Markov Chain

1 Introduction

Markov chains are fundamental stochastic processes, whose characteristics can be summarized as "future depends on today but not the past". In this chapter, we will restrict ourselves to discrete time Markov chains with a countable state space $\mathcal{S}$.

Definition: A matrix $\mathcal{P} = [P_{ij}]_{i,j\in\mathcal{S}}$ is said to be a *stochastic matrix* on state space $\mathcal{S}$ if

$$P_{ij} \geq 0, \qquad \sum_{j\in\mathcal{S}} P_{ij} = 1.$$

for all $i \in \mathcal{S}$.

Definition: A stochastic process $X = \{X_0, X_1, \ldots\}$ is said to be a **Markov chain** with *transition probability matrix* $\mathcal{P}$ if

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) & \\ = \ & \mathbb{P}(X_{n+1} = j | X_n = i) \\ = \ & P_{ij} \end{aligned}$$

for all n and any $i_0, \ldots, i_{n-1}, i, j \in \mathcal{S}$.

Note that the transition probability matrix $\mathcal{P}$ is necessarily a stochastic matrix, and P_{ij} is the *transition probability* from i to j. Since the transition probability does not depend on the time step n, such Markov chains are said to be *time homogeneous.*

It follows immediately from definition that the distribution of a Markov chain X is completely determined by its *initial distribution*, i.e., the distribution of X_0, and the transition probability matrix $\mathcal{P}$. Indeed,

$$\begin{aligned} & \mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) \\ = \ & \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \\ & \qquad \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) \\ = \ & \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) \\ = \ & \mathbb{P}(X_0 = i_0) \cdot P_{i_0 i_1} \cdots P_{i_{n-1} i_n}. \end{aligned}$$

Lemma 1.1. *Let X be a Markov chain with transition probability matrix $\mathcal{P} = [P_{ij}]$. Then for any $n \geq 1$ and $m \geq 0$*

$$\begin{aligned}\mathbb{P}(X_{m+n} = j_n, \ldots, X_{m+1} = j_1 | X_m = i, X_{m-1} = i_{m-1}, \ldots, X_0 = i_0) \\ = \mathbb{P}(X_{m+n} = j_n, \ldots, X_{m+1} = j_1 | X_m = i) \\ = P_{ij_1} P_{j_1 j_2} \cdots P_{j_{n-1} j_n}.\end{aligned}$$

The proof of this lemma is simple and left as an exercise.

1.1 Examples

Example: We call a state $i \in \mathcal{S}$ an **absorbing** state if $P_{ii} = 1$. The gambler's ruin problem can be regarded as a Markov chain with two absorbing states. Suppose that the starting wealth is x. Let

$$X_n \doteq x + \sum_{i=1}^{n} Y_i,$$

where $\{Y_1, Y_2, \ldots\}$ is a sequence of iid random variables with

$$\mathbb{P}(Y_i = 1) = p, \quad \mathbb{P}(Y_i = -1) = q = 1 - p.$$

The gambler stops playing the game when the total wealth X reaches either b or 0. The state space is $\mathcal{S} = \{0, 1, \cdots, b\}$ and the transition probability matrix $\mathcal{P} = [P_{ij}]$ is determined by

$$P_{ij} = \begin{cases} p & \text{if } j = i + 1, \\ q & \text{if } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

for all $1 \leq i \leq b - 1$, and $P_{00} = 1$, $P_{bb} = 1$. In other words, the transition probability matrix is

$$\mathcal{P} = \begin{Vmatrix} 1 & 0 & 0 & 0 \cdots 0 & 0 \\ q & 0 & p & 0 \cdots 0 & 0 \\ 0 & q & 0 & p \cdots 0 & 0 \\ & \vdots & & \vdots \cdots \vdots & \\ 0 & 0 & 0 & 0 \cdots 0 & p \\ 0 & 0 & 0 & 0 \cdots 0 & 1 \end{Vmatrix}$$

The two states 0 and b are the absorbing states. It is often visually helpful to use graphs to represent Markov Chains.

Example: Simple random walk is a Markov chain with infinite but countable state space $\{0, \pm 1, \ldots\}$, with transition probability matrix $\mathcal{P} = [P_{ij}]$ where

$$P_{ij} = \begin{cases} p & \text{if } j = i+1, \\ q & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Example: Suppose a binary message 0 or 1 is transmitted through a sequence of channels, and the transmission through each channel is subject to a fixed probability of error p. Let X_n be the signal received after n-th channel. Then $X = \{X_n\}$ is a Markov chain with state space $\mathcal{S} = \{0, 1\}$ and transition probability matrix

$$\mathcal{P} = \left\| \begin{matrix} p & q \\ q & p \end{matrix} \right\|.$$

Example: A classical model to describe diffusions through a membrane is the *Ehrenfest urn model.* Imagine two containers, say A and B, containing a total $2a$ balls. At each step, a ball is selected at random (equally likely) from these $2a$ balls at moved to the other container. This process can be modeled by a Markov chain. Let X_n be the number of balls in urn A. Then $X = \{X_n\}$ is a Markov chain with a finite state space $\mathcal{S} = \{0, 1, \ldots, 2a\}$ and transition probability matrix $\mathcal{P} = [P_{ij}]$ where

$$P_{ij} = \begin{cases} 1 - \dfrac{i}{2a} & \text{if } j = i+1, \\ \dfrac{i}{2a} & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Example: A classical model for stock price is the binomial tree. Suppose that the stock price fluctuates as follows. If the stock price at time n is S, then at time $n+1$, with probability p the stock price increases to uS and with probability $q = 1 - p$ the stock price decreases to dS. Here u and d are given constants with $d < 1 < u$ and $ud = 1$. Let X_n be the stock price at time n, then $X = \{X_n\}$ is a Markov chain.

2 Chapman-Kolmogorov Equation

Consider a Markov chain X with transition probability matrix $\mathcal{P} = [P_{ij}]$. The matrix $\mathcal{P}$ describes the one-step transition of the Markov chain. The question is, what about the n-step transition in general.

Theorem 2.1. *For any $n \geq 1$ and $m \geq 0$,*

$$\mathbb{P}(X_{m+n} = j | X_m = i, X_{m-1} = i_{m-1}, \ldots, X_0 = i_0) = [\mathcal{P}^n]_{ij}$$

where $\mathcal{P}^n$ is the n-th power of $\mathcal{P}$.

Proof: We first observe that by Lemma 1.1,

$$\mathbb{P}(X_{m+n} = j | X_m = i, X_{m-1} = i_{m-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{m+n} = j | X_m = i).$$

We will argue by induction on n. The assertion is trivial for $n = 1$. Suppose that it is true for n. Then

$$\begin{aligned}
\mathbb{P}(X_{m+n+1} = j | X_m = i) &= \sum_k \mathbb{P}(X_{m+n+1} = j, X_{m+n} = k | X_m = i) \\
&= \sum_k \mathbb{P}(X_{m+n+1} = j | X_{m+n} = k, X_m = i) \cdot \mathbb{P}(X_{m+n} = k | X_m = i) \\
&= \sum_k \mathbb{P}(X_{m+n+1} = j | X_{m+n} = k) \cdot \mathbb{P}(X_{m+n} = k | X_m = i) \\
&= \sum_k P_{kj} [\mathcal{P}^n]_{ik} \\
&= \sum_k [\mathcal{P}^n]_{ik} \mathcal{P}_{kj} \\
&= [\mathcal{P}^{n+1}]_{ij}.
\end{aligned}$$

This completes the proof. ■

If we let $P_{ij}^{(n)}$ be the probability that the Markov chain will be at state j after n steps if it is currently at state i. Then by the preceding theorem

$$P_{ij}^{(n)} = [\mathcal{P}^n]_{ij},$$

and the **Chapman-Kolmogorov Equation**

$$P_{ij}^{(n+m)} = \sum_{k \in \mathcal{S}} P_{ik}^{(n)} P_{kj}^{(m)}.$$

holds.

Exercise: Argue that $\mathcal{P}^n$ is a stochastic matrix.

Exercise: Consider a Markov chain with transition probability matrix $\mathcal{P}$ and initial distribution $\Lambda = [\Lambda_i]_{i\in\mathcal{S}}$, that is, $\mathbb{P}(X_0 = i) = \Lambda_i$. Show that for every $n \geq 0$ and $i \in \mathcal{S}$

$$\mathbb{P}(X_n = i) = [\Lambda \cdot \mathcal{P}^n]_i \, .$$

2.1 examples

Example: For a two state $I = \{0, 1\}$ Markov chain with initial distribution $\Lambda = (\Lambda_0, \Lambda_1)$ where $\Lambda_i = \mathbb{P}(X_0 = i)$, and transition probability matrix

$$\mathcal{P} = \left\| \begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right\|,$$

with $0 < a, b < 1$. Compute $\mathcal{P}^n$, $\lim_n \mathcal{P}^n$ and the limit distribution of X_n.

Solution: To calculate $\mathcal{P}^n$, we first calculate the eigenvalues of $\mathcal{P}$. The characteristic equation is

$$0 = \det(\lambda I - \mathcal{P}) = [\lambda - (1-a)] \cdot [\lambda - (1-b)] - ab,$$

which yields two eigenvalues

$$\lambda_1 = 1, \quad \lambda_2 = 1 - (a+b).$$

The corresponding eigenvectors are

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -b \\ a \end{bmatrix}.$$

Therefore, we have decomposition

$$\mathcal{P} = Q \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} Q^{-1}$$

with

$$Q = [v_1, v_2] = \begin{bmatrix} 1 & -b \\ 1 & a \end{bmatrix}, \quad Q^{-1} = \frac{1}{a+b} \begin{bmatrix} a & b \\ -1 & 1 \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\mathcal{P}^n &= Q \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} Q^{-1} \\ &= \frac{1}{a+b} \begin{bmatrix} a + b(1-a-b)^n & b - b(1-a-b)^n \\ a - a(1-a-b)^n & b + a(1-a-b)^n \end{bmatrix}.\end{aligned}$$

Since $-1 < a + b < 1$, we have

$$\lim_{n\to\infty} \mathcal{P}^n = \frac{1}{a+b} \begin{bmatrix} a & b \\ a & b \end{bmatrix}.$$

For any initial distribution Λ, the limit distribution is

$$\pi \doteq \lim_{n\to\infty} \begin{bmatrix} \mathbb{P}(X_n = 0) \\ \mathbb{P}(X_n = 1) \end{bmatrix} = \lim_{n\to\infty} \Lambda \cdot \mathcal{P}^n = \frac{1}{a+b} \begin{bmatrix} a \\ b \end{bmatrix}$$

is *independent* of the initial distribution Λ. ■

Exercise: In the previous example, show that if the initial distribution for the Markov chain is π, then the distribution of X_n is always π for every n.

