Introduction to Markov Chains

Chapter 5

STATS 102C: Introduction to Monte Carlo Methods

Introduction

- Our main goal in using Monte Carlo methods thus far has been to simulate iid random variables to estimate integration.
- **Markov Chain Monte Carlo (MCMC)** methods use Markov chains to simulate correlated samples that are (approximately) from a target distribution.
- We will only consider Markov chains with countable or finite state spaces (i.e., discrete-state discrete-time Markov chains).

Markov chain

$$X_0 \to \boxed{X_1} \to \boxed{X_2} \to \cdots \to X_n \to X_{n+1} \to \cdots$$

$\{0,1\}$ $\{0,1\}$ $\{0,1\}$

A (discrete-time) Markov chain $\{X_t : t = 0, 1, 2, \ldots\}$ is a stochastic process (i.e., a sequence of random variables) that satisfies the **Markov property**:

$$P(X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \underline{P(X_{n+1} = j | X_n = i)} \tag{1}$$

- The Markov property means that the probability that the chain moves to state j on the next step only depends on the current state i, not on where the chain has been previously.
- The **state space** of a Markov chain is the collection of all possible values for $X_0, X_1, \ldots$.
- Without loss of generality, we can write the state space as $\underline{\{0, 1, 2, \ldots\}}$ (if countable) or $\underline{\{0, 1, 2, \ldots, N\}}$ (if finite).

today's weather tomorrow's weather

Transition Probabilities (1)

$X_n \to X_{n+1}$

The **(one-step) transition probabilities** for a Markov chain $\{X_t : t = 0, 1, 2, \ldots\}$ are defined as the conditional probabilities

$$P_{ij} := P(X_{n+1} = j | X_n = i), \tag{2}$$

for all n and all states i and j in the state space.

Transition Probabilities (2)

If the stat space is finite, the transition probabilities P_{ij} can be represented by a transition matrix

$P_{ij} = P(X_{n+1} = j \mid X_n = i)$

$$\underset{N\times N}{\mathbb{P}} = \begin{array}{c|c} & X_{n+1} \\ \hline X_n & \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots & P_{0N} \\ P_{10} & P_{11} & P_{12} & \cdots & \\ \vdots & \vdots & \vdots & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots & \\ \vdots & \vdots & \vdots & & \\ P_{N0} & & & & P_{NN} \end{bmatrix} \end{array},$$

$\sum_{j=1}^{N} P_{ij} = 1$

$\int P_{ij} = 1$

where

1. All the entries are non-negative:

$$P_{ij} \geq 0, \text{ for all } i, j. \tag{3}$$

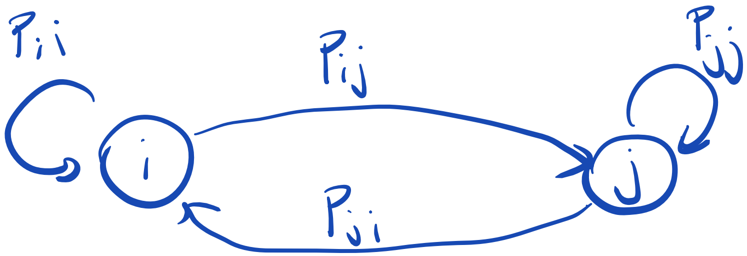
2. The sum of each row is 1:

$$\sum_{j=0}^{\infty} P_{ij} = \sum_{j=0}^{\infty} P(X_{n+1} = j | X_n = i) = 1, \text{ for all } i. \tag{4}$$

Transition Probabilities (3)

Another way to visualize the transition probabilities of a Markov chain is with a **transition state diagram**:

- Each state in the state space is represented by a node/vertex.
- Each nonzero transition probability P_{ij} is represented by an arrow from vertex i to vertex j.

