)(1-a) + \pi(1)a;$$
$$\pi(1) = \pi(0)a + \pi(1)(1-a).$$

Each of these two equations is equivalent to

$$\pi(0) = \pi(1).$$

Thus, the two equations are redundant. If we add the condition that the components of π add up to one, we find that the only solution is $\begin{bmatrix} \pi(0) & \pi(1) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, which is not surprising in view of symmetry.

For the five-state Markov chain, the balance equations are

$$\begin{bmatrix} \pi(A) & \pi(B) & \pi(C) & \pi(D) & \pi(E) \end{bmatrix} = \begin{bmatrix} \pi(A) & \pi(B) & \pi(C) & \pi(D) & \pi(E) \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Once again, these five equations in the five unknowns are redundant: they do not determine π uniquely. However, if we add the condition that the components of π add up to one, then we find that the solution is unique and given by (see the Appendix for the calculations):

$$\begin{bmatrix}\pi(A) & \pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix} = \frac{1}{39}\begin{bmatrix}12 & 9 & 10 & 6 & 2\end{bmatrix}. \tag{8}$$

Thus, in this web-browsing example, in the long term page A is visited most often, then page C, then page B. A Google search would return the pages in order of most frequent visits, i.e., in the order A, C, B, D, E. This ranking of the pages is called *PageRank* and can be determined by solving the balance equations. (In fact, the actual ranking by Google combines the estimate of π with other factors.)

How many invariant distributions does a Markov chain have? We have seen that for the two examples, the answer was one. However, that is not always the case. For instance, consider the two-state Markov chain with $a = 0$ instead of $0 < a < 1$ as we assumed previously. This Markov chain does not change state. Its transition probability matrix is $P = I$ where I denotes the identity matrix. Since $\pi I = \pi$ for any vector π, we see that *any* distribution is invariant for this Markov chain.

However, we will see in the next section that two simple conditions guarantee the uniqueness of the invariant distribution.

6 Hitting Time

Consider the Markov chain in Figure 3. Assume it starts in state A. What is the average number of steps until it reaches state E for the first time? To calculate this average time, for $i \in \{A,B,C,D,E\}$ define $\beta(i)$ to be the average time until the Markov chain reaches state E given that it starts from state i.

Thus, $\beta(E) = 0$ since it takes 0 steps to reach E when starting in state E. We want to calculate $\beta(A)$. However, it turns out that to calculate $\beta(A)$, one also has to calculate $\beta(B), \ldots, \beta(D)$. We do this by finding equations that these quantities satisfy and then solving these equations.

We claim that

$$\beta(A) = 1 + \frac{1}{2}\beta(B) + \frac{1}{2}\beta(D). \tag{9}$$

To see this, note that when the Markov chain starts in state A, it stays there for one step. Then, with probability $\frac{1}{2}$ it moves to state B. In that case, the average time until it reaches E is $\beta(B)$. With probability $1/2$, the Markov chain moves to state D and then takes $\beta(D)$ steps (on average) to reach E. Thus, the time to reach E starting from state A is 1 step plus an average of $\beta(B)$ steps with probability $\frac{1}{2}$ and an average of $\beta(D)$ steps with probability $\frac{1}{2}$. Equation (9) captures this decomposition.

An identity similar to (9) can be written for every starting state. We find

$$\begin{aligned}
\beta(A) &= 1 + \frac{1}{2}\beta(B) + \frac{1}{2}\beta(D) \\
\beta(B) &= 1 + \beta(C) \\
\beta(C) &= 1 + \beta(A) \\
\beta(D) &= 1 + \frac{1}{3}\beta(A) + \frac{1}{3}\beta(B) + \frac{1}{3}\beta(E) \\
\beta(E) &= 0
\end{aligned}$$

These equations are called the *first step equations*.

Solving these equations, we find (see the Appendix for the calculations):

$$\beta(A) = 17 \qquad \beta(B) = 19 \qquad \beta(C) = 18 \qquad \beta(D) = 13 \qquad \beta(E) = 0 \tag{10}$$

Let us now consider a general finite Markov chain with transition probability matrix P on the state space $\mathscr{X}$. Let $A \subset \mathscr{X}$ be a set of states. For each $i \in \mathscr{X}$, let $\beta(i)$ be the average number of steps until the Markov chain enters one of the states in A, given that it starts in state i.

Then one has the first step equations

$$\begin{aligned}
\beta(i) &= 0 \quad \text{if } i \in A \\
\beta(i) &= 1 + \sum_{j \in \mathscr{X}} P(i,j)\beta(j) \quad \text{otherwise.}
\end{aligned}$$

As another example, consider the Markov chain in Figure 1. Let $\beta(i)$ be the average number of steps until the Markov chain enters state 1, starting in state i. The first step equations are

$$\begin{aligned}
\beta(0) &= 1 + (1-a)\beta(0) + a\beta(1) \\
\beta(1) &= 0
\end{aligned}$$

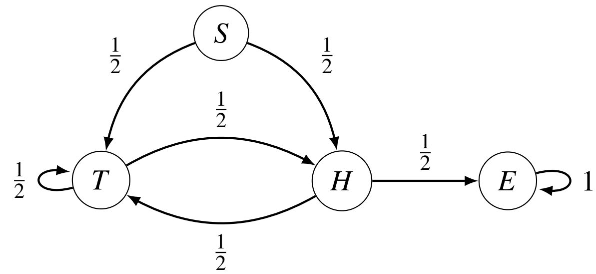

Figure 5: Flipping a fair coin until two heads in row.

Solving, we find $\beta(0) = \frac{1}{a}$. Note that the time to enter state 1 starting from state 0 is the number of times one has to flip a biased coin with $\mathbb{P}[H] = a$ until the first heads. This number of steps has a geometric distribution with parameter a. Thus, we have rediscovered the fact that the mean value of a Geometric(a) random variable is $\frac{1}{a}$.

Now suppose you flip a fair coin repeatedly until you get two heads in a row. How many times do you have to flip the coin, on average? Figure 5 shows a state transition diagram that corresponds to this situation. The Markov chain starts in state S. The state is H or T if the last coin flip was H or T, respectively, except that the state is E if the last two flips were heads. This state E is *absorbing*, i.e., $P(E,E) = 1$ (the chain never leaves it). For $i \in \{S,T,H,E\}$, let $\beta(i)$ be the average number of steps until the Markov chain enters state E. The first step equations are

$$\begin{aligned}
\beta(S) &= 1 + \frac{1}{2}\beta(T) + \frac{1}{2}\beta(H)\\
\beta(T) &= 1 + \frac{1}{2}\beta(T) + \frac{1}{2}\beta(H)\\
\beta(H) &= 1 + \frac{1}{2}\beta(T) + \frac{1}{2}\beta(E)\\
\beta(E) &= 0
\end{aligned}$$

Solving, we find

$$\beta(S) = 6 \tag{11}$$

(See the Appendix for the calculations.)

Consider now the 20-rung ladder. A man starts on the ground. At each step, he moves up one rung with probability p and falls back to the ground otherwise. Let $\beta(i)$ be the average number of steps needed to reach the top rung, starting from rung $i \in \{0,1,\ldots,20\}$ where rung 0 refers to the ground. The first step equations are

$$\begin{aligned}
\beta(i) &= 1 + (1-p)\beta(0) + p\beta(i+1), \quad i = 0,\ldots,19\\
\beta(20) &= 0
\end{aligned}$$

Solving, we find

$$\beta(0) = \frac{p^{-20} - 1}{1-p} \tag{12}$$

(See the Appendix for the calculations.) For instance, if $p = 0.9$, then $\beta(0) \approx 72$. Also, if $p = 0.8$, then $\beta(0) \approx 429$. The moral of the story is that you should be careful on a ladder!

7 Probability of A before B

Let X_n be a finite Markov chain with state space $\mathscr{X}$ and transition probability matrix P. Let also A and B be two disjoint subsets of $\mathscr{X}$. We want to determine the probability $\alpha(i)$ that, starting in state i, the Markov chain enters one of the states in A before one of the states in B.

The first step equations for $\alpha(i)$ are

$$\begin{aligned} \alpha(i) &= \sum_j P(i,j)\alpha(j), \quad \forall i \notin A \cup B \\ \alpha(i) &= 1, \quad \forall i \in A \\ \alpha(i) &= 0, \quad \forall i \in B \end{aligned}$$

To see why the first set of equations hold, we observe that the event that the Markov chain enters A before B starting from i is partitioned into the events that it does so by first moving to state j, for all possible value of j. Now, the probability that it enters A before B starting from i after moving first to j is the probability that it enters A before B starting from j, because the Markov chain is amnesic. The second and third sets of equations are obvious.

As an illustration, suppose we play a game of heads-or-tails with a coin such that $\mathbb{P}[H] = p$. For every heads, your fortune increases by 1 and for every tails, it decreases by 1. Your initial fortune is n. You stop playing when either you go bankrupt (your fortune reaches zero), or your fortune reaches some target value $M > n$. We want to calculate the probability that your fortune reaches M before you go bankrupt. Call this probability $\alpha(n)$. As usual, we will set up a system of equations to compute $\alpha(n)$ for all n.

The first step equations are

$$\begin{aligned} \alpha(n) &= (1-p)\alpha(n-1) + p\alpha(n+1), \quad 0 < n < M \\ \alpha(M) &= 1 \\ \alpha(0) &= 0 \end{aligned}$$

Solving these equations, we find

$$\alpha(n) = \frac{1-\rho^n}{1-\rho^M} \tag{13}$$

where $\rho := (1-p)p^{-1}$. (See the Appendix for the calculations.) Note that $\rho < 1$ under the reasonable assumption that $p < 0.5$ (so that the casino has an advantage). For instance, with $p = 0.48$ and $M = 100$, we find that $\alpha(10) \approx 4 \times 10^{-4}$, which is sobering when contemplating a trip to Las Vegas. Note that for each gambler who plays this game, the Casino makes \$10.00 with probability $1 - 4 \times 10^{-4}$ and loses \$990.00 with probability 4×10^{-4}, so that the expected gain of the Casino per gambler is approximately $(1 - 4 \times 10^{-4}) \times \$10.00 - 4 \times 10^{-4} \times \$990.00 \approx \$9.60$. Observe that the probability of winning in one step is 48%, so that if the gambler did bet everything on a single game and stopped after one step, the Casino would only make $0.52 \times \$10.00 - 0.48 \times \$10.00 = \$0.40$ on average per gambler, instead of \$9.60. (Of course, if $p > 0.5$ then $\rho > 1$ and you as the gambler have the advantage; similar conclusions then hold with the roles of you and the casino reversed.)

Appendix: Calculations

This section presents the details of the calculations of this note. The actual calculations are not very important and are included here for completeness.

Equation (6)

By symmetry, we can write

$$P^n = \begin{bmatrix} 1-\alpha_n & \alpha_n \\ \alpha_n & 1-\alpha_n \end{bmatrix}$$

for some α_n that we determine below. Note that $\alpha_1 = a$. Also,

$$P^{n+1} = \begin{bmatrix} 1-\alpha_{n+1} & \alpha_{n+1} \\ \alpha_{n+1} & 1-\alpha_{n+1} \end{bmatrix} = PP^n = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix} \begin{bmatrix} 1-\alpha_n & \alpha_n \\ \alpha_n & 1-\alpha_n \end{bmatrix}.$$

Consequently, by looking at component $(0,1)$ of this product,

$$\alpha_{n+1} = (1-a)\alpha_n + a(1-\alpha_n) = a + (1-2a)\alpha_n.$$

Let us try a solution of the form $\alpha_n = b + c\lambda^n$. We need

$$\begin{aligned} \alpha_{n+1} &= b + c\lambda^{n+1} \\ &= a + (1-2a)\alpha_n \\ &= a + (1-2a)(b + c\lambda^n) \\ &= a + (1-2a)b + (1-2a)c\lambda^n. \end{aligned}$$

Matching the terms, we see that this identity holds if

$$b = a + (1-2a)b \quad \text{and} \quad \lambda = 1-2a.$$

The first equation gives $b = \frac{1}{2}$. Hence, $\alpha_n = \frac{1}{2} + c(1-2a)^n$. To find c, we use the fact that $\alpha_1 = a$, so that $\frac{1}{2} + c(1-2a) = a$, which yields $c = -\frac{1}{2}$.

Hence,

$$\alpha_n = \frac{1}{2} - \frac{1}{2}(1-2a)^n.$$

Equation (8)

The balance equations are $\pi = \pi P$.

We know that the equations do not determine π uniquely. Let us choose arbitrarily $\pi(A) = 1$. We then solve for the other components of π and we renormalize later. We can ignore any equation we choose. Let us ignore the first one. The new equations are

$$\begin{bmatrix} \pi(B) & \pi(C) & \pi(D) & \pi(E) \end{bmatrix} = \begin{bmatrix} 1 & \pi(B) & \pi(C) & \pi(D) & \pi(E) \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Equivalently,

$$\begin{bmatrix}\pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2} & 0\end{bmatrix} + \begin{bmatrix}\pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix}\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ \frac{1}{3} & 0 & 0 & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{bmatrix}.$$

By inspection, we see that $\pi(D) = \frac{1}{2}$, then $\pi(E) = \frac{1}{3}\pi(D) = \frac{1}{6}$, then

$$\pi(B) = \frac{1}{2} + \frac{1}{3}\pi(D) + \frac{1}{2}\pi(E) = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} = \frac{3}{4}.$$

Finally,

$$\pi(C) = \pi(B) + \frac{1}{2}\pi(E) = \frac{3}{4} + \frac{1}{12} = \frac{5}{6}.$$

The components $\pi(A) + \cdots + \pi(E)$ add up to $1 + \frac{3}{4} + \frac{5}{6} + \frac{1}{2} + \frac{1}{6} = \frac{39}{12}$. To normalize, we multiply each component by $\frac{12}{39}$ and we get

$$\pi = \begin{bmatrix}\frac{12}{39} & \frac{9}{39} & \frac{10}{39} & \frac{6}{39} & \frac{2}{39}\end{bmatrix}.$$

We could have proceeded differently and observed that our identity implies that

$$\begin{bmatrix}\pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix}\left(I - \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ \frac{1}{3} & 0 & 0 & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{bmatrix}\right) = \begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2} & 0\end{bmatrix}$$

$$\begin{bmatrix}\pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix}\begin{bmatrix}1 & -1 & 0 & 0\\ 0 & 1 & 0 & 0\\ -1/3 & 0 & 1 & -1/3\\ -1/2 & -1/2 & 0 & 1\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2} & 0\end{bmatrix}$$

Hence,

$$\begin{bmatrix}\pi(B) & \pi(C) & \pi(D) & \pi(E)\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2} & 0\end{bmatrix}\begin{bmatrix}1 & -1 & 0 & 0\\ 0 & 1 & 0 & 0\\ -\frac{1}{2} & 0 & 1 & -\frac{1}{3}\\ -\frac{1}{2} & -\frac{1}{2} & 0 & 1\end{bmatrix}^{-1}.$$

This procedure is a systematic way to solve the balance equations by computer.

Equation (10)

Using the third equation in the second, we find $\beta(B) = 2 + \beta(A)$. The fourth equation then gives

$$\beta(D) = 1 + \frac{1}{3}\beta(A) + \frac{1}{3}(2 + \beta(A)) = \frac{5}{3} + \frac{2}{3}\beta(A).$$

The first equation then gives

$$\beta(A) = 1 + \frac{1}{2}(2 + \beta(A)) + \frac{1}{2}\left(\frac{5}{3} + \frac{2}{3}\beta(A)\right) = \frac{17}{6} + \frac{5}{6}\beta(A).$$

Hence, $\frac{1}{6}\beta(A) = \frac{17}{6}$, so that $\beta(A) = 17$. Consequently, $\beta(B) = 19$ and $\beta(D) = \frac{5}{3} + \frac{34}{3} = 13$. Finally, $\beta(C) = 18$.

Equation (11)

The last two equations give $\beta(H) = 1 + \frac{1}{2}\beta(T)$. If we substitute this expression in the second equation, we get

$$\beta(T) = 1 + \frac{1}{2}\beta(T) + \frac{1}{2}\left(1 + \frac{1}{2}\beta(T)\right) = \frac{3}{2} + \frac{3}{4}\beta(T).$$

Hence, $\beta(T) = 6$. Consequently, $\beta(H) = 1 + \frac{1}{2}\cdot 6 = 4$. Finally, $\beta(S) = 1 + \frac{1}{2}\cdot 6 + \frac{1}{2}\cdot 4 = 6$.

Equation (12)

Let us look for a solution of the form $\beta(i) = a + b\lambda^i$. Then

$$a + b\lambda^i = 1 + (1-p)(a+b) + p[a + b\lambda^{i+1}] = 1 + (1-p)(a+b) + pa + bp\lambda^{i+1}.$$

This identity holds if

$$a = 1 + (1-p)(a+b) + pa \quad \text{and} \quad \lambda = p^{-1},$$

i.e.,

$$b = -(1-p)^{-1} \quad \text{and} \quad \lambda = p^{-1}.$$

Then,

$$\beta(i) = a - (1-p)^{-1}p^{-i}.$$

Since $\beta(20) = 0$, we need

$$0 = a - (1-p)^{-1}p^{-20},$$

so that $a = (1-p)^{-1}p^{-20}$ and

$$\beta(i) = (1-p)^{-1}p^{-20} - (1-p)^{-1}p^{-i} = \frac{p^{-20} - p^{-i}}{1-p}.$$

Equation (13)

We again look for a solution of the form $\alpha(n) = a\lambda^n + b$ for suitable constants a, b, λ. Plugging this into the first step equations gives

$$a\lambda^n + b = (1-p)(a\lambda^{n-1} + b) + p(a\lambda^{n+1} + b),$$

which simplifies to

$$\lambda^n = (1-p)\lambda^{n-1} + p\lambda^{n+1}.$$

Canceling a factor of λ^{n-1} and rearranging gives the quadratic equation

$$p\lambda^2 - \lambda + (1-p) = 0,$$

whose solutions are

$$\lambda = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} = \frac{1 \pm (1-2p)}{2p}.$$

One solution is $\lambda = 1$, which is not interesting, so we take $\lambda = \frac{1-p}{p} =: \rho$. To compute a and b, we use the boundary conditions $\alpha(M) = 1$ and $\alpha(0) = 0$, which become

$$\begin{aligned} a\rho^M + b &= 1 \\ a + b &= 0. \end{aligned}$$

Solving gives us $a = -\frac{1}{1-\rho^M}$ and $b = \frac{1}{1-\rho^M}$. This yields our final answer

$$\alpha(n) = \frac{1-\rho^n}{1-\rho^M}.$$