3 First step analysis

First step analysis is a very useful technique. After seeing its preliminary applications in simple random walks, we will start this section by giving more examples.

3.1 Examples

Example: Consider a particle moves around on a circle with 4 nodes. At each node, the particle is equally likely to move to two adjacent nodes. Compute the probability that the particle hits node 2 before node 3, starting from node 1.

Solution: For each node $x = 1, 2, 3, 4$, define

$$h(x) = \mathbb{P}(\text{hits node 2 before node 3} \mid \text{starting from node } x).$$

Observe that $h(2) = 1$ and $h(3) = 0$ by definition. Conditioning on the first move and by law of total probability

$$h(1) = \frac{1}{2}h(2) + \frac{1}{2}h(4) = \frac{1}{2} + \frac{1}{2}h(4)$$

and

$$h(4) = \frac{1}{2}h(1) + \frac{1}{2}h(3) = \frac{1}{2}h(1).$$

Solving these two equations we have

$$h(1) = \frac{2}{3}, \quad h(4) = \frac{1}{3}.$$

The probability we are interested is $h(1) = 2/3$. ■

Example: Consider a gambler whose wealth each day can be described as high, low, or broke. Denote 0 = broke, 1 = low, and 2 = high. Let X_n be the wealth of the gambler at day n. Assume that $X = \{X_n\}$ is a Markov chain with transition probability matrix

$$\mathcal{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{Vmatrix}.$$

Define T to be the time to broke, that is,

$$T \doteq \inf\{n \geq 0 : X_n = 0\}.$$

Compute the expectation

$$h(x) = E^x \left[\sum_{n=0}^{T-1} U(X_n) \right]$$

where U is an arbitrary given function.

Solution: Clearly $h(0) = 0$. By conditioning on X_1 and tower property

$$h(1) = \frac{1}{2}h(0) + \frac{1}{2}h(2) + U(1) = \frac{1}{2}h(2) + U(1)$$

and

$$h(2) = \frac{1}{2}h(0) + \frac{1}{2}h(1) + U(2) = \frac{1}{2}h(1) + U(2).$$

Solve the system of equations we have

$$h(1) = \frac{4}{3}U(1) + \frac{2}{3}U(2), \quad h(2) = \frac{2}{3}U(1) + \frac{4}{3}U(2).$$

Note that when $U(1) = U(2) = 1$ we obtain the expected time until broke

$$h(x) = E^x[T] = 2$$

for $x = 1$ or 2. ■

Example: In the previous example, compute

$$H(x) = E^x \left[\sum_{n=0}^{T-1} \beta^n U(X_n) \right]$$

where $0 < \beta < 1$ is a given positive constant called *discount factor*.

Solution: Similar consideration yields $H(0) = 0$ and

$$H(1) = \frac{1}{2}\beta H(0) + \frac{1}{2}\beta H(2) + U(1), \quad H(2) = \frac{1}{2}\beta H(0) + \frac{1}{2}\beta H(1) + U(2).$$

Solving the equations yields

$$\begin{aligned} H(1) &= \frac{1}{4-\beta^2}\left[4U(1) + 2\beta U(2)\right] \\ H(2) &= \frac{1}{4-\beta^2}\left[2\beta U(1) + 4U(2)\right]. \quad \blacksquare \end{aligned}$$

Exercise: Use the previous example to argue that for all $0 < \beta < 1$

$$E^x[\beta^T] = \beta/(2-\beta).$$

3.2 General hitting times

By now it is clear that first step analysis is a powerful method for studying hitting times. But the next example will show that the first step analysis alone can sometimes be a bit troublesome.

Example: Consider a simple random walk on the positive half-line absorbing at state 0, that is, the state space $\mathcal{S} = \{0, 1, 2, \cdots\}$, and

$$S_n \doteq x + \sum_{i=1}^{n} Y_i,$$

where $x \geq 0$ and $\{Y_1, Y_2, \ldots\}$ is a sequence of iid random variables with

$$\mathbb{P}(Y_i = 1) = p, \quad \mathbb{P}(Y_i = -1) = q = 1 - p.$$

The only difference from the standard simple random walk is that this process is absorbing at 0. We are interested in

$$h(x) \doteq \mathbb{P}^x(\text{absorbed eventually})$$

under the assumption that $p > q$.

Solution: We actually know the answer. The probability is the same as that of a simple random walk hitting 0, starting from $x \geq 0$. Therefore $h(x) = (q/p)^x$. However, the approach we used to derive this result is to first compute the probability of hitting 0 before hitting level b and then let b go to infinity. What if we use the first step analysis directly?

By conditioning on Y_1 and law of total probability, for every $x > 0$,

$$h(x) = ph(x+1) + qh(x-1),$$

with boundary condition $h(0) = 1$. The general solution to this difference equation is

$$h(x) = A + B(q/p)^x.$$

Plugging $h(0) = 1$, we have $A + B = 1$ and

$$h(x) = 1 + B[(q/p)^x - 1].$$

However, since $h(x)$ has to be between 0 and 1, we must have $0 \leq B \leq 1$. The trouble is that B cannot be uniquely determined — every such B corresponds to a solution to the difference equation that satisfies the boundary condition. What now? ■

From the preceding example, we know that the first step analysis may lead to equations that have multiple solutions. In order to establish a criterion to systematically characterize the true solution, we should restrict our attention to hitting times for clarity, even though similar results can be established in much more general situations.

Consider a Markov chain with state space $\mathcal{S}$ and transition probability matrix $\mathcal{P}$. Let A be an arbitrary subset of the state space $\mathcal{S}$. The hitting time of A is defined as

$$T^A \doteq \inf\{n \geq 0 : X_n \in A\}$$

with the convention $\inf\{\emptyset\} \doteq \infty$. We are interested in

$$\begin{aligned} h^A(x) &\doteq \mathbb{P}^x(T^A < \infty) \\ d^A(x) &\doteq E^x[T^A] \end{aligned}$$

Proposition 3.1. *The vector* $\left\{h^A(x) : x \in \mathcal{S}\right\}$ *is the minimal nonnegative solution of the system of linear equations:*

$$h(x) = \sum_{y \in \mathcal{S}} P_{xy} h(y), \quad x \notin A$$

and $h(x) = 1$ *for all* $x \in A$.

Similarly, the vector $\left\{d^A(x) : x \in \mathcal{S}\right\}$ *is the minimal nonnegative solution of the system of linear equations:*

$$d(x) = 1 + \sum_{y \in \mathcal{S}} P_{xy} d(y), \quad x \notin A$$

and $d(x) = 0$ *for all* $x \in A$.

Proof: By law of total probability it is clear that $\{h^A(x) : x \in \mathcal{S}\}$ is a solution to the system of linear equations. Now suppose $\{h(x) : x \in \mathcal{S}\}$ is an arbitrary nonnegative solution. We want to show that $h(x) \geq h^A(x)$ for all $x \in \mathcal{S}$. This is trivial for $x \in A$ where $h(x) = h^A(x) = 1$. Consider $x \notin A$.

$$\begin{aligned}
h(x) &= \sum_{y \in \mathcal{S}} P_{xy} h(y) \\
&= \sum_{y \in A} P_{xy} + \sum_{y \notin A} P_{xy} h(y) \\
&= \sum_{y \in A} P_{xy} + \sum_{y \notin A} P_{xy} [\sum_{z \in A} P_{yz} + \sum_{z \notin A} P_{yz} h(z)] \\
&= \sum_{y \in A} P_{xy} + \sum_{y \notin A} \sum_{z \in A} P_{xy} P_{yz} + \sum_{y \notin A} \sum_{z \notin A} P_{xy} P_{yz} h(z) \\
&= \mathbb{P}^x(T^A = 1) + \mathbb{P}^x(T^A = 2) + \sum_{y \notin A} \sum_{z \notin A} P_{xy} P_{yz} h(z) \\
&= \cdots
\end{aligned}$$

This implies that $h(x) \geq \mathbb{P}^x(T^A \leq n)$ for every n. Letting $n \to \infty$ we obtain

$$h(x) \geq \mathbb{P}^x(T^A < \infty) = h^A(x).$$

As to $\left\{d^A(x)\right\}$, by tower property of conditional expectation, it is a solution to the corresponding system of equations. Now let $\{d(x) : x \in \mathcal{S}\}$ is an arbitrary nonnegative solution. We wish to show $d(x) \geq d^A(x)$. Again it is trivial for $x \in A$. For $x \notin A$, observe that

$$\begin{aligned}
d(x) &= 1 + \sum_{y \notin A} P_{xy} d(y) \\
&= 1 + \sum_{y \notin A} P_{xy} [1 + \sum_{z \notin A} P_{yz} h(z)] \\
&= 1 + \sum_{y \notin A} P_{xy} + \sum_{y \notin A} \sum_{z \notin A} P_{xy} P_{yz} d(z) \\
&= \mathbb{P}^x(T^A \geq 1) + \mathbb{P}^x(T^A \geq 2) + \sum_{y \notin A} \sum_{z \notin A} P_{xy} P_{yz} d(z) \\
&= \cdots
\end{aligned}$$

This yields that, for all N,

$$d(x) \geq \sum_{n=1}^{N} \mathbb{P}^x(T^A \geq n)$$

Letting $N \to \infty$ and recalling

$$E^x[T^A] = \sum_{n=1}^{\infty} \mathbb{P}^x(T^A \geq n)$$

since T^A is a nonnegative integer-valued random variable, we complete the proof. ■

Example: Going back to the previous example, and observe that all the nonnegative solutions have form

$$h(x) = 1 + B[(q/p)^x - 1]$$

for all $0 \leq B \leq 1$. Since $q < p$, the minimal solution is with $B = 1$ or

$$h(x) = (q/p)^x$$

which is the true solution as noted. ■

Example: Consider the following Markov chain X with state space $\mathcal{S} = \{0, 1, \ldots\}$ and absorbing state 0. For all $x > 0$,

$$P_{xy} = \begin{cases} 1/2 & \text{if } y = 0 \\ 1/4 & \text{if } y = x+1 \\ 1/4 & \text{if } y = x-1 \\ 0 & \text{otherwise.} \end{cases}$$

Compute the expected duration until absorption, that is,

$$d(x) = E^x[T^0],$$

where T^0 is the time to absorption.

