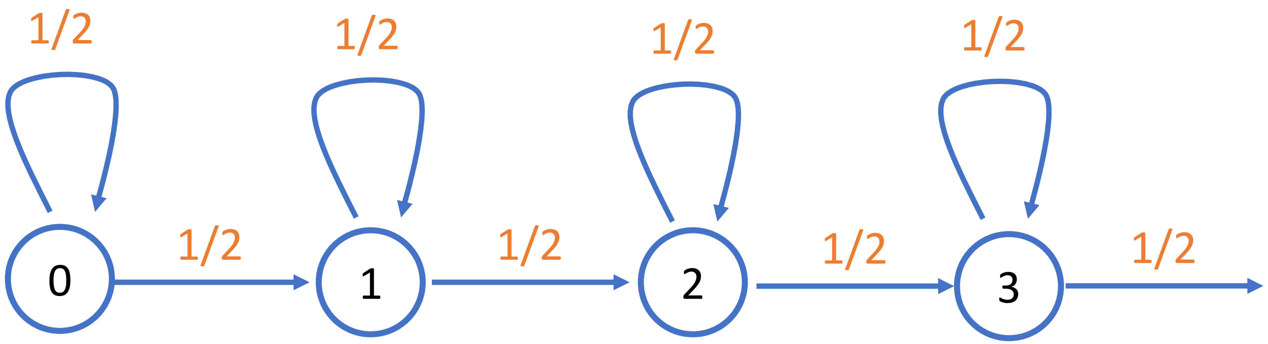
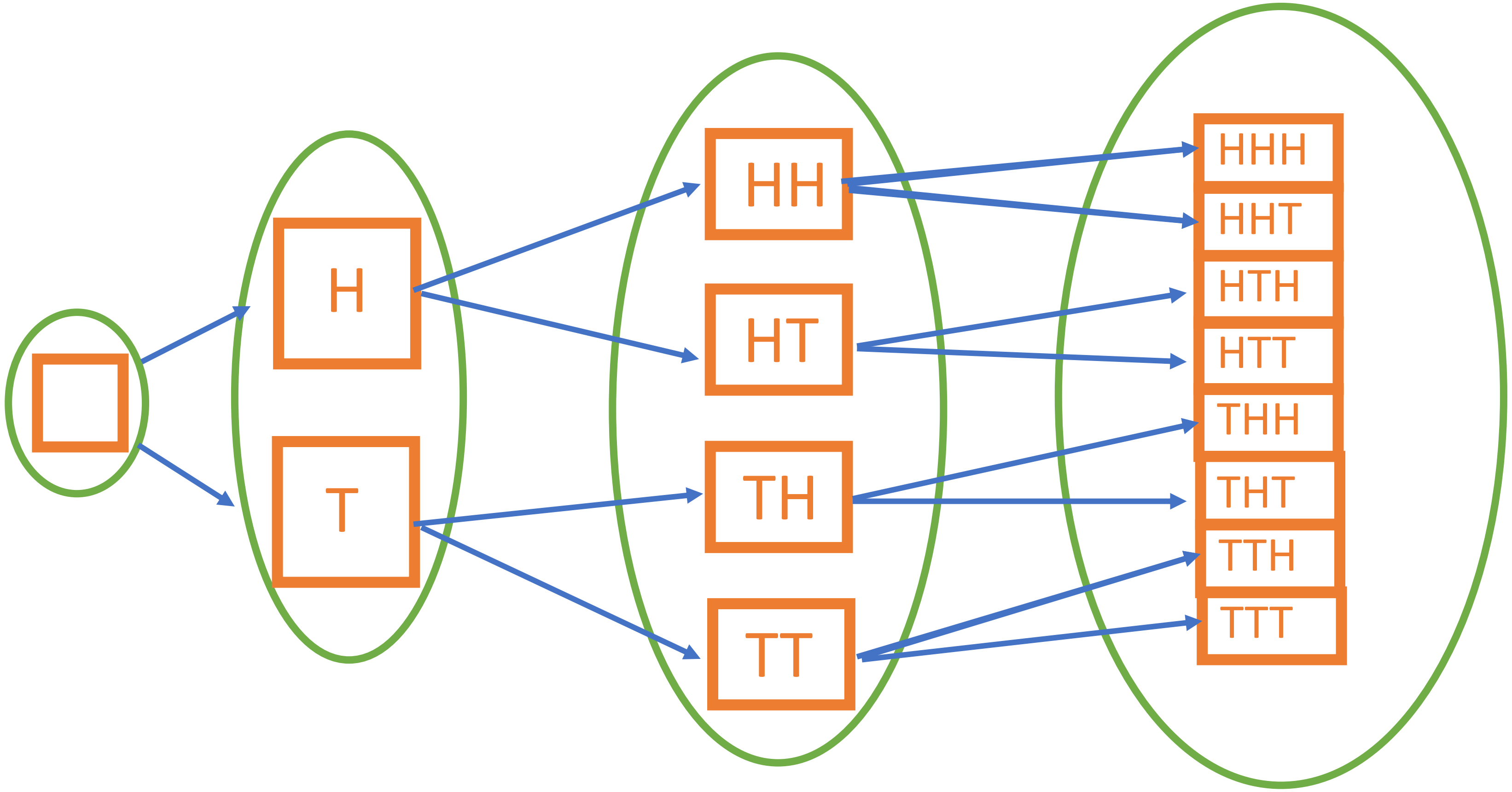
Lecture 21: Markov Chain - I



Recap: Random process

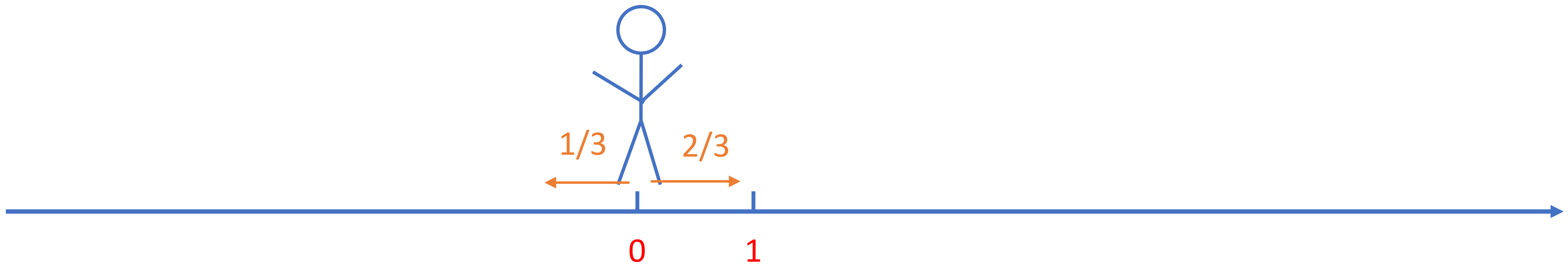
Probability Space

We can intuitively think of random process as evolving possibilities.



Examples of random processes

Random walk on a line:



Examples of random processes

Coupon Collector

- If you buy a bag, you get a uniformly random card from n cards.

- How many bag in expectation do you need to buy to collect all cards?

Examples of random processes

Vitamin problem:

Q6.3 Part 3

0.1 Points

Rafael tosses a fair coin repeatedly. Let T be the number of tosses until Rafael gets 3 heads.

- Uniform
- Bernoulli
- Binomial
- Geometric
- Poisson
- None

Examples of random processes

Shuffle Cards:

In every step, with prob ½, pick two random cards and swap them.
with prob ½, do nothing.

Can we unify all of them?

What is shared by all of them

- State

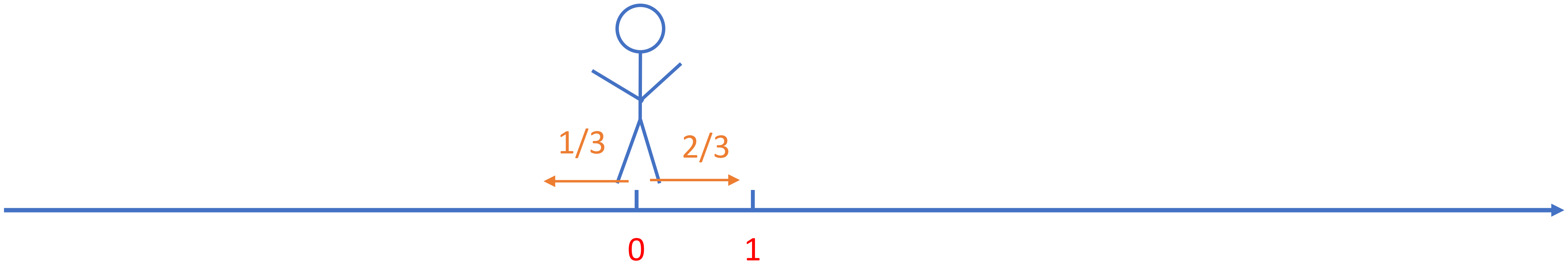
- Transition between States

Can we unify all of them?

What is shared by all of them : Random Walk

- State : Position x

- Transition between States: with 1/3 prob x-1, with 2/3 prob x+1.



Can we unify all of them?

What is shared by all of them: Coupon collector

- State : The number of card x you already have

- Transition between States: with (n-i+1)/n prob x+1, with remaining prob stay x

Can we unify all of them?

What is shared by all of them: Vitamin Problem – The third head

- State : The number of head x you already have

- Transition between States: with ½ prob x+1, with ½ prob stay x

Q6.3 Part 3
0.1 Points

Rafael tosses a fair coin repeatedly. Let T be the number of tosses until Rafael gets 3 heads.

- ○ Uniform
- ○ Bernoulli
- ○ Binomial
- ○ Geometric
- ○ Poisson
- ◉ None

Can we unify all of them?

What is shared by all of them: Shuffle Cards

- State : The order of cards (a permutation)

- Transition between States:

  With probability 1/n^2 select cards (i,j) and swap them

Can we unify all of them?

A general process that rules them all: Markov Chain

Definition: A Markov Chain $(\mathcal{X}, P)$ contains

- A finite/infinite set of states $\mathcal{X}$.

