Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Nov. 26 2020

Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Chong Han

Outline

- 4. Random Process-II
  - Introduction to Markov Processes
  - Classifications of States and MCs
  - Computing State Probabilities
  - Continuous-time MC
  - Ergodicity Theorems
  - Series Expansions

No problem for non-steady state

Steady State Probability
(P8 之后 uncovered)

Not covered

Introduction to Markov Processes

Ve501 2020-2021 Fall

Introduction to Markov Processes

A Markov Process is a RP such that its future is independent of its past, if its present value is given.

This means that if $t_1 < t_2$, then

$$P\left(X\left(t_2\right) \le x_2 \mid X\left(t\right), t \le t_1\right) = P\left(X\left(t_2\right) \le x_2 \mid X\left(t_1\right)\right)$$

Classes of MPs

- Time can be discrete or continuous
- State can be discrete or continuous
- The discrete-time random walk is discrete-time, discrete-state Markov process
  - also known as Discrete Time Markov Chain (DTMC)
- In this course, we focus on DTMC and CTMC

General Properties

for & DTMC CTMC

These five properties hold for all Markov Processes, (although they are stated here for DTMC).

Let X_n denote the state of the DTMC at time n.

1. $$P\left(X_n = i_n \mid X_0 = i_0, X_1 = i_1, \ldots X_{n-1} = i_{n-1}\right)$$
   $$= P\left(X_n = i_n \mid X_{n-1} = i_{n-1}\right)$$
   This is just the Markovian (memoryless) property stated for a DTMC.
2. $$E\left[X_n \mid X_0, X_1, \ldots, X_{n-1}\right] = E\left[X_n \mid X_{n-1}\right]$$
3. A Markov Process still has the Markovian property if time is reversed.
   $$P\left(X_n = i_n \mid X_{n+1} = i_{n+1}, \ldots, X_{n+k} = i_{n+k}\right)$$
   $$= P\left(X_n = i_n \mid X_{n+1} = i_{n+1}\right)$$

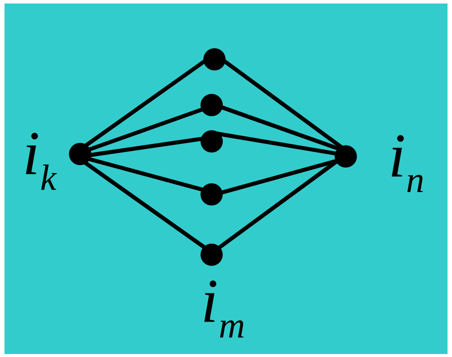
General Properties, Cont'd

4. Past and future are independent, given the present. Let $k < m < n$, then

$$P\left(X_k = i_k \cap X_n = i_n \mid X_m = i_m\right)$$
$$= P\left(X_k = i_k \mid X_m = i_m\right) P\left(X_n = i_n \mid X_m = i_m\right)$$

5. Chapman-Kolmogorov Equation: Let $k < m < n$,

$$P\left(X_n = i_n \mid X_k = i_k\right)$$
$$= \sum_{\text{all } i_m} P\left(X_n = i_n \mid X_m = i_m\right) P\left(X_m = i_m \mid X_k = i_k\right)$$



All possible states

Proof of Chapman-Kolmogorov Equation

Assume $k < m < n$,

$$P(X_n = i_n \mid X_k = i_k)$$

$$= \sum_{\text{all } i_m} P(X_n = i_n \cap X_m = i_m \mid X_k = i_k)$$

Getting a marginal

By definition of conditional prob.

$$= \sum_{\text{all } i_m} P(X_n = i_n \mid X_m = i_m \cap X_k = i_k) P(X_m = i_m \mid X_k = i_k)$$

By Markov Property

$$= \sum_{\text{all } i_m} P(X_n = i_n \mid X_m = i_m) P(X_m = i_m \mid X_k = i_k) \quad \checkmark$$

Proof of Chapman-Kolmogorov

This property is proven using the Total Probability Theorem:

$$P(A \mid C) = \sum_{B} P(A \mid B, C) P(B \mid C)$$

$$P(X_n = i_n \mid X_k = i_k) = \sum_{\text{all } i_m} P(X_n = i_n \mid X_m = i_m, X_k = i_k) P(X_m = i_m \mid X_k = i_k)$$

The Markov property eliminates this term

Transition Probabilities

Conditional probabilities of the form

$$P\left(X_n = i_n \mid X_m = i_m\right)$$

for n > m are called transition probabilities.

