Finite State-Space Markov Chains

Mark Gales

Lent 2019

Stochastic Processes: Handout 1

IIA Module 3M1: Mathematical Methods

Overview

- This part of the course will focus of Markov Chains and related applications
- Course has three parts
  - finite-space Markov chains: 2 lectures
  - continuous state-space systems: 1.5 lectures
  - Monte Carlo Markov chains: 1.5 lectures
- Handouts available from `3M1` web-site
- Insufficient time to examine all topics in detail
  - useful starting point for 4th year modules

Applications of Markov Chains

- Markov Chains under-pin the work in many areas
  - Google Page ranker (very large Markov chain!)
  - Information theory (Entropy)
  - Speech and Language Processing (acoustic models/language models)
  - Physics (statistical mechanics/thermodynamics)
  - Economics and finance (asset pricing)
  - Queueing theory

Pushkin's "Eugene Onegin"

А. А. Марков (1886).

- Andrej Markov (left) analysed Alexander Pushkin's novel Eugene Onegin
  - probability of vowels following consonants (and vice-versa)
  - first analysis of chains (sequences) of (stochastic) events

Language Modelling

- Use current word and previous words (state) to predict the next word

Unigram - No information

- Every enter now severally so, let
- Will rash been and by I the me loves gentle me not slavish page, the and hour; ill let

Bigram - Current word

- What means, sir. I confess she? then all sorts, he is trim, captain.
- The world shall- my lord!

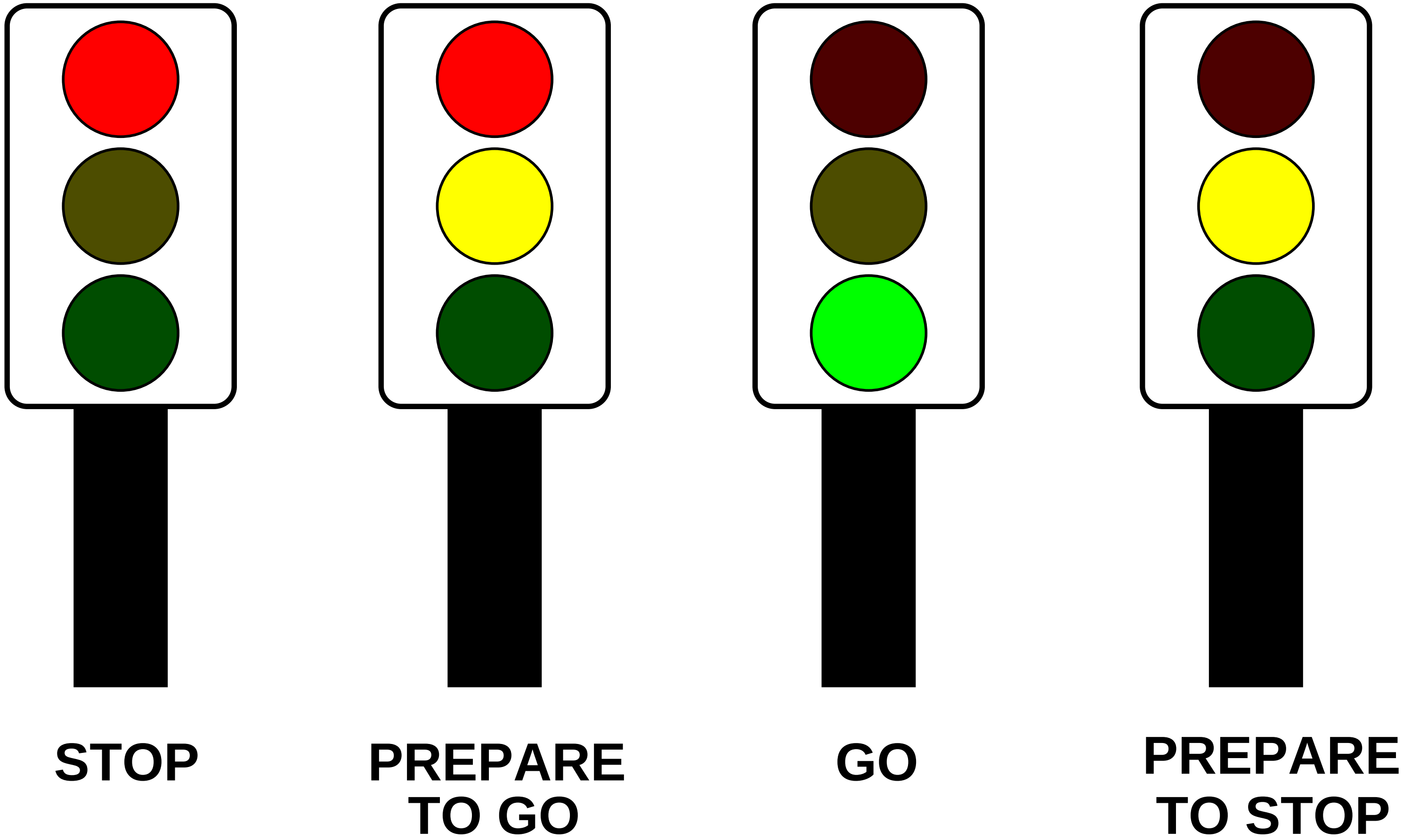
Trigram - Current and previous word

- Indeed the duke; and had a very good friend.
- Sweet prince, Fallstaff shall die. Harry of Monmouth's grave.