  These are "vertices".

- Transitions P.

  $P_{i,j}$ is the probability that from state i, we transition to state j in the next step.

  $\forall i, \ \sum_j P_{i,j} = 1.$

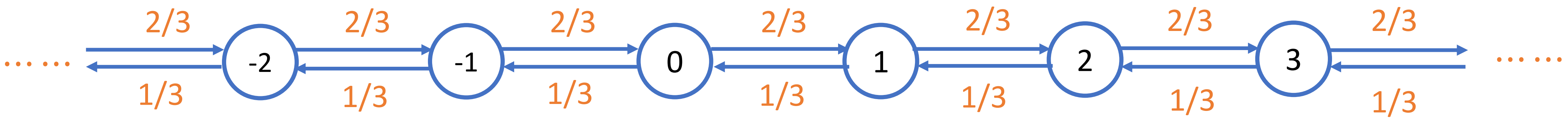
  These are "edges".

Can we unify all of them?

What is shared by all of them : Random Walk

- State : Position x

- Transition between States: with 1/3 prob x-1, with 2/3 prob x+1.

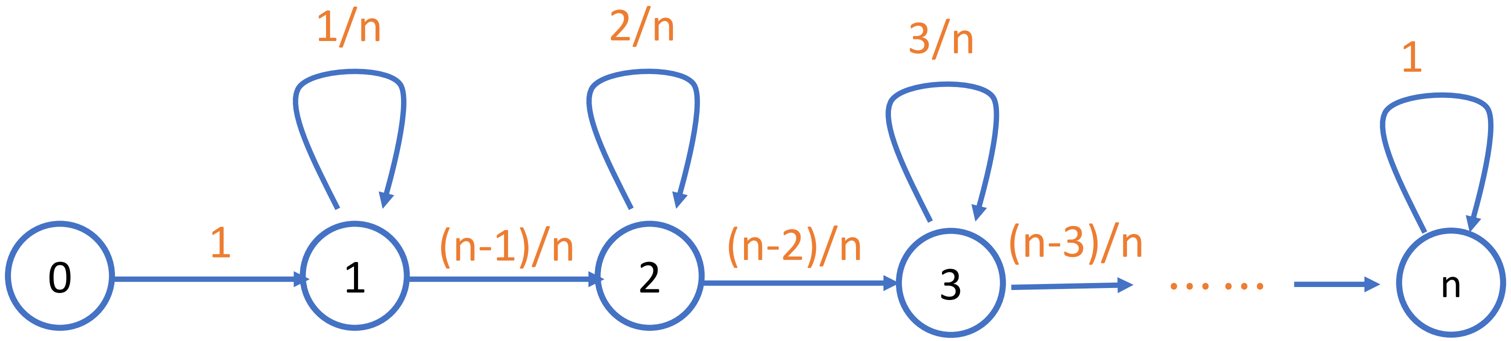


Can we unify all of them?

What is shared by all of them: Coupon collector

- State : The number of card x you already have

- Transition between States: with (n-i+1)/n prob x+1, with remaining prob stay x

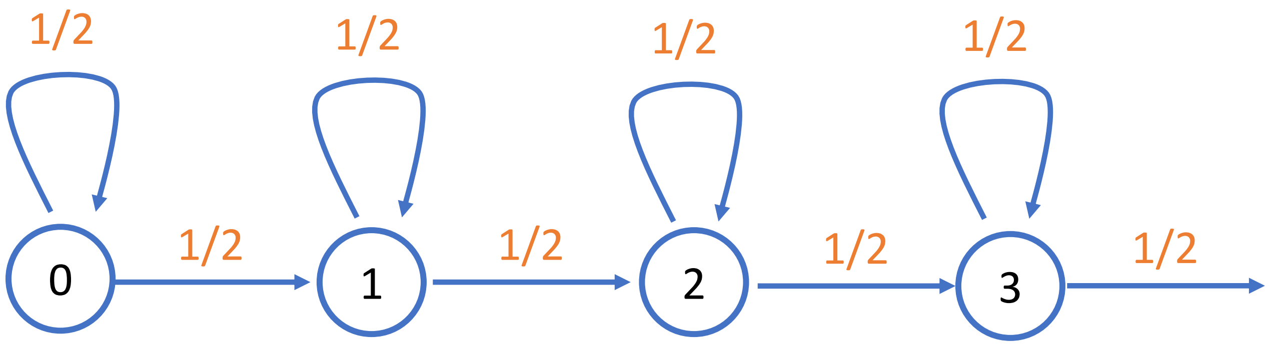


Can we unify all of them?

What is shared by all of them: Vitamin Problem – The third head

- State : The number of head x you already have

- Transition between States: with ½ prob x+1, with ½ prob stay x



Q6.3 Part 3
0.1 Points

Rafael tosses a fair coin repeatedly. Let T be the number of tosses until Rafael gets 3 heads.

- ○ Uniform
- ○ Bernoulli
- ○ Binomial
- ○ Geometric
- ○ Poisson
- ◉ None

Can we unify all of them?

What is shared by all of them: Shuffle 3 Cards JQK

- State : The order of cards (a permutation)

- Transition between States:

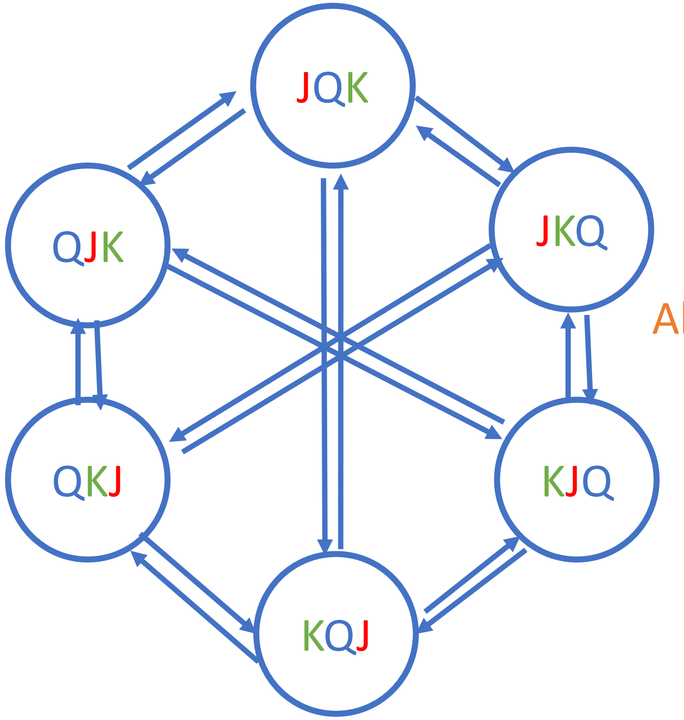
  With probability 1/n^2 select cards (i,j) and swap them

Can we unify all of them?

What is shared by all of them: Shuffle 3 Cards JQK

- State : The order of cards (a permutation)
- Transition between States:

With probability 1/n^2 select cards (i,j) and swap them



All edges have prob 1/3

Random Walk on Markov Chain

Definition:

A Random walk on a Markov Chain $(\mathcal{X}, P)$ is a sequence of states $X_0, X_1, X_2, \dots,$
It starts at a state X_0. Then for each step t, X_t is sampled according to

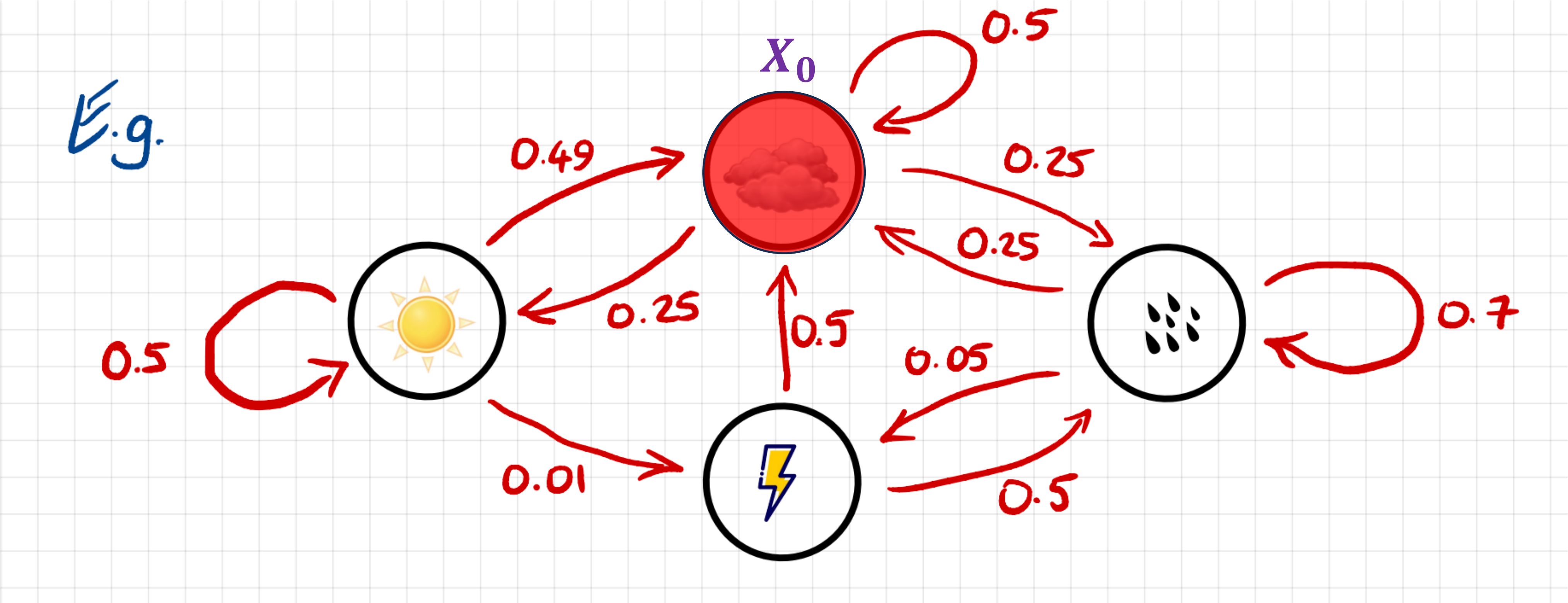
$$\mathbb{P}[X_t = j \mid X_{t-1} = i, X_{t-2}, X_{t-3}, \dots, X_1] = \mathbb{P}[X_t = j \mid X_{t-1} = i] = P_{i,j}.$$

Remark:

- $\mathbb{P}[X_t = j \mid X_{t-1} = i, X_{t-2}, X_{t-3}, \dots, X_1] = \mathbb{P}[X_t = j \mid X_{t-1} = i]$ is called the Markovian property (also called memoryless property.)