Solution: The equation that $d(x)$ satisfies is

$$d(x) = 1 + \frac{1}{4}h(x+1) + \frac{1}{4}h(x-1)$$

with boundary condition $d(0) = 0$. The characteristic equation for this difference equation is

$$s = \frac{1}{4}s^2 + \frac{1}{4}$$

which has two distinct roots

$$s_1 = 2 + \sqrt{3}, \quad s_2 = 2 - \sqrt{3}.$$

Furthermore a particular solution is just the constant solution 2. Therefore

$$d(x) = As_1^x + Bs_2^x + 2 = A(2+\sqrt{3})^x + B(2-\sqrt{3})^x + 2.$$

Since $d(0) = 0$ we have $A + B = -2$ and

$$d(x) = A\left[(2+\sqrt{3})^x - (2-\sqrt{3})^x\right] + 2\left[1 - (2-\sqrt{3})^x\right].$$

Note that $d(x)$ is nonnegative if and only if $A \geq 0$. The minimal solution is just by taking $A = 0$ and thus

$$d(x) = 2\left[1 - (2-\sqrt{3})^x\right]. \quad \blacksquare$$

Example: Consider a stochastic model for population growth such that each individual in the population *independently* reproduces a random number Z of offspring with probability distribution

$$\mathbb{P}(Z = j) = p_j, \quad j = 0, 1, \ldots, \quad \sum_{j=0}^{\infty} p_j = 1.$$

Let X_n denote the population size at the n-th generation, and with no loss of generality, we assume $X_0 = 1$. Compute the probability of extinction.

Solution: We can write that

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i$$

where $\{Z_1, Z_2, \cdots\}$, representing the number of offsprings from each individual, is a sequence of iid random variable with distribution $\mathbb{P}(Z_m = j) = p_j$ for all j. It is easy to show that $X = \{X_n\}$ is a Markov chain. Let T^0 be the time to extinction, i.e.,

$$T^0 \doteq \inf\{n \geq 0 : X_n = 0\}$$

and the probability of extinction is $\mathbb{P}^1(T_0 < \infty)$. Define for each $i \geq 0$,

$$H(i) \doteq \mathbb{P}^i(T_0 < \infty).$$

The probability of interest is $H(1)$. It follows that H is the minimal nonnegative solution to the system of equations

$$h(i) = \sum_{j=0}^{\infty} P_{ij} h(j)$$

for all $i \geq 1$ and $h(0) = 1$. The key observation is that this system of equation can be reduced to a *single* equation! Indeed, it is not difficult to to see that for all $j \geq 0$,

$$H(j) = [H(1)]^j \doteq \pi^j.$$

Therefore, we expect that π satisfies

$$\pi = \sum_{j=0}^{\infty} P_{1j} \pi^j = \sum_{j=0}^{\infty} p_j \pi^j. \tag{3.1}$$

Furthermore, for every π that satisfies the preceding equation automatically satisfies

$$\pi^i = \left[\sum_{j=0}^{\infty} P_{1j} \pi^j \right]^i = \sum_{j=0}^{\infty} P_{ij} \pi^j.$$

Therefore, π must be the minimal nonnegative solution to the equation (3.1). Denoting $\mu \doteq E[Z]$, it is not difficult to show that,

1. If $p_0 = 0$ then $\pi = 0$.
2. If $p_0 > 0$ and $\mu \leq 1$ then $\pi = 1$.
3. If $p_0 > 0$ and $\mu > 1$ then $0 < \pi < 1$. ■

4 Strong Markov Property

People often use strong Markov property subconsciously as if it were part of the definition of Markov chain, without realizing that it is indeed a theorem. Fortunately, for discrete time Markov chains, strong Markov property always holds. This section is to prove the strong Markov property and fill the gap between applications and theory.

Consider a Markov chain with transition probability matrix $\mathcal{P}$ and any fixed n. Then given $X_n = i$, the chain $\{X_n, X_{n+1}, \ldots\}$ is again a Markov chain with transition probability matrix $\mathcal{P}$. This new chain starts at state i, and is independent of $\{X_0, X_1, \ldots, X_{n-1}\}$. In other words, at any time n, the Markov chain starts afresh from X_n.

The strong Markov property basically asserts that the above properties not only hold for fixed time n, but also for any stopping time. In other words, at any stopping time T, the Markov chain starts afresh from X_T. The rigorous formulation of strong Markov property is presented in the next theorem. But for the students, it is best to understand the strong Markov property intuitively so as to avoid the heavy notational burden that is often associated with the strong Markov property.

Theorem 4.1. *Suppose that* $X = \{X_n\}$ *is a Markov chain with transition probability matrix* $\mathcal{P}$*, and that* T *is an arbitrary stopping time. Conditional on* $\{T < \infty, X_T = i\}$*, the process* $\{X_T, X_{T+1}, X_{T+2}, \cdots\}$ *is a Markov Chain starting from* i*, with transition probability matrix* $\mathcal{P}$*, and independent of* $\{X_0, X_1, \cdots, X_{T-1}\}$*.*

Proof: Throughout the proof we assume that $\mathbb{P}(T < \infty) = 1$. The proof for the general case is almost verbatim and omitted. First we prove the independence and the Markov property, that is

$$\begin{aligned} &\mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1, X_{T-1} = j_{T-1}, \cdots, X_0 = j_0 \mid X_T = i) \\ &\quad = \mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1 | X_T = i) \\ &\qquad \cdot \mathbb{P}(X_{T-1} = j_{T-1}, \cdots, X_0 = j_0 | X_T = i), \end{aligned} \tag{4.2}$$

and

$$\mathbb{P}(X_{T+n} = i_n | X_{T+n-1} = i_{n-1}, \cdots, X_{T+1} = i_1, X_T = i) = P_{i_{n-1} i_n}. \tag{4.3}$$

To prove these two identities it suffices to show that

$$\begin{aligned} &\mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1, X_{T-1} = j_{T-1}, \cdots, X_0 = j_0, X_T = i) \\ &\quad = \left(P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n}\right) \cdot \mathbb{P}(X_{T-1} = j_{T-1}, \cdots, X_0 = j_0, X_T = i). \end{aligned} \tag{4.4}$$

Actually, summing over $\{j_0, \cdots, j_{T-1}\}$, we obtain

$$\mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1, X_T = i) = \left(P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n}\right) \cdot \mathbb{P}(X_T = i),$$

or

$$\mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1 | X_T = i) = P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n},$$

which easily implies the Markov property (4.3), and the independence (4.2). In order to show (4.4), observe that for any $m \geq 0$

$$\begin{aligned} &\mathbb{P}(X_{T+n} = i_n, \cdots, X_{T+1} = i_1, X_{T-1} = j_{T-1}, \cdots, X_0 = j_0, X_T = i, T = m) \\ &\quad = \mathbb{P}(X_{m+n} = i_n, \cdots, X_{m+1} = i_1, X_{m-1} = j_{m-1}, \cdots, X_0 = j_0, X_m = i, T = m) \\ &\quad = \mathbb{P}(X_{m+n} = i_n, \cdots, X_{m+1} = i_1 \mid X_{m-1} = j_{m-1}, \cdots, X_0 = j_0, X_m = i, T = m) \\ &\qquad\quad \cdot \mathbb{P}(X_{m-1} = j_{m-1}, \cdots, X_0 = j_0, X_m = i, T = m) \\ &\quad = \mathbb{P}(X_{m+n} = i_n, \cdots, X_{m+1} = i_1 | X_m = i) \\ &\qquad\quad \cdot \mathbb{P}(X_{m-1} = j_{m-1}, \cdots, X_0 = j_0, X_m = i, T = m) \\ &\quad = P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n} \cdot \mathbb{P}(X_{m-1} = j_{m-1}, \cdots, X_0 = j_0, X_m = i, T = m) \end{aligned}$$

Here the third equality comes from Markov property and the fact that $\{T = m\}$ is determined by $\{X_0, X_1, \cdots, X_m\}$. Summing over m, we complete the proof. ■

4.1 Examples

Example: Consider a simple random walk starting from 1, with

$$X_n = 1 + \sum_{i=1}^{n} Y_i$$

where $\{Y_1, Y_2, \ldots\}$ is a sequence of iid random walks with

$$\mathbb{P}(Y_i = 1) = p, \qquad \mathbb{P}(Y_i = -1) = 1 - p.$$

Define

$$T \doteq \min\{n \geq 0 : \ X_n = 0\}.$$

Find the moment generating function

$$g(s) = g(s;p) \doteq E[s^T]$$

for $0 < s < 1$.

Solution: By conditioning on Y_1, we have

$$g(s) = pE[s^T | Y_1 = 1] + qE[s^T | Y_1 = -1] = pE[s^T | Y_1 = 1] + qs$$

Define

$$T' \doteq \min\{n \geq 1 : \ X_n = 1\}, \qquad T'' \doteq \min\{n \geq T' : \ X_n = 0\}.$$

They are both stopping times, and on $\{Y_1 = 1\}$, we have

$$T = 1 + T' + T''.$$

Therefore

$$g(s) = psE[s^{T'+T''} | Y_1 = 1] + qs = psE^1[s^{T'+T''}] + qs.$$

However, given $T' < \infty$, by strong Markov property T' and T'' are independent, and T'' has the same distribution as T. Therefore

$$\begin{aligned} E^1[s^{T'+T''} | T' < \infty] &= E^1[s^{T'} | T' < \infty] \cdot E^1[s^{T''} | T' < \infty] \\ &= E^1[s^{T'} | T' < \infty] \cdot g(s). \end{aligned}$$

Furthermore, since $0 < s < 1$, on $\{T' = \infty\}$, $s^{T'} = s^{T'+T''} = 0$, it follows that

$$\begin{aligned} E^1[s^{T'+T''}] &= E^1[s^{T'+T''} | T' < \infty]\mathbb{P}^1(T' < \infty) \\ E^1[s^{T'}] &= E^1[s^{T'} | T' < \infty]\mathbb{P}^1(T' < \infty). \end{aligned}$$

Therefore

$$E^1[s^{T'+T''}] = E^1[s^{T'}] \cdot g(s) = [g(s)]^2.$$

Therefore

$$g(s) = ps[g(s)]^2 + q(s)$$

which yields

$$g(s) = \frac{1 \pm \sqrt{1-4pqs^2}}{2ps}.$$

But we should have $g(0+) = 0$, and thus

$$g(s) = \frac{1 - \sqrt{1-4pqs^2}}{2ps}.$$

Note that

$$\mathbb{P}(T < \infty) = g(1-) = \frac{1-\sqrt{1-4pq}}{2p} = \frac{1}{2p}(1 - |p-q|) = \begin{cases} 1 & \text{if } p \leq 1/2, \\ q/p & \text{if } p > 1/2. \end{cases}$$

This coincides with the result that we have obtained in the chapter of "Simple Random Walk". ■

Exercise: For a simple random walk starting from 0, define R to be the first return time to 0. Use the result from the previous example, show that

$$E[s^R] = 1 - \sqrt{1-4pqs^2}.$$

for all $0 < s < 1$.

5 Classification of States

The classification of Markov chain states deals with the basic structure of the Markov chain. It is important since it helps characterize the behavior of the Markov chain in the long run. There are three distinct categories of classification: irreducibility, periodicity, and recurrence and transience.

5.1 Irreducible Markov chains

Let $X = \{X_n\}$ be a Markov chain with state space $\mathcal{S}$ and transition probability matrix $\mathcal{P}$. A state j is said to be **accessible** from state i, denote by $i \to j$ if

$$\mathbb{P}^i(X_n = j, \text{for some } n \geq 0) = \mathbb{P}(\cup_{n=0}^{\infty}\{X_n = j\} | X_0 = i) > 0$$

That is, $i \to j$ if there is positive probability that the Markov chain will reach state j, starting from state i. Two states i and j that are accessible to each other are said to **communicate**, denote by $i \leftrightarrow j$. If all the states communicate with each other, the Markov Chain is said to be **irreducible**.

Lemma 5.1. *For $i \neq j \in \mathcal{S}$, the following statements are equivalent:*

1. *$i \to j$.*
2. *$P_{ij}^{(n)} > 0$ for some $n \geq 0$.*
3. *$P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} j} > 0$ for some n and $i_1, \cdots, i_{n-1} \in \mathcal{S}$.*

Proof: We want to show that (1) ⇒ (2) ⇒ (3) ⇒ (1).

(1) ⇒ (2) : Suppose that $i \to j$. Then

$$\begin{aligned} 0 &< \mathbb{P}(\cup_{n=0}^{\infty}\{X_n = j\} | X_0 = i) \\ &\leq \sum_{n=0}^{\infty} \mathbb{P}(X_n = j | X_0 = i) \\ &= \sum_{n=0}^{\infty} P_{ij}^{(n)}. \end{aligned}$$

(2) ⇒ (3) : By Chapman-Kolmogorov Equation, we have

$$0 < P_{ij}^{(n)} = [\mathcal{P}^n]_{ij} = \sum_{i_1, i_2, \cdots, i_{n-1} \in \mathcal{S}} P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} j}.$$

(3) ⇒ (1) : Observe that

$$\begin{aligned} \mathbb{P}(\cup_{k=0}^{\infty}\{X_k = j\} | X_0 = i) &\geq \mathbb{P}(X_n = j | X_0 = i) \\ &= P_{ij}^{(n)} \\ &\geq P_{ii_1} P_{i_1 i_2} \cdots P_{i_{n-1} j} \\ &> 0. \end{aligned}$$

This completes the proof. ■

Lemma 5.2. *If $i \to j$ and $j \to k$, then $i \to k$.*

Lemma 5.3. *Communication relation is an equivalence relation. That is*

1. *(reflexivity) $i \leftrightarrow i$.*
2. *(symmetry) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$.*
3. *(transitivity) $i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k$.*

Thanks to Lemma 5.3, we can partition the state space into disjoint communication classes such that any two states in the same class communicate with each other, while any two states from different classes do not communicate. It is possible for a Markov chain to start from one class and enter another class with positive probability, but in this case it is not possible for the Markov chain to return to the initial class [why?]. Note that an irreducible Markov chain has only one communication class.

Example: Consider a Markov chain with transition probability matrix

$$\mathcal{P} = \begin{Vmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{Vmatrix}$$

Partition the Markov Chain according to the communication relation.

Solution: The classes are $\{1, 2, 3\}$, $\{4\}$ and $\{5, 6\}$. ■

5.2 Periodicity

The **period** of state i, denoted by $d(i)$, is defined as the *greatest common divisor* of all those $n \geq 1$ such that $P_{ii}^{(n)} > 0$. If $P_{ii}^{(n)} = 0$ for all $n \geq 1$, we define $d(i) = 0$. State i is said to be **aperiodic** if $d(i) = 1$.

Example: Every state in a simple random walk has period 2.

Example: For the example in the last section, $d(5) = d(6) = 2$, $d(1) = d(4) = 1$, and $d(1) = d(2) = d(3) = 1$.

Proposition 5.4. *Periodicity is a class property, that is, if* $i \leftrightarrow j$*, then*

$$d(i) = d(j).$$

Proof: Assume that $i \neq j$. By definition, there exist m and n such that $P_{ij}^{(m)} > 0$ and $P_{ji}^{(n)} > 0$. Fix an arbitrary k such that $P_{ii}^{(k)} > 0$. Since

$$P_{jj}^{(m+n)} \geq P_{ji}^{(n)} P_{ij}^{(m)} > 0, \qquad P_{jj}^{(m+n+k)} \geq P_{ji}^{(n)} P_{ii}^{(k)} P_{ij}^{(m)} > 0,$$

it follows that both $m+n$ and $m+n+k$ are divisible by $d(j)$, hence k is divisible by $d(j)$. But k is arbitrary as long as $P_{ii}^{(k)} > 0$, and thus $d(j)$ is a common divisor of such k's. By definition, $d(i)$ is the greatest common divisor of all such k's. Therefore

$$d(j) \leq d(i).$$

Similarly, we have $d(i) \leq d(j)$. This completes the proof. ■

Remark 5.5. An irreducible Markov Chain is said to be **aperiodic**, if some states (whence all states) have period 1.

The following lemma is a basic property of period. The proof, which requires some knowledge from number theory, is presented only for the sake of completeness.

Lemma 5.6. *Suppose that the state* i *has period* $d(i)$*. Then there exists an* N *such that for* $n \geq N$,

$$P_{ii}^{(nd(i))} > 0.$$

Proof. We should only give the proof for the case of $d(i) = 1$. The general case is essentially the same. By definition of $d(i)$, it is not difficult to see that there exists a finite collection of integers $\{n_1, \ldots, n_k\}$ such that

$$P_{ii}^{(n_j)} > 0$$

for all $j = 1, \ldots, k$ and the greatest common divisor of $\{n_1, \ldots, n_k\}$ is $d(i) = 1$. It follows from number theory that there exist $\{a_1, a_2, \cdots, a_k\} \subseteq \mathbb{Z}$ such that

$$1 = d(i) = a_1 n_1 + a_2 n_2 + \cdots + a_k n_k = n^+ - n^-,$$

where n^+ and n^- are the positive and negative part respectively. Let $N = (n^-)^2$. For any $n \geq N$, we can write

$$n = (n^-)^2 + an^- + b,$$

where $0 \leq b < n^-$. It follows that $n = n^-(n^- + a - b) + bn^+$ and

$$P_{ii}^{(n)} \geq \left[P_{ii}^{(n^-)}\right]^{n^- + a - b} \cdot \left[P_{ii}^{(n^+)}\right]^{b} > 0.$$

We complete the proof. ■

5.3 Recurrence and transience

Consider a Markov chain $X = \{X_n\}$ with transition probability matrix $\mathcal{P}$. For each state i, define

$$f_i \doteq \mathbb{P}^i(R_i < \infty) = \mathbb{P}(R_i < \infty \mid X_0 = i),$$

where R_i is the return time to state i, that is,

$$R_i \doteq \inf\{n \geq 1 : X_n = i\}.$$

If $f_i = 1$, the state i is said to be **recurrent**. If $f_i < 1$, the state i is said to be **transient**.

Theorem 5.7. *A state i is recurrent if and only if*

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty$$

and is transient if and only if

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} < \infty.$$

Furthermore, recurrence and transience are both class properties. That is, if $i \leftrightarrow j$, then i and j are either both recurrent or both transient.

Proof. Let

$$V_i \doteq \sum_{n=0}^{\infty} 1_{\{X_n = i\}} = \{\text{number of visits to state } i\}.$$

Observe that

$$E^i[V_i] = E^i\left[\sum_{n=0}^{\infty} 1_{\{X_n=i\}}\right] = \sum_{n=0}^{\infty} E^i\left[1_{\{X_n=i\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}^i(X_n = i) = \sum_{n=0}^{\infty} P_{ii}^{(n)}.$$

If the state i is recurrent, then the Markov chain will return to state i with probability one, starting at state i. Once it returns to state i, by strong Markov property, the Markov chain will start afresh and will once again return to i with probability one. Repeating this argument, it is not difficult to see that the Markov chain will visit state i infinitely often, or

$$\mathbb{P}^i(V_i = \infty) = 1.$$

This implies that

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = E^i[V_i] = \infty.$$

On the other hand, if state i is transient, then the Markov chain will return to state i with probability f_i and never return with probability $1 - f_i$, starting at state i. Once it return to state i, the Markov Chain will start afresh. Therefore, the Markov Chain will visit state i exactly n times with probability $f_i^{n-1}(1 - f_i)$. That is, for all $n \geq 1$,

$$\mathbb{P}^i(V_i = n) = f_i^{n-1}(1 - f_i).$$

In particular,

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = E^i[V_i] = \frac{1}{1 - f_i} < \infty.$$

It remains to show that recurrence and transience is a class property. It suffices to show that if $i \leftrightarrow j$ and i is transient, then j is transient too. To this end, observe that by the definition of $i \leftrightarrow j$, there exist m and k such that $P_{ij}^{(m)} > 0$, $P_{ji}^{(k)} > 0$. It follows that for all n

$$P_{ii}^{(m+n+k)} \geq P_{ij}^{(m)} P_{jj}^{(n)} P_{ji}^{(k)},$$

which implies that

$$\sum_{n=0}^{\infty} P_{jj}^{(n)} \leq \frac{1}{P_{ij}^{(m)} \cdot P_{ji}^{(k)}} \sum_{n=0}^{\infty} P_{ii}^{(m+n+k)} < \infty,$$

or state j is transient. ■

Lemma 5.8. *Every irreducible Markov chain with finite state space is recurrent.*

Proof: We will argue by contradiction. Suppose that the state space is $\mathcal{S} = \{1, 2, \cdots, n\}$, and the Markov chain is transient. Recall the definition of V_i:

$$V_i \doteq \sum_{n=0}^{\infty} 1_{\{X_n = i\}} = \{\text{number of visits to state } i\}.$$

By transience, for every state i

$$\mathbb{P}^i(V_i < \infty) = 1.$$

For any $1 \leq i \leq n$, define

$$T_i \doteq \inf\{n \geq 0 : X_n = i\}.$$

It follows that, for every i

$$\begin{aligned}
\mathbb{P}(V_i < \infty) &= \mathbb{P}(V_i < \infty, T_i = \infty) + \mathbb{P}(V_i < \infty | T_i < \infty)\mathbb{P}(T_i < \infty) \\
&= \mathbb{P}(T_i = \infty) + \mathbb{P}^i(V_i < \infty) \cdot \mathbb{P}(T_i < \infty) \\
&= \mathbb{P}(T_i = \infty) + \mathbb{P}(T_i < \infty) \\
&= 1.
\end{aligned}$$

However, note that

$$\sum_{i=1}^{n} V_i = \infty.$$

Therefore,

$$1 = \mathbb{P}(V_1 < \infty, \ldots, V_n < \infty) = \mathbb{P}\left(\sum_{i=1}^{n} V_i < \infty\right) = 0,$$

a contradiction. ■

Lemma 5.9. *Suppose that state i is recurrent and state j is accessible from i, that is $i \to j$. Then $i \leftrightarrow j$ and*

$$\mathbb{P}^i(T_j < \infty) = 1, \qquad T_j \doteq \inf\{n \geq 0 : X_n = j\}.$$

Proof: It suffices to show that $\mathbb{P}^i(T_j < \infty) = 1$. Indeed, if this is the case, by the recurrence of state i,

$$1 = \mathbb{P}^i(V_i = \infty) = \mathbb{P}^i(V_i = \infty, T_j < \infty),$$

where V_i is again defined as the number of visits to state i. This necessarily implies that $j \to i$ or $i \leftrightarrow j$ because otherwise

$$\mathbb{P}^i(V_i = \infty, T_j < \infty) = 0,$$

a contradiction.