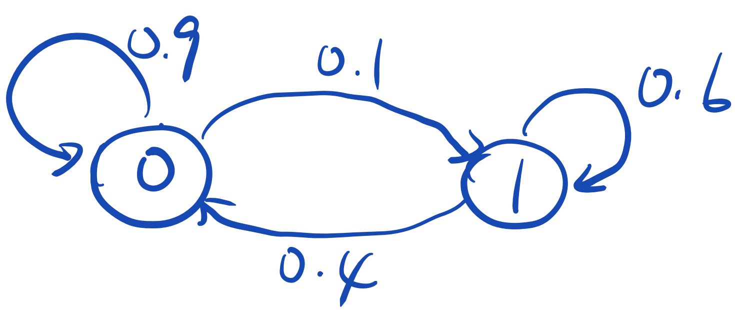


Example: Chance of Rain

Suppose the weather on any given day in Los Angeles is either sunny or rainy. The chance of rain tomorrow depends on whether or not it is raining today and not on past weather conditions. We can model the weather as a two-state Markov chain, with state space $\{0, 1\}$, where 0 = sunny" and 1 =rainy”. Suppose if it is sunny today, then it will be sunny tomorrow with probability 90%; and if it rains today, then it will rain tomorrow with probability 60%. Show its transition matrix.

today: n tomorrow: $n+1$

$X = \{0: \text{sunny}, 1: \text{rainy}\}$

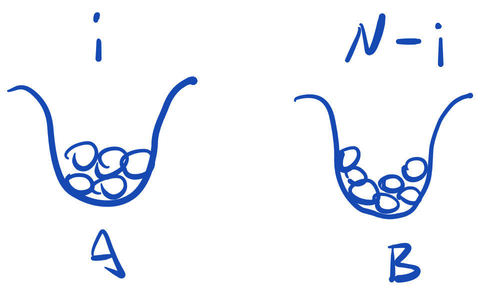


$$P = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

Example: The Ehrenfest Urn Model

Suppose we have two urns labeled A and B that contain a total of N balls. Initially, some of these balls are in urn A and the rest are in urn B. At each step, a ball is randomly chosen from the N balls and moved to the other urn. Let $\boxed{X_n}, n \geq 0$ be the number of balls in urn A at step n. Hence $X_n, n \neq 0$ is a sequence of random variables taking values in $\{0, 1, 2, \ldots, N\}$. What are the transition probabilities?

at step n



a ball was chosen from A

chosen from B

if $X_n = i$, then $X_{n+1} \in \{ \underset{\uparrow}{i-1}, \underset{\uparrow}{i+1} \}$

$$P(X_{n+1} = i-1 \mid X_n = i) = \frac{i}{N}$$

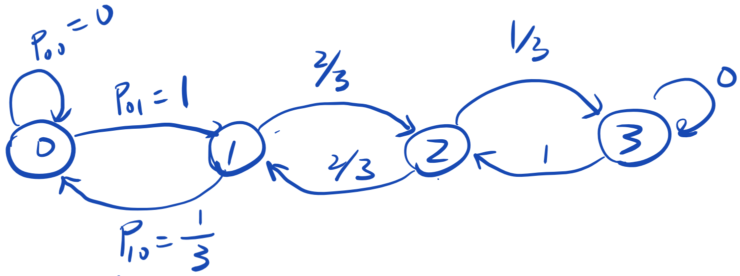
$$P(X_{n+1} = i+1 \mid X_n = i) = \frac{N-i}{N} = 1 - \frac{i}{N}$$

$$P(X_{n+1} = j \mid X_n = i) = 0 \quad j \in \quad i-1 \quad i+$$

E.g.

$N = 3$

At n

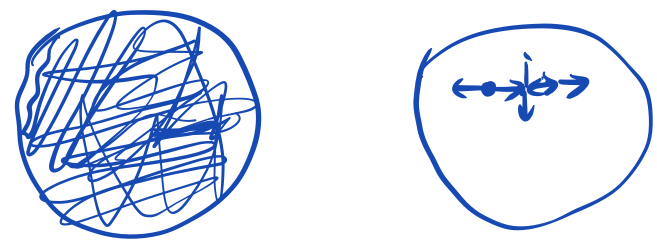


$$
\begin{array}{c|cccc}
n \backslash n+1 & 0 & 1 & 2 & 3 \\
\hline
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \multicolumn{4}{c}{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}}
\end{array}
$$

$\backslash\{\ ,\ 1\}$

Random Walk Model (1)

- A **random walk** is a discrete-time stochastic process that is widely used to model the path an object or particle takes as it moves through space.