- Whether a process is memoryless or not, depends on how you define the states. It is memoryless when the state contains all the info that's useful for next step.

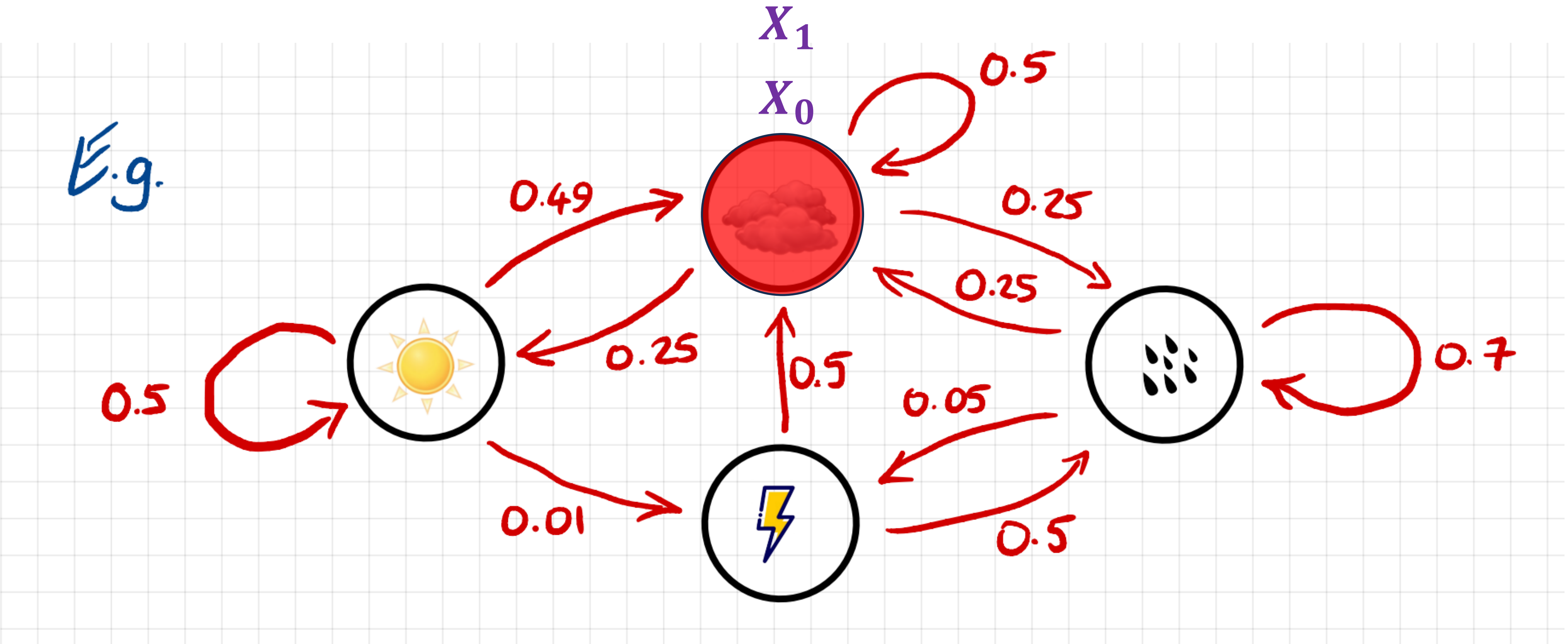
Random Walk on Markov Chain - Example



$\mathbb{P}(X_0 = \text{cloud}) = 1.$

Picture is from Prof. Sinclair's slides in Spring 2024

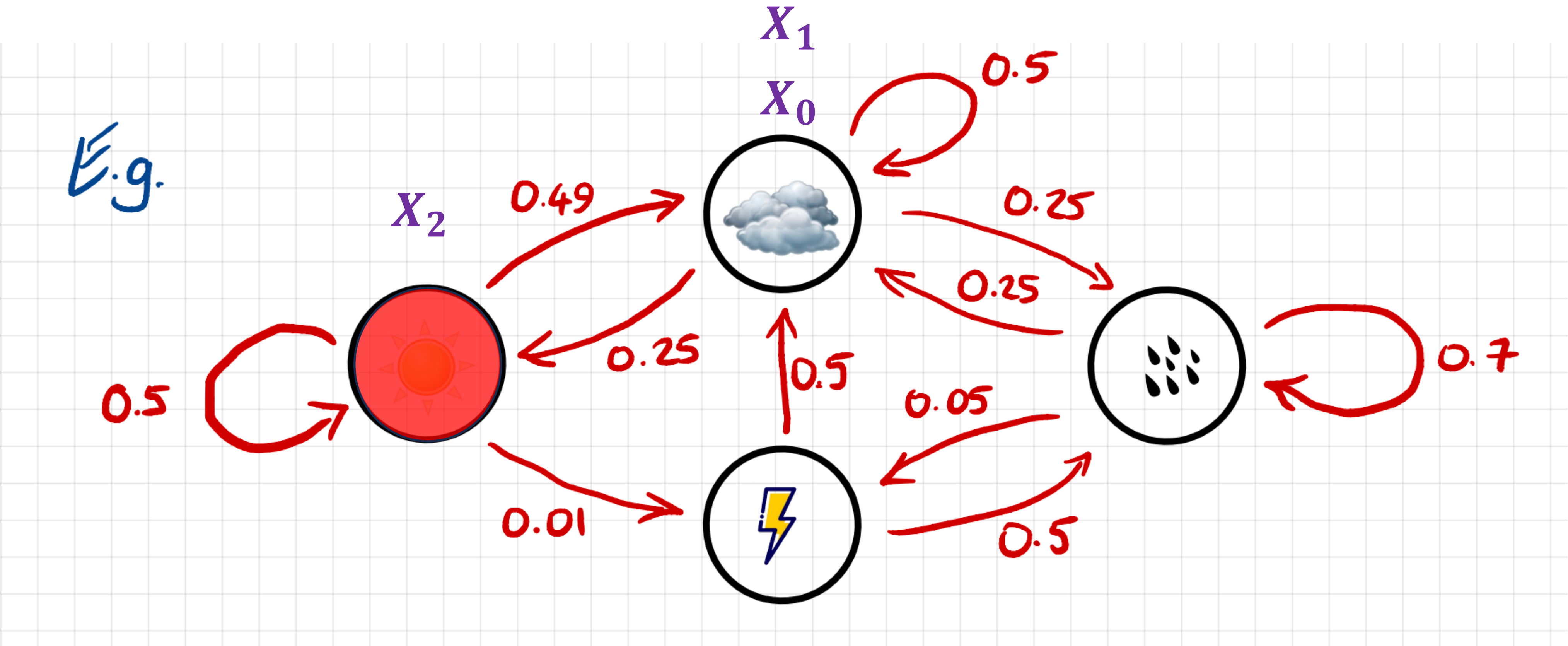
Random Walk on Markov Chain - Example



$$\mathbb{P}(X_0 X_1 = \text{cloud}, \text{cloud}) = \mathbb{P}(X_0 = \text{cloud}) \ * P_{\text{cloud},\text{cloud}} = 1 \ * 0.5 = 0.5$$