In order to proceed, observe that there exists $k > 0$ such that $\alpha = P_{ij}^{(k)} > 0$. Define a sequence of stopping times $\{\sigma_0, \sigma_1, \ldots\}$ recursively by

$$\begin{aligned}
\sigma_0 &\doteq 0 \\
\sigma_m &\doteq \inf\left\{n \geq \sigma_{m-1} + k : X_n = i\right\}.
\end{aligned}$$

Since state i is recurrent, it follows that

$$\mathbb{P}^i(\sigma_m < \infty) = 1$$

for all m. It follows from strong Markov property that the events $\{X_{\sigma_m + k} \neq j\}$ are independent and for each m

$$\mathbb{P}^i(X_{\sigma_m + k} \neq j) = 1 - \alpha < 1.$$

Therefore

$$\begin{aligned}\mathbb{P}^i(X_{\sigma_m+k}=j,\ \text{for some } m) &= 1-\mathbb{P}^i(X_{\sigma_m+k}\neq j,\ \text{for all } m)\\ &= 1-(1-\alpha)^\infty\\ &= 1.\end{aligned}$$

This implies that $\mathbb{P}^i(T_j<\infty)=1$ and completes the proof. ■

5.3.1 Examples

Example: For the example in Section 5.1, class $\{5,6\}$ is recurrent. Class $\{4\}$ is transient since the chain will eventually leave 4 to 5 and never come back. Class $\{1,2,3\}$ is also transient, since it will eventually leave 3 and never come back. ■

Example: The simple random walk on $\mathbb{Z}$ is clearly a irreducible Markov Chain. Is this Markov chain recurrent or transient?

Solution: The solution depends on the so-called **Stirling formula**, which states that for any positive integer $n\geq 2$,

$$n! = \sqrt{2\pi n}\cdot n^n e^{-n+\varepsilon_n}, \qquad \frac{1}{12n+1}<\varepsilon_n<\frac{1}{12n}.$$

In particular, when n is large, we have

$$n!\approx\sqrt{2\pi n}n^n e^{-n}.$$

The transition probabilities of this Markov chain are

$$P_{00}^{(2n)}=\binom{2n}{n}(pq)^n, \qquad P_{00}^{2n+1}=0.$$

By Stirling formula

$$\binom{2n}{n}=\frac{(2n)!}{n!n!}\approx\frac{\sqrt{4\pi n}\cdot(2n)^{2n}e^{-2n}}{\left(\sqrt{2\pi n}n^n e^{-n}\right)^2}=\frac{4^n}{\sqrt{\pi n}},$$

and thus when n is large

$$P_{00}^{(2n)}\approx\frac{(4pq)^n}{\sqrt{\pi n}}$$

It follows that

$$\sum_{n=0}^{\infty}P_{00}^{(n)}=\infty \text{ if and only if } p=q=\frac{1}{2}.$$

Therefore the simple random walk on $\mathbb{Z}$ is recurrent if and only if it is symmetric. This coincides with the results obtained in the chapter of "Simple Random Walk." ■

6 Stationary Distribution

Suppose $\{X_0, X_1, X_2, \cdots\}$ is a Markov chain with transition probability matrix $\mathcal{P}$ and state space $\mathcal{S}$. Recall the return time to state i,

$$R_i \doteq \inf\{n \geq 1 : X_n = i\}.$$

A recurrent state i is said to be **positive recurrent** if $E^i[R_i]$ is finite, and **null recurrent** if otherwise.

Stationary Distribution: A *stationary distribution* of a Markov chain with transition probability matrix $\mathcal{P}$ is a (row) vector $\pi = \{\pi_i\}_{i \in \mathcal{S}}$ such that

$$\pi_i \geq 0, \quad \sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi\mathcal{P} = \pi.$$

Theorem 6.1. *Let $X = \{X_n\}$ be an irreducible Markov chain with transition probability matrix $\mathcal{P}$. The following statements are equivalent:*

1. *All the states are positive recurrent.*
2. *Some state is positive recurrent.*
3. *There exists a stationary distribution π, i.e., system of equations*

$$\pi_i \geq 0, \quad \sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi\mathcal{P} = \pi$$

admits a solution.

Furthermore, if X is positive recurrent, then the stationary distribution π is unique and satisfies

$$\pi_i = \frac{1}{E^i[R_i]}.$$

Proof: For equivalence, it suffices to argue that "(2) ⇒ (3) ⇒ (1)".

"(2) ⇒ (3)". Suppose that state k is positive recurrent. So state k is at least recurrent. Since the chain is irreducible and recurrence is a class property, every state of the chain is recurrent. Recall the definition of the return time R_k. Note that the recurrence of state k implies that $P^k(R_k < \infty) = 1$. Let

$$\nu_i \doteq E^k\left[\sum_{n=1}^{R_k} 1_{\{X_n = i\}}\right].$$

for every $i \in \mathcal{S}$. It follows that

$$\nu_k = 1, \qquad \sum_i \nu_i = E^k \left[\sum_i \sum_{n=1}^{R_k} 1_{\{X_n = i\}} \right] = E^k[R_k] < \infty.$$

We claim that the (row) vector $\nu = \{\nu_i\}_{i \in \mathcal{S}}$ satisfies the equation

$$\nu \cdot \mathcal{P} = \nu.$$

Indeed, note the following identity

$$\sum_{n=1}^{R_k} 1_{\{X_n = i\}} = \sum_{n=1}^{\infty} 1_{\{X_n = i, R_k \geq n\}},$$

and whence

$$\begin{aligned}
\nu_j &= E^k \left[\sum_{n=1}^{\infty} 1_{\{X_n = j, R_k \geq n\}} \right] \\
&= E^k \left[\sum_{n=1}^{\infty} \sum_i 1_{\{X_n = j, X_{n-1} = i, R_k \geq n\}} \right] \\
&= \sum_i \sum_{n=1}^{\infty} \mathbb{P}(X_n = j, X_{n-1} = i, R_k \geq n) \\
&= \sum_i \sum_{n=1}^{\infty} \mathbb{P}(X_n = j | X_{n-1} = i, R_k \geq n) \cdot \mathbb{P}(X_{n-1} = i, R_k \geq n) \\
&= \sum_i \sum_{n=1}^{\infty} P_{ij} \cdot \mathbb{P}(X_{n-1} = i, R_k \geq n) \\
&= \sum_i P_{ij} E^k \left[\sum_{n=1}^{R_k} 1_{\{X_{n-1} = i\}} \right] \\
&= \sum_i \nu_i P_{ij}
\end{aligned}$$

Define $\pi = \{\pi_i\}_{i \in \mathcal{S}}$ by

$$\pi_i \doteq \frac{\nu_i}{E^k[R_k]}.$$

Then π is a stationary distribution.

"(3) $\Rightarrow$ (1)". Suppose $\pi = \{\pi_i\}$ is an arbitrary stationary distribution. We first show that $\pi_i > 0$ for all i. Indeed, there exists a j such that $\pi_j > 0$. But $i \leftrightarrow j$, therefore there exist $m > 0$ such that $P_{ji}^{(m)} > 0$. Since

$$\pi = \pi \cdot \mathcal{P} = \pi \cdot \mathcal{P}^2 = \cdots = \pi \cdot \mathcal{P}^m,$$

it follows that

$$\pi_i = \sum_k \pi_k P_{ki}^{(m)} \geq \pi_j P_{ji}^{(m)} > 0.$$

Arbitrarily fix an $k \in \mathcal{S}$ and define a vector $z = \{z_i : i \in \mathcal{S}\}$ by $z_i \doteq \pi_i/\pi_k$. It follows that z satisfies the equation $z = z\mathcal{P}$ and $z_k = 1$. Therefore, for every state j,

$$\begin{aligned} z_j &= \sum_i z_i P_{ij} \\ &= P_{kj} + \sum_{i \neq k} z_i P_{ij} \\ &= P_{kj} + \sum_{i \neq k} (P_{ki} + \sum_{l \neq k} z_l P_{li}) P_{ij} \\ &= P_{kj} + \sum_{i \neq k} P_{ki} P_{ij} + \sum_{i \neq k} \sum_{l \neq k} z_l P_{li} P_{ij} \\ &= \cdots \end{aligned}$$

However, observing that

$$P_{kj} = \mathbb{P}^k(X_1 = j) = \mathbb{P}^k(X_1 = j, R_k \geq 1)$$

$$\sum_{i \neq k} P_{ki} P_{ij} = \sum_{i \neq k} \mathbb{P}^k(X_1 = i, X_2 = j) = \mathbb{P}^k(X_2 = j, R_k \geq 2).$$

Repeating the argument, we have that

$$z_j \geq \sum_{n=1}^{\infty} \mathbb{P}^k(X_n = j, R_k \geq n) = E^k \left[\sum_{n=1}^{R_k} 1_{\{X_n = j\}} \right]$$

for every j. It follows that

$$E^k[R_k] = \sum_j E^k \left[\sum_{n=1}^{R_k} 1_{\{X_n = j\}} \right] \leq \sum_j z_j = \frac{1}{\pi_k} \sum_j \pi_j = \frac{1}{\pi_k} < \infty.$$

Therefore state k is positive recurrent. But k is arbitrary, thus (1) holds.

It remains to show that, when the chain is positive recurrent, $\hat{\pi} = \{\hat{\pi}_i\}$ where

$$\hat{\pi}_i = \frac{1}{E^i[R_i]}$$

is the unique stationary distribution. Suppose that $\pi = \{\pi_i\}$ is an arbitrary stationary distribution. We wish to show that $\pi = \hat{\pi}$. To this end, fix an arbitrary state k and let $z = \{z_i\}$ where $z_i = \pi_i/\pi_k$. Also recall the definition of vector

$\nu = \{\nu_i\}$ from the proof "(2) $\Rightarrow$ (3)" and that $\nu = \nu\mathcal{P}$. The proof of "(3) $\Rightarrow$ (1)" implies that

$$z_j \geq E^k \left[\sum_{n=1}^{R_k} 1_{\{X_n = j\}} \right] = \nu_j.$$

for every j. Let $u \doteq z - \nu$. It follows that $u_i \geq 0$ for all i and $u\mathcal{P} = u$. Moreover,

$$u_k = z_k - \nu_k = 1 - 1 = 0.$$

Now for any state j, there exists m such that

$$P_{jk}^{(m)} > 0.$$

Therefore, since $u = u\mathcal{P}^m$ and $u \geq 0$,

$$0 = u_k \geq u_j P_{jk}^{(m)} \geq 0,$$

which implies $u_j = 0$. Since j is arbitrary, we have $u = 0$ or $z = \nu$. In particular,

$$\frac{1}{\pi_k} = \sum_j z_j = \sum_j \nu_j = E^k[R_k] = \frac{1}{\hat{\pi}_k}.$$

Since k is arbitrary, we complete the proof. ■

Corollary 6.2. *Positive recurrence and null recurrence are both class properties.*

Proof: It suffices to show that positive recurrence is a class property. Suppose $i \leftrightarrow j$ and state i is positive recurrent. Define $\mathcal{S}^* = \{k \in \mathcal{S} : i \to k\}$. Thanks to Lemma 5.9, it follows that every pair of states in $\mathcal{S}^*$ communicate. Moreover, if $k \notin \mathcal{S}^*$, k is not accessible from i. Consequently, if the Markov chain starts from i, it will stay is $\mathcal{S}^*$, which implies that the chain $\{X_0, X_1, X_2, \cdots\}$ is an irreducible Markov chain with state space $\mathcal{S}^*$. Furthermore, it is positive recurrent. From Theorem 6.1, every state in $\mathcal{S}^*$ is positive recurrent. In particular, state j is. This complete the proof. ■

Finally we give an necessary and sufficient condition for transience to complete the classification.

Theorem 6.3. *Let $X = \{X_n\}$ be an irreducible Markov chain with transition probability matrix $\mathcal{P}$ and infinite state space $\mathcal{S}$. Then the following statements are equivalent:*

1. *The Markov chain is transient.*

2. *For any state, say i^*, the system of equations*

$$\sum_{j \in \mathcal{S}} P_{ij} x_j = x_i, \quad \forall i \neq i^*$$

   admits a bounded non-constant solution.

3. *For some state, say i^*, the system of equations*

$$\sum_{j \in \mathcal{S}} P_{ij} x_j = x_i, \quad \forall i \neq i^*$$

   admits a bounded non-constant solution.

Proof: "(1) $\Rightarrow$ (2)". Fix any state $i^* \in \mathcal{S}$. Let $T^* \doteq T_{i^*}$ the first hitting time to the state i^*, i.e.

$$T^* \doteq \min\{n \geq 0: \; X_n = i^*\},$$

and define for every $i \in \mathcal{S}$

$$f_i \doteq \mathbb{P}^i(T^* < \infty).$$

We have $f_i \leq 1$ for every i and $f_{i^*} = 1$. By the law of total probability, $f \doteq \{f_i\}$ is a solution of the system equation. It suffices to show that $f_i < 1$ for some i. This is trivial since otherwise we have

$$\begin{aligned} \mathbb{P}^{i^*}(\text{ever return to state } i^*) &= \sum_i P_{i^*i} \cdot \mathbb{P}^i(\text{ever return to state } i^*) \\ &= \sum_i P_{i^*i} \\ &= 1, \end{aligned}$$

which implies that the state i^* is recurrent, a contradiction.

Since "(2) $\Rightarrow$ (3)" is trivial, it remains to show "(3) $\Rightarrow$ (1)". Define a new transition probability matrix $\bar{\mathcal{P}} = [\bar{P}_{ij}]$ by

$$\bar{P}_{ij} \doteq \begin{cases} P_{ij} & \text{if } i \neq i^* \\ 1 & \text{if } i = j = i^* \\ 0 & \text{if } i = i^*, j \neq i^*. \end{cases}$$

In other words, under this new transition probability matrix, the Markov chain follows the original dynamics except that state i^* now is an absorbing state. Suppose $x = \{x_i\}$ is a bounded non-constant solution. Regarding x as a column vector, we have

$$\bar{\mathcal{P}} x = x$$

and therefore

$$\bar{\mathcal{P}}^n \cdot x = x$$

for all $n \geq 1$. We will now argue by contradiction and assume that the original Markov chain is recurrent, which by Lemma 5.9 implies that

$$1 = P^i(T^* < \infty) = \lim_n \mathbb{P}^i(T^* \leq n)$$

for all state i. It is not difficult to see that

$$\bar{P}^{(n)}_{ii^*} \geq \mathbb{P}^i(T^* \leq n),$$

and whence

$$\lim_n \bar{P}^{(n)}_{ii^*} = 1.$$

Assume all x_i's are bounded by M. It follows that

$$\sum_{j \neq i^*} \bar{P}^{(n)}_{ij} x_j \leq M \sum_{j \neq i^*} \bar{P}^{(n)}_{ij} = M(1 - \bar{P}^{(n)}_{ii^*}) \to 0,$$

which yields

$$x_i = \sum_{j \neq i^*} \bar{P}^{(n)}_{ij} x_j + \bar{P}^{(n)}_{ii^*} x_{i^*} \to x_{i^*}.$$

Therefore $x_i \equiv x_{i^*}$ for every i, a contradiction. ■

Remark 6.4. Theorems 6.1 and 6.3 completely characterize the recurrence and transience properties of an irreducible Markov chain in terms of the existence of solutions to suitable systems of linear equations. Note that if the chain is not irreducible, one can first divide the states into disjoint classes according to the communication relation. If a class, say $\mathcal{A}$, has some positive probability to move to another class, all the states in class $\mathcal{A}$ are transient. Otherwise, we can restrict the Markov chain to the states in $\mathcal{A}$ (whence it becomes an irreducible chain) and use Theorems 6.1 and 6.3 to classify the states in $\mathcal{A}$.

7 Convergence of Markov chains

In this section we discuss the convergence of the distributions of Markov chains. The most important result is the convergence of irreducible aperiodic positive recurrent chains to stationary distributions, i.e., *ergodicity*. The convergence results on null recurrent and transient chains are included for completeness.

7.1 Convergence for positive recurrent chains

Theorem 7.1. *Let $X = \{X_n\}$ be an irreducible aperiodic positive recurrent Markov chain with transition probability matrix $\mathcal{P}$ and arbitrary initial distribution Λ. Denote by π the stationary distribution of X. Then*

$$\lim_n \mathbb{P}[X_n = j] = \pi_j$$

for every j. In particular,

$$\lim_n P_{ij}^{(n)} = \pi_j$$

for all states i and j.

The proof of this theorem uses a technique called *coupling.* We need the following lemma.

Lemma 7.2. *Suppose that $X = \{X_n\}$ and $Y = \{Y_n\}$ are two independent irreducible aperiodic recurrent Markov chains. Then the process $Z_n \doteq (X_n, Y_n)$ is an irreducible aperiodic Markov chain. If in addition X and Y are both positive recurrent. Then Z is also positive recurrent.*

Proof. Let $[P_{ik}]$ and $[Q_{jl}]$ denote the transition probability matrix for X and Y respectively. The Markov property of Z is clear, and the new transition probabilities are

$$R_{(i,j)(k,l)} = P_{ik} Q_{jl}$$

We now show that Z is irreducible and aperiodic. Fix any state of Z, say (i, j). Thanks to Lemma 5.6, there exists N such that

$$P_{ii}^{(n)} > 0, \quad Q_{jj}^{(n)} > 0$$

for all $n \geq N$. Thus

$$R_{(i,j)(i,j)}^{(n)} = P_{ii}^{(n)} Q_{jj}^{(n)} > 0$$

for all $n \geq N$. In particular, the state (i, j) has period 1. Furthermore, for an arbitrary state (k, l), by the irreducibility of X and Y, there exist m_1 and m_2 such that

$$P_{ik}^{(m_1)} > 0, \quad Q_{jl}^{(m_2)} > 0.$$

Therefore, for

$$P_{ik}^{(N+m_1+m_2)} \geq P_{ii}^{(N+m_2)} P_{ik}^{(m_1)} > 0$$

and

$$Q_{jl}^{(N+m_1+m_2)} \geq Q_{jj}^{(N+m_1)} Q_{jl}^{(m_2)} > 0,$$

which implies that

$$R_{(i,j)(k,l)}^{(N+m_1+m_2)} > 0.$$

In other words, $(i,j) \to (k,l)$. Therefore the chain Z is irreducible.

It remains to show that Z is positive recurrent if both X and Y are. Assume that the stationary distributions for X and Y are $\mu = \{\mu_i\}$ and $\nu = \{\nu_j\}$ respectively. Define $\pi = \{\pi_{(i,j)} = \mu_i \nu_j\}$. Observe that π is a stationary distribution of Z since

$$\sum_{(i,j)} \pi_{(i,j)} = 1, \qquad \sum_{(i,j)} \pi_{(i,j)} R_{(i,j)(k,l)} = \sum_i \mu_i P_{ik} \sum_j \nu_j Q_{jl} = \mu_k \nu_l = \pi_{(k,l)}$$

By Theorem 6.1, we complete the proof. ■

Proof of Theorem 7.1. Introduce a new Markov chain $Y = \{Y_n\}$ that is independent of X and has the same transition probability matrix $\mathcal{P}$ but with initial distribution π. Therefore the distribution of Y_n is always π. Define $Z_n \doteq (X_n, Y_n)$. Lemma 7.2 implies that Z is itself an irreducible aperiodic positive recurrent Markov chain. Define the stopping time

$$T \doteq \inf \{ n \geq 0 : X_n = Y_n \} .$$

Clearly $\mathbb{P}(T < \infty) = 1$ by the recurrence of the chain Z. Define a new chain

$$W_n \doteq \begin{cases} X_n & \text{if } n < T, \\ Y_n & \text{if } n \geq T. \end{cases}$$

We show that $W = \{W_n\}$ is a Markov chain with transition probability matrix $\mathcal{P}$ and initial distribution Λ. Indeed, we can see that

$$\begin{aligned} \mathbb{P}(W_{n+1} = j | W_n = i, W_{n-1}, \cdots, W_0, T \leq n) &= \mathbb{P}(Y_{n+1} = j | Y_n = i) = P_{ij} \\ \mathbb{P}(W_{n+1} = j | W_n = i, W_{n-1}, \cdots, W_0, T > n) &= \mathbb{P}(X_{n+1} = j | X_n = i) = P_{ij}, \end{aligned}$$

which easily implies the Markov property

$$\mathbb{P}(W_{n+1} = j | W_n = i, W_{n-1}, \cdots, W_0) = P_{ij}.$$

It follows that X_n and W_n have the same distribution, or

$$\mathbb{P}(X_n = j) = \mathbb{P}(W_n = j) = \mathbb{P}(X_n = j, n < T) + \mathbb{P}(Y_n = j, n \geq T)$$

and

$$\begin{aligned} |\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \\ &= |\mathbb{P}(X_n = j, n < T) - \mathbb{P}(Y_n = j, n < T)| \\ &\leq \mathbb{P}(n < T) \\ &\to 0 \end{aligned}$$

Therefore $\mathbb{P}(X_n = j) \to \pi_j$. ■

Theorem 7.3. *Suppose $X = \{X_n\}$ is an irreducible aperiodic positive recurrent Markov chain with state space $\mathcal{S}$ and stationary distribution π. Then for any function h defined on the state space $\mathcal{S}$,*

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} h(X_k) = \sum_{i \in \mathcal{S}} h(i)\pi_i$$

with probability one.

Proof: Without loss of generality we assume $h \geq 0$. For general case we can write $h = h^+ - h^-$ and show for h^+ and h^- respectively. Fix an arbitrary state say $i^* \in \mathcal{S}$. Since we are only concerned with the long run average and the chain will eventually hits state i^*, we can assume that $X_0 = i^*$.

Define a sequence of strictly increasing stopping times by

$$\begin{aligned} R_0^* &\doteq 0 \\ R_m^* &\doteq \inf\left\{n > R_{m-1}^* : X_n = i^*\right\}. \end{aligned}$$

In other words, R_m^* is the m-th return time to the state i^*. It follows from the recurrence of X and strong Markov property that

$$\mathbb{P}(R_m^* < \infty) = 1$$

for all m, that the sequence $\{R_1^* - R_0^*, R_2^* - R_1^*, \cdots\}$ is iid, and that the sequence $\{Y_m : m \geq 0\}$ is also iid where

$$Y_m \doteq \sum_{n=R_m^*+1}^{R_{m+1}^*} h(X_n).$$

Observe that for all $m \geq 1$, by Theorem 6.1

$$E[R_{m+1}^* - R_m^*] = E^{i^*}[R_1^*].$$

Furthermore,

$$E[Y_m] = E^{i^*}\left[\sum_{n=1}^{R_1^*} \sum_j h(j) 1_{\{X_n = j\}}\right] = \sum_j h(j) E^{i^*}\left[\sum_{n=1}^{R_1^*} 1_{\{X_n = j\}}\right].$$

However, by the proof of Theorem 6.1,

$$E^{i^*}\left[\sum_{n=1}^{R_1^*} 1_{\{X_n = j\}}\right] = \nu_j = \pi_j E^{i^*}[R_1^*]$$

and therefore

$$E[Y_m] = E^{i^*}[R_1^*] \sum_j h(j)\pi_j.$$

For each $n \geq 1$, let σ_n be such that

$$R_{\sigma_n}^* \leq n < R_{\sigma_n+1}^*,$$

or equivalently,

$$\sum_{m=0}^{\sigma_n-1} [R_{m+1}^* - R_m^*] \leq n < \sum_{m=0}^{\sigma_n} [R_{m+1}^* - R_m^*]$$

By law of large numbers and that $\sigma_n \to \infty$ as $n \to \infty$, it is easy to see that

$$\lim_n \frac{\sigma_n}{n} = \frac{1}{E^{i^*}[R_{m+1}^* - R_m^*]} = \frac{1}{E^{i^*}[R_1^*]}.$$

However, by the nonnegativity of h,

$$\frac{1}{n} \sum_{m=0}^{\sigma_n-1} Y_m \leq \frac{1}{n} \sum_{k=1}^{n} h(X_k) \leq \frac{1}{n} \sum_{m=0}^{\sigma_n} Y_m.$$

By law of large numbers again,

$$\lim_n \frac{1}{\sigma_n} \sum_{m=0}^{\sigma_n-1} Y_m = \lim_n \frac{1}{\sigma_n} \sum_{m=0}^{\sigma_n} Y_m = E[Y_m] = E^{i^*}[R_1^*] \sum_j h(j)\pi_j.$$

Therefore,

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} h(X_k) = \lim_n \frac{\sigma_n}{n} E^{i^*}[R_1^*] \sum_j h(j)\pi_j = \sum_j h(j)\pi_j.$$

This completes the proof. ■

Remark 7.4. A special case of Theorem 7.3 is with

$$h(x) \doteq 1_{\{x=i\}}$$

where i is an arbitrarily fixed state. It follows that

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} 1_{\{X_k=i\}} = \pi_i.$$

That is, in the long run the Markov chain spend π_i proportion of time at state i.

7.2 Convergence for null recurrent and transient chains

Theorem 7.5. *Let $X = \{X_n\}$ be an irreducible aperiodic Markov chain that is either transient or null recurrent. Then*

$$\lim_n \mathbb{P}^i[X_n = j] = \lim_n P_{ij}^{(n)} = 0$$

for all states i and j.

Proof: We first show for the case where X is transient. Fix any two states i and j. Since $j \to i$, there exists m such that $P_{ji}^{(m)} > 0$. Therefore, by theorem 5.7,

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} P_{ji}^{(m)} \leq \sum_{n=1}^{\infty} P_{ii}^{(n+m)} < \infty,$$

which implies that

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} < \infty,$$

for all i and j. The claim follows readily.

Now assume the Markov chain is null-recurrent. We will use the coupling method once again here. Let $Y = \{Y_n\}$ be an independent Markov chain with the same transition probability matrix $\mathcal{P}$. Consider the process $Z_n \doteq (X_n, Y_n)$. It follows from Lemma 7.2 that Z_n is itself an irreducible aperiodic Markov chain. If W is transient, then we have for any states i and j,

$$0 = \lim_n P_{(i,i)(j,j)}^{(n)} = \lim_n [P_{ij}^{(n)}]^2,$$

and we completes the proof.

Suppose for now that Z is recurrent. We will argue by contradiction and assume that there exist two states i^* and j^* such that

$$\lim_n P_{i^*j^*}^{(n)} = 0$$

does *not* hold. Define the vector $v^{(n)} \doteq \{P_{i^*j}^{(n)} : j \in \mathcal{S}\}$. A standard Cantor-diagonal method yields that there exists a sequence, say $\{n_1, n_2, n_3, \cdots\} \in \mathbb{N}$, such that

$$\lim_k v^{(n_k)} = v \neq 0.$$

Assume that $X_0 = i^*$ and Y_0 has the same distribution as X_1, that is,

$$\mathbb{P}(Y_0 = j) = P_{i^*j}.$$

Define

$$T = \inf\{n \geq 0 : X_n = Y_n\}.$$

By the recurrence of Z, T is finite with probability one. The coupled process is $W = \{X_0, X_1, \cdots, X_{T-1}, Y_T, Y_{T+1}, \cdot\}$, which is a Markov chain with initial condition $W_0 = i^*$ and the same transition probability matrix $\mathcal{P}$. Furthermore, we can similarly argue that

$$\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j) \to 0$$

for every state j. In particular

$$\mathbb{P}(X_{n_k} = j) - \mathbb{P}(Y_{n_k} = j) \to 0.$$

However, it is clear the distribution of X_{n_k} is $v^{(n_k)}$ and that of Y_{n_k} is $v^{(n_k+1)} = v^{(n_k)}\mathcal{P}$. Therefore

$$\lim_k v^{(n_k)}\mathcal{P} = \lim_k v^{(n_k)} = v.$$

This implies that $v\mathcal{P} \le v$ by Fatou Lemma. It follows that

$$\sum_i v_i \ge \sum_i [v\mathcal{P}]_i = \sum_i \sum_j v_j P_{ji} = \sum_j v_j$$

which in turn implies that $v = v\mathcal{P}$. Since v is non-zero and by Fatou Lemma

$$\sum_i v_i \le 1,$$

the vector

$$\frac{v}{\sum_i v_i}$$

defines a stationary distribution. This is a contradiction. ■

Corollary 7.6. *An irreducible Markov Chain with finite state space is positive recurrent.*

Proof: Note that an irreducible Markov chain with finite state space is recurrent by Lemma 5.8. So we only need to show that it is not null recurrent. We will argue by contradiction and assume that the Markov chain is null recurrent.

First we assume that the chain is aperiodic. Theorem 7.5 implies that

$$\lim_n P_{ij}^{(n)} = 0$$

for all i and j, and thus

$$1 = \sum_j P_{ij}^{(n)} \to 0,$$

a contradiction.

Assume for now that the chain has period d. Consider the process $Y = \{Y_n\}$ with $Y_n = X_{nd}$. The process Y is again a Markov chain, but with transition probability matrix $\mathcal{P}^n$. It is easy to see that Y is aperiodic and every state is still null recurrent. However, Y may not be irreducible. Fix an arbitrary state, say i^*.

Let $\mathcal{C}$ be the communicating class that contains i^*. Assume for now that $Y_0 = i^*$. Since every state in $\mathcal{C}$ is null recurrent, $Y_n \in \mathcal{C}$ for all n, thanks to Lemma 5.9. Therefore $Y = \{Y_n\}$ with state space $\mathcal{C}$ is an irreducible aperiodic Markov chain with finite state space. A contradiction. ■

Lemma 7.7. (Fatou Lemma). *Suppose that* $a_i^{(n)} \geq 0$ *for every* i *and every* n. *Then*

$$\liminf_n \sum_{i=1}^{\infty} a_i^{(n)} \geq \sum_{i=1}^{\infty} \liminf_n a_i^{(n)}$$

Proof: For any $N > 0$, we have

$$\liminf_n \sum_{i=1}^{\infty} a_i^{(n)} \geq \liminf_n \sum_{i=1}^{N} a_i^{(n)} = \sum_{i=1}^{N} \liminf_n a_i^{(n)}.$$

Letting $N \to \infty$ we complete the proof. ■

8 Examples

Example: Suppose that $X = \{X_n\}$ is a Markov chain with $\mathcal{S} = \{1, 2, 3\}$ with transition probability matrix

$$\mathcal{P} = \left\| \begin{array}{ccc} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{array} \right\|$$

1. Classify this Markov chain.
2. Find the stationary distribution.
3. Suppose the chain start in state 1. What is the expected time until it is in state 1 again?
4. What is the proportion of time in the long run that the chain is in state 1?

Solution: This is a simple example.

1. The chain is irreducible and aperiodic with finite state space, whence positive recurrent.
2. The stationary distribution $\pi = (1/3, 1/3, 1/3)$ by symmetry.
3. The expected time is just $1/\pi_1 = 3$. One can also use first step analysis to obtain the expected time.
4. The proportion of time in the long run that the chain is at state 1 is just $\pi_1 = 1/3$. ■

Example: Let $X = \{X_n\}$ be a Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ and transition probability matrix

$$\mathcal{P} = \left\| \begin{array}{ccccc} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 4/5 \\ 0 & 0 & 0 & 2/5 & 3/5 \\ 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{array} \right\|$$

1. Classify the Markov chain.
2. Find the stationary distribution.
3. Suppose the chain starts in state 1. What is the expected time until it returns to state 1 again?
4. Suppose the chain starts in state 1. What is the expected number of steps until it is in state 4?
5. Suppose the chain starts in state 1. What is the probability that the chain will enter state 5 before it enters state 3?

Solution:

1. The chain is irreducible, aperiodic, and positive recurrent.

2. Solve the system of equations
$$\pi \mathcal{P} = \pi, \qquad \sum_i \pi_i = 1$$
to obtain the stationary distribution $\pi = \{\pi_i\}$ where
$$\pi_1 = \frac{10}{37}, \quad \pi_2 = \pi_3 = \frac{5}{37}, \quad \pi_4 = \frac{3}{37}, \quad \pi_5 = \frac{14}{37}$$
3. The expected return time is just $1/\pi_1 = 37/10$.
4. The expected return time to state 4 starting at 4 is $1/\pi_4 = 37/3$. But from state 4, after first step the Markov chain always hits state 1. Therefore the expected time of hitting state 4 starting at 1 is $37/3 - 1 = 34/3$. One can also use first step analysis to solve this problem.
5. We will use first step analysis. Define
$$h_i \doteq \mathbb{P}^i(\text{The Markov chain hits state 5 before it hits state 3}).$$
Hence $h_5 = 1$, $h_3 = 0$. By conditioning on the first step, we have:
$$\begin{aligned} h_1 &= \frac{1}{2} h_2, \\ h_2 &= \frac{1}{5} h_4 + \frac{4}{5}, \\ h_4 &= h_1, \end{aligned}$$
which yields $h_1 = 4/9$, $h_2 = 8/9$, $h_4 = 4/9$. ■

Example: Consider a simple random walk on the half-line $\mathcal{S} = \{0, 1, 2, \cdots\}$ with a reflecting barrier at 0. The transition probability matrix is
$$\mathcal{P} = \left\| \begin{array}{cccccc} 0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & 0 & p_2 & 0 & \cdots \\ 0 & 0 & q_3 & 0 & p_3 & \cdots \\ & \cdots & & \cdots & & \cdots \end{array} \right\|$$
where $p_0 = 1$ and for every $i \geq 1$,
$$p_i + q_i = 1, \quad p_i > 0, \ q_i > 0.$$
Classify this Markov chain.

Answer: Clearly the chain is irreducible and aperiodic. The Markov Chain would be positive recurrent if the following system of equations

$$\pi \mathcal{P} = \pi, \quad \sum_i \pi_i = 1, \quad \pi_i \geq 0$$

admits a solution. Plugging the form of $\mathcal{P}$, the equation of $\pi \mathcal{P} = \pi$ amounts to

$$\pi_0 = \pi_1 q_1, \quad \pi_j = \pi_{j-1} p_{j-1} + \pi_{j+1} q_{j+1}, \quad j \geq 1.$$

Rewrite this equation to arrive at

$$\pi_0 = \pi_1 q_1, \quad \pi_{j+1} q_{j+1} - \pi_j q_j = \pi_j p_j - \pi_{j-1} p_{j-1}, \quad j \geq 1.$$

Summing over j, we have

$$\pi_{j+1} q_{j+1} - \pi_1 q_1 = \pi_j p_j - \pi_0 p_0 = \pi_j p_j - \pi_1 q_1$$

or equivalently $\pi_{j+1} q_{j+1} = \pi_j p_j$ which yields

$$\pi_{j+1} = \pi_j \frac{p_j}{q_{j+1}} = \cdots = \pi_0 \frac{p_0 p_1 \cdots p_j}{q_1 q_2 \cdots q_{j+1}}.$$

Therefore, the chain will be positive recurrent if and only if

$$\sum_{j=1}^{\infty} \frac{p_0 p_1 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} < \infty.$$

Now pick a state, say 0. Consider the system of equations

$$\sum_{j=0}^{\infty} P_{ij} x_j = x_i, \quad i \geq 1,$$

which amounts to

$$x_i = x_{i+1} p_i + x_{i-1} q_i, \quad i \geq 1.$$

This implies that

$$x_{i+1} - x_i = \frac{q_i}{p_i}(x_i - x_{i-1}) = \cdots = \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1} \cdot (x_1 - x_0)$$

and therefore

$$x_{i+1} - x_0 = (x_1 - x_0) \cdot \sum_{j=0}^{i} \frac{q_1 q_2 \cdots q_j}{p_1 p_2 \cdots p_j}.$$

Therefore, the system of equation has a non-constant bounded solution if and only if

$$\sum_{j=0}^{\infty} \frac{q_1 q_2 \cdots q_j}{p_1 p_2 \cdots p_j} < \infty,$$

in which case the chain is transient.

To summarize, we have the following complete classification of the Markov chain.

1. The chain is transient if

$$\sum_{j=0}^{\infty} \frac{q_1 q_2 \cdots q_j}{p_1 p_2 \cdots p_j} < \infty.$$

2. The chain is recurrent if

$$\sum_{j=0}^{\infty} \frac{q_1 q_2 \cdots q_j}{p_1 p_2 \cdots p_j} = \infty.$$

   (a) The chain is positive recurrent if

$$\sum_{j=1}^{\infty} \frac{p_0 p_1 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} < \infty.$$

   (b) The chain is null-recurrent if

$$\sum_{j=1}^{\infty} \frac{p_0 p_1 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} = \infty.$$

Example: Suppose that during each time period, every member of a population independently dies with probability $p > 0$, and also that the number of new members that join the population in each time period is an independent Poisson random variable with rate λ. If we let X_n be the population size at the beginning of time period n, then $X = \{X_n\}$ is a Markov chain with state space $\mathcal{S} = \{0, 1, \ldots\}$. Classify this Markov chain and compute the stationary distribution if there exists one.

Solution: The Markov chain is clearly irreducible and aperiodic. Consider the case where X_n has a Poisson distribution with parameter α. It follows that X_{n+1} is a Poisson random variable with parameter $\alpha(1-p) + \lambda$. If we set

$$\alpha = \alpha(1-p) + \lambda \quad \text{or} \quad \alpha = \lambda/p,$$

then the distribution of X_n coincides with that of X_{n+1}. Therefore Poisson distribution with parameter λ/p is a stationary distribution. It follows that X is positive recurrent and Poisson with parameter λ/p is the unique stationary distribution.■

Example: Assuming the coin is fair, what is the expected number of tosses needed to obtain the pattern HHT?

Solution: Construct a Markov chain with four states $\mathcal{S} = \{0, H, HH, HHT\}$. 0 represents a starting point which says all the previous tosses are irrelevant in the quest of HHT; H represents a single heads on the last toss; HH represents two successive heads on the last two tosses; HHT is the pattern we are looking for. For each sequence of tosses, let X_n be the state at n-th toss. For example, suppose that the outcome of the first 8 tosses is

$$HTTHHHHT\cdots$$

Then the corresponding state process $X = \{X_n\}$ is

$$\begin{array}{ccccccccc}
 & H & T & T & H & H & H & H & T \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
0 & H & 0 & 0 & H & HH & HH & HH & HHT \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
X_0 & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8
\end{array}$$

Clearly $X = \{X_n\}$ is a Markov chain with $X_0 = 0$ and transition probability matrix

$$\mathcal{P} = \left\| \begin{array}{cccc}
1/2 & 1/2 & 0 & 0 \\
1/2 & 0 & 1/2 & 0 \\
0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1
\end{array} \right\| .$$

Note that HHT is an absorbing state. The expected number of tosses needed to obtain pattern HHT is the same as the expected time for the Markov chain X to reach state HHT from state 0.

To compute this expected value, we use first step analysis. Let

$$d_i = E^i[\text{number of steps to reach } HHT].$$

Then conditioning on X_1,

$$\begin{aligned}
d_0 &= \frac{1}{2}d_0 + \frac{1}{2}d_H + 1 \\
d_H &= \frac{1}{2}d_0 + \frac{1}{2}d_{HH} + 1 \\
d_{HH} &= \frac{1}{2}d_{HH} + 1
\end{aligned}$$

Solve this system to obtain $d_0 = 8$. ■

Example: Suppose that the weather on any given day depends on the weather conditions for the previous two days. There are only two categories of weather conditions: sunny and cloudy.

1. If it was sunny today and yesterday, it will be sunny tomorrow with probability 0.8.

2. If it was sunny today and cloudy yesterday, it will be sunny tomorrow with probability 0.6.
3. If it was cloudy today and sunny yesterday, it will be sunny tomorrow with probability 0.4.
4. If it was cloudy today and yesterday, it will be sunny tomorrow with probability 0.1.

What is the fraction of days that are sunny in the long run?

Solution: We consider a Markov chain whose state is the weather conditions yesterday and today $\mathcal{S} = \{(s,s),(c,s),(s,c),(c,c)\}$. The transition probability matrix is then

$$\mathcal{P} = \left\| \begin{array}{cccc} 0.8 & 0 & 0.2 & 0 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.1 & 0 & 0.9 \end{array} \right\|.$$

This Markov chain is irreducible, aperiodic, and positive recurrent. Its stationary distribution, by solving

$$\pi\mathcal{P} = \pi, \quad \sum_i \pi_i = 1, \quad \pi_i \geq 0,$$

is

$$\pi = (3/11, 1/11, 1/11, 6/11).$$

The long run fraction of sunny days is

$$\pi(s,s) + \pi(c,s) = \frac{3}{11} + \frac{1}{11} = \frac{4}{11}. \quad \blacksquare$$