For continuous-state Markov Process, we use transition densities

$$f_{X(t_n)|X(t_m)}\left(x \mid y\right)$$

A Markov Process, therefore, is completely specified by its transition probabilities (densities)

$$P\left(X_n = i\right) \quad \left(\text{or } f_{X(t_n)}(x)\right)$$

State Enumeration

In the previous slides, we indicated the value of the DTMC at time n by i_n

$$P\left(X_n = i_n \mid X_m = i_m\right)$$

For the rest of this module, we will assume that the state values are {0,1,2,..}, and write

$$P\left(X_n = j \mid X_m = i\right)$$

For DTMCs that do not have these values, we can still use this notation by simply enumerating the values. In other words, there would be a one-to-one mapping between the real state values and the state indices

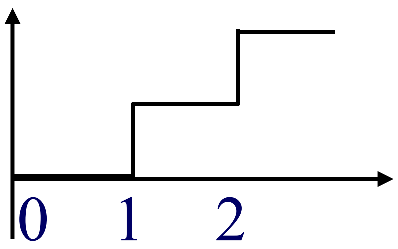
Homogeneity

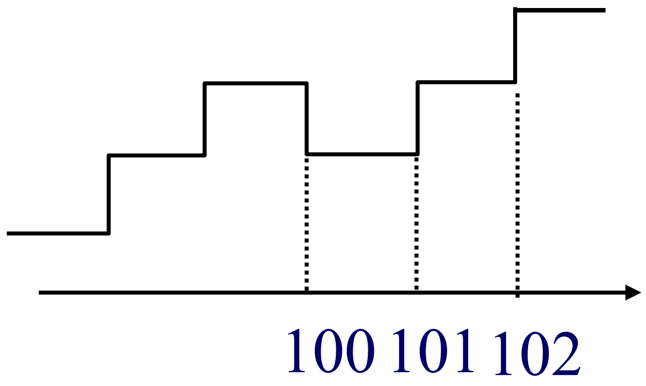
A Markov Process X_n is homogenous if the transition probability is invariant to a shift of the time origin:

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

for all n, m, k and i and j in the state space.

Ex: discrete-time random walk.





Notation for the DTMC

Assume the DTMC is homogenous.

Define the n–step transition probability.

$$P_{ij}^{(n)} = P\left(X_{m+n} = j \mid X_m = i\right)$$

$n = 1$: 1-step transition

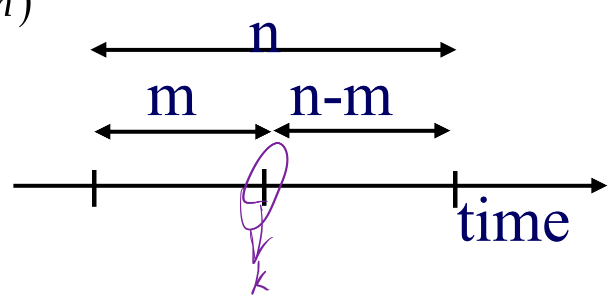
$n \geq 1$: n-step transition

"Homogeneity" means this expression does not depend on m.

The Chapman-Kolmogorov Equation becomes:

$$P_{ij}^{(n)} = \sum_{k} P_{ik}^{(m)} P_{kj}^{(n-m)}$$

k possible states



for time m such that m<n

Transition Probability Matrix

Let P_{ij}, be the one-step transition probability $(n = 1)$. For finite-state spaces, these can be collected into a transition probability matrix

Sum = 1

$$\mathbf{\Pi} = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0N} \\ P_{10} & & & P_{1N} \\ \vdots & & \ddots & \vdots \\ P_{N0} & & \cdots & P_{NN} \end{bmatrix}$$

Each row sums to unity.

Observe that $P_{ij}^{(n)} = P\left(X_{m+n} = j \mid X_m = i\right)$ is the ij[th] element of $\prod^n$.

State Probability Vector

Let $$\pi_j^{(n)} = P\left(X_n = j\right)$$

For an N-state DTMC, the state probability vector is defined as the row vector:

$$\boldsymbol{\pi}^{(n)} = \left[\pi_1^{(n)} \quad \pi_2^{(n)} \quad \pi_3^{(n)} \quad \cdots \quad \pi_N^{(n)}\right]$$

Then the Total Probability Theorem implies:

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-k)}\boldsymbol{\Pi}^k$$

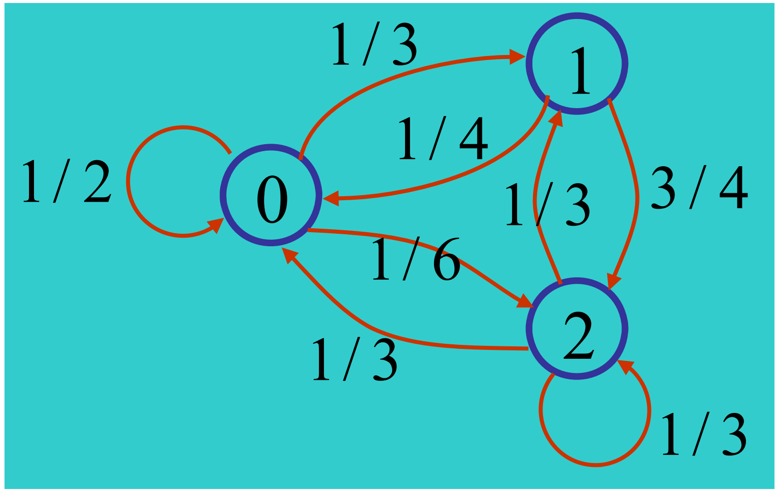
State Diagram

The states and the transition probabilities can be expressed graphically in the state diagram

Ex:

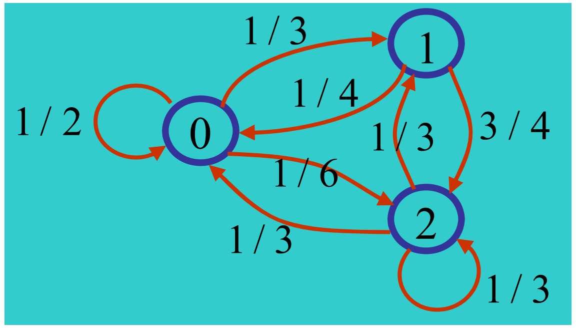
$$N = 3 \qquad \mathbf{\Pi} = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 0 & 3/4 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Label the states $0, 1, 2.$



Example Calculation

$$\mathbf{\Pi} = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 0 & 3/4 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$



Suppose the DTMC starts in state 0. Therefore

$$\boldsymbol{\pi}^{(0)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Calculate the state probability vectors at times 1 and 2.

$$\boldsymbol{\pi}^{(1)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 0 & 3/4 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/3 & 1/6 \end{bmatrix}$$

$$\boldsymbol{\pi}^{(2)} = \begin{bmatrix} 1/2 & 1/3 & 1/6 \end{bmatrix} \mathbf{\Pi} = \begin{bmatrix} 7/18 & 2/9 & 7/18 \end{bmatrix}$$

Example Calculation, Cont'd

For certain cases, the state probability vector approaches a limit, or a steady state, as time goes on, regardless of the initial state. For example:

$\pi^{(0)}$	1	0	0
$\pi^{(1)}$	0.5000	0.3333	0.1667
$\pi^{(2)}$	0.3889	0.2222	0.3889
$\pi^{(3)}$	0.3796	0.2593	0.3611
$\pi^{(4)}$	0.3750	0.2469	0.3781
$\pi^{(5)}$	0.3737	0.2510	0.3737
$\pi^{(6)}$	0.3750	0.2497	0.3754
$\pi^{(7)}$	0.3750	0.2501	0.3749
$\pi^{(8)}$	0.3750	0.2500	0.3750

Time

No matter what the initial state is.

In these cases,

$$\lim_{n\to\infty} \mathbf{\Pi}^{(n)} = \begin{bmatrix} \pi^{\infty} \\ \pi^{\infty} \\ \pi^{\infty} \end{bmatrix}$$

where, every row is π^{∞}, the steady state probability.

Techniques are given for how to find the SS probability vector directly in another module.

Summary

- For a Markov process, knowing the present value of the process tells you all you need to know about the future; past values are irrelevant

- The Chapman-Kolmogorov equation gives a way to express a transition probability in terms of MC values at an intermediate point in time

- The transition probability matrix and state probability vectors enable concise matrix representations of the MC, especially for homogeneous MCs

Classifications of States and MCs

Classification of States

The behavior of a Markov Chain can be better understood by classifying its states in one of these categories:

Transient	Recurrent			
	Recurrent Null		Recurrent Non-null	
	Periodic	Aperiodic	Periodic	Aperiodic (Ergodic)

If All State Ergodic

✓ Steady State Can Exist

Recurrence Probabilities

- Classification depends on the probability that the Markov Process returns to a state after leaving it.

- Let

2 steps

$$f_j^{(n)} = P(\text{first return to state } j \text{ occurs } n \text{ steps after leaving it})$$

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)} = P(\text{ever returning to state } j)$$

If $f_j = 1$, state j is called recurrent

guaranteed to return

If $f_j < 1$, state j is called transient

There are some prob you leave state j and never return

Recurrence Formula

Observe that $f_j^{1} = [\Pi]_{jj}$ ← The jjth element of Π

$f_j^{(n)}$ may be calculated from the recursion

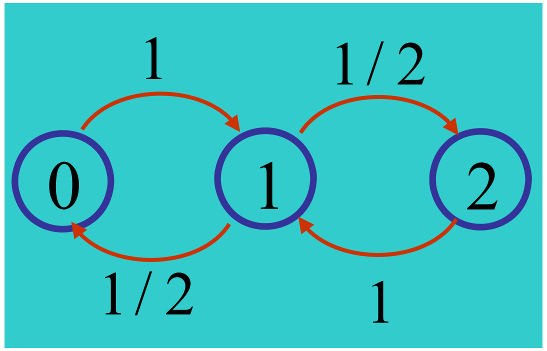
$$[\Pi^n]_{jj} = \sum_{k=1}^{n} f_j^{(k)} [\Pi^{n-k}]_{jj}$$

The jjth element of Π^n

Recurrence Example I

- Consider state 0.
- In order for the process to leave state 0 and never come back, it must cycle through states 1 and 2 an infinite number of times, with probability

$$\left(1/2\right)^{\infty} = 0$$



- Therefore state 0 is a recurrent state
- Alternatively, the formula yields the sequence

$$\left\{f_0^1, f_0^2, f_0^3, f_0^4, \ldots\right\} = \left\{0, \frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{8}, \ldots\right\} \Rightarrow f_0 = \sum_{n=1}^{\infty} f_0^{(1)} = 1$$

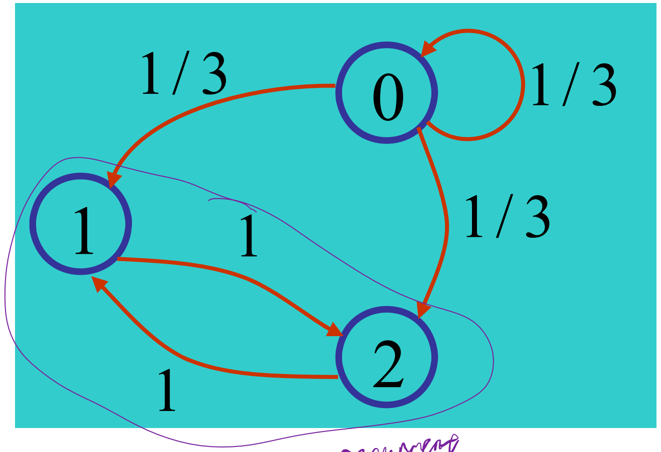
Recurrence Example II

We observe that

$$f_0^{(1)} = \frac{1}{3}$$

$$f_0^{(n)} = 0, \quad n = 2, 3, \ldots$$

$$f_0 = \sum_{n=1}^{\infty} f_0^{(n)} = \frac{1}{3}$$



recurrent

Since $f_0 < 1$, state 0 is transient.

States 1 and 2 are recurrent.

Mean Recurrence Time

For recurrent states, the mean recurrence time for state j is:

$$M_j = \sum_{n=1}^{\infty} n f_j^{(n)} = \text{the average time to return to state j}$$

总会回来
但 expectation of steps = +∞

If $M_j = +\infty$, state j is recurrent null.

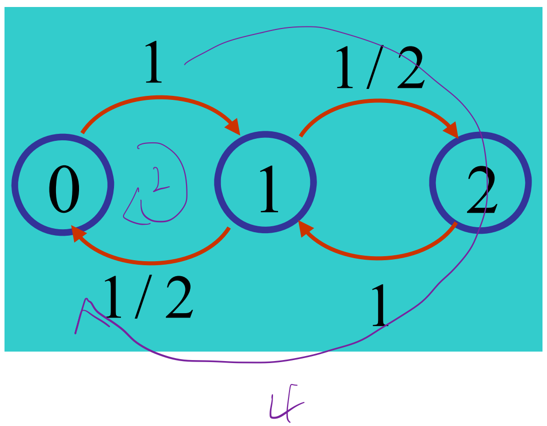
If $M_j < +\infty$, state j is recurrent non-null.

$M_j = +\infty$ can happen when there are an infinite number of states, such as in the discrete-time random walk.

Periodic States

If the only possible numbers of steps for returning to state j are integer multiples of some integer γ, where γ is the largest such integer, then state j is periodic with period γ .

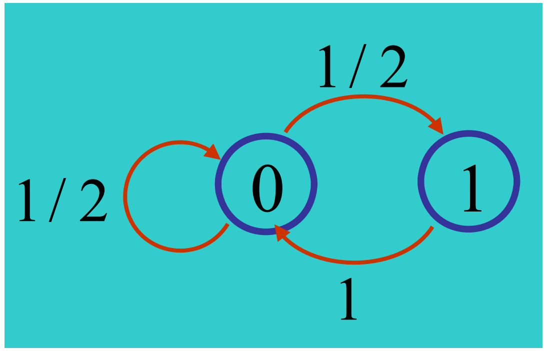
Previous Example:

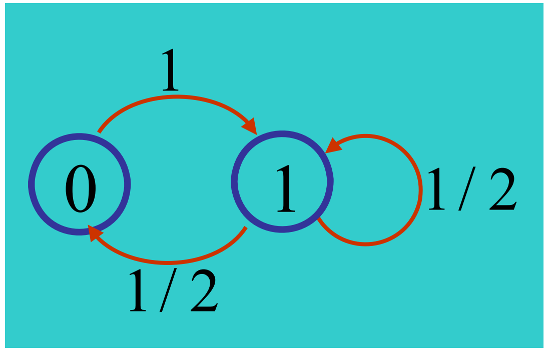


State 0 is periodic with period 2.

Aperiodic States

If $\gamma = 1$, the state is aperiodic.





In both examples above, state 0 is aperiodic.

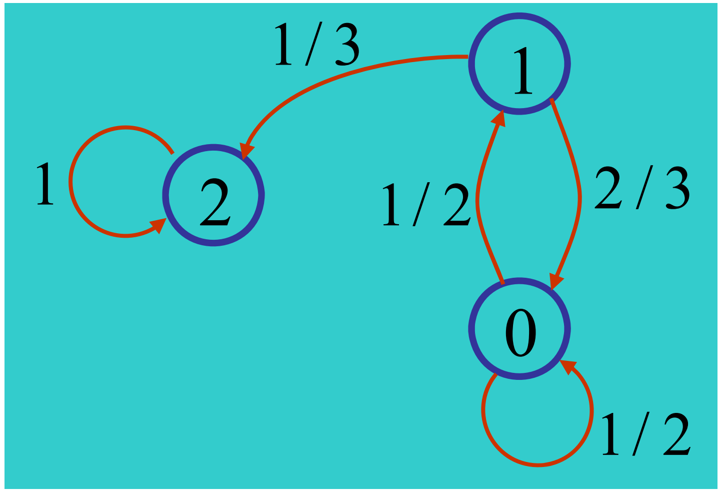
Ergodicity

- A state is ergodic if it is aperiodic and recurrent non-null

- Interpretation:
  - If the state is ergodic, then you can estimate its steady-state probability by observing, over a long window of time, the fraction of time the Markov process occupies this state.

Absorbing

A state is absorbing, if the process never leaves it once the process has entered it.



State 2 is absorbing.

Summary

- States can be classified into the following categories, based on the probabilities of returning for the first time in n steps:

<table>
<tr><td rowspan="3">Transient</td><td colspan="4">Recurrent</td></tr>
<tr><td colspan="2">Recurrent Null</td><td colspan="2">Recurrent Non-null</td></tr>
<tr><td>Periodic</td><td>Aperiodic</td><td>Periodic</td><td>Aperiodic (Ergodic)</td></tr>
</table>

Classification of Homogenous Markov Chains

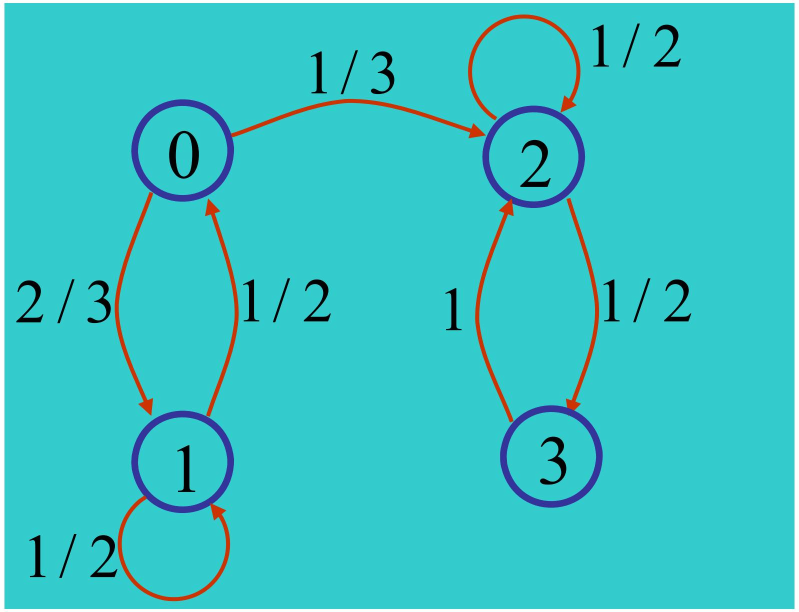
- We have classified states. Now we classify whole Markov Chains

- Knowing how a Markov Chain is classified can simplify its analysis
  - Reducible OR Irreducible
  - Periodic OR Aperiodic
  - Ergodic OR Non-ergodic

Communication Between States

- Two states, i and j, are said to communicate if i is reachable from j and vice versa, or in other words, if there is a sequence of transitions from i to j and back again that occurs with nonzero probability

- States 0 and 1 communicate
- States 1 and 2 do not communicate.



Equivalence Classes

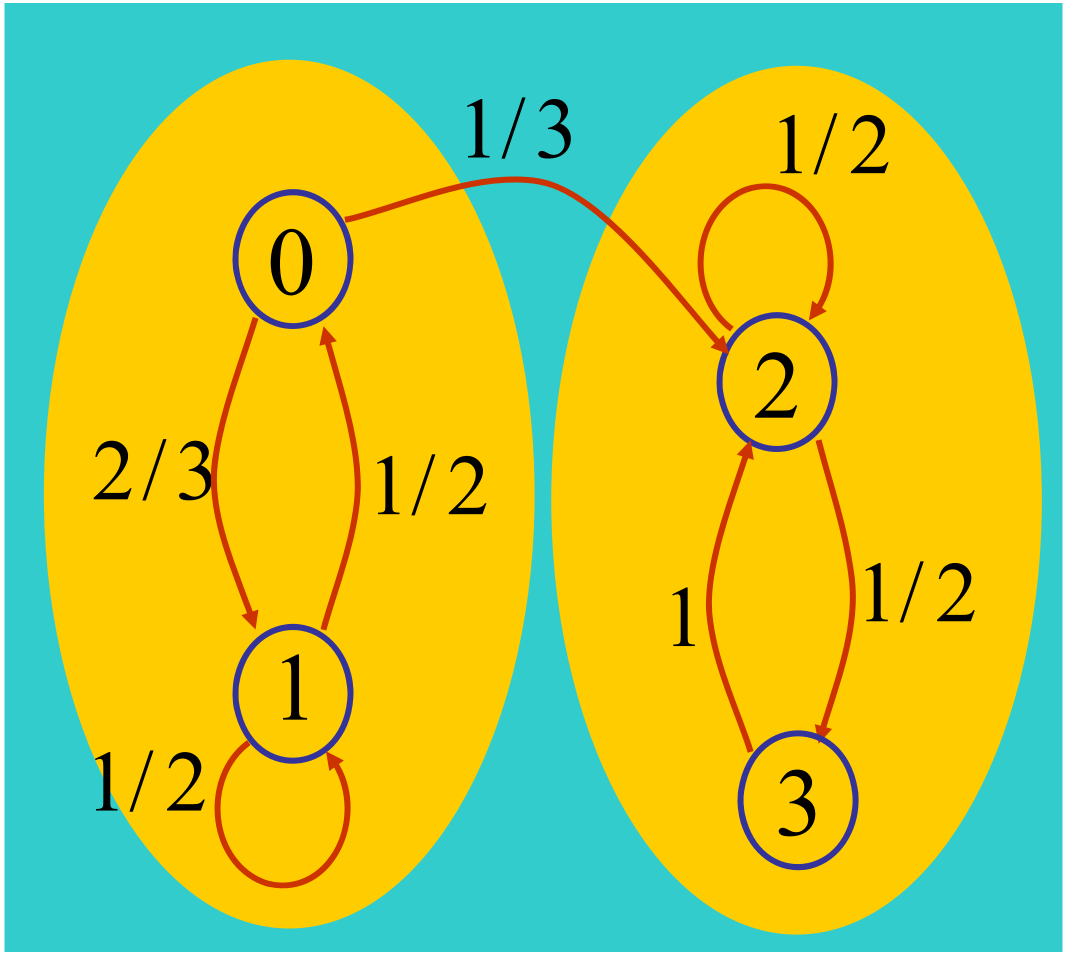
- Two states belong to the same equivalence class (or just class) if they communicate with each other

- Two different classes are disjoint
  - Otherwise, there would be a two-way path between them

- The equivalence classes of a Markov Chain form a partition of the states

Classes Example

- In the previous example, there are two classes indicated by the two ovals



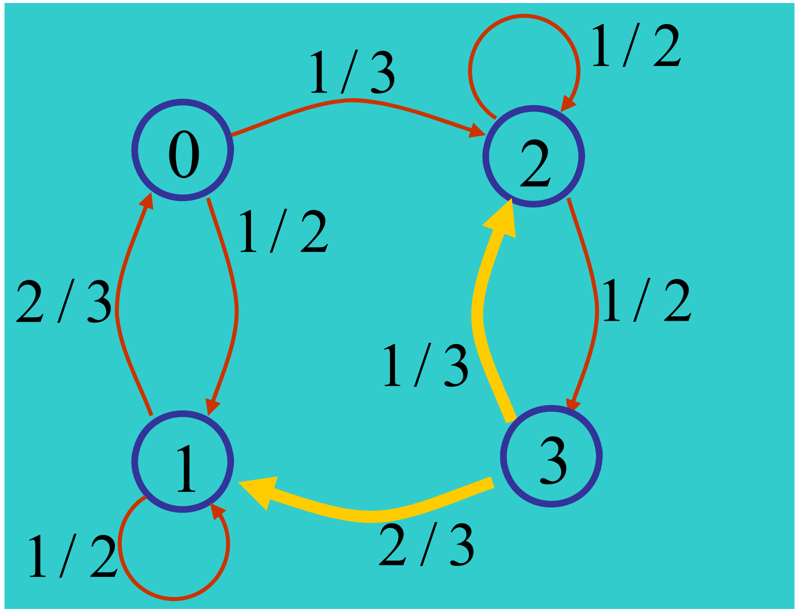
Recurrence Within a Class

- Recall recurrence: A state is recurrent if the MC is guaranteed (i.e. with probably one) to eventually return to the state
- As a MC returns again and again to a particular recurrent state, all the other states in the same class are eventually visited
- Therefore, recurrence is a class property

Reducibility

- A Markov Chain is irreducible if all of its states are in the same equivalence class
- Otherwise, the Markov Chain is reducible

The yellow paths represent a modification of the previous example that makes the new MC irreducible



Transience Within a Class

- Since a state cannot be both transient and recurrent, then there will never be a mix of transient and recurrent states in a class
  - Therefore, transience is a class property

- All states in a finite-state, irreducible (i.e. single class) MC are recurrent non-null
  - Because they can't all be transient

Periodicity

- If a state is periodic, that is, if the state can be revisited only at multiples of some non-unity number of steps, then all states in the class are periodic
- For the random walk, a state recurs when there are an equal number of +1s and −1s, implying that the period is 2

Ergodicity

- Recall that a state is ergodic if it is aperiodic and recurrent non-null

- If all states of a Markov Chain (MC) are ergodic, then the MC is ergodic. 1 ergodic => this class ergodic

- Theorem: All states of a finite, aperiodic, irreducible MC are ergodic.
  - Interpretation - any statistic of the MC, such as the steady state probability vector $\pi^{(\infty)}$, can be estimated by observing one outcome over a long time window.

Uniformity

- Theorem: The states of an irreducible MC are either
  - All recurrent non-null, or
  - All recurrent null

- If the states are periodic, then they all have the same period.

- The random walk is an example of an irreducible MC with all states being recurrent null

Stationarity

The homogenous MC is stationary if it is inititalized with a state probability vector

$$\boldsymbol{\pi}^{(0)} = \boldsymbol{\rho}$$

such that

$$\boldsymbol{\rho} = \boldsymbol{\rho}\boldsymbol{\Pi}^{n} \quad \forall n > 0.$$

The state probability vector of a stationary MC is called a stationary distribution

Steady State Existence

Theorem: In an irreducible and aperiodic homogenous MC, the limiting probabilities

$$\boldsymbol{\pi}^{(\infty)} = \lim_{n\to\infty} \boldsymbol{\pi}^{(n)}$$

always exist and are independent of $\pi^{(0)}$. Moreover, either

a) All states are transient or recurrent null, in which cases $\pi_j^{(\infty)} = 0$ for all j and there exists no stationary distribution, or

b) All states are recurrent non-null, in which cases $\pi_j^{(\infty)} > 0$ for all j and the vector $\pi^{(\infty)}$ is a stationary distribution

Summary

- Markov chain classifications follow from
  - The uniformity of states within equivalence classes
  - The property of having only one class (irreducible MC)
- Stationarity and steady state of MCs

Thank You!