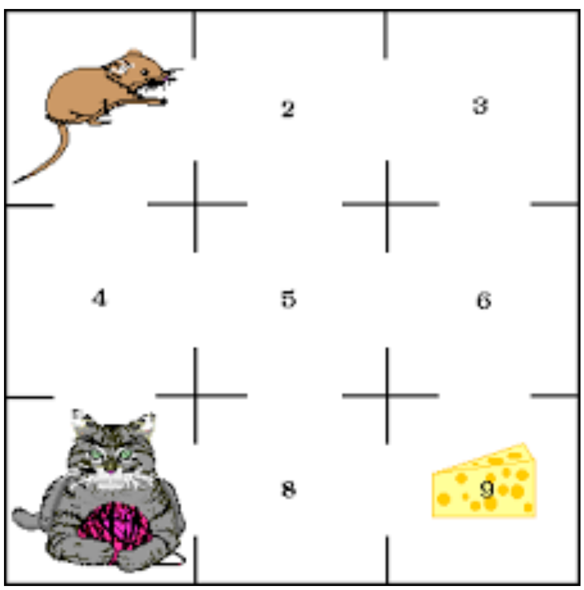
Week 2

Markov chains: transition probabilities, Chapman-Kolmogorov equations, classification of states.

Motivation examples

Example 2.1: A mouse starts from cell 1 in the maze shown below. A cat is hiding patiently in cell 4 and there is a piece of chess in cell 9. In the absence of learning, when the mouse is in the given cell, it will choose the next cell to visit with probability $1/k$, where the k is the number of adjoining cells. Assume that once the mouse finds either the piece of chess or the cat, it will understandably stay there forever.



Let $X_0 = 1$ and $X_n, n \geq 1$, be the number of the cell where the mouse is located after the n-th movements. $\{X_n\}_{n\geq 0}$ is a discrete process with state space $S = \{1, 2, ..., 9\}$.

Things of interest:

- the probability of reaching the cheese first?
- the probability distribution for the time in reaching the cat before reaching the chess?

Example 2.2: (**Gambler's ruin**) A gambler starts with $\$N$, win and loss $1 with prob 1/2 respectively. He must leave when he goes broke or he reaches $\$K > N$.

Things of interest:

- what is the probability that the gambler eventually goes broke (or win $\$(K - N)$)?
- on average, how many games he can play before he goes broke?
- how many times he can have $k < K$ dollars before the game ends?

In this example, $\{X_n\}_{n \geq 0}$ is a discrete process with the state space $S = \{0, 1, ..., K\}$, where X_n is the money the gambler has after the n-th game.

Example 2.3: A salesman lives in Town A and is responsible for the sales consisting of Towns A, B and C. Each week he is required to visit a different town. When he is in his own town, it makes no difference which town he visits next so he slips a coin and if it is heads he goes to B and if tails he goes to C. However, after spending a week away from home, he has a slight preference for going home so when he is in either towns B or C, he has a slight preference for going home so when he is in either towns B or C, he flips two coins. If two heads occur, then he goes to the other town; otherwise he goes to A.

The successive towns that the salesman visits form a discrete process with state space A, B, C.

Let 1 = "Town A"; 2 = "Town B"; 3 = "Town C".

Let $X_n, n \geq 0$, be the number of the Town where the salesman is located during n-th week.

$\{X_n\}_{n\geq 0}$ is a discrete process with state space $S = \{1, 2, 3\}$. .

Things of interest:

- $P(X_n = j)$?
- $P(X_n = 2, X_{n+1} = 3)$?
- In fifteen years, what is the average times that the salesman has visited Town B and/or C ?

4 Markov Chain: Transition probabilities

A discrete random process $\{X_n, n \geq 0\}$ is called a **Markov Chain** (MC) if, for all $n \geq 1$, all $i_0, i_1, ..., i_{n-1}, i, j \in S$,

$$\begin{aligned} & P(X_{n+1} = j \mid X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) \\ = \; & P(X_{n+1} = j \mid X_n = i). \end{aligned}$$

Namely, the probability distribution of future states of a MC conditional on both the past and present states depends only on the present state. This is usually referred to **Markov property** of a process.

Equivalent definition: a discrete random process $\{X_n, n \geq 0\}$ is a MC if and only if for all $n \geq 1$, $A_0, A_1, ..., A_{n-1} \subset S, i, j \in S$,

$$\begin{aligned} & P(X_{n+1} = j \mid X_0 \in A_0, ..., X_{n-1} \in A_{n-1}, X_n = i) \\ = \; & P(X_{n+1} = j \mid X_n = i). \end{aligned}$$

By the definition, for a MC $\{X_n, n \geq 0\}$, the conditional distribution of X_{n+1} depends only on X_n, the most recently known state value and we have: for any $k \geq 1$,

$$\begin{aligned} & P(X_{n+k} = j \mid X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) \\ = \; & P(X_{n+k} = j \mid X_n = i). \end{aligned}$$

$P(X_{n+1} = j \mid X_n = i)$ is called **one-step transition probability** from state i to state j.

When one-step transition probabilities are independent of n, we say that the MC has stationary transition probabilities or the MC is **Homogeneous.**

This course mainly focus on homogeneous MCs. In this case, we define

$$p_{ij} = P(X_1 = j \mid X_0 = i), \qquad i, j \in S,$$

the one-step transition probability from state i to state j.

Clearly, we have the following result for the one-step transition probability,

$$p_{ij} \geq 0 \quad \text{for all } i, j \in S \quad \text{and} \quad \sum_{j \in S} p_{ij} = 1 \quad \text{for each } i \in S.$$

It is customary to arrange these numbers p_{ij} as a matrix, i.e., a (finite or infinite) square array:

$$P \;=\; \begin{array}{c} \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ r \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & r \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{pmatrix} \end{array}$$

where $r = \#S$, or simply

$$P = (p_{ij})_{i,j \in S}$$

We say that P is a Markov matrix or transition matrix of the process.

Examples 2.1 revisited: Transition probability and transition matrix

$$P \; = \; \begin{array}{c c} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Examples 2.2 revisited: Transition probability and transition matrix

Note that $p_{KK} = p_{00} = 1$, $p_{ij} = 1/2$ if $j = i+1, i-1$, $0 < i < K-1$, $p_{ij} = 0$, otherwise.

$$P \;=\; \begin{bmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \\ 0 & 0 & 0 & . & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & ... & 0 & 0 & 1 \end{bmatrix}$$

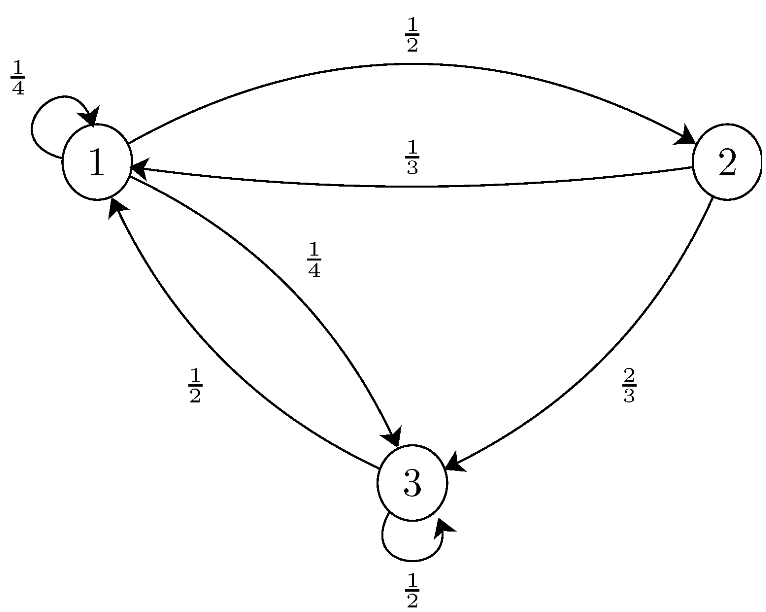
A Markov chain is usually shown by a **state transition diagram**.

Consider a Markov chain with three possible states 1, 2, and 3 and the following transition probabilities

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

Figure below shows the state transition diagram for the above Markov chain. In this diagram, there are three possible states 1, 2, and 3, and the arrows from each state to other states show the transition probabilities p_{ij}. When there is no arrow from state i to state j, it means that $p_{ij} = 0$.

A state transition diagram



5 Markov Chain: Chapman-Kolmogorov equation

Example 2.3. (revisited). A salesman lives in Town A and is responsible for the sales consisting of Towns A, B and C. Each week he is required to visit a different town. When he is in his own town, it makes no difference which town he visits next so he slips a coin and if it is heads he goes to B and if tails he goes to C. However, after spending a week away from home, he has a slight preference for going home so when he is in either towns B or C, he has a slight preference for going home so when he is in either towns B or C, he flips two coins. If two heads occur, then he goes to the other town; otherwise he goes to A.

Problem:

- Suppose this salesman is in Tower A this week. What is the probability for this salesman in Tower C five weeks later?

Let 1 = "Town A"; 2 = "Town B"; 3 = "Town C".

Let $X_n, n \geq 0$, be the number of the Town where the salesman is located during n-th week.

X_n is a discrete process with state space $S = \{1, 2, 3\}$. The transition matrix is given by

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

We need to find $P(X_5 = 3 \mid X_0 = 1)$.

n-Step Transition Probabilities

Consider a Markov chain $\{X_n\}_{n\geq 0}$, where the state space $S = \{1, 2, \cdots, r\}$.

$p_{ij} = P(X_1 = j \mid X_0 = i)$ gives us the probability of going from state i to state j in one step.

Now suppose that we are interested in finding the probability of going from state i to state j in two steps, i.e.,

$$p_{ij}^{(2)} = P(X_2 = j | X_0 = i).$$

The probability can be calculated by applying the law of total probability. Since X_1 can take one of the possible values in S, we can write

$$\begin{aligned} p_{ij}^{(2)} &= P(X_2 = j | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) P(X_1 = k | X_0 = i) \quad \text{(by Markov property)} \\ &= \sum_{k \in S} p_{kj} p_{ik}. \end{aligned}$$

We conclude

$$p_{ij}^{(2)} = P(X_2 = j | X_0 = i) = \sum_{k \in S} p_{ik} p_{kj} \tag{1}$$

The above formula can be explained as follows. In order to get to state j, we need to pass through some intermediate state k. The probability of this event is $p_{ik}p_{kj}$. To obtain $p_{ij}^{(2)}$, we sum over all possible intermediate states.

Accordingly, we can define the two-step transition matrix as follows:

$$P^{(2)} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & \dots & p_{1r}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & \dots & p_{2r}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1}^{(2)} & p_{r2}^{(2)} & \dots & p_{rr}^{(2)} \end{bmatrix}.$$

Note that $p_{ij}^{(2)}$ is in fact the element in the ith row and jth column of the matrix

$$P^2 = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix} \cdot \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix}.$$

Namely, we have

$$P^{(2)} = P^2.$$

In general, we can define the n-step transition probabilities $p_{ij}^{(n)}$ as

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i), \quad \text{for } n = 0, 1, 2, \cdots, \tag{2}$$

and the n-step transition matrix, $P^{(n)}$, as

$$P^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \dots & p_{1r}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \dots & p_{2r}^{(n)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1}^{(n)} & p_{r2}^{(n)} & \dots & p_{rr}^{(n)} \end{bmatrix}. \tag{3}$$

Note that, $p_{ii}^{(0)} = 1$, $p_{ij}^{(0)} = 0$ for $i \neq j$, by convention; and $p_{ij}^{(1)} = p_{ij}$.

Let m and n be two positive integers. In order to get to state j from i in $(m+n)$ steps, the chain will be at some intermediate state k from i in m steps, and then get to state j from k in n steps.

To obtain $p_{ij}^{(m+n)}$, we sum over all possible intermediate states:

$$\begin{aligned} p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\ &= \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}. \end{aligned}$$

The above equation is called the **Chapman-Kolmogorov (C-K) equation**.

Similar to the case of two-step transition probabilities, we can show that

$$P^{(n)} = P^n, \quad \text{for } n = 1, 2, 3, \cdots.$$

In summary, the **Chapman-Kolmogorov equation** can be written as

$$\begin{aligned} p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\ &= \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}. \end{aligned} \tag{4}$$

In matrix theory, (4) is equivalent to

$$P^{(m+n)} = P^{(m)} P^{(n)}, \tag{5}$$

where the n-step transition matrix $P^{(n)}$ satisfies that

$$P^{(n)} = P^n, \quad \text{for } n = 1, 2, 3, \cdots . \tag{6}$$

Example 2.3 (continued). The 5 step transition probability matrix of the samples man example is

$$P^{(5)} \;=\; P^5 = \begin{bmatrix} .293 & .3535 & .3535 \\ .5330 & .2344 & .235 \\ .5303 & .2354 & .2344 \end{bmatrix}$$

Hence $P(X_5 = 3 \mid X_0 = 1) = 0.3535.$

Probability distribution of a MC

Consider a Markov chain $\{X_n\}_{n\geq 0}$, where the state space $S = \{1, 2, \cdots, r\}$.

Suppose that the probability distribution of the initial variable X_0 is given. How can we obtain the probability distribution of X_1, X_2, $\cdots$?

Define, for $n = 0, 1, 2, ...$,

$$\pi^{(n)} = \left[P(X_n = 1), \;\; P(X_n = 2), \;\; \cdots, \;\; P(X_n = r)\right], \tag{7}$$

If $\pi^{(0)} = \left(\pi_1^{(0)}, ..., \pi_k^{(0)}\right)$ is given, i.e., the distribution of X_0 is given, the law of total probability yields

$$\begin{aligned} P(X_n = j) &= \sum_{k\in S} P(X_n = j \mid X_0 = k)P(X_0 = k) \\ &= \sum_{k\in S} \pi_k^{(0)} p_{kj}^{(n)}, \end{aligned}$$

the j-th element of the vector $\pi^{(0)} \cdot P^{(n)}$. In general, in matrix theory, we have

$$\pi^{(n)} \;=\; \pi^{(0)}P^{(n)} = \pi^{(0)}P^n, \quad \text{for } n = 0, 1, 2, \cdots . \tag{8}$$

Note that

$$P^n = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix} \cdot \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix} \cdots \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{bmatrix}$$

The facts above, together with more detailed analysis, show that a homogeneous MC is completely defined by its one-step transition matrix P and the distribution of X_0: $\pi_k^{(0)} = P(X_0 = k), k = 1, ..., r$.

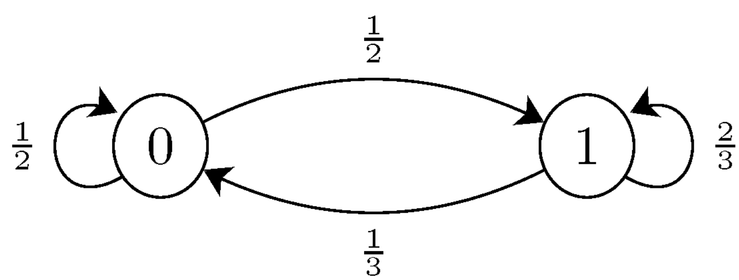
Example 2.4. Consider a system that can be in one of two possible states, $S = \{0, 1\}$. In particular, suppose that the transition matrix is given by

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

Suppose that the system is in state 0 at time $n = 0$, i.e., $X_0 = 0$ with probability one.

- Draw the state transition diagram.
- Find the probability that the system is in state 1 at time $n = 3$.

Solution:
The state transition diagram is shown as follows:



Here, we know

$$\begin{aligned}\pi^{(0)} &= \left[P(X_0 = 0), \quad P(X_0 = 1)\right] \\ &= \left[1, \;\; 0\right].\end{aligned}$$

Thus,

$$\begin{aligned}\pi^{(3)} &= \pi^{(0)} P^3 \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}^3 \\ &= \begin{bmatrix} \frac{29}{72} & \frac{43}{72} \end{bmatrix}.\end{aligned}$$

Thus, the probability that the system is in state 1 at time $n = 3$ is $\frac{43}{72}$.

6 Markov Chains: Classifications of states

To better understand Markov chains, we have to introduce some principles of classifying states of a MC.

Definitions:

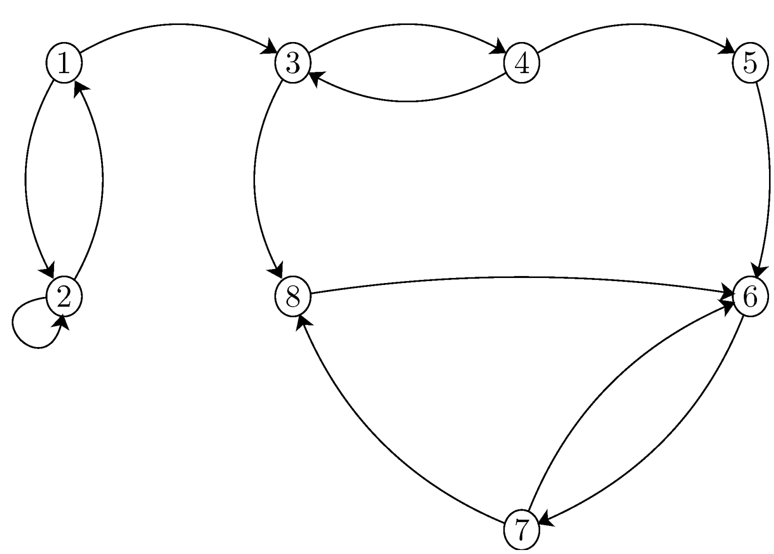
We say that a state j is **accessible** from state i, written as $i \to j$, if $p_{ij}^{(m)} > 0$ for some $m \geq 0$. We assume every state is accessible from itself since $p_{ii}^{(0)} = 1$ by the convention.

Two states i and j are said to **communicate**, written as $i \longleftrightarrow j$, if they are **accessible** from each other. In other words,

$$i \longleftrightarrow j \text{ means } i \to j \text{ and } j \to i.$$

If two states i and j do not communicate, then for all $m \geq 0$

$$p_{ij}^{(m)} = 0 \quad \text{or} \quad p_{ji}^{(m)} = 0.$$



Communication is an **equivalence** relation. That means that

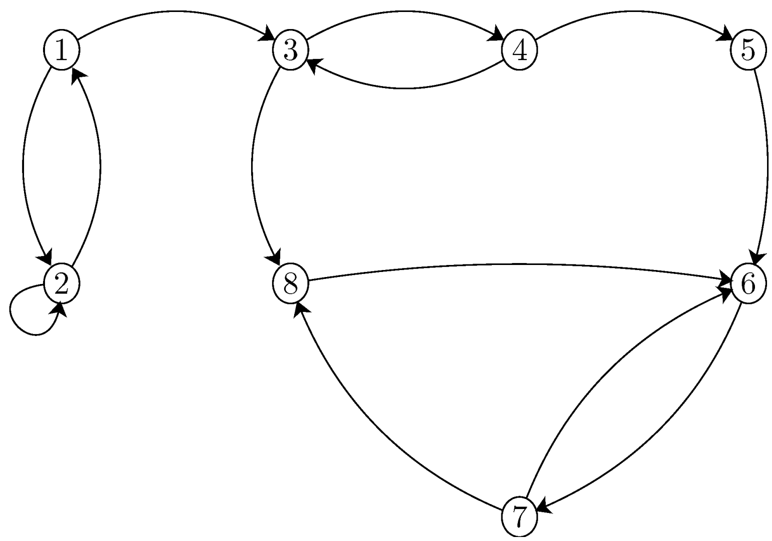
- every state communicates with itself, $i \longleftrightarrow i$;
- if $i \longleftrightarrow j$, then $j \longleftrightarrow i$;
- if $i \longleftrightarrow j$ and $j \longleftrightarrow k$, then $i \longleftrightarrow k$.

Therefore, the state space S of any MC can be partitioned into disjoint communicating **classes**: to do this, choose a number of S (0, say) and let $C_0 = \{j \in S : \ j \longleftrightarrow 0\}$. Then, choose a $k \in S - C_0$ and let $C_k = \{j \in S : \ j \longleftrightarrow k\}$, and so on.

In fact, let $C_i = \{j \in S : \ j \longleftrightarrow i\}$, $i \in S$. It can be shown that for any $i \neq j$,

$$C_i = C_j \quad \text{or} \quad C_i \cap C_j = \phi \text{ (empty set)}.$$

Example 2.4: Consider a Markov chain having the following state transition diagram:



It is assumed that when there is an arrow from state i to state j, then $p_{ij} > 0$. Find the equivalence classes for this Markov chain.

Solution: There are four communicating classes in this Markov chain. We notice that states 1 and 2 communicate with each other, but they do not communicate with any other nodes in the graph. Similarly, nodes 3 and 4 communicate with each other, but they do not communicate with any other nodes in the graph. State 5 does not communicate with any other states, so it by itself is a class. Finally, states 6, 7, and 8 construct another class. Thus, here are the classes:

$$\begin{aligned}
&\text{Class } 1 = \{\text{state } 1, \text{state } 2\},\\
&\text{Class } 2 = \{\text{state } 3, \text{state } 4\},\\
&\text{Class } 3 = \{\text{state } 5\},\\
&\text{Class } 4 = \{\text{state } 6, \text{state } 7, \text{state } 8\}.
\end{aligned}$$

We say a MC is **irreducible** if the MC has only one class, i.e., all states of the MC communicate.

We say a class C is **closed** if no state outside C can be reached from any state in C, i.e. $p_{ij} = 0$ whenever $i \in C$ and $j \notin C$.

A state j is called **absorbing** if $\{j\}$ is a closed class, i.e., $p_{jj} = 1$.

Example 2.5: Consider a MC with transition matrix:

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