Picture is from Prof. Sinclair's slides in Spring 2024

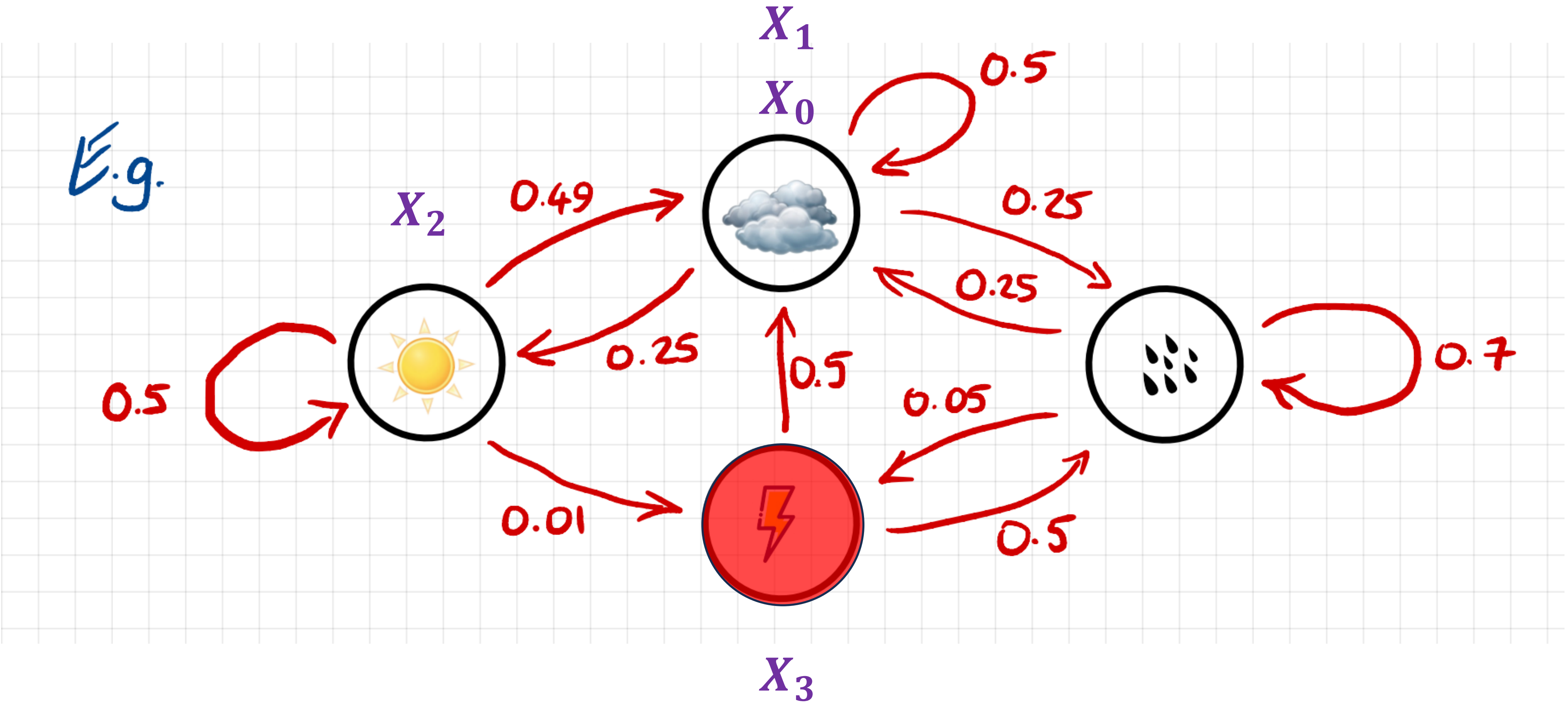
Random Walk on Markov Chain - Example



$$\mathbb{P}(\boldsymbol{X_0 X_1 X_2} = \text{cloud, cloud, sun})$$
$$= \mathbb{P}(\boldsymbol{X_0 X_1} = \text{cloud, cloud}) * \mathrm{P}_{\text{cloud,sun}} = 0.5 * 0.25 = 0.125$$

Picture is from Prof. Sinclair's slides in Spring 2024

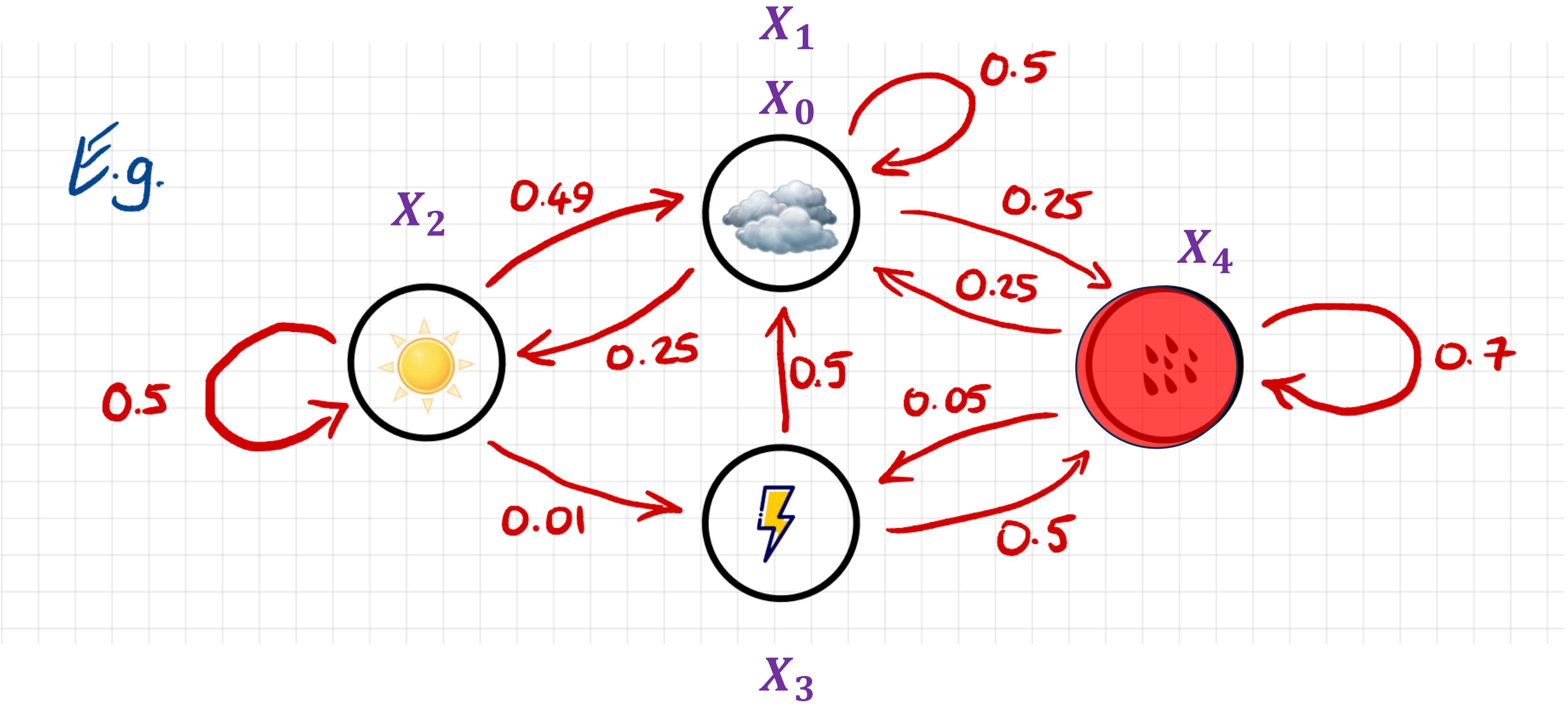
Random Walk on Markov Chain - Example



$\mathbb{P}(X_0 X_1 X_2 X_3 = \text{cloud, cloud, sun, thunder}) = 0.125 * 0.01 = 0.00125$

Picture is from Prof. Sinclair's slides in Spring 2024

Random Walk on Markov Chain - Example



$\mathbb{P}(X_0X_1X_2X_3X_4 = \text{cloud}, \text{cloud}, \text{sun}, \text{thunder}, \text{rain}) = 0.00125 \ * 0.5 = 0.000625$

Picture is from Prof. Sinclair's slides in Spring 2024

Distribution of random walk

Definition:

Let π_0 be the distribution of X_0; ($\pi_0(j)$ denotes the probability of starting at j)

π_1 be the distribution of X_1;

π_2 be the distribution of X_2

$\cdots\cdots$

Transition

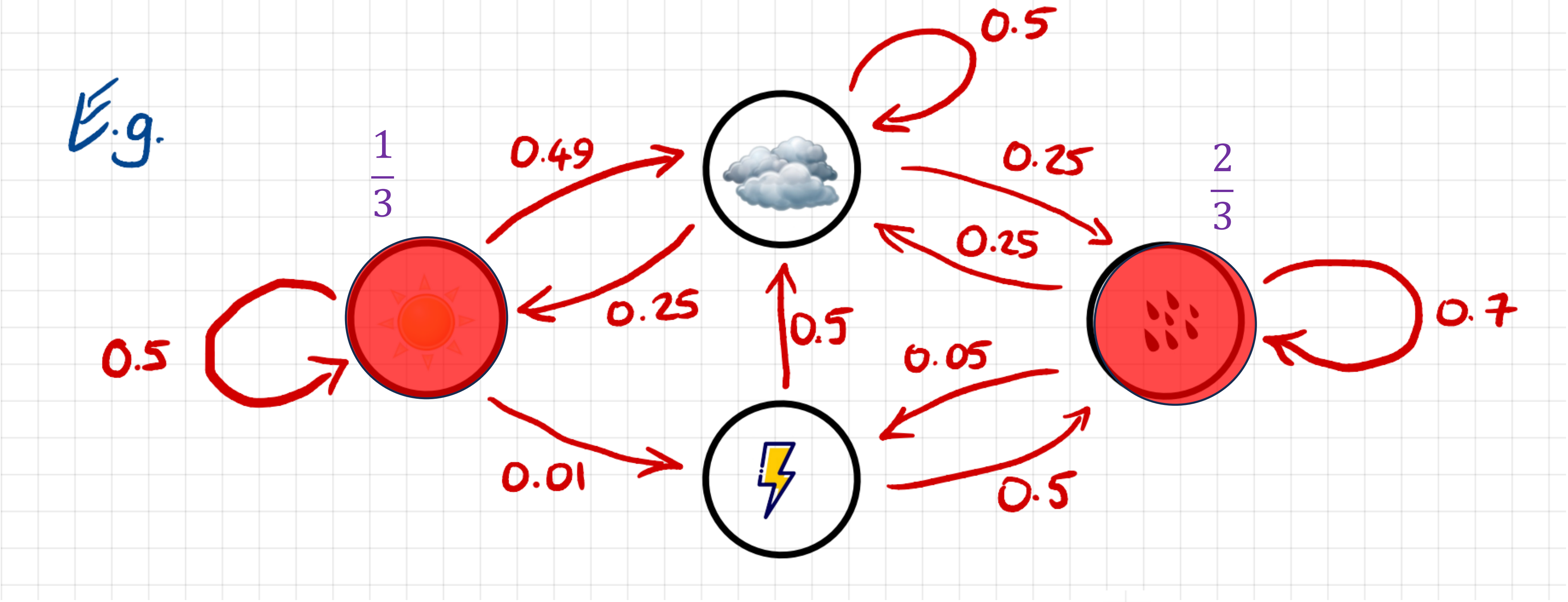
$$\pi_t(j) = \mathbb{P}[X_t = j]$$

$$= \sum_{i \in \mathcal{X}} \mathbb{P}[X_{t-1} = i] \cdot \mathbb{P}[X_t = j \mid X_{t-1} = i]$$