4-gram - - Current and previous two words

- It cannot be but so.
- Enter Leonato's brother Antonio, and the rest, but seek the weary beds of people sick.

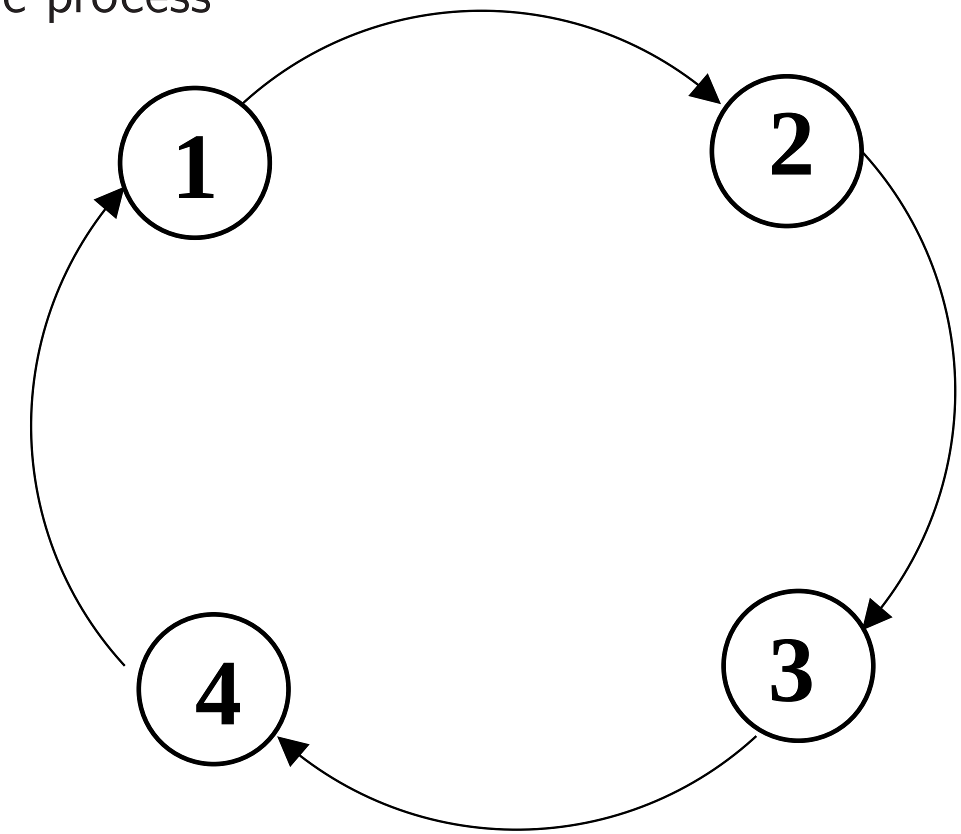
Traffic Lights



- Simple deterministic process:
  - four distinct states (numbered 1 to 4) - with associated meaning!

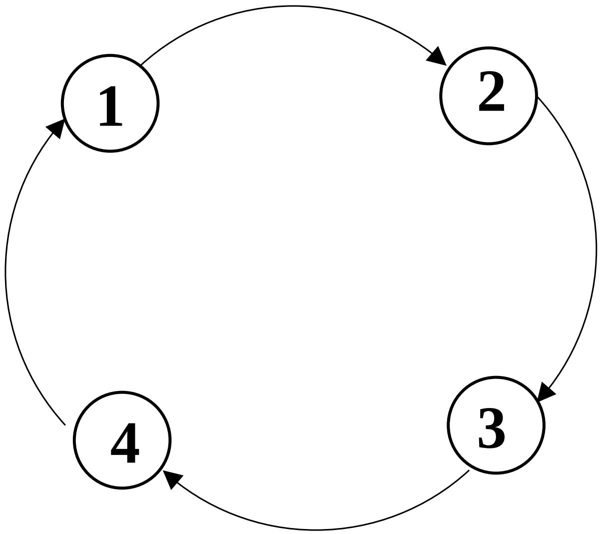
Traffic Lights

- Can describe the process as a state-diagram with transitions
  - a deterministic process



The current state completely determines the next state

Transition Matrix Description



$$\begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

- The transition matrix for this process is very simple:

$$\texttt{red} \rightarrow \texttt{red/amber} \rightarrow \texttt{green} \rightarrow \texttt{amber} \rightarrow \texttt{red}$$
$$\text{state 1} \rightarrow \text{state 2} \rightarrow \text{state 3} \rightarrow \text{state 4} \rightarrow \text{state 1}$$