- Some applications:
  - The path a particle takes as it moves through a liquid or gas (this is a continuous-time process called **Brownian motion**)
  - The path an animal takes as it searches for food
  - Polymer configurations (self-avoiding walks)
  - A gambler's winnings/losings
  - Stock prices

Random Walk Model (2)

A **random walk** is a stochastic process:

$$\{X_0, X_1, X_2, \ldots, X_n, \ldots\}, \tag{5}$$

defined on the integers $\mathbb{Z}$, such that:

1. The walk starts at 0: $X_0 = 0$.
2. At each step, the random walk moves to the right 1 unit with probability p and moves to the left 1 unit with probability q (so $p + q = 1$).

$$i-1 \xleftarrow{q} i \xrightarrow{p} i+1$$

$$\$i-1 \xleftarrow{q} \$i \xrightarrow{p} \$i+1$$

$X_0 = 0$

$X_1 = \begin{cases} 1 & \text{w/ } p \\ -1 & \text{w/ } q \end{cases}$

$X_n = i$

$X_{n+1} = \begin{cases} i+1 & \text{w/ } p \\ i-1 & \text{w/ } q \end{cases}, \quad q = 1-p$

$X_0 \to X_1 \to X_2 \to \cdots \to X_n$

$X_{n+1} = \begin{cases} X_n + 1 & \text{w/ } p \\ X_n - & \text{w/ } \end{cases}$