(law of total probability)

$$= \sum_{i \in \mathcal{X}} \pi_{t-1}(i) \cdot P_{i,j}$$

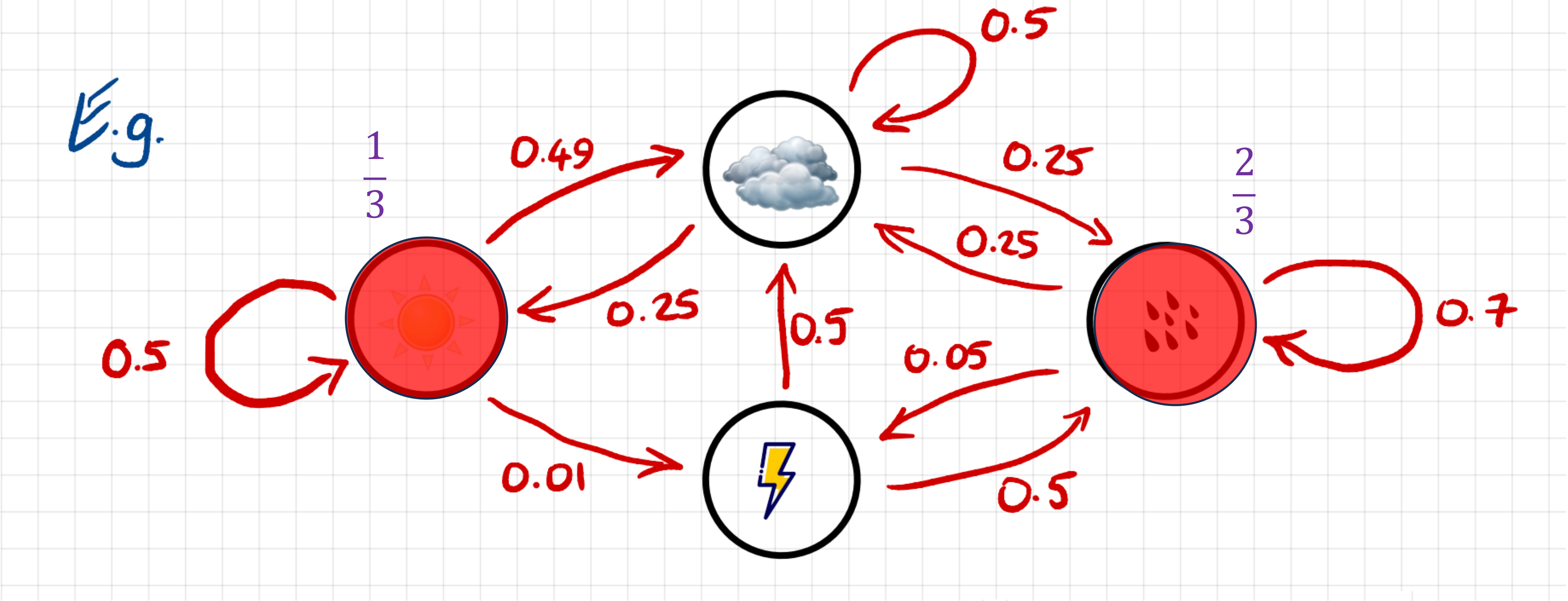
Distribution of random walk - Example



$$\pi_0 = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right)$$

Picture is from Prof. Sinclair's slides in Spring 2024

Distribution of random walk - Example



$$\pi_1 = \left(\frac{1}{3} * 0.49 + \frac{2}{3} * 0.25, \quad \frac{1}{3} * 0.5, \quad \frac{2}{3} * 0.7, \quad \frac{1}{3} * 0.01 + \frac{2}{3} * 0.05\right)$$

Picture is from Prof. Sinclair's slides in Spring 2024

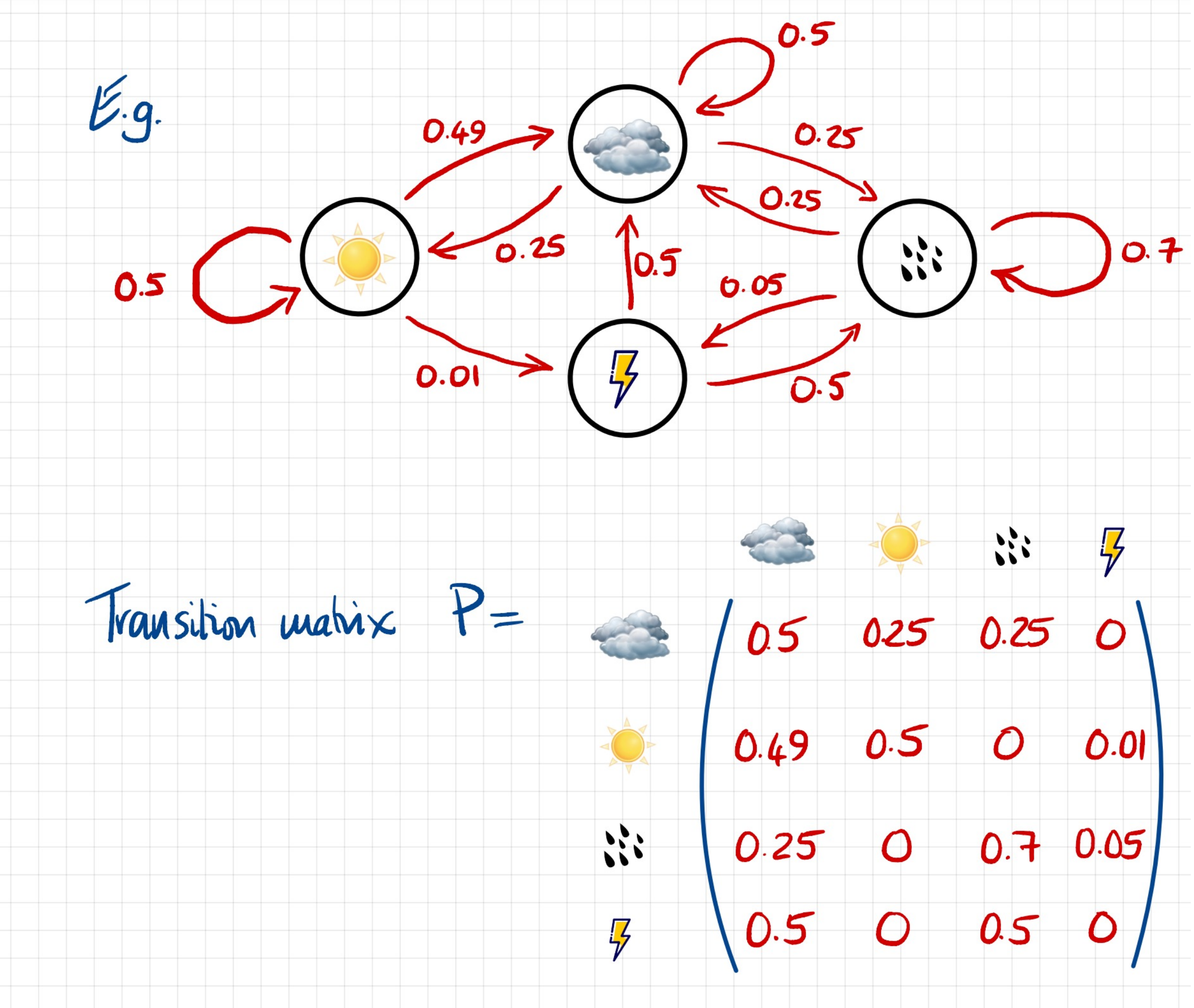
Matrix-vector representation

Transition between states:

$$\mathbb{P}[X_t = j \mid X_{t-1} = i] = P_{i,j}.$$

We can write P as a matrix.

This is called the transition matrix.



Picture is from Prof. Sinclair's slides in Spring 2024

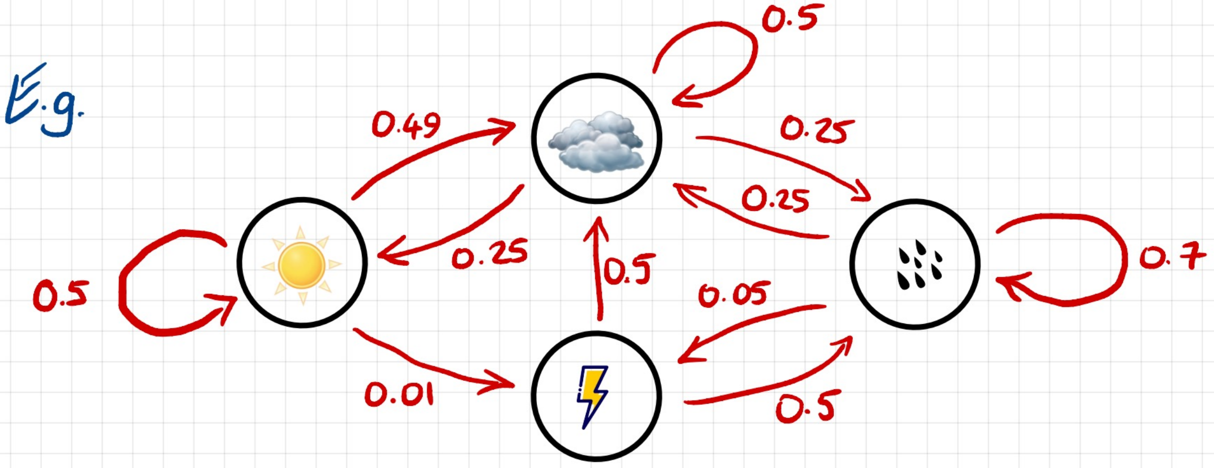
Matrix-vector representation

Distribution π_t.

For the distribution π_t of states at step t, we can write it as a row vector.

$$\boldsymbol{\pi_0} = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right)$$

E.g.



Transition matrix $P = \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.49 & 0.5 & 0 & 0.01 \\ 0.25 & 0 & 0.7 & 0.05 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$

Picture is from Prof. Sinclair's slides in Spring 2024

Matrix-vector representation

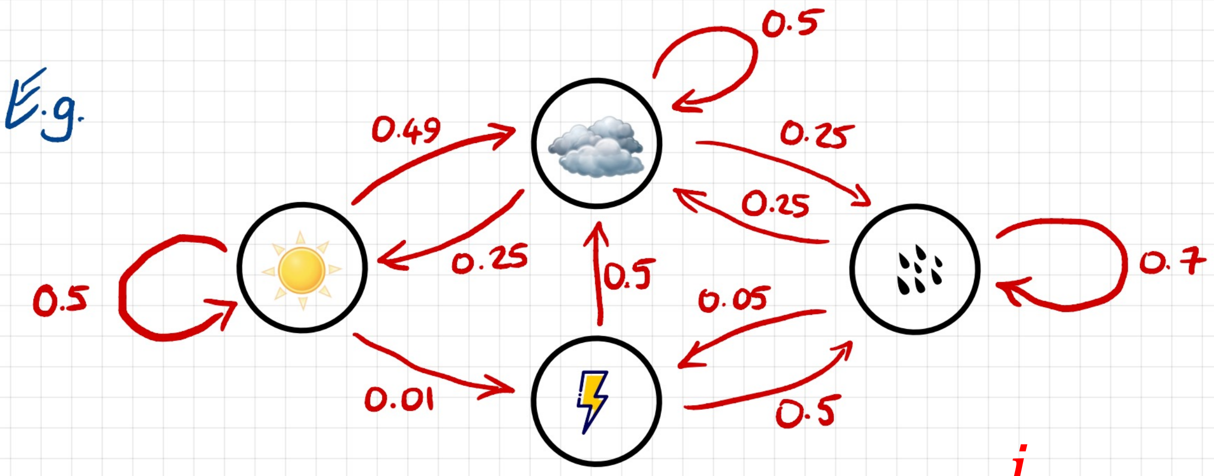
From π_{t-1} to π_t .

$$\pi_t(j) = \sum_{i \in \mathcal{X}} \pi_{t-1}(i) \cdot P_{i,j}$$

$$\boxed{\pi_0 = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right)}$$

i

E.g.



Transition matrix $P =$

$$\begin{array}{cccccc} & & & j & & \\ & 0.5 & 0.25 & 0.25 & 0 & \leftarrow i \\ & 0.49 & 0.5 & 0 & 0.01 & \\ & 0.25 & 0 & 0.7 & 0.05 & \\ & 0.5 & 0 & 0.5 & 0 & \end{array}$$

$\times$

Picture is from Prof. Sinclair's slides in Spring 2024

Matrix-vector representation

From π_{t-1} to π_t .

$$\pi_t = \pi_{t-1} \cdot P$$

$$\begin{array}{c} \hphantom{\boldsymbol{\pi_1} = (\pi_1(1),\pi_1(2),}\ j \hphantom{(3),\pi_1(4))} \\ \boldsymbol{\pi_1} = (\pi_1(1), \pi_1(2), \pi_1(3), \pi_1(4)) \end{array}$$

$$P = \begin{array}{cc} \begin{array}{cccc} \hphantom{0.49} & \hphantom{0.25} & j & \hphantom{0.05} \end{array} & \\ \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.49 & 0.5 & 0 & 0.01 \\ 0.25 & 0 & 0.7 & 0.05 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} & \begin{array}{c} i \\ \\ \\ \\ \end{array} \end{array}$$

$$\times$$

$$\boxed{\boldsymbol{\pi_0} = \left(0, \frac{1}{3}, \frac{2}{3}, 0\right)}$$

$$\begin{array}{c} i \end{array}$$

Picture is from Prof. Sinclair's slides in Spring 2024

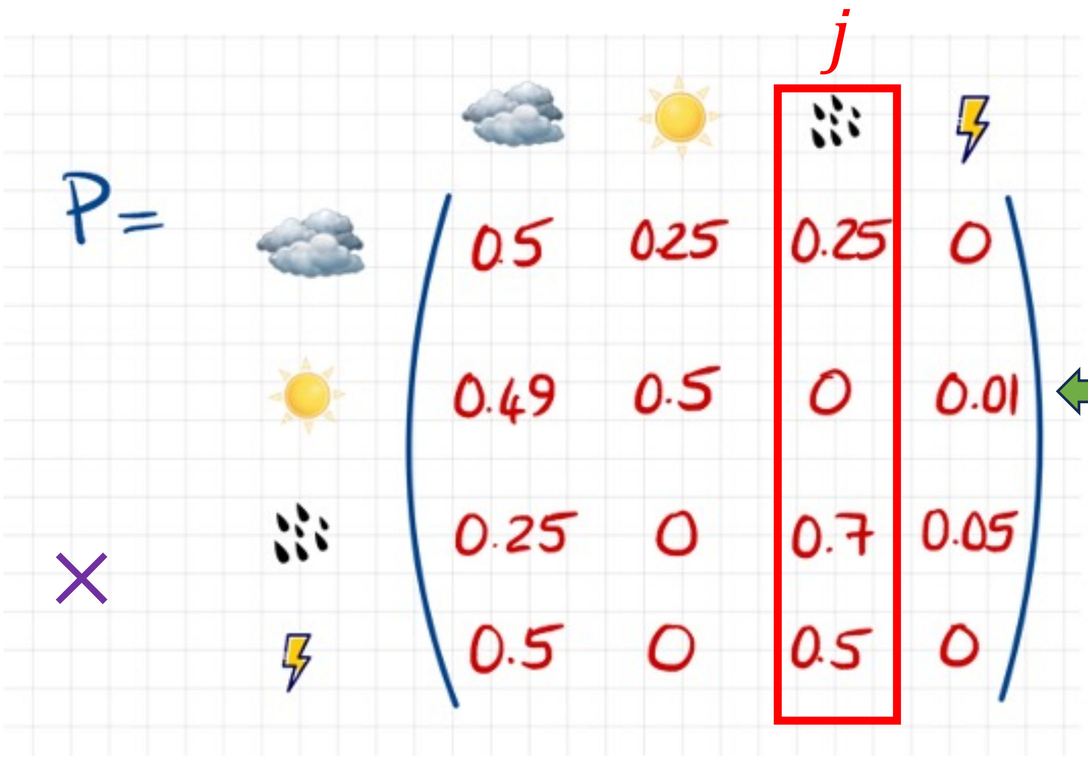
Matrix-vector representation

From π_{t-1} to π_t .

$$\pi_t = \pi_{t-1} \cdot P$$

$$j$$

$$\boldsymbol{\pi_1} = (\pi_1(1), \pi_1(2), \pi_1(3), \pi_1(4))$$



i

$$\boldsymbol{\pi_0} = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right)$$

i

Picture is from Prof. Sinclair's slides in Spring 2024

Matrix-vector representation

From π_{t-1} to π_t .

$$\pi_t = \pi_{t-1} \cdot P$$

$$\boldsymbol{\pi_0} = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right)$$

i

j

$$\boldsymbol{\pi_1} = (\pi_1(1), \pi_1(2), \pi_1(3), \pi_1(4))$$

$$\begin{array}{c} \\ \times \end{array} \quad P = \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.49 & 0.5 & 0 & 0.01 \\ 0.25 & 0 & 0.7 & 0.05 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} \begin{array}{c} \\ \\ i \\ \\ \end{array}$$

Picture is from Prof. Sinclair's slides in Spring 2024

Matrix-vector representation

From π_{t-1} to π_t .

$$\pi_t = \pi_{t-1} \cdot P$$

$$\begin{array}{c} \boldsymbol{\pi_0} = \left(0, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad 0\right) \\ \qquad\qquad\qquad\qquad\qquad\quad i \end{array}$$

$$\begin{array}{c} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad j \\ \boldsymbol{\pi_1} = (\pi_1(1), \pi_1(2), \pi_1(3), \pi_1(4)) \end{array}$$

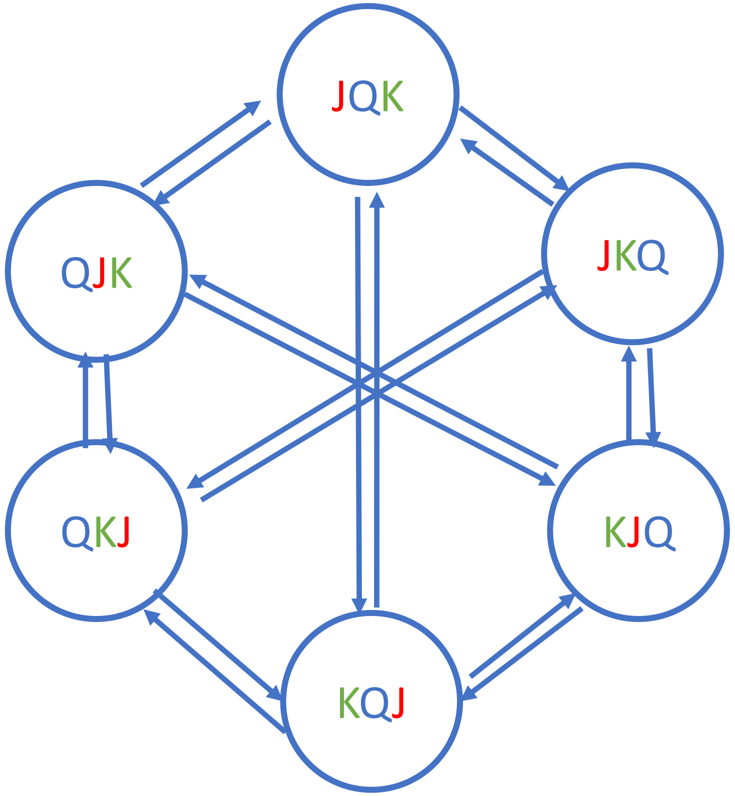
$$\times \quad P = \begin{array}{c} \begin{array}{cccc} & & j & \end{array} \\ \left(\begin{array}{cccc} 0.5 & 0.25 & \boxed{0.25} & 0 \\ 0.49 & 0.5 & 0 & 0.01 \\ 0.25 & 0 & 0.7 & 0.05 \\ 0.5 & 0 & 0.5 & 0 \end{array}\right) \begin{array}{c} \\ \\ \\ \leftarrow i \end{array} \end{array}$$

Picture is from Prof. Sinclair's slides in Spring 2024

Convergence

What is the distribution π_t as $t \to \infty$?

Example: When shuffling cards, will π_t converge?

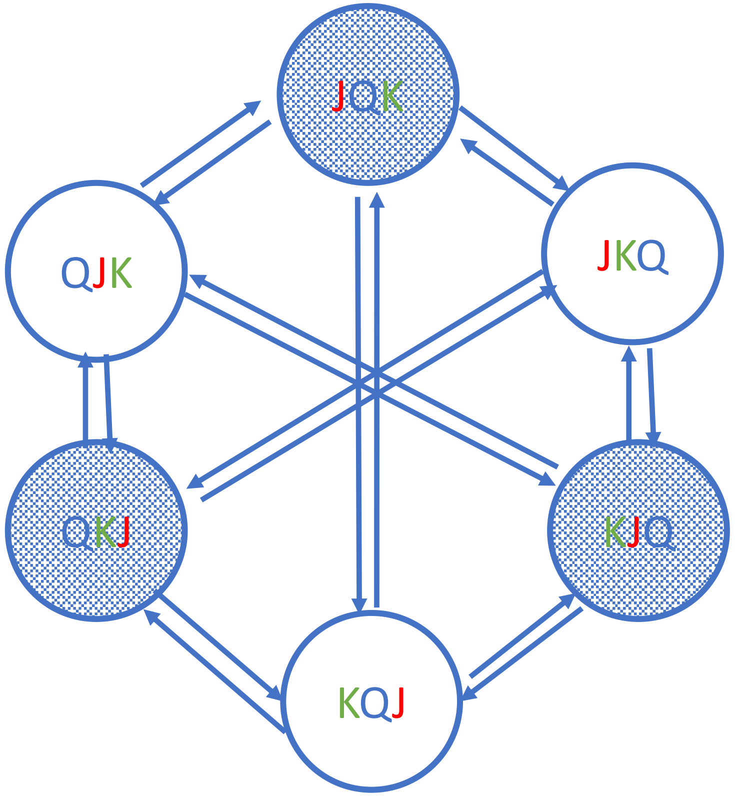


All edge has prob 1/3

Convergence

What is the distribution π_t as $t \to \infty$?

Example: When shuffling cards, will π_t converge? NO!

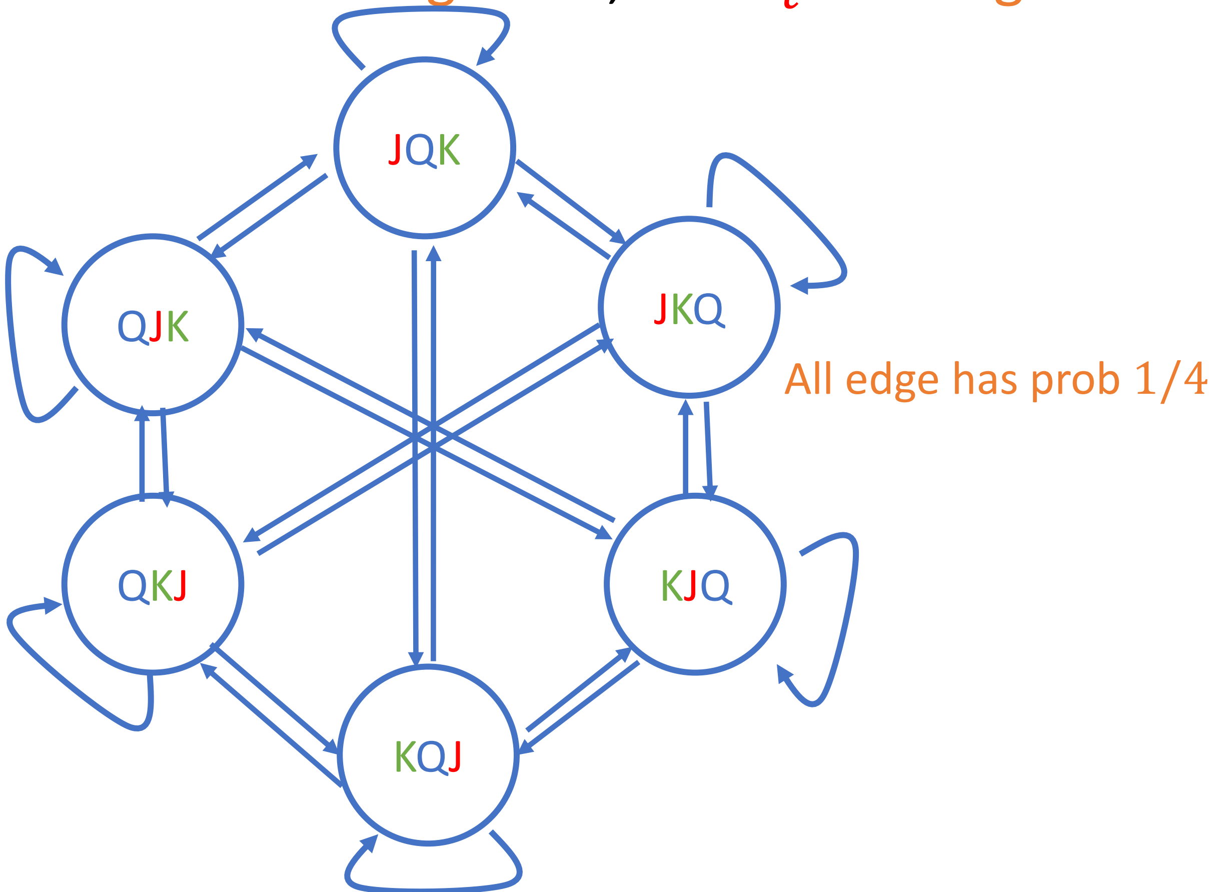


All edge has prob 1/3

Convergence

What is the distribution π_t as $t \to \infty$?

Example: When shuffling cards, will π_t converge? Yes if add self-loops.



Invariant Distribution
(aka Stationary Distribution)

If it does converge, to which distribution?

Definition:

The invariant distribution of a Markov chain is the distribution π that satisfies the balance equation: $\pi = \pi \cdot P$

"If the distribution is already π, after one step of transition, it is still π."

Note: If π is invariant distribution, When every $\pi_t = \pi$ for all π_{t+n} we have

$$\pi_{t+n} = \pi_t \cdot P^n = \pi$$

Find Invariant Distribution: Solve balance equ

$\pi = \pi \cdot P$ gives us $|X|$ equations to solve for π.

$$[\pi(1), \pi(2), \pi(3), \pi(4)] \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.49 & 0.5 & 0 & 0.01 \\ 0.25 & 0 & 0.7 & 0.05 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} = [\pi(1), \pi(2), \pi(3), \pi(4)]$$

$$\Rightarrow \quad \begin{aligned} 0.5\,\pi(1) + 0.49\,\pi(2) + 0.25\,\pi(3) + 0.5\,\pi(4) &= \pi(1) \\ 0.25\,\pi(1) + 0.5\,\pi(2) &= \pi(2) \\ 0.25\,\pi(1) + 0.7\,\pi(3) + 0.5\,\pi(4) &= \pi(3) \\ 0.01\,\pi(2) + 0.05\,\pi(3) &= \pi(4) \end{aligned}$$

Picture is from Prof. Sinclair's slides in Spring 2024

Find Invariant Distribution: Solve balance equ

$\pi = \pi \cdot P$ gives us $|X|$ equations to solve for π.

$$\Rightarrow \begin{aligned} 0.5\,\pi(1) + 0.49\,\pi(2) + 0.25\,\pi(3) + 0.5\,\pi(4) &= \pi(1) \\ 0.25\,\pi(1) + 0.5\,\pi(2) &= \pi(2) \\ 0.25\,\pi(1) + 0.7\,\pi(3) + 0.5\,\pi(4) &= \pi(3) \\ 0.01\,\pi(2) + 0.05\,\pi(3) &= \pi(4) \end{aligned}$$

However, this is not enough. Because these are homogenous equations.

If π is a solution, any $\alpha \cdot \pi$ will also be a solution.

Picture is from Prof. Sinclair’s slides in Spring 2024

Find Invariant Distribution: Solve balance equ

$\pi = \pi \cdot P$ gives us $|X|$ equations to solve for π.

$$
\Rightarrow \quad
\begin{aligned}
0.5\,\pi(1) + 0.49\,\pi(2) + 0.25\,\pi(3) + 0.5\,\pi(4) &= \pi(1) \\
0.25\,\pi(1) + 0.5\,\pi(2) &= \pi(2) \\
0.25\,\pi(1) + 0.7\,\pi(3) + 0.5\,\pi(4) &= \pi(3) \\
0.01\,\pi(2) + 0.05\,\pi(3) &= \pi(4)
\end{aligned}
$$

However, this is not enough. Because these are homogenous equations.

If π is a solution, any $\alpha \cdot \pi$ will also be a solution.

Add $\sum_{i \in X} \pi(i) = 1$.

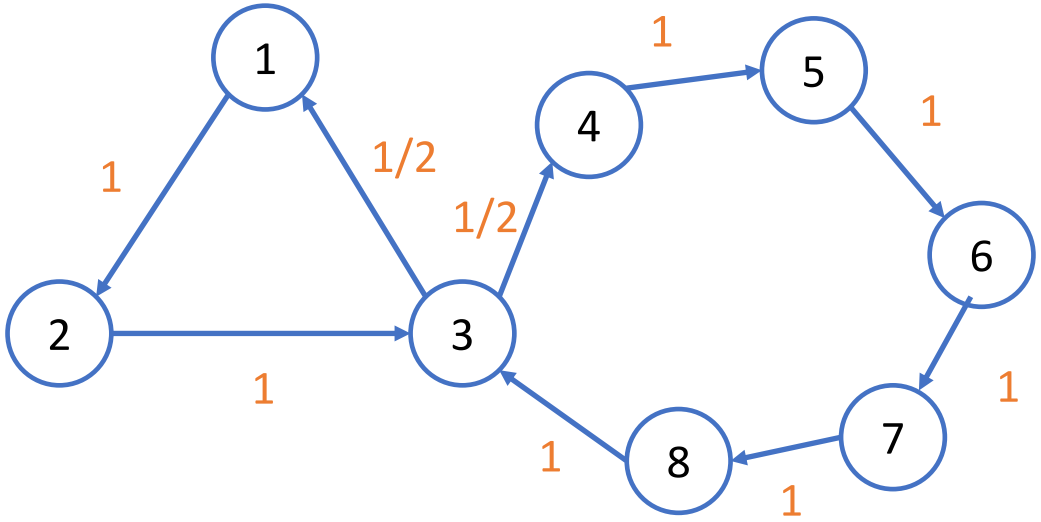
Picture is from Prof. Sinclair's slides in Spring 2024

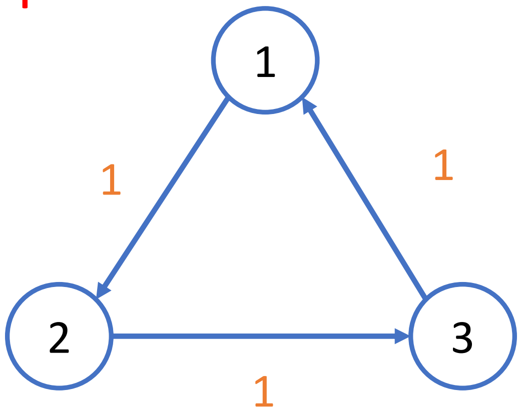
When do we converge to stationary?

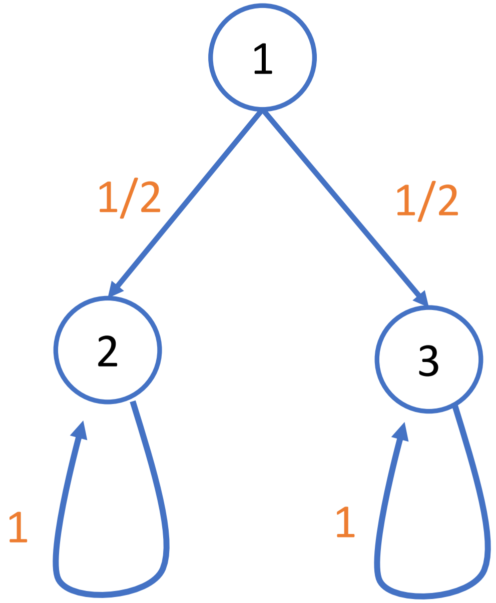
Which Markov chains converge?

Is the stationary distribution always unique?

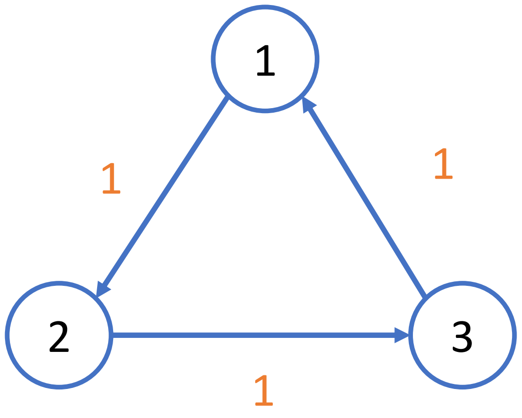
Counterexamples:

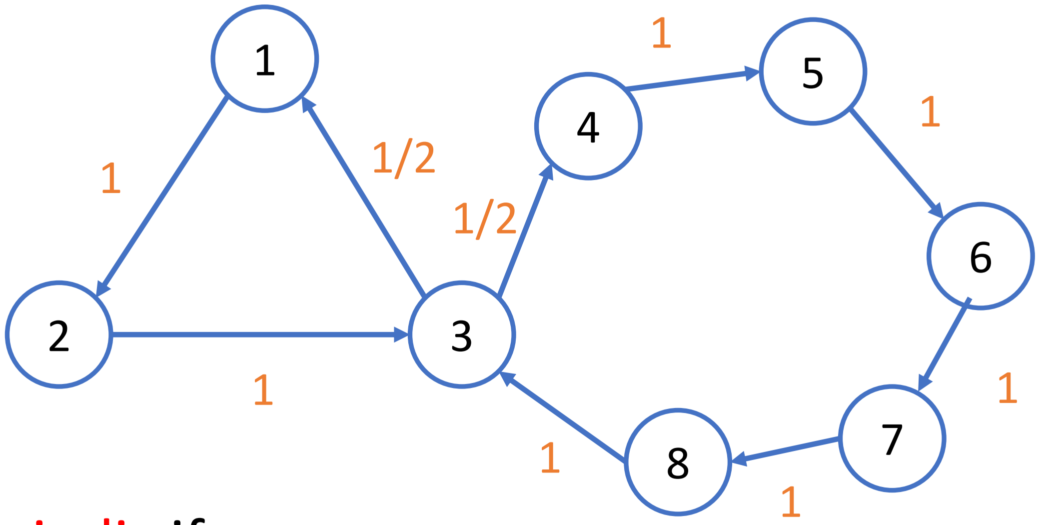






Non-convergent Case1: Periodic



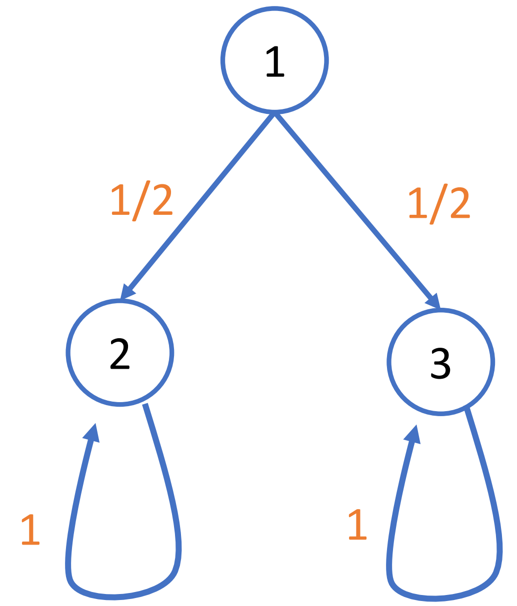


Definition: We say a Markov chain is aperiodic if

For any two states $i, j \in \mathcal{X}$, $\gcd\{n: P^n(i,j) > 0\} = 1$.

That is, the lengths of all the paths $i \to j$ does not have a nontrivial common period.

Non-convergent Case2: reducible

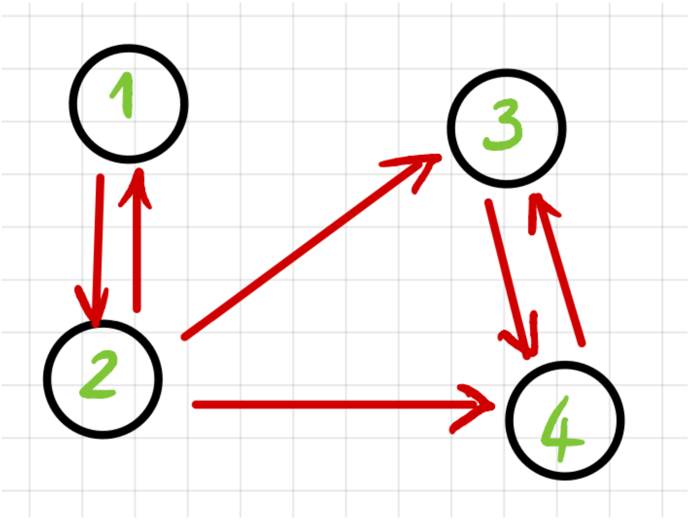


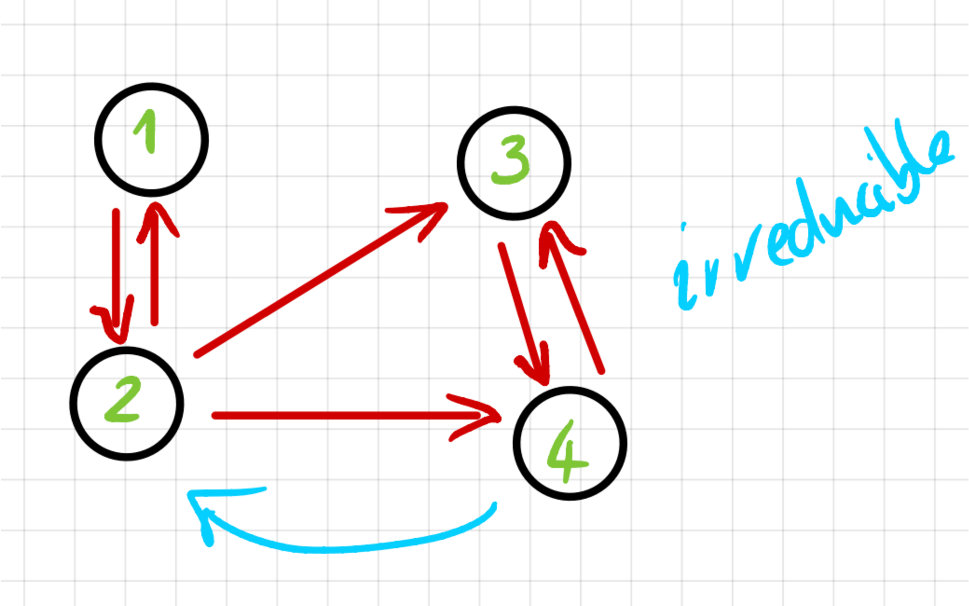
Definition: We say a Markov chain is irreducible if

For any two states $i, j \in \mathcal{X}$, $\exists\, n.\ s.t.\ \ P^n(i,j) > 0.$

That is, there exists a path $i \to j$.

Non-convergent Case2: reducible





Definition: We say a Markov chain is irreducible if

For any two states $i, j \in \mathcal{X}$, $\exists\, n.\ s.t.\ \ P^n(i,j) > 0.$

That is, there exists a path $i \to j$.

Fundamental Theorem for Markov Chains.

Theorem:

If a finite Markov chain $(\mathcal{X}, P)$ is irreducible & aperiodic, then it has a unique Invariant distribution π with $\pi(i) > 0$ for all $i \in \mathcal{X}$.

Also for any initial distribution π_0, as $n \to \infty$, we have $\pi_n = \pi_0 \cdot P^n$ converges to π.

That is, as for all $i \in \mathcal{X}$, as $n \to \infty$, we have $\mathbb{P}[X_n = i] \to \pi(i)$.