- More generally each element of the transition matrix $\mathbf{P}$ indicates:

$$P_{j,k} = \text{Prob}(j \rightarrow k), \quad \sum_k P_{j,k} = 1, \quad P_{j,k} \geq 0$$

Traffic Lights in a City

- Consider a (large) city with a large number of traffic lights
- At any instance in time traffic lights (drivers) can be split into
  - `red` (driving stopped)
  - `red/amber` (driver preparing to go)
  - `green` (driver going)
  - `amber` (driver preparing to stop)
- The initial distribution of traffic lights, $\mathbf{x}^{(0)}$, will be written as

$$\mathbf{x}^{(0)} = \left[\; P(X_0 = \texttt{red}) \quad P(X_0 = \texttt{red/amber}) \quad P(X_0 = \texttt{green}) \quad P(X_0 = \texttt{amber}) \;\right]$$

  - at any instance in time a particular traffic light will be in one of the states
  - the transition matrix will describe the state movement of each light

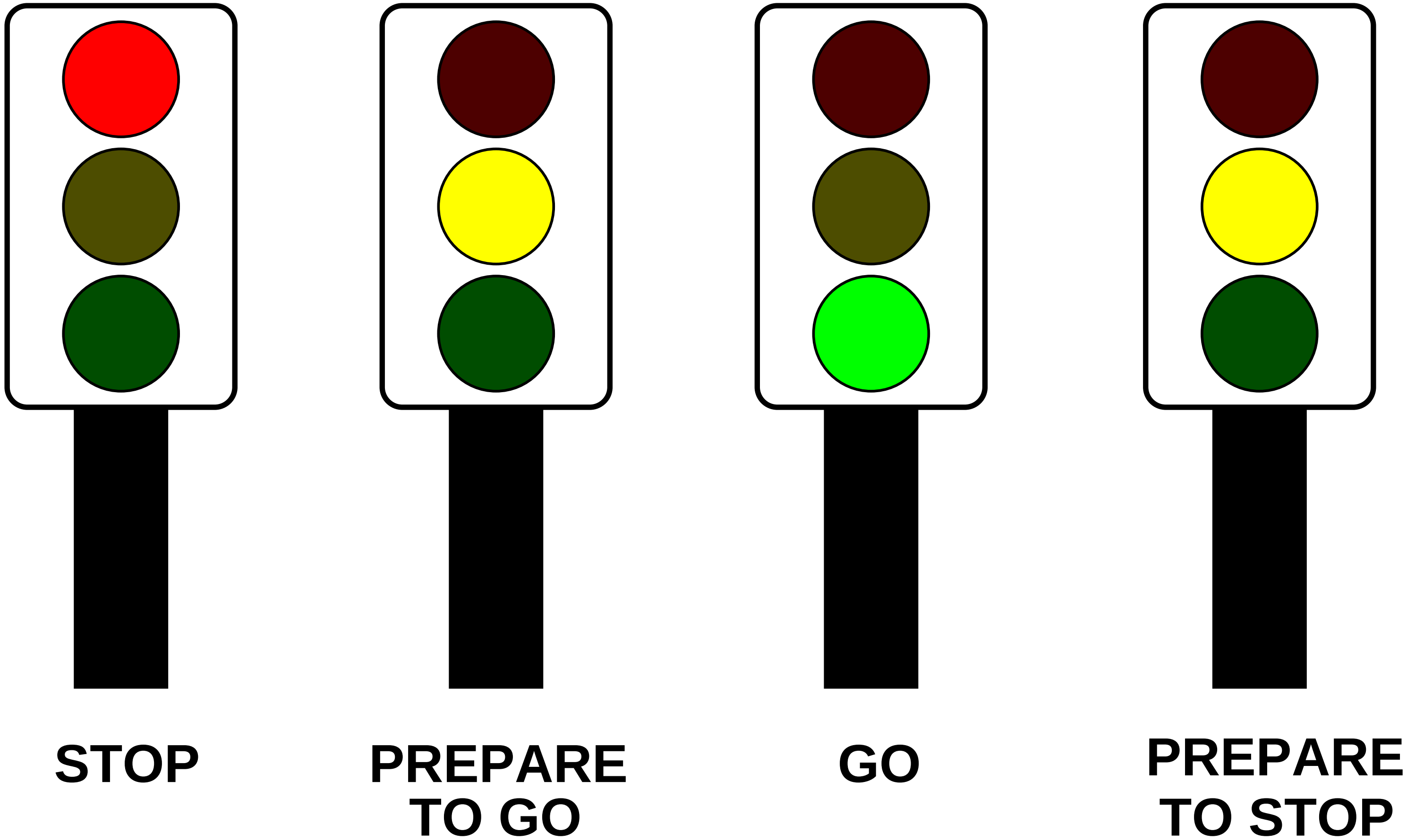
Markov Chain (Formal Definition)