$X_1 = X_0 + Y_1$

$Y_i \sim \text{Ber}\left(P(Y_i = 1) = P, \quad P(Y_i = -1) = q \right)$

$X_2 = X_1 + Y_2$

$\vdots$

$X_n = X_{n-1} + Y_n$

$Y = (-1, 1, \dots \quad)$

$\Rightarrow \quad \boxed{X_0} + \sum_{i=1}^{n} Y_i$

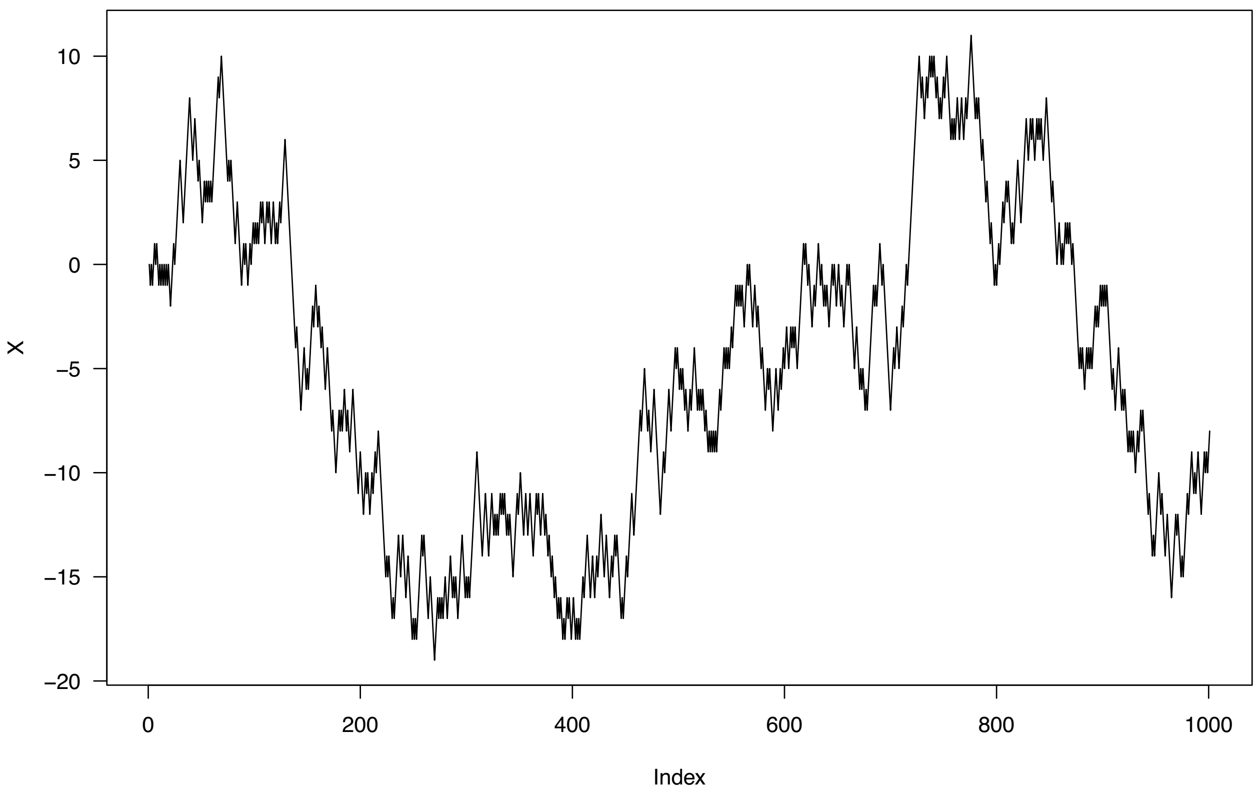
$P(X_{n+1} = j \mid X_n = i)$

$P(X_{n+k} = j \mid X_n = i)$

1 $\quad$ q

R Code to generate a random walk

```r
set.seed(999) # for reproduceability
n <- 1000 # specify length of random walk
p <- 0.5 # specify P(Y = 1)
Y <- sample(c(1, -1), size = n, replace = TRUE,
            prob = c(p, 1 - p)) # Generate n iid samples from Y
X <- c(0, cumsum(Y)) # Compute the random walk X
plot(X, type = "l", las = 1) # Plot the random walk over time
```



Chapman-Kolmogorov Equations

We have defined the one-step transition probabilities P_{ij}. We now define the probability that the chain moves from state i to state j in n steps is $P_{ij}^{(n)}$.

$$P_{ij}^{(n)} = P(X_{n+k} = j | X_k = i),\ n \neq 0,\ i, j \neq 0$$

The Chapman-Kolmogorov equations provide a method for computing these n-step transition probabilities:

$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)} \text{ for all, } n, m \neq 0, \text{ all } i, j. \tag{6}$$

Example

Consider a two-state Markov chain, with state space $\{0, 1\}$ and the transition matrix

$$\mathbb{P} = \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array}.$$

$P(X_2 = 0 \mid X_0 = 0) = P_{00}^{(2)}$

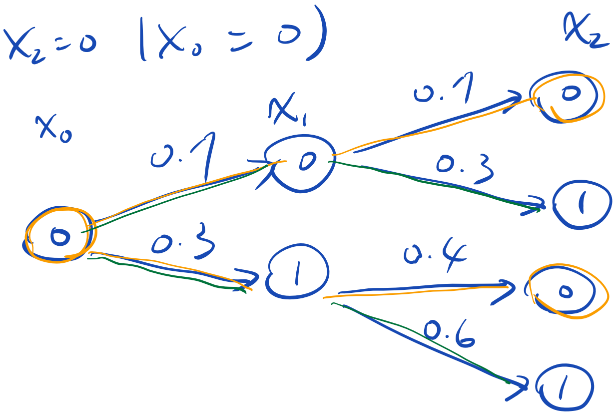
Suppose we are given that the chain starts at $0 : X_0 = 0$. Calculate the probability that the state 0 in ~~3~~ 2 steps, $\cancel{P(X_3 = 0) = P^{(3)}_{00}}$.

If we generate $X_1, X_2, \ldots, X_n$, for some large n (for example, $n = 10000$). How often is the Markov chain in state 0? How often is the Markov chain in state 1?

$X_0 = 0$

$$\underset{\{0,1\}}{X_0} \longrightarrow \underset{\{0,1\}}{X_1} \longrightarrow \underset{\{0,1\}}{X_2}$$

$P(X_2=0 \mid X_0=0)$



$$P_{00}^{(2)} = \boxed{0.7 * 0.7 + 0.3 * 0.4}$$

$$P_{01}^{(2)} = 0.7 * 0.3 + 0.3 * 0.6$$

$$\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} \boxed{0.7*0.7 + 0.3*0.4} & 0.7*0.3 + 0.3*0.6 \\ \bigcirc & \bigcirc \end{bmatrix}$$

$$P^{(n)} = P^n$$

R code to generate two-state Markov chain

```
n <- 1000
P <- matrix(c(0.7,0.4,0.3,0.6), nrow = 2)
P2 <- P
for(i in 1:(n - 1)){
    P2 <- P2 %*% P
}
P2
```

$\pi_0 = $ [,1], $\pi_1 = $ [,2]

```
##              [,1]      [,2]
## [1,] 0.5714286 0.4285714
## [2,] 0.5714286 0.4285714
```

$$P(X_n = 0 \mid X_0 = 0) = P(X_n = 0 \mid X_0 = 1) = 0.57 = \pi_0$$

$$P(X_n = 1 \mid X_0 = 0) = P(X_n = 1 \mid X_0 = 1) = 0.43 = \pi_1$$

$n * \pi_0$

$n * \pi_1$

The Limiting Distribution

Let $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ be a probability distribution on the state space $\{0, 1, 2, \ldots, N\}$. We say π is the **limiting distribution** of a Markov chain $\{X_0, X_1, X_2, \ldots\}$ if

$$\lim_{n\to\infty} P(X_n = j | X_0 = i) = \pi_j, \tag{7}$$

for all $i, j \in \{0, 1, 2, \ldots, N\}$.

The Limiting Distribution

If the limiting distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ exists, then π_j represents:

- The probability that X_n is in state j, independent of the initial state X_0.
- The long run mean (i.e., expected) fraction of time that the Markov chain $\{X_t : t = 0, 1, 2, \ldots\}$ spends in state j.

If the limiting distribution exists, how do we find it?

What conditions will guarantee the existence of the limiting distribution?

Assume the limiting dist. exists. Then for n large enough

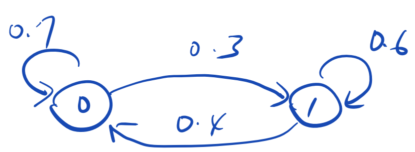
$p(X_n = k \mid X_0 = i) = \pi_k$ and $p(X_{n+1} = j \mid X_0 = i) = \pi_j$

find $\pi = \{\pi_0, \pi_1, \dots, \pi_N\}$

$$
\begin{aligned}
\pi_j &= p(X_{n+1} = j \mid X_0 = i) \\
&= \sum_{k=0}^{N} p(X_{n+1} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k=0}^{N} p(X_{n+1} = j \mid X_n = k, X_0 = i) \cdot \underline{p(X_n = k \mid X_0 = i)} \\
&= \sum_{k=0}^{N} p(X_{n+1} = j \mid X_n = k)\, \pi_k \\
&= \sum_{k=0}^{N} P_{kj}\, \pi_k = \sum_{k=0}^{N} \pi_k P_{kj}
\end{aligned}
$$

$$
\begin{cases}
\pi_j = \underline{\sum_{k=0}^{N} \pi_k P_{kj}} \quad \cdots\cdots \quad ① \qquad \underset{N\times N}{P}\ \underset{N\times 1}{\pi} = \underset{N\times 1}{\pi} \\
\pi_0 + \pi_1 + \cdots + \pi_N = 1 \quad \cdots \quad ②
\end{cases}
$$

E.g. $P = \underline{\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}}$



$\pi_0 = \pi_0 P_{00} + \pi_1 P_{10}$

$\pi_0 = \pi_0\, 0.7 + \pi_1\, 0.4$

$\pi_1 = \pi_0 P_{01} + \pi_1 P_{11}$

$\quad = \pi_0\, 0.3 + \pi_1\, 0.6$

$\pi_0 + \pi_1 = 1$

$0.3\pi_0 = 0.4\pi_1 \Rightarrow \pi_1 = \frac{3}{4}\pi_0$

$\pi_0 + \frac{3}{4}\pi_0 = 1 \Rightarrow \pi_0 = \frac{4}{7}$

$\pi_1 = 1 - \frac{4}{7} = \frac{3}{7}$

$\pi = \left\{ \frac{4}{7}, \frac{3}{7} \right\}$

Suppose $\underline{X_0 \sim \pi}$ we can show $\underline{X_1 \sim \pi}$

$P(X_0 = 0) = \pi_0 \quad P(X_0 = 1) = \pi_1$

$$P(X_1 = 0) = \sum_{k=0}^{1} P(X_1 = 0 \mid X_0 = k) \cdot P(X_0 = k)$$

$$= P_{00}\pi_0 + P_{10}\pi_1$$

$$= 0.7\,\frac{4}{7} + 0.4\,\frac{3}{7} = \frac{4}{7} = P(X_0 = 0) = \pi_0$$

7 7

Stationary Distributions

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain, with state space $\{0, 1, 2, \ldots, N\}$ and transition matrix $\mathbb{P} = [P_{ij}]$. The (row) vector $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ is called a **stationary distribution** of the Markov chain if it satisfies:

- $$\pi_j = \sum_{k=0}^{N} \pi_k P_{kj}, \quad \text{for } j = 0, 1, 2, \ldots, N,$$
  or, equivalently (in matrix notation), $\pi = \pi\mathbb{P}$.
- $$\sum_{i=0}^{N} \pi_i = 1.$$

Conditions for the Limiting Distribution

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $\{0, 1, 2, \ldots, N\}$ and transition matrix $\mathbb{P}$. The Markov chain has a limiting distribution if:

1. There is a solution to the system of equations defined by

$$\begin{cases} \pi^T = \pi^T \mathbb{P} \\ \sum_{i=0}^{N} \pi_i = 1 \end{cases}$$

   i.e., there is a **stationary distribution**.

2. The Markov chain is **irreducible**.
3. The Markov chain is **aperiodic**.

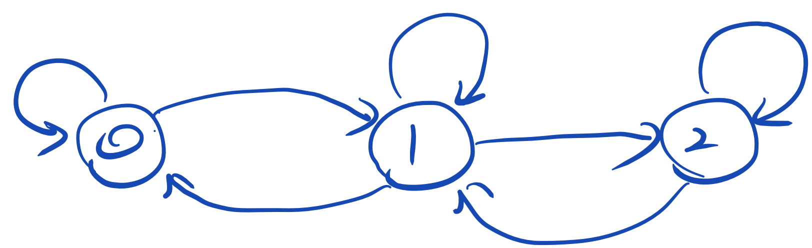
ergotic

Irreducible Markov Chains (1)

Definitation:

1. State j is **accessible** from state i, denoted by $i \to j$, if there is a path from state i to state j.
2. Two states i and j **communicate**, denoted by $i \leftrightarrow j$, if each state is accessible from the other. In other words, if $i \to j$ and $j \to i$, then $i \leftrightarrow j$.
3. A Markov chain is **irreducible** if all states communicate with each other. A Markov chain is **reducible** if it is not irreducible.

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0 & 0.6 & 0.4 \end{bmatrix} \end{array}$$



Irreducible Markov Chains (2)

Accessibility can be defined more rigorously using n-step transition probabilities:

Let $\{X_0, X_1, X_2, \dots\}$ be a Markov chain with state space $\{0, 1, 2, \dots, N\}$. The **n-step transition probability** from state i to j is defined by

$$P_{ij}^{(n)} := P(X_n = j | X_0 = i). \tag{8}$$

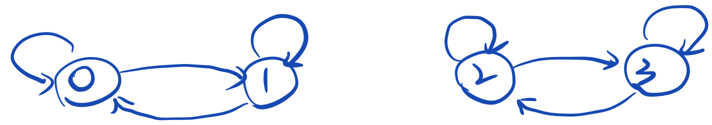
Note that when $n = 1$, $P_{ij}^{(1)} = P_{ij}$.

State j is **accessible** from state i, denoted by $i \to j$, if there is positive probability that state j can be reached from state i in a finite number of transitions. In other words, $P_{ij}^{(n)} > 0$ for some n.

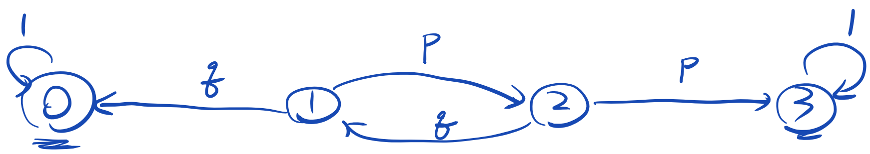
Reducible

e.g.

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1/4 & 3/4 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$



e.g



absorbing states

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$0 < p, q < 1$

$p + q = 1$

$d(0) = \gcd\{1, 2, 3, 4, \dots\} = 1$

$d(3) = 1$

$d(1) = ?$

$1 \to 2 \to 1$ 2 transitions

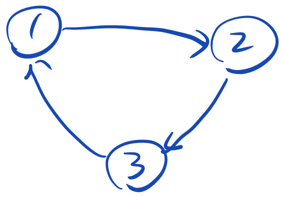
$1 \to 2 \to 1 \to 2 \to 1$ 4 transitions

$\gcd(2, 4, 6, 8, \dots\} = 2 \Rightarrow d(1) = 2$

$d(2) = 2$

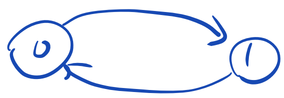
$$\boxed{\text{If } i \leftrightarrow j, \text{ then } d(i) = d(j)}$$

E.g.



$d(1) = \gcd\{3, 6, 9, \dots\} = 3$

$d(1) = d(2) = d(3) = 3$



$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\pi = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$\pi_0 + \pi_1 = 1$

$P\pi = \pi$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$P^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$P^3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} P = IP = P$$

$$P^n = P$$

Periodicity of a Markov Chain

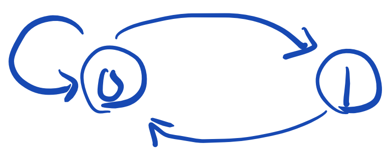
- The **period** of a state i, denoted by $d(i)$, is the greatest common divisor (gcd) of all integers $n \geq 1$ for which $P_{ii}^{(n)} > 0$.
- To find the period of a state i, we consider the number of steps it takes for any path starting from state i to return to state i.

$d(i) = 1 \quad \forall i$

- A Markov chain is called **aperiodic** if each state has period 1.

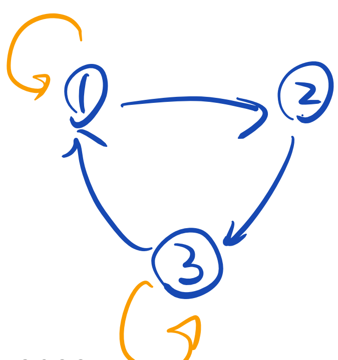
state i after $n = \{k+1, k+2, k+3, \dots\}$ return i

- Some ways to check if a Markov chain is aperiodic:
  If $i \leftrightarrow j$ and $P_{ii} > 0$, then $d(j) = 1$



$d(0) = 1$

$P_{00} > 0, \quad 0 \leftrightarrow 1 \Rightarrow d(1) = 1$



$1 \leftrightarrow 1, \; k+1, k+2 \dots \; k+t$

$d(2)$

$P_{11} > 0$

$P_{33} > 0$

$2 \leftrightarrow 1$ ✓

$2 \leftrightarrow 3$ ✓

$\Rightarrow d(2) = 1$

$d(2) = d(1) = d(3) = 1$

The Basic Limit Theorem of Markov Chains

If a Markov chain $\{X_0, X_1, X_2, \ldots\}$ is irreducible, aperiodic, and has a stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$, then

$$\lim_{n\to\infty} P(X_n = j | X_0 = i) = \pi_j.$$

That is, π is the limiting distribution of the Markov chain, and π is uniquely determined by the system of equations $\begin{cases} \pi = \pi\mathbb{P} \\ \sum_{i=0}^{N} \pi_i = 1 \end{cases}$ where $\mathbb{P}$ is the transition matrix of the Markov chain.