Chapter 5

Markov Chains

A *Markov chain* is a discrete-time Markov process. In particular, if $X(0), X(1), \ldots$ is a sequence of random variables, then this sequence is a Markov chain if

$$\begin{aligned} P(X(t+1) = i_{t+1} | X(t) = i_t, X(t-1) = i_{t-1}, \ldots, X(0) = i_0) \\ = P(X(t+1) = i_{t+1} | X(t) = i_t) \end{aligned}$$

for all nonnegative integers t and all values $i_{t+1}, i_t, \ldots, i_0$ assumed by $X(t+1), X(t), \ldots, X(0)$, respectively.

We consider only *finite-state* Markov chains, i.e., the values assumed by the random variables form a finite set, $\mathcal{S}$, which we take to be $\{1, 2 \ldots, N\}$.

Typically, t denotes time. Then, in a Markov chain, given the present, the future is independent of the past.

The elements of $\mathcal{S}$ are called *states*. If $X(t) = i$, then we say that the state at time t is i. The random variable $X(0)$ is called the *initial state* of the Markov chain.

The Markov chain is *homogeneous* if $P(X(t+1) = j | X(t) = i)$ does not depend on t. We assume that this is the case in the following. Let $p_{i,j} = P(X(t+1) = j | X(t) = i)$ and

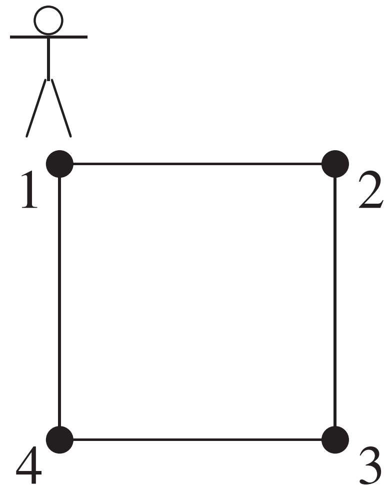
$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{pmatrix}.$$

The matrix $\mathbf{P}$ is called the *(one-step) transition probability matrix*.

The transition probability matrix is a *stochastic matrix*, i.e., all its elements are nonnegative and the sum of elements in any row is 1.

The transition probability matrix can be represented by a *state transition diagram* which is a directed graph with N vertices representing the possible states $1, 2, \ldots, N$ and where an oriented edge with the label $p_{i,j}$ exists from state i to state j whenever $p_{i,j} > 0$.

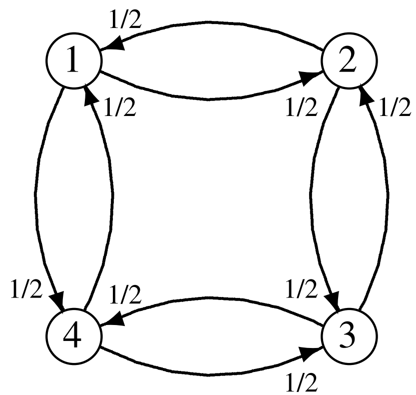
Example 1 A small town consists of four streets and four street corners, numbered 1, 2, 3, and 4.



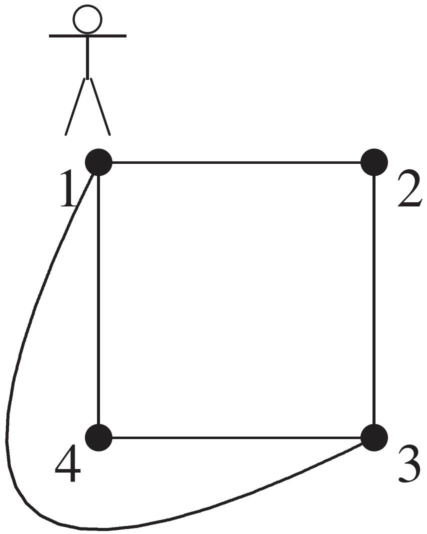
A walker stands at a corner and flips a fair coin. If it shows a head, he walks clockwise to the next corner and if it shows a tail, he walks counter clockwise to the next corner. This is repeated at each corner he arrives at. Let $X(t)$, for $t = 0, 1, \ldots$, be the corner at which the walker is at after the t^{th} movement.

Then the sequence $X(0), X(1), \ldots$ forms a Markov chain with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$ and state transition diagram



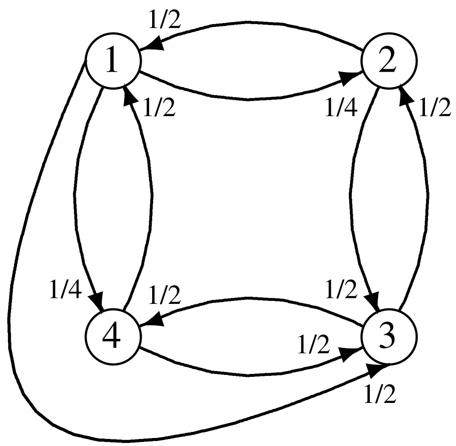
Example 2 Our walker discovered that there is a bus from corner 1 to corner 3. He decided to change his scheme only if he is at corner 1. He tosses the coin twice. If two heads show, he walks to corner 2. If two tails show, he walks to corner 4. Otherwise, if he gets two different faces in the two tosses, he takes the bus to corner 3. Again, let $X(t)$, for $t = 0, 1, \ldots$, be the corner at which the walker is at after the t^{th} movement.



Then the sequence $X(0), X(1), \ldots$ forms a Markov chain with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

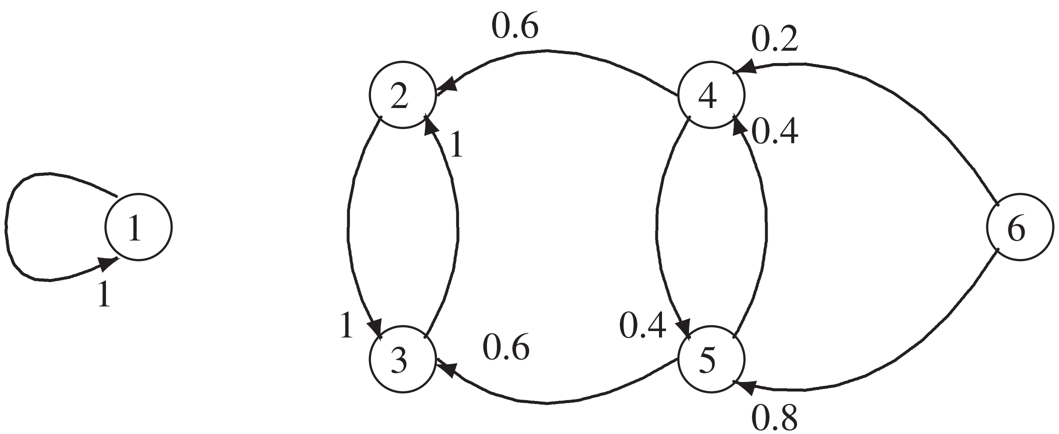
and state transition diagram



Example 3 The transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 \end{bmatrix}$$

can be represented by the state diagram



The m-Step Transition Probability We define the m-step transition probability $p_{i,j}^{(m)}$, for $m \geq 1$, to be the probability that $X(t+m) = j$ given that $X(t) = i$, i.e., $p_{i,j}^{(m)} = P(X(t+m) = j | X(t) = i)$. This probability does not depend on t.

In Example 3, let us compute $p_{4,2}^{(3)}$. We have

$$\begin{aligned} p_{4,2}^{(3)} &= P(X(3) = 2 | X(0) = 4) \\ &= \sum_{i_2 \in \{1,2,\ldots,6\}} P(X(3) = 2, X(2) = i_2 | X(0) = 4) \\ &= \sum_{i_2 \in \{1,2,\ldots,6\}} \Big(P(X(3) = 2 | X(2) = i_2, X(0) = 4) \\ &\qquad \times P(X(2) = i_2 | X(0) = 4) \Big) \end{aligned}$$

$$
\begin{aligned}
p_{4,2}^{(3)} &= \sum_{i_2 \in \{1,2,\ldots,6\}} \Big(P(X(3) = 2 | X(2) = i_2) \\
&\qquad \times P(X(2) = i_2 | X(0) = 4) \Big) \\
&= \sum_{i_2 \in \{1,2,\ldots,6\}} \Big(P(X(3) = 2 | X(2) = i_2) \\
&\qquad \times \sum_{i_1 \in \{1,2,\ldots,6\}} P(X(2) = i_2, X(1) = i_1 | X(0) = 4) \Big) \\
&= \sum_{i_2 \in \{1,2,\ldots,6\}} \Big(P(X(3) = 2 | X(2) = i_2) \\
&\qquad \times \sum_{i_1 \in \{1,2,\ldots,6\}} \big(P(X(2) = i_2 | X(1) = i_1, X(0) = 4) \\
&\qquad \times P(X(1) = i_1 | X(0) = 4) \big) \Big)
\end{aligned}
$$

$$
\begin{aligned}
p_{4,2}^{(3)} &= \sum_{i_2 \in \{1,2,\ldots,6\}} \Big(P(X(3) = 2 | X(2) = i_2) \\
&\qquad \times \sum_{i_1 \in \{1,2,\ldots,6\}} \big(P(X(2) = i_2 | X(1) = i_1) \\
&\qquad \times P(X(1) = i_1 | X(0) = 4) \big) \Big) \\
&= \sum_{i_1, i_2 \in \{1,2,\ldots,6\}} p_{4,i_1} p_{i_1,i_2} p_{i_2,2} \\
&= 0.6 \times 1 \times 1 + 0.4 \times 0.6 \times 1 + 0.4 \times 0.4 \times 0.6 \\
&= 0.936 = \frac{117}{125}.
\end{aligned}
$$

The m-*step transition probability matrix*[a], $\mathbf{P}^{(m)}$, is defined as

$$\mathbf{P}^{(m)} = \left[p_{i,j}^{(m)}\right]_{i,j\in\{1,2,\ldots,N\}}.$$

Notice that $\mathbf{P} = \mathbf{P}^{(1)}$.

The Chapman-Kolmogorov equations give a systematic way to compute $\mathbf{P}^{(m)}$ and, in particular, $p_{i,j}^{(m)}$ for all $i, j \in \{1, 2, \ldots, N\}$ and $m \geq 1$.

[a] In textbook, $p_{i,j}^{(m)}$ is denoted as $f_{ij}(m)$ and $\mathbf{P}^{(m)}$ as $\mathbf{F}(m)$.

The Chapman-Kolmogorov Equations:
$p_{i,j}^{(m)} = \sum_{k \in \mathcal{S}} p_{i,k}^{(r)} p_{k,j}^{(m-r)}$ for every $r = 1, 2, \ldots, m-1$. Indeed,

$$\begin{aligned} p_{i,j}^{(m)} &= P(X(t+m) = j | X(t) = i) \\ &= \sum_{k \in \mathcal{S}} P(X(t+m) = j | X(t+r) = k, X(t) = i) \\ &\qquad P(X(t+r) = k | X(t) = i) \\ &= \sum_{k \in \mathcal{S}} P(X(t+m) = j | X(t+r) = k) \\ &\qquad P([X(t+r) = k | X(t) = i) \\ &= \sum_{k \in \mathcal{S}} p_{i,k}^{(r)} p_{k,j}^{(m-r)}. \end{aligned}$$

The Chapman-Kolmogorov equations can be written in compact form as $\mathbf{P}^{(m)} = \mathbf{P}^m$.

In Example 3, the 3-step transition probability matrix is

$$\mathbf{P}^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 117/125 & 0 & 0 & 8/125 & 0 \\ 0 & 0 & 117/125 & 8/125 & 0 & 0 \\ 0 & 84/125 & 21/125 & 4/125 & 16/125 & 0 \end{bmatrix}.$$

State Distribution The PMF of $X(t)$ is represented by the vector $\boldsymbol{\pi}(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_N(t))$ in which $\pi_i(t) = P(X(t) = i)$. This vector is the *state distribution* of the Markov chain $X(t)$ at time t.

The state distribution $\boldsymbol{\pi}(0)$ of the initial state $X(0)$ is called the *initial distribution* of the chain.

We have

$$\boldsymbol{\pi}(t+1) = \boldsymbol{\pi}(t)\mathbf{P}.$$

Indeed, for $i \in \mathcal{S}$,

$$\begin{aligned}
\pi_j(t+1) &= P(X(t+1) = j) \\
&= \sum_{i \in \mathcal{S}} P(X(t+1) = j | X(t) = i) P(X(t) = i) \\
&= \sum_{i \in \mathcal{S}} p_{i,j} \pi_i(t).
\end{aligned}$$

It follows that for all $m \geq 1$,

$$\boldsymbol{\pi}(t+m) = \boldsymbol{\pi}(t)\mathbf{P}^m = \boldsymbol{\pi}(t)\mathbf{P}^{(m)}.$$

Stationary Distribution A distribution $\boldsymbol{\pi}$ is said to be *stationary* if $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

Stationary distributions can be obtained by solving $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ for the N elements $\pi_1, \pi_2, \ldots, \pi_N$ of $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$ with the condition that they are nonnegative and their sum is 1. The stationary distributions are left eigenvectors of the transition probability matrix $\mathbf{P}$ with eigenvalues equal to 1.

The Markov chain in Example 1 has one stationary distribution, namely, $(1/4, 1/4, 1/4, 1/4)$ and the Markov chain in Example 2 has one stationary distribution, namely, $(2/9, 2/9, 1/3, 2/9)$. The Markov chain in Example 3 has an infinite number of stationary distributions: $(p, (1-p)/2, (1-p)/2, 0, 0, 0)$ where $0 \leq p \leq 1$.

Theorem 1 *Every Markov chain has at least one stationary distribution.*

Proof. [a] Let $\mathbf{P} = [p_{i,j}]_{1 \le i,j \le N}$ be an $N \times N$ transition probability matrix of a Markov chain and $\mathbf{u} = (1, 1, \ldots, 1)^\mathsf{T}$, be the column vector of N 1's. Then, as the sum of elements in each row is 1, it follows that $\mathbf{Pu} = \mathbf{u}$. Let $\mathbf{I}$ be the $N \times N$ identity matrix. Then, we have $\mathbf{Pu} = \mathbf{Iu}$ or, equivalently, $(\mathbf{P} - \mathbf{I})\mathbf{u} = 0$. This implies that the columns of $\mathbf{P} - \mathbf{I}$ are linearly dependent, and so are the rows. Hence, there exists a nonzero row vector $\mathbf{v}$ such that $\mathbf{v}(\mathbf{P} - \mathbf{I}) = 0$ or, equivalently, $\mathbf{vP} = \mathbf{vI} = \mathbf{v}$.

[a] This theorem follows directly from Brouwer's fixed-point theorem as stated in Section 5.2 in textbook. Here we give a proof that does not resort to this theorem.

If all the elements in $\mathbf{v}$ are nonnegative, then we normalize it by dividing every element in $\mathbf{v}$ by the sum of its elements to produce a vector $\boldsymbol{\pi}$. Then, $\boldsymbol{\pi}$ is a probability distribution, i.e., its elements are nonnegative and sum to 1, satisfying $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Hence, $\boldsymbol{\pi}$ is a stationary distribution of the Markov chain.

The same can be done if all the elements in $\mathbf{v}$ are nonpositive as the normalization also results in a stationary distribution $\boldsymbol{\pi}$.

Now, suppose that $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ has an element or more which is positive and an element or more which is negative. Let $\mathbf{v}^{+} = (v_1^{+}, v_2^{+}, \ldots, v_n^{+})$ where $v_i^{+} = \max\{v_i, 0\}$ and $\mathbf{v}^{-} = (v_1^{-}, v_2^{-}, \ldots, v_n^{-})$ where $v_i^{-} = -\min\{v_i, 0\}$. Then $\mathbf{v}^{+}$ and $\mathbf{v}^{-}$ are nonzero vectors with nonnegative elements and $\mathbf{v} = \mathbf{v}^{+} - \mathbf{v}^{-}$.

We will argue that $\mathbf{v}^{+}\mathbf{P} = \mathbf{v}^{+}$ and, therefore, by normalizing $\mathbf{v}^{+}$, we obtain a stationary distribution for the Markov chain.

Let $\mathbf{w} = \mathbf{v}^+\mathbf{P}$. As all the elements of $\mathbf{P}$ are nonnegative, it follows that all the elements of $\mathbf{w}$ are nonnegative and, in particular, the j^{th} element of $\mathbf{w}$ is

$$w_j = \sum_{i=1}^{N} v_i^+ p_{i,j} \geq \sum_{i=1}^{N} (v_i^+ - v_i^-) p_{i,j} = \sum_{i=1}^{N} v_i p_{i,j} = v_j,$$

where the inequality follows from the fact that $v_i^- \geq 0$ and $p_{i,j} \geq 0$ for all $i, j \in \mathcal{S}$. In case $v_j \geq 0$, then $v_j = v_j^+$ and $w_j \geq v_j^+$. In case $v_j < 0$, then $v_j^+ = 0$ and $w_j \geq v_j^+$. We conclude that $w_j \geq v_j^+$ for all $j \in \mathcal{S}$. On the other hand,

$$\sum_{j=1}^{N} w_j = \sum_{j=1}^{N} \sum_{i=1}^{N} v_i^+ p_{i,j} = \sum_{i=1}^{N} v_i^+ \sum_{j=1}^{N} p_{i,j} = \sum_{i=1}^{N} v_i^+.$$

Since $w_j \geq v_j^+$ for all j, it follows that $\mathbf{w} = \mathbf{v}^+$ and $\mathbf{v}^+\mathbf{P} = \mathbf{v}^+$. $\square$

Limiting Distribution A Markov chain is said to have a *limiting distribution* if for every initial distribution $\boldsymbol{\pi}(0)$, the limit $\lim_{t\to\infty} \boldsymbol{\pi}(t)$ exists and is independent of $\boldsymbol{\pi}(0)$. This limit will be denoted by $\boldsymbol{\pi}(\infty)$. In this case, we say that $\boldsymbol{\pi}(\infty)$ is a limiting distribution of the chain.

Hence, a Markov chain has a limiting distribution if and only if for any initial distribution $\boldsymbol{\pi}(0)$, $\lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t$ exists and is independent of $\boldsymbol{\pi}(0)$.

In Example 1, the Markov chain has no limiting distribution. Indeed, suppose $\boldsymbol{\pi}(0) = (1/2, 0, 1/2, 0)$, then $\boldsymbol{\pi}(t) = (0, 1/2, 0, 1/2)$ if t is odd and $\boldsymbol{\pi}(t) = (1/2, 0, 1/2, 0)$ if t is even. Hence, the limit does not exist and the Markov chain has no limiting distribution.

The Markov chain in Example 3 does not have a limiting distribution since if the initial distribution is $\boldsymbol{\pi}(0) = (1, 0, 0, 0, 0, 0)$, then $\boldsymbol{\pi}(t) = (1, 0, 0, 0, 0, 0)$ for all t and if the initial distribution is $\boldsymbol{\pi}(0) = (0, 1/2, 1/2, 0, 0, 0)$, then $\boldsymbol{\pi}(t) = (0, 1/2, 1/2, 0, 0, 0)$ for all t. This is because both distributions are stationary.

It can be shown that the Markov chain in Example 2 has a limiting distribution given by

$$\boldsymbol{\pi}(\infty) = (2/9, 2/9, 1/3, 2/9) \approx (0.222, 0.222, 0.333, 0.222).$$

The table below shows the evolution of $\boldsymbol{\pi}(t)$ for different vectors $\boldsymbol{\pi}(0)$. The entries are decimal approximations of rationals.

t	$\boldsymbol{\pi}(t)$	$\boldsymbol{\pi}(t)$	$\boldsymbol{\pi}(t)$
0	$(1.000, 0.000, 0.000, 0.000)$	$(0.250, 0.250, 0.250, 0.250)$	$(0.250, 0.000, 0.000, 0.750)$
1	$(0.000, 0.250, 0.500, 0.250)$	$(0.250, 0.188, 0.375, 0.188)$	$(0.375, 0.063, 0.500, 0.063)$
5	$(0.000, 0.250, 0.500, 0.250)$	$(0.234, 0.215, 0.336, 0.215)$	$(0.273, 0.191, 0.344, 0.191)$
10	$(0.220, 0.224, 0.333, 0.224)$	$(0.221, 0.223, 0.333, 0.222)$	$(0.219, 0.224, 0.333, 0.224)$

A Markov chain with transition probability matrix $\mathbf{P}$ has a limiting distribution if and only if $\lim_{t\to\infty} \mathbf{P}^t$ exists and the limit is a matrix, denoted by $\mathbf{P}^\infty$, all of its rows are identical. The limiting distribution is the common row in $\mathbf{P}^\infty$.

Indeed, if the Markov chain has a limiting distribution, then $\lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t$ exists and is independent of $\boldsymbol{\pi}(0)$. If $\boldsymbol{\pi}(0) = (\pi(1), \pi(2), \ldots, \pi(N))$, where $\pi(i) = 1$ for some $i = 1, 2 \ldots, N$, and $\pi(j) = 0$ for $j \neq i$, then $\boldsymbol{\pi}(\infty) = \lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t$ is the i^{th} row in $\lim_{t\to\infty} \mathbf{P}^t$. Since $\boldsymbol{\pi}(\infty)$ does not depend on i, it follows that $\lim_{t\to\infty} \mathbf{P}^t$ exists and the limit is a matrix all of its rows are identical.

Conversely, if $\lim_{t\to\infty} \mathbf{P}^t$ exists and the limit is a matrix, $\mathbf{P}^\infty = [p_{i,j}^{(\infty)}]_{1\le i,j\le N}$, all of its rows are identical, then for any initial distribution $\boldsymbol{\pi}(0)$, we have

$$\lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t \quad = \quad \boldsymbol{\pi}(0) \lim_{t\to\infty} \mathbf{P}^t.$$

The j^{th} element in $\lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t$ is

$$\sum_{i=1}^{N} \pi_i(0) p_{i,j}^{(\infty)} = p_{i,j}^{(\infty)} \sum_{i=1}^{N} \pi_i(0) = p_{i,j}^{(\infty)}$$

as $p_{i,j}^{(\infty)}$ does not depend on i. Hence, $\lim_{t\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^t$ exists and equals the common row in $\mathbf{P}^\infty$. This proves that the Markov chain has a limiting distribution which equals the common row.

Clearly if a Markov chain has a limiting distribution, then this distribution is unique.

On the other hand, if the initial distribution $\boldsymbol{\pi}(0)$ is a stationary distribution, then $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)$ for all $t \geq 1$. Hence, if a Markov chain has a limiting distribution, then it has a unique stationary distribution and the limiting distribution is that unique stationary distribution.

Classification of States We say that state j is *accessible* from state i if $p_{i,j}^{(m)} > 0$ for some $m \geq 0$, i.e., it is possible to reach state j from state i in some nonnegative number of steps. This is denoted by $i \to j$. By convention, $p_{i,i}^{(0)} = 1$ and $p_{i,j}^{(0)} = 0$ if $i \neq j$.

We say that states i and j *communicate* if j is accessible from i and i is accessible from j. This is denoted by $i \leftrightarrow j$.

Communicativity defines an *equivalence relation*:

1. $i \leftrightarrow i$.
2. $i \leftrightarrow j$ implies $j \leftrightarrow i$.
3. $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$.

Two states that communicate with each other are said to be in the same *communicating class*. From the above, any two communicating classes are either disjoint or identical.

We say that the Markov chain is *irreducible* if it has only one communicating class, i.e., all states in the Markov chain communicate with each other. Hence, the Markov chain is irreducible if for all $i, j \in \mathcal{S}$, $p_{i,j}^{(m)} > 0$ for some $m \geq 0$.

The Markov chains in Examples 1 and 2 are irreducible as each has only one communicating class. The Markov chain in Example 3 has four communicating classes, namely, $\{1\}$, $\{2, 3\}$, $\{4, 5\}$, and $\{6\}$. It is not irreducible.

The *period* of a state i is defined as

$$d_i = \text{gcd}\{m \geq 1 : p_{i,i}^{(m)} > 0\}$$

if there is a positive integer m for which $p_{i,i}^{(m)} > 0$, i.e., if starting from state i, with positive probability it is possible to revisit state i at some future time. In words, d_i is the greatest common divisor of the lengths of all paths starting from state i and returning to state i. Otherwise, i.e., if $p_{i,i}^{(m)} = 0$ for all $m \geq 1$, then we define $d_i = \infty$.

In Example 1, every state has period 2 while in Example 2, every state has period 1. In Example 3, $d_1 = 1$, $d_2 = d_3 = 2$, $d_4 = d_5 = 2$, and $d_6 = \infty$.

If two distinct states i and j communicate, i.e., $i \leftrightarrow j$, then they have the same period, i.e., $d_i = d_j$. Indeed, if i and j communicate, then $p_{i,j}^{(m_{i,j})} > 0$ for some $m_{i,j} \geq 1$, and $p_{j,i}^{(m_{j,i})} > 0$ for some $m_{j,i} \geq 1$. This implies that $p_{i,i}^{(m_{i,j}+m_{j,i})} > 0$. Hence, d_i is finite and divides $m_{i,j} + m_{j,i}$. The same can be said about d_j. In particular, $p_{j,j}^{(m)} > 0$ for some positive integer m. We conclude that $p_{i,i}^{(m_{i,j}+m_{j,i}+m)} > 0$ and, therefore, d_i divides $m_{i,j} + m_{j,i} + m$. Together with the statement that d_i divides $m_{i,j} + m_{j,i}$, it follows that d_i divides m. This is true for every positive integer m such that $p_{j,j}^{(m)} > 0$. As d_i divides all such positive integers, it divides their greatest common divisor d_j. By symmetry, we can also conclude that d_j divides d_i. Hence, $d_i = d_j$.

It follows that all states in the same communicating class have the same period.

We say that a Markov chain is *aperiodic* if every state has period equal to 1. The Markov chain in Example 2 is aperiodic but not those in Examples 1 and 3.

Theorem 2 *Every irreducible aperiodic Markov chain has a limiting distribution.*

Proof. The proof of this result is composed of two parts:

Part 1 We will show that if $\mathbf{P}$ is a transition probability matrix of an irreducible aperiodic Markov chain, then there is some positive integer m such that all the elements of $\mathbf{P}^m$ are positive.

Part 2 We will show that if $\mathbf{P}$ is a stochastic matrix in which all the elements are positive, then $\lim_{t\to\infty} \mathbf{P}^t$ exists and is a matrix in which all the rows are identical.

Once these two parts are established, the result follows. Clearly, it suffices to consider the case $N \geq 2$ which we will assume in the following.

Part 1 of the Proof Since $\mathbf{P} = [p_{i,j}]_{1 \leq i,j \leq N}$ is the transition probability matrix of an irreducible aperiodic Markov chain, then for any pair (i, j), $1 \leq i, j \leq N$, there is a positive integer $m_{i,j}$ such that $p_{i,j}^{(m_{i,j})} > 0$. What we need to show is that there is a positive integer m such that $p_{i,j}^{(m)} > 0$ for all pairs (i, j), $1 \leq i, j \leq N$.

The proof is based on Bézout Identity explained next.

Bézout Identity Let $m_1, m_2, \ldots, m_h$ be integers, the greatest common divisor of which is 1, i.e., there is no integer greater than 1 that divides all of them. Then, every integer ℓ can be written as $\ell = a_1 m_1 + a_2 m_2 + \cdots + a_h m_h$ for some integers $a_1, a_2, \ldots, a_h$.

Example 4 For example, let $m_1 = 12, m_2 = 20, m_3 = 15$. Then,

$$\begin{aligned} 1 &= -2 \times 12 - 1 \times 20 + 3 \times 15, \\ 2 &= -4 \times 12 - 2 \times 20 + 6 \times 15, \\ -3 &= 6 \times 12 + 3 \times 20 - 9 \times 15, \end{aligned}$$

etc.

To prove Bézout identity, consider the set $\mathcal{B}$ of all integers of the form $a_1 m_1 + a_2 m_2 + \cdots + a_h m_h$ where $a_1, a_2, \ldots, a_h$ are integers. Let $\ell_{\min}$ be the minimum positive integer in $\mathcal{B}$. Then, we claim that every integer ℓ in $\mathcal{B}$ is divisible by $\ell_{\min}$. Indeed, let r be the remainder obtained by dividing ℓ by $\ell_{\min}$. Then, $\ell = q\ell_{\min} + r$, where q is the quotient, which is an integer, and r is the remainder, which is an integer satisfying $0 \leq r < \ell_{\min}$. Since $\ell_{\min}$ is in $\mathcal{B}$, $q\ell_{\min}$ is also in $\mathcal{B}$, and $\ell - q\ell_{\min}$ is in $\mathcal{B}$. We conclude that r is in $\mathcal{B}$. As $0 \leq r < \ell_{\min}$, this contradicts the choice of $\ell_{\min}$ unless $r = 0$, i.e., ℓ is divisible by $\ell_{\min}$. Since $m_1, m_2, \ldots, m_h$ are in $\mathcal{B}$, $\ell_{\min}$ divides all of them. This implies that $\ell_{\min} = 1$. Hence, we can write $1 = a_1 m_1 + a_2 m_2 + \cdots + a_h m_h$ for some integers $a_1, a_2, \ldots, a_h$ which implies that $\ell = (\ell a_1) m_1 + (\ell a_2) m_2 + \cdots + (\ell a_h) m_h$ for every integer ℓ.

If $\ell = a_1m_1 + a_2m_2 + \cdots + a_hm_h$, then we say that $a_1, a_2, \ldots, a_h$ are *Bézout coefficients* of ℓ. The Bézout coefficients may not be unique.

For example, in Example 4, in addition to

$$1 = -2 \times 12 - 1 \times 20 + 3 \times 15,$$

we can also write

$$1 = -2 \times 12 + 5 \times 20 - 5 \times 15.$$

We will argue in the following that if $m_1, m_2, \ldots, m_h$ are positive and ℓ is sufficiently large, then ℓ can be expressed in terms of nonnegative Bézout coefficients.

Suppose $1 = a_1 m_1 + a_2 m_2 + \cdots + a_h m_h$. Let $1^+ = \sum_{i:a_i \geq 0} a_i m_i$ and $1^- = \sum_{i:a_i<0}(-a_i)m_i = -\sum_{i:a_i<0} a_i m_i$. Then, $1^+ - 1^- = 1$. Clearly, 1^- is nonnegative. If $1^- = 0$, then 1 can be expressed in terms of nonnegative Bézout coefficients which implies, by multiplying these coefficients by ℓ, that every positive integer ℓ can also be expressed in terms of nonnegative Bézout coefficients. So, we assume in the following that $1^- > 0$. Let $\ell \geq (1^-)^2$, we obtain by dividing ℓ by 1^-, $\ell = q1^- + r$, where $q \geq 1^-$ is the quotient and r, $0 \leq r < 1^-$, is the remainder. In particular, $q > r$. Noticing that $1^+ - 1^- = 1$, we get

$$\begin{aligned} \ell &= q1^- + r = q1^- + r(1^+ - 1^-) = r1^+ + (q-r)1^- \\ &= r \sum_{i:a_i \geq 0} a_i m_i + (q-r) \sum_{i:a_i<0} (-a_i) m_i . \end{aligned}$$

As $q - r > 0$, this gives an expression for each $\ell \geq (1^-)^2$ with nonnegative Bézout coefficients.

Now we are in a position to prove Part 1. Let $\mathbf{P} = [p_{i,j}]_{1 \le i,j \le N}$ be the transition probability matrix of an irreducible aperiodic Markov chain. Since the Markov chain is aperiodic, for each $i = 1, 2, \ldots, N$, there are positive numbers $m_{i,i;1}, m_{i,i;2}, \ldots, m_{i,i;h_i}$ such that $p^{(m_{i,i;1})}, p^{(m_{i,i;2})}, \ldots, p^{(m_{i,i;h_i})} > 0$ and their greatest common divisor is 1. Hence, there is a positive integer ℓ'_i such that for any integer $\ell \ge \ell'_i$, we can write

$$\ell = a_{i;1} m_{i,i;1} + a_{i;2} m_{i,i;2} + \cdots + a_{i;h_i} m_{i,i;h_i}$$

where the Bézout coefficients $a_{i;1}, a_{i;2}, \ldots, a_{i;h_i}$ are nonnegative integers. This implies that starting from state i, we can follow a path of length $m_{i,i;1}$ a number of times equal to $a_{i;1}$ followed by a path of length $m_{i,i;2}$ a number of times equal to $a_{i;2}$ and so on until we follow a path of length $m_{i,i;h_i}$ a number of times equal to $a_{i;h_i}$ to end at state i after ℓ steps. Hence, $p_{i,i}^{(\ell)} > 0$.

Furthermore, since the chain is irreducible, for each pair (i, j), $1 \leq i, j \leq N$, there exists a positive integer m such that $p_{i,j}^{(m)} > 0$. Let $m_{i,j}$ be the minimum m for which this holds. Then there is a path from state i to state j of length $m_{i,j}$.

Let m be an integer satisfying

$$m \geq \max_{1 \leq i \leq N} \ell_i' + \max_{1 \leq i,j \leq N} m_{i,j}.$$

We will argue that all the elements $p_{i,j}^{(m)}$ of $\mathbf{P}^m = [p_{i,j}^{(m)}]_{1 \leq i,j \leq N}$ are positive. Consider the pair (i, j), $1 \leq i, j \leq N$. Since $m - m_{i,j} \geq \ell_i'$, there is a path of length $m - m_{i,j}$ from state i to state i. Following this with a path from state i to state j of length $m_{i,j}$, we deduce that there is a path from state i to state j in m steps. Hence, $p_{i,j}^{(m)} > 0$. This is true for all $1 \leq i, j \leq N$ and concludes the proof of Part 1.

Part 2 of the Proof Let $\mathbf{P}$ be a stochastic matrix all of its elements are positive. We want to show that $\lim_{t\to\infty} \mathbf{P}^t$ exists and is a matrix in which all the rows are identical. Let q be the minimum element in $\mathbf{P}$. Since $N \geq 2$, then $0 < q \leq 1/2$.

Let $\mathbf{u}(0) = (u_1, u_2, \ldots, u_N)^\mathsf{T}$ be a column vector of nonnegative elements and define $\mathbf{u}(t) = \mathbf{P}\mathbf{u}(t-1)$ for $t = 1, 2, \ldots$. Then, the i^{th} element of $\mathbf{u}(t)$ is obtained by taking the weighted sum of the elements in $\mathbf{u}(t-1)$ with the elements in the i^{th} row of $\mathbf{P}$ as the weights. These weights sum to 1 and its minimum element is at least q. For $t = 0, 1, \ldots$, let $u_{\min}(t)$ and $u_{\max}(t)$ denote the minimum and maximum values of $\mathbf{u}(t)$, respectively.

Under the conditions that the minimum element of $\mathbf{P}$ is q and the minimum and maximum values of $\mathbf{u}(t-1)$ are $u_{\min}(t-1)$ and $u_{\max}(t-1)$, respectively, it follows that the smallest element in $\mathbf{u}(t)$ is at least $qu_{\max}(t-1)+(1-q)u_{\min}(t-1)$. This bound is obtained by noticing that under the stated conditions, the minimum is attained when one element in $\mathbf{u}(t-1)$ has value $u_{\max}(t-1)$ and each of the remaining elements has value $u_{\min}(t-1)$. Suppose that the element in $\mathbf{u}(t-1)$ of value $u_{\max}(t-1)$ is multiplied by q', where $q \leq q' \leq 1$. Then,

$$u_{\min}(t) \geq \min_{q \leq q' \leq 1} \{q' u_{\max}(t-1) + (1-q') u_{\min}(t-1)\}.$$

Since $u_{\min}(t-1) \leq u_{\max}(t-1)$, the minimum is attained if $q' = q$ and we conclude that

$$u_{\min}(t) \geq q u_{\max}(t-1) + (1-q) u_{\min}(t-1).$$

Similarly, under the stated conditions, it follows that the largest element in $\mathbf{u}(t)$ is at most $qu_{\min}(t-1)+(1-q)u_{\max}(t-1)$. This bound is obtained by noticing that under the stated conditions, the maximum is attained when one element in $\mathbf{u}(t-1)$ has value $u_{\min}(t-1)$ and each of the remaining elements has value $u_{\max}(t-1)$. Suppose that the element in $\mathbf{u}(t-1)$ of value $u_{\min}(t-1)$ is multiplied by q'', where $q \leq q'' \leq 1$. Then,

$$u_{\max}(t) \leq \max_{q \leq q'' \leq 1} \{q'' u_{\min}(t-1) + (1-q'')u_{\max}(t-1)\}.$$

Since $u_{\min}(t-1) \leq u_{\max}(t-1)$, the maximum is attained if $q'' = q$ and we conclude that

$$u_{\max}(t) \leq qu_{\min}(t-1) + (1-q)u_{\max}(t-1).$$

From the above bounds on $u_{\min}(t)$ and $u_{\max}(t)$, we deduce that

$$u_{\max}(t) - u_{\min}(t) \leq (1-2q)(u_{\max}(t-1) - u_{\min}(t-1)).$$

As t assumes the values $1, 2, \ldots$, we obtain

$$u_{\max}(t) - u_{\min}(t) \leq (1-2q)^t (u_{\max}(0) - u_{\min}(0)).$$

Since $0 < q \leq 1/2$, $u_{\max}(t) - u_{\min}(t) \to 0$ as $t \to \infty$ and $\mathbf{u}(t)$ converges to a vector all of its elements are equal.

Now, for each $i = 1, 2, \ldots, N$, let $\mathbf{u}(0)$ be the column vector in which all elements are zero except for the i^{th} element which equals 1. Then, for $t = 1, 2, \ldots$, $\mathbf{u}(t)$ is the i^{th} column in $\mathbf{P}^t$ and it converges to a vector all of its elements are equal. We conclude that as $t \to \infty$, $\mathbf{P}^t$ converges to a matrix, $\mathbf{P}^\infty$, all of its rows are equal. This concludes the proof of Part 2. $\square$

Recurrent and Transient States We say that state i is *recurrent* if it accessible from every state j which is accessible from i, i.e., $i \to j$ implies $j \to i$.

A state i is *transient* if it is not recurrent, i.e., there exists at least one state k that can be reached from i but from which there is no return to i, i.e., for some state k, $i \to k$ but $k \not\to i$.

Notice that no transient state is accessible from a recurrent state. Indeed, suppose $i \to j$, where i is a recurrent state and j is a transient state. Since j is a transient state, then there is a state k such that $j \to k$ but $k \not\to j$. However, as $j \to k$, then $i \to k$ which implies $k \to i$ as state i is recurrent. Hence, $k \to j$ leading to a contradiction.

It follows that all states in a communicating class containing a recurrent state are recurrent and a chain starting from a recurrent state remains forever in the communicating class containing that state.

Notice that a (finite-state) Markov chain has at least one recurrent state. Indeed if all states are transient, then we can find an infinite sequence of states $i_0, i_1, \ldots$ such that $i_t \to i_{t+1}$ but $i_{t+1} \not\to i_t$ for all $t = 0, 1, \ldots$. No two of these states are identical since if $i_t = i_{t'}$ with $t' > t$, then $i_{t+1} \to i_{t'} = i_t$. The same argument shows that for every transient state there is at least one recurrent state accessible from it otherwise, starting from that transient state, we can form such an infinite sequence of distinct transient states.

Hence, all states in an irreducible (finite-state) Markov chain are recurrent since there exists at least one recurrent state and no transient state, if any, can communicate with a recurrent state.

In Examples 1 and 2 all states are recurrent. In Example 3 states 1, 2, and 3 are recurrent while states 4, 5, and 6 are transient.

Let $f_{i,j}^{(m)}$, where $m \geq 0$, be the probability that, starting from state i, the first transition to state j occurs at time m. Formally,

$$f_{i,j}^{(0)} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and

$$f_{i,j}^{(m)} = P(X(m) = j, X(m-1) \neq j, \ldots, X(1) \neq j | X(0) = i)$$

for $m \geq 1$.

Then, $f_{i,j} = \sum_{m=1}^{\infty} f_{i,j}^{(m)}$ is the probability of accessing state j given that the chain starts in state i, in a positive number of steps.

We will show that $f_{i,i} = 1$ if i is a recurrent state and $f_{i,i} < 1$ if i is a transient state.

Indeed, suppose state i is recurrent. A chain starting at state i remains forever in the communicating class containing that state. Since the number of states is finite, there is a state j in the communicating class containing state i which is visited an infinite number of times. Since states i and j are in the same communicating class, then $f_{j,i}^{(m)} > 0$ for some positive integer m. Let $m_{j,i}$ be the minimum such positive integer. Then, each time the chain is at state j, the event that the chain is not at state i after $m_{j,i}$ steps has probability $1 - f_{j,i}^{(m_{j,i})} < 1$. As state j is visited an infinite number of times, the probability that state i is never revisited is zero.

On the other hand, if state i is transient, then there is a positive probability that starting from state i, the chain reaches a state from which it never goes back to state i. Hence, $f_{i,i} < 1$.

Next, we will show that starting from a state, the expected number of visits to the same state is infinite if the state is recurrent and finite if the state is transient.

Indeed, if state i is recurrent, then $f_{i,i} = 1$ and starting from state i, with probability 1, it will be revisited again. Using the Borel-Cantelli Lemma and the fact that the visits are independent due to the Markov property, it follows that, starting from state i, with probability 1, state i will be revisited an infinite number of times. Hence, the expected number of visits is infinite.

On the other hand, if state i is transient, then starting from state i, with probability $f_{i,i} < 1$ it will be revisited again. Hence, the probability that state i will be visited exactly n number of times is $f_{i,i}^{n-1}(1 - f_{i,i})$ for $n \geq 1$. Hence, the number of visits is a geometric random variable with parameter $1 - f_{i,i}$. The mean of this random variable is $1/(1 - f_{i,i}) < \infty$.

Next, we will show that state i is recurrent if $\sum_{m=1}^{\infty} p_{i,i}^{(m)} = \infty$ and transient if $\sum_{m=1}^{\infty} p_{i,i}^{(m)} < \infty$.

Indeed, let us define the random variable

$$I^{(m)} = \begin{cases} 1 & \text{if } X(m) = i \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\sum_{m=0}^{\infty} I^{(m)}$ is the number of visits to state i. The expected number of such visits if the chain starts at state i is then

$$\mathrm{E}\left[\sum_{m=0}^{\infty} I^{(m)} | X(0) = i\right] = \sum_{m=0}^{\infty} \mathrm{E}\left[I^{(m)} | X(0) = i\right] = \sum_{m=0}^{\infty} p_{i,i}^{(m)}.$$

The result follows from the fact that a state is recurrent if this average is infinite and transient if otherwise.

Absorbing State and Absorbing Markov Chain A state i is *absorbing* if $p_{i,i} = 1$, i.e., once the chain reaches state i, it never leaves that state. A Markov chain is said to be *absorbing* if it has at least one absorbing state and from any state, there is at least one accessible absorbing state. Thus the Markov chain is absorbing if and only if every recurrent state is absorbing.

Let j be an absorbing state in an absorbing Markov chain. Let A_j be the event that the Markov chain is eventually absorbed in state j. Then $f_{i,j}$ is the probability of this event given that the Markov chain starts at the transient state i.

We have

$$f_{i,j} = p_{i,j} + \sum_{k \in \mathcal{S}_{\mathrm{Tr}}} p_{i,k} f_{k,j},$$

where $\mathcal{S}_{\mathrm{Tr}}$ is the set of transient states.

Indeed,

$$\begin{aligned}
f_{i,j} &= P(A_j | X(0) = i) \\
&= \sum_{k \in \mathcal{S}} P(A_j | X(1) = k, X(0) = i)\ P(X(1) = k | X(0) = i) \\
&= \sum_{k \in \mathcal{S}} P(A_j | X(1) = k)\ P(X(1) = k | X(0) = i) \\
&= \sum_{k \notin \mathcal{S}_{\mathrm{Tr}}} P(A_j | X(1) = k)\ P(X(1) = k | X(0) = i) \\
&\qquad + \sum_{k \in \mathcal{S}_{\mathrm{Tr}}} P(A_j | X(1) = k)\ P(X(1) = k | X(0) = i).
\end{aligned}$$

If $k \notin \mathcal{S}_{\mathrm{Tr}}$, then k is a recurrent state and, hence, an absorbing state. Therefore,

$$P(A_j|X(1)=k) = \begin{cases} 0 & k \neq j \\ 1 & k = j. \end{cases}$$

On the other hand, if $k \in \mathcal{S}_{\mathrm{Tr}}$, then $P(A_j|X(1)=k) = f_{k,j}$. Hence,

$$\begin{aligned} f_{i,j} &= \sum_{k \notin \mathcal{S}_{\mathrm{Tr}}} P(A_j|X(1)=k)\ P(X(1)=k|X(0)=i) \\ &\quad + \sum_{k \in \mathcal{S}_{\mathrm{Tr}}} P(A_j|X(1)=k)\ P(X(1)=k|X(0)=i) \\ &= p_{i,j} + \sum_{k \in \mathcal{S}_{\mathrm{Tr}}} f_{k,j} p_{i,k}. \end{aligned}$$

Let $T_{\rm absorb}$ be the time it takes to be absorbed in some absorbing state and μ_i be the average of this time if the Markov chain starts at the transient state i. We have

$$\mu_i = 1 + \sum_{j \in \mathcal{S}_{\rm Tr}} p_{i,j} \mu_j .$$

Indeed,

$$\begin{aligned}
\mu_i &= \mathrm{E}[T_{\rm absorb} | X(0) = i] \\
&= \sum_{j \in \mathcal{S}} \mathrm{E}[T_{\rm absorb} | X(1) = j, X(0) = i] P(X(1) = j | X(0) = i) \\
&= \sum_{j \in \mathcal{S}} \mathrm{E}[T_{\rm absorb} | X(1) = j] P(X(1) = j | X(0) = i)
\end{aligned}$$

$$
\begin{aligned}
\mu_i &= \sum_{j \notin \mathcal{S}_{\mathrm{Tr}}} \mathrm{E}[T_{\mathrm{absorb}}|X(1)=j]P(X(1)=j|X(0)=i) \\
&\qquad + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} \mathrm{E}[T_{\mathrm{absorb}}|X(1)=j]P(X(1)=j|X(0)=i) \\
&= \sum_{j \notin \mathcal{S}_{\mathrm{Tr}}} \mathrm{E}[T_{\mathrm{absorb}}|X(1)=j]p_{i,j} + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} \mathrm{E}[T_{\mathrm{absorb}}|X(1)=j]p_{i,j} \\
&= \sum_{j \notin \mathcal{S}_{\mathrm{Tr}}} 1 \times p_{i,j} + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} (1 + \mathrm{E}[T_{\mathrm{absorb}}|X(0)=j])p_{i,j}
\end{aligned}
$$

since $\mathrm{E}[T_{\mathrm{absorb}}|X(1)=j]$ is the time it takes from time 0 to be absorbed if at time 1 the Markov chain is at state j. This is one more than the time it takes to be absorbed from time 0 if at time 0 the Markov chain is at state j. Hence,

$$
\mu_i = \sum_{j \notin \mathcal{S}_{\mathrm{Tr}}} 1 \times p_{i,j} + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} (1 + \mu_j)p_{i,j} = 1 + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} p_{i,j}\mu_j .
$$

Random Walk with Absorbing Barriers A *random walk* is the path traced by a particle which repeatedly takes a step of unit length in some randomly selected direction, independently of the current position of the particle or the directions of previous steps.

We consider a one-dimensional random walk in which the particle is confined to positions $1, 2, \ldots, N$, where $N \geq 2$, as absorbing barriers are placed at positions 1 and N. Once the particle reaches one of these two positions, it stays there. Otherwise, if the particle is at position $i \in \{2, \ldots, N-1\}$, then it moves to position $i+1$ with probability p and to position $i-1$ with probability $q = 1 - p$.

We model the random walk as a Markov chain with states $1, 2, \ldots, N$ corresponding to the positions of the particle. States 1 and N are absorbing and all other states are transient.

Then, $p_{1,1} = p_{N,N} = 1$, $p_{i,i+1} = p$, $p_{i,i-1} = q$ for $i \in \{2, \ldots, N-1\}$, and $p_{i,j} = 0$ for all other choices of i and j.

The transition probability matrix is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & q & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Probability of Absorption at 1 and N

From $f_{i,j} = p_{i,j} + \sum_{k \in \mathcal{S}_{\mathrm{Tr}}} p_{i,k} f_{k,j}$, it follows that

$$\begin{aligned} f_{2,N} &= p f_{3,N} \\ f_{i,N} &= q f_{i-1,N} + p f_{i+1,N}, \text{ for } i = 3, \ldots, N-2 \\ f_{N-1,N} &= p + q f_{N-2,N}. \end{aligned}$$

For convenience, we define $f_{1,N} = 0$ and $f_{N,N} = 1$. Then, for $i = 2, \ldots, N-1$,

$$f_{i,N} = q f_{i-1,N} + p f_{i+1,N}.$$

Since $p + q = 1$, then $f_{i,N} = p f_{i,N} + q f_{i,N}$ and we can write

$$p(f_{i+1,N} - f_{i,N}) = q(f_{i,N} - f_{i-1,N})$$

for $i = 2, \ldots, N-1$.

Hence, for $i = 2, \ldots, N-1$,

$$f_{i+1,N} - f_{i,N} = (q/p)(f_{i,N} - f_{i-1,N}).$$

This recursion gives

$$f_{i+1,N} - f_{i,N} = (q/p)^{i-1}(f_{2,N} - f_{1,N}) = (q/p)^{i-1} f_{2,N}$$

since $f_{1,N} = 0$. Therefore, for $k = 2, \ldots, N-1$,

$$\begin{aligned} f_{k+1,N} - f_{2,N} &= \sum_{i=2}^{k} (f_{i+1,N} - f_{i,N}) \\ &= \sum_{i=2}^{k} (q/p)^{i-1} f_{2,N} \end{aligned}$$

This gives for $k = 2, \ldots, N-1$,

$$
\begin{aligned}
f_{k+1,N} &= f_{2,N} + \sum_{i=2}^{k} (q/p)^{i-1} f_{2,N} \\
&= \sum_{i=1}^{k} (q/p)^{i-1} f_{2,N} \\
&= \begin{cases} \frac{1-(q/p)^k}{1-(q/p)} f_{2,N} & p \neq q \\ k f_{2,N} & p = q. \end{cases}
\end{aligned}
$$

To find $f_{2,N}$, we let $k = N-1$ and recall that $f_{N,N} = 1$ to get

$$
f_{2,N} = \begin{cases} \frac{1-(q/p)}{1-(q/p)^{N-1}} & p \neq q \\ \frac{1}{N-1} & p = q. \end{cases}
$$

Hence, we can write for all $i = 1, \ldots, N$,

$$f_{i,N} = \begin{cases} \frac{1-(q/p)^{i-1}}{1-(q/p)^{N-1}} & p \neq q \\ \frac{i-1}{N-1} & p = q \end{cases}$$

as the probability that a particle starting at position i is absorbed at position N.

A similar argument shows that for all $i = 1, \ldots, N$,

$$f_{i,1} = \begin{cases} \frac{1-(p/q)^{N-i}}{1-(p/q)^{N-1}} & p \neq q \\ \frac{N-i}{N-1} & p = q \end{cases}$$

as the probability that a particle starting at position i is absorbed at position 1.

Notice that $f_{i,1} + f_{i,N} = 1$ for all $i = 1, \ldots, N$, i.e., it is certain that starting from any position, the particle will eventually be absorbed.

Average Time to Absorption From $\mu_i = 1 + \sum_{j\in\mathcal{S}_{\mathrm{Tr}}} p_{i,j}\mu_j$, it follows that for $i = 2,\ldots,N-1$,

$$\mu_i = 1 + \sum_{j=1}^{N} p_{i,j}\mu_j, \text{ i.e., } \mu_i = 1 + q\mu_{i-1} + p\mu_{i+1},$$

where $\mu_1 = \mu_N = 0$.

Since $p + q = 1$, we can write

$$\mu_{i+1} - \mu_i = \frac{q}{p}(\mu_i - \mu_{i-1}) - \frac{1}{p}$$

for $i = 2,\ldots,N$. This recursion gives

$$\begin{aligned} \mu_{i+1} - \mu_i &= (q/p)^{i-1}\mu_2 - \frac{1}{p}\left(1 + (q/p) + \cdots + (q/p)^{i-2}\right) \\ &= \begin{cases} (q/p)^{i-1}\mu_2 - \frac{1-(q/p)^{i-1}}{p-q} & p \neq q \\ \mu_2 - 2(i-1) & p = q. \end{cases} \end{aligned}$$

Therefore, by summing over $i = 2, \ldots, k$, it follows that for $k = 2, \ldots, N-1$,

$$\mu_{k+1} - \mu_2 = \begin{cases} \sum_{i=2}^{k-1} (p/q)^{i-1} \left(\mu_2 + \frac{1}{p-q}\right) - \frac{k-1}{p-q} & p \neq q \\ (k-1)\mu_2 - 2(1 + 2 + \cdots + (k-1)) & p = q \end{cases}$$

$$= \begin{cases} \left(\frac{1-(q/p)^k}{1-(q/p)} - 1\right) \left(\mu_2 + \frac{1}{p-q}\right) - \frac{k-1}{p-q} & p \neq q \\ (k-1)\mu_2 - k(k-1) & p = q \end{cases}$$

Hence, for $k = 2, \ldots, N-1$,

$$\mu_{k+1} = \begin{cases} \frac{1-(q/p)^k}{1-q/p}\left(\mu_2 + \frac{1}{q-p}\right) - \frac{k}{p-q} & p \neq q \\ k(\mu_2 - k + 1) & p = q \end{cases}$$

To find μ_2, we let $k = N-1$ and recall that $\mu_N = 0$ to get

$$\mu_2 = \begin{cases} \frac{N-1}{p(1-(q/p)^{N-1})} - \frac{1}{p-q} & p \neq q \\ N-2 & p = q \end{cases}$$

Hence, we can write for all $i = 1, \ldots, N$,

$$\mu_i = \begin{cases} \frac{1}{p-q}\left(\frac{1-(q/p)^{i-1}}{1-(q/p)^{N-1}}(N-1) - i + 1\right) & p \neq q \\ (i-1)(N-i) & p = q. \end{cases}$$

The Gambler's Ruin Problem Two players PA and PB starting with \$a and \$b, respectively, where a and b are positive numbers, gamble by flipping a coin. If the outcome is a head, which happens with probability p, PB gives PA \$1. Otherwise, if the outcome is a tail, which happens with probability $q = 1 - p$, PA gives PB \$1. The two players continue this gambling game until one of them is bankrupt in which case the game stops and the other gambler ends up with \$a+\$b and is declared the winner of the game.

The game can be represented as a Markov chain with $a + b + 1$ states. The state at time $t = 0, 1 \ldots$ is the amount of money plus 1 PA has after the t^{th} gamble. Thus, if the state is 1, then PA is bankrupt and if the state is $a + b + 1$, then PA has \$a+\$b and PB is bankrupt. This can be modeled as a random walk with two absorbing barriers at 1 and $N = a + b + 1$.

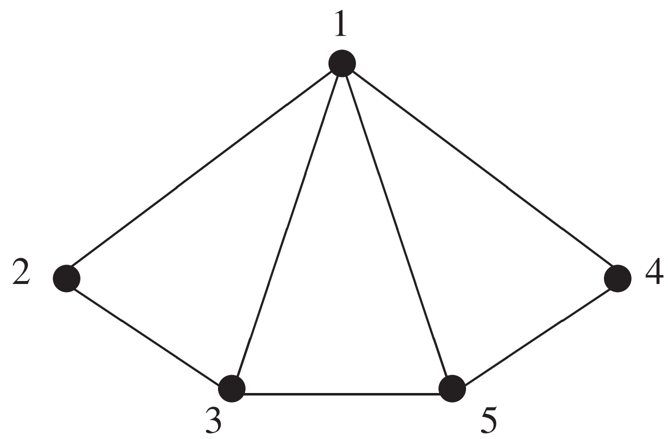
Then, $f_{a+1,a+b+1}$ is the probability that PA is the winner and $f_{a+1,1}$ is the probability that PB is the winner. Also, μ_{a+1} is the average number of gambles before the game ends.

Notice that if we consider the limit as $N \to \infty$, we get

$$f_{a+1,\infty} \quad = \quad \begin{cases} 1 - (q/p)^a & p > q \\ 0 & p \leq q, \end{cases}$$

i.e., if PA is playing against an infinitely rich player, then the fortune of PA can increase without bounds if $p > q$ and this happens with probability $1 - (q/p)^a$. However, if $p \leq q$, then with probability 1, PA will be bankrupt.

Example 5 A particle performs a random walk over the triangulazied pentagon shown. From any given node, the particle chooses an edge with equal probability and takes a step by tracing the chosen edge. [a]



[a] Example 5.4 in textbook.

The random walk can be represented as a Markov chain with states $1, 2, 3, 4, 5$ corresponding to nodes $1, 2, 3, 4, 5$, respectively. The transition probability matrix of this Markov chain is

$$\begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \end{bmatrix}.$$

We will find the expected number of steps μ_i, $i = 2, 3, 4, 5$, for the particle starting at state i to reach state 1.

To find these numbers, we transform state 1 into an absorbing state to form an absorbing Markov chain with transition probability matrix

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

Then, states $2, 3, 4, 5$ are transient.

We use

$$\mu_i = 1 + \sum_{j \in \mathcal{S}_{\mathrm{Tr}}} p_{i,j}\mu_j$$

to get

$$\begin{aligned}
\mu_2 &= 1 + \frac{1}{2}\mu_3 \\
\mu_3 &= 1 + \left(\frac{1}{3}\mu_2 + \frac{1}{3}\mu_5\right) \\
\mu_4 &= 1 + \frac{1}{2}\mu_5 \\
\mu_5 &= 1 + \left(\frac{1}{3}\mu_3 + \frac{1}{3}\mu_4\right).
\end{aligned}$$

By symmetry $\mu_2 = \mu_4$ and $\mu_3 = \mu_5$. The solution of the above system of equations is $\mu_2 = \mu_4 = 7/3$ and $\mu_3 = \mu_5 = 8/3$.