A Stochastic Process $X_0, X_1, X_2, \ldots$ is a Markov chain if and only if for all times $i \geq 1$

$$P(X_{i+1}|X_0 = j_0, X_1 = j_1, \ldots, X_i = j_i) = P(X_{i+1}|X_i = j_i)$$

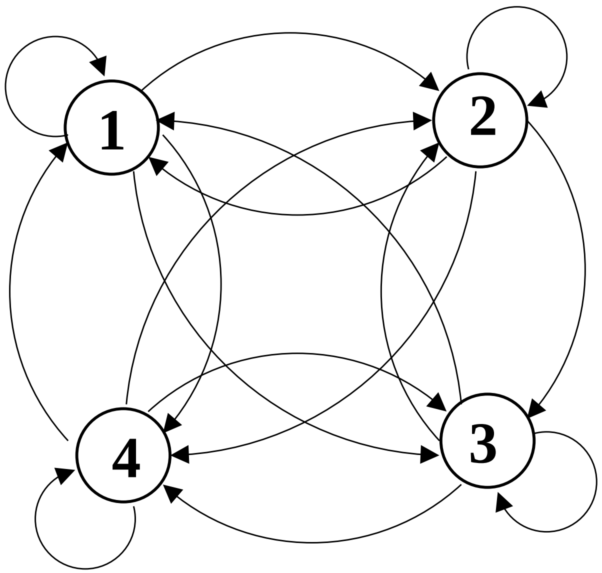
- There are a set of random variables $X_0, X_1, X_2, \ldots$ indexed by time
  - each random variable takes a value from the state-space, $\mathcal{S}$.

Traffic Lights - Ambiguous



- Simple deterministic process:
  - but not a Markov chain
  - amber state ambiguity

Faulty Traffic Lights

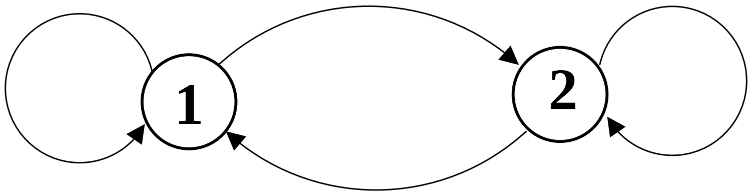


$$\begin{bmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/2 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/2 \end{bmatrix}$$

- How would we want to characterise these sorts of processes:
  - what happens after n steps?
  - what happens if the system is run for a long-time?
  - if light in `red`, how long until light expected to be in `green`?
  - will lights end up in a periodic cycle?

Population of California

- Simple model of population change in California: each year
  - 1/10 of the population outside California moves in
  - 2/10 of the population inside California moves out



$$\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

- Two states:
  - state 1 outside California
  - state 2 inside California

What is the probability of an individual's location?

What happens to the population of California in the log-run?

"Path" Probability

- What is the probability of the following sequence for an individual

$$\texttt{inside} \rightarrow \texttt{outside} \rightarrow \texttt{inside} \rightarrow \texttt{inside}$$

  – initial distribution $\mathbf{x}^{(0)} = \left[\begin{array}{cc} 0.5 & 0.5 \end{array}\right]$

- Need to compute

$$\begin{aligned}
&P(X_0 = \texttt{inside}, X_1 = \texttt{outside}, X_2 = \texttt{inside}, X_3 = \texttt{inside}) \\
&\quad = \quad P(X_0 = \texttt{inside})P(X_1 = \texttt{outside}|X_0 = \texttt{inside}) \\
&\quad \qquad \times P(X_2 = \texttt{inside}|X_1 = \texttt{outside})P(X_3 = \texttt{inside}|X_2 = \texttt{inside}) \\
&\quad = \quad x_2^{(0)} \times 0.2 \times 0.1 \times 0.8 = 0.5 \times 0.2 \times 0.1 \times 0.8 \\
&\quad = \quad 0.008
\end{aligned}$$

Population of California after n Years

- Initial distribution of population at year zero (row vector):

$$\begin{bmatrix} x_1^{(0)} & x_2^{(0)} \end{bmatrix}$$

- After 1 year population:

$$\begin{aligned} \begin{bmatrix} 0.9x_1^{(0)} + 0.2x_2^{(0)} & 0.1x_1^{(0)} + 0.8x_2^{(0)} \end{bmatrix} &= \begin{bmatrix} x_1^{(0)} & x_2^{(0)} \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \\ &= \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \end{bmatrix} \end{aligned}$$

- After n years population (by simple recursion):

$$\begin{bmatrix} x_1^{(n)} & x_2^{(n)} \end{bmatrix} = \begin{bmatrix} x_1^{(n-1)} & x_2^{(n-1)} \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} x_1^{(0)} & x_2^{(0)} \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^n$$

Theorem

For any time n

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)}\mathbf{P}^n$$

where $\mathbf{x}^{(n)}$ is the vector denoting the distribution of X_n.

- Possible to get the transition matrix for n-steps:
  - raise the 1-step matrix to the power n

Chapman-Kolmogorov Equations

- From the previous discussion we can say

$$P(X_{i+n} = k|X_i = j) = (\mathbf{P}^n)_{j,k} = P^n_{j,k}$$

- Chapman-Kolmogorov equations: possible to write as

$$\begin{aligned} P(X_{i+n} = k|X_i = j) &= P^n_{j,k} \\ &= \left(\mathbf{P}^s\mathbf{P}^{n-s}\right)_{j,k} \\ &= \sum_l P^s_{j,l}P^{n-s}_{l,k} \\ &= \sum_l P(X_{i+s} = l|X_i = j)P(X_{i+n} = k|X_{i+s} = l) \end{aligned}$$

Limiting and Stationary Distributions

- For the faulty traffic lights, what happens after a very large number of steps

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/2 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/2 \end{bmatrix} \qquad \mathbf{P}^{\infty} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

- So for an initial distribution $\mathbf{x}^{(0)}$, the "final" distribution is

$$\mathbf{x}^{(\infty)} = \mathbf{x}^{(0)}\mathbf{P}^{\infty} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

  – whatever initial distribution of start states, after "long-enough" equally likely to be in any state!

Limiting Distributions/Regular Ergodic (Definition)

For a Markov Process with transition matrix $\mathbf{P}$, a limiting distribution, $\mathbf{x}^{(\infty)}$, is one which for any initial distribution $\mathbf{x}^{(0)}$ satisfies $\mathbf{x}^{(0)}\mathbf{P}^n = \mathbf{x}^{(\infty)}$ as $n \to \infty$

- For any limiting distribution:

$$\mathbf{x}^{(\infty)}\mathbf{P} = \mathbf{x}^{(\infty)}$$

  – any distribution that satisfies this expression is a stationary distribution

A Markov Chain is regular ergodic if there exists a unique stationary distribution which is the limit point for all initial distributions and which puts positive mass on every element of the state-space $\mathcal{S}$.

- For an $n \times n$ transition matrix $\mathbf{P}$ then

1. $\lambda = 1$ is an eigenvalue of $\mathbf{P}$
2. all eigenvalues satisfy $|\lambda| \leq 1$

Transition Matrix Eigenvalues (Proof)

- For one to be an eigenvalue: $\det(\mathbf{P} - \mathbf{I}) = 0$

$$\mathbf{P} - \mathbf{I} = \begin{bmatrix} P_{1,1} - 1 & P_{1,2} & \cdots & P_{1,N} \\ P_{2,1} & P_{2,2} - 1 & \cdots & P_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N,1} & P_{N,2} & \cdots & P_{N,N} - 1 \end{bmatrix}$$

  – summing first $N - 1$ columns yields (negative) last column
  – not a full-rank matrix, determinant is zero, one is an eigenvalue

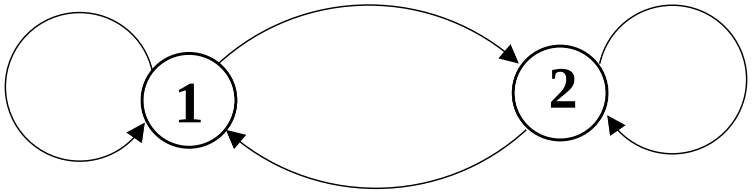
- Select an eigenvalue λ and (right) eigenvector $\boldsymbol{v}$ of $\mathbf{P}$

  – select largest (magnitude) element of eigenvector v_k $(|v_k| \geq |v_i|, \forall i)$
  – consider equality $\mathbf{P}\boldsymbol{v} = \lambda\boldsymbol{v}$ and element k

$$\begin{aligned} |\lambda||v_k| &= |\lambda v_k| = |P_{k,1}v_1 + \ldots + P_{k,N}v_N| \\ &\leq P_{k,1}|v_1| + \ldots + P_{k,N}|v_N| \\ &\leq P_{k,1}|v_k| + \ldots + P_{k,N}|v_k| = (P_{k,1} + \ldots + P_{k,N})|v_k| = |v_k| \end{aligned}$$

California Population

- Interested in the stationary distribution for California population



$$\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

  - in the limit

$$\mathbf{P}^{\infty} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}^{\infty} = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}$$

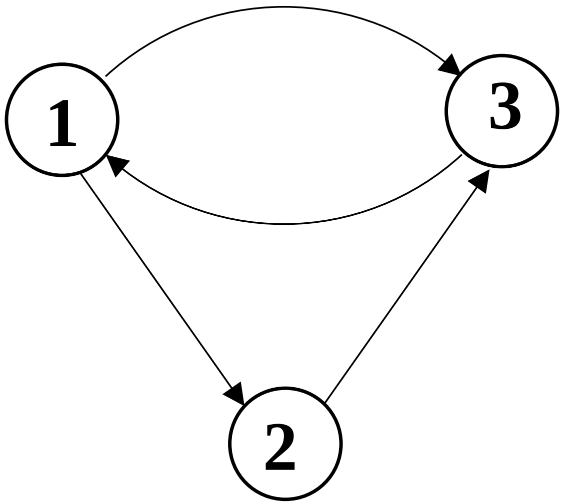
  - stationary (limiting) distribution

$$\mathbf{x}^{(\infty)} = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}$$

  - California has a third of the population of the US!
  - does not depend on the initial population distribution.

Random Surfer

- Consider an individual who surfs the web in the following fashion
  - d fraction of the time randomly selects a link on the current page
  - $(1-d)$ fraction of the time randomly selects a web-page
- This is a Markov Process - simple 3 web-page example



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

When used with random surfer

$$\tilde{\mathbf{P}} = \frac{(1-d)}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + d\mathbf{P}$$

What is the stationary distribution?

PageRank

- Pagerank ($\texttt{Pr}$) important component for original Google search
  - there is a total of N webpages
  - $\mathcal{L}_i$ set of pages that link to i, $\texttt{Ln}(j)$ number of links from j

$$\texttt{Pr}(i) = (1-d)/N + d\left(\sum_{j\in\mathcal{L}_i}\texttt{Pr}(j)/\texttt{Ln}(j)\right)$$

- Consider stationary distribution and element i for random surfer

$$x_i^{(\infty)} = (\mathbf{x}^{(\infty)}\mathbf{P})_i = (1-d)/N + d\left(\sum_{j\in\mathcal{L}_i} x_j^{(\infty)}P_{j,i}\right)$$

  - $P_{j,i} = 1/\texttt{Ln}(j)$ - probability of random move from j to i
  - $x_i^{(\infty)}$ is the PageRank of page i

Stationary Distributions and (Left) EigenVectors

- To find a stationary distribution consider

$$\mathbf{P}^{\mathrm{T}}\mathbf{x}^{(\infty)\mathrm{T}} = \mathbf{x}^{(\infty)\mathrm{T}}$$

  - for the transition matrix $|\lambda| \leq 1$,
  - need to find (right) eigenvector(s) of $\mathbf{P}^{\mathrm{T}}$ when $\lambda = 1$

- For California distribution

$$\mathbf{P}^{\mathrm{T}} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.894 & -0.707 \\ 0.447 & 0.707 \end{bmatrix} \begin{bmatrix} 1.0 & 0 \\ 0 & 0.7 \end{bmatrix} \begin{bmatrix} 0.894 & -0.707 \\ 0.447 & 0.707 \end{bmatrix}^{-1}$$

  - eigenvector for $\lambda = 1$ is $\begin{bmatrix} 0.894 & 0.447 \end{bmatrix}^{\mathrm{T}}$
  - need to satisfy sum to one constraint:

$$\mathbf{x}^{(\infty)} = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}$$

EigenValues/EigenVectors (Reminder)

- Possible to express $\mathbf{P}$ in terms of eigenvectors/eigenvalues (note not symmetric)
  - set-up as (standard) right eigenvectors

$$\mathbf{P}^{\mathsf{T}} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$$

  - $\mathbf{V}$ is the matrix of eigenvectors
  - $\mathbf{D}$ is the diagonal matrix of eigenvalues

- Interested in finding $\mathbf{P}^{\mathsf{T}\infty}$

$$\mathbf{P}^{\mathsf{T}\infty} = \mathbf{V}\mathbf{D}^{\infty}\mathbf{V}^{-1}$$

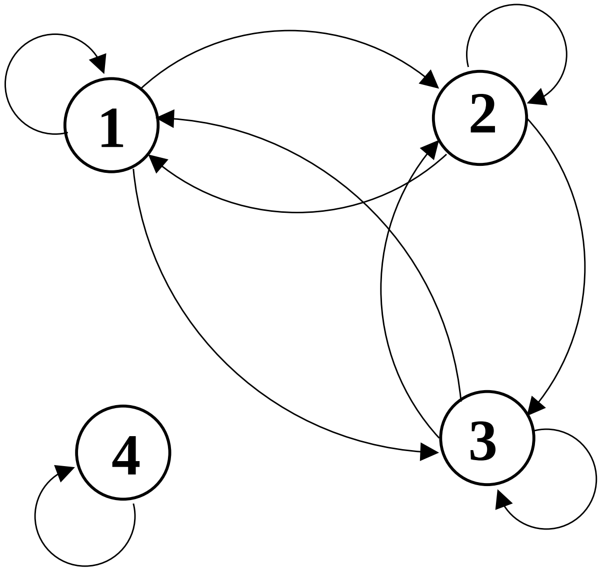
  - $\mathbf{D}^{\infty}$ will only be non-zero for $|\lambda| = 1$ (note $|\lambda| \leq 1$)

- Simple to find eigenvector for $\lambda = 1$ solving

$$\mathbf{P}^{\mathsf{T}}\mathbf{v} = \mathbf{v}$$

Modified Faulty Traffic Lights

- The traffic lights are now faulty with the following Markov Chain:



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

- The stationary distributions for this Markov Chain are:

$$\mathbf{x}^{(\infty)} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \end{bmatrix}, \quad \mathbf{x}^{(\infty)} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$$

Modified Faulty Traffic Lights (cont)

- Clearly the process is not regular ergodic:
  - two stationary distributions;
  - the stationary distribution depends on the initial distribution;
- Consider an initial distribution of the form

$$\mathbf{x}^{(0)} = \left[\begin{array}{cccc} a & 0.0 & 0.0 & 1-a \end{array}\right]$$

  - the final distribution of traffic lights is

$$\mathbf{x}^{(\infty)} = \left[\begin{array}{cccc} a/3 & a/3 & a/3 & 1-a \end{array}\right]$$

Communicate (Definition)

Two states of a Markov chain j and k communicate if, and only if, there exists integers m and n such that

$$P^m_{j,k} > 0 \text{ and } P^n_{k,j} > 0$$

- Intuitively this means:
  - there must be a route from state j to k (in m steps)
  - and there must be a route from state k to j (in n steps)

Recurrent Set (Definition)

Let $\mathcal{S}$ denote the set of all possible states of a Markov chain. A subset $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ is a recurrent set if

1. all pairs of states in $\tilde{\mathcal{S}}$ communicate
2. if $j \in \tilde{\mathcal{S}}$ and $k \notin \tilde{\mathcal{S}}$ then $P^i_{j,k} = 0$ for all $i \geq 0$

- Intuitively this means:
  - it is possible to move between all states in a recurrent set
  - cannot move from a member of a recurrent state to not a member

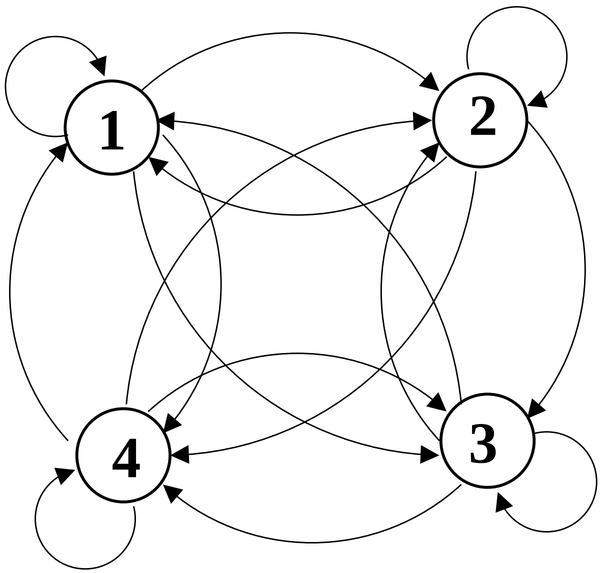
Recurrent and Transient / Irreducible (Definitions)

If a state belongs to a recurrent set then it is recurrent, otherwise it is transient

- Possible to show
  1. if state k is recurrent then $P(\text{return to state } k|X_o = k) = 1$
  2. if we start in a transient state then eventually we will leave that state and never return to it

If all states in a Markov chain communicate with each other then the Markov chain is irreducible

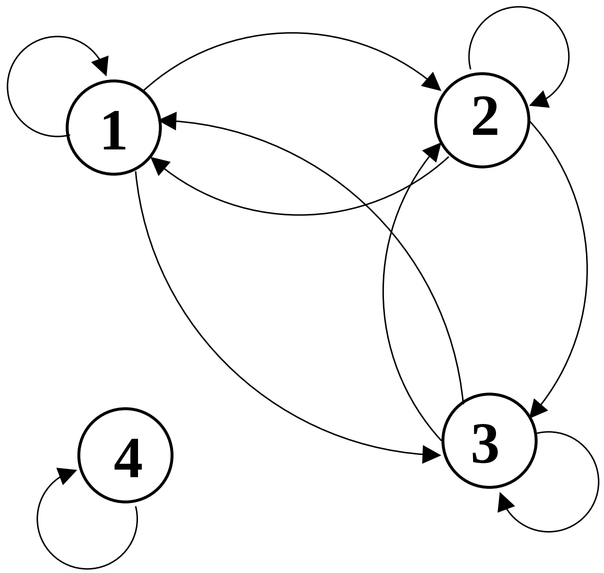
Faulty Traffic Lights



$$\begin{bmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/2 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/2 \end{bmatrix}$$

- This Markov chain is irreducible
  - all states communicate with each other

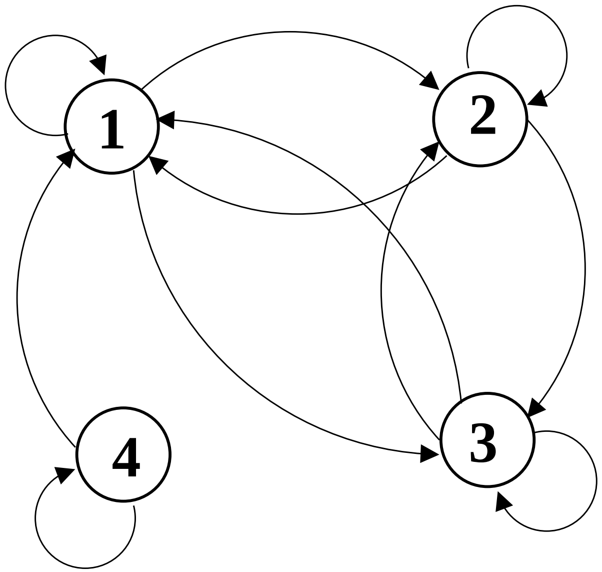
Modified Faulty Traffic Lights



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

- States 1, 2 and 3 form a recurrent set
- State 4 forms a recurrent set

Modified Faulty Traffic Lights (2)



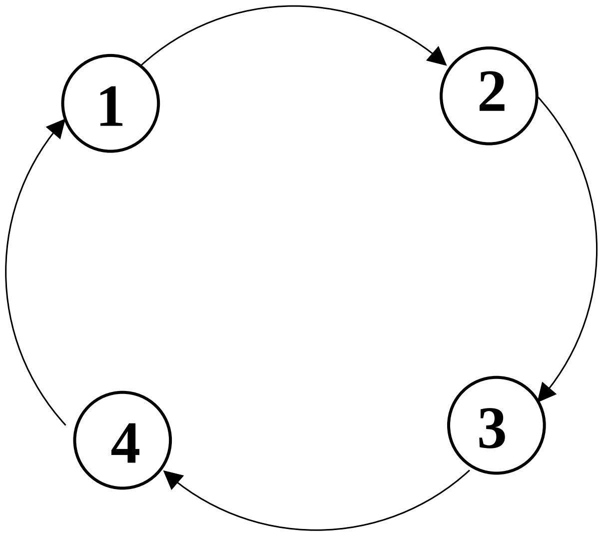
$$\begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}$$

- States 1, 2 and 3 form a recurrent set
- State 4 is a transient state
  - eventually state 4 is left and never returned to

Waiting Time

- I arrive when the traffic lights are at red: how long do I expect to wait

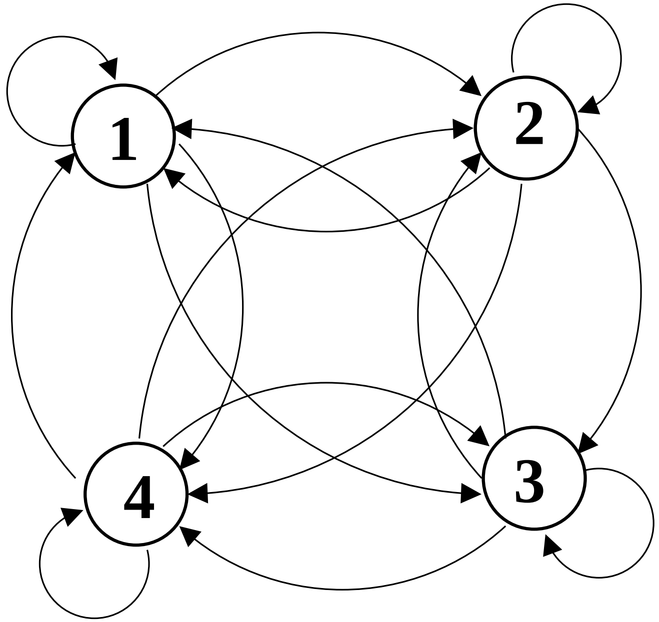
$$q_R = \mathcal{E}[\text{time to wait until first green}|X_0 = \texttt{red}]$$

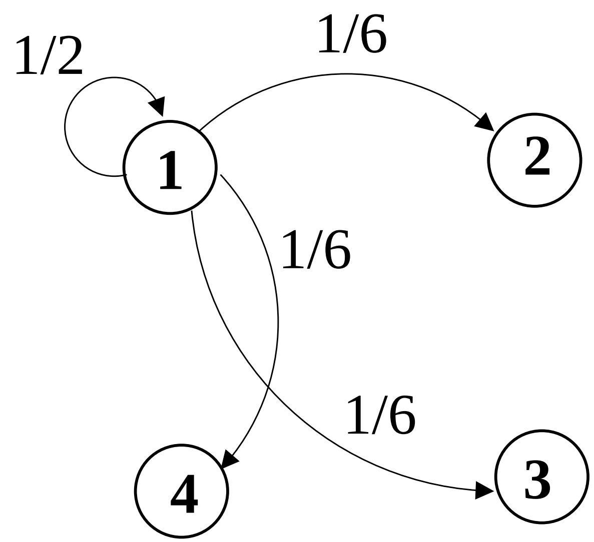


$$\begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

- For standard traffic lights, simple - 2!
  - state 1 is `red`, state 3 is `green`

Waiting Time - Faulty Traffic Lights





- Starting in `red` - possible transitions on the right
  - if light in `red`, then still expect to wait q_R
  - if light goes to `red/amber` then expect to wait q_{RA}
  - if light goes to `amber` then expect to wait q_A
  - if light goes to `green` then wait finished!!

Waiting Time - Faulty Traffic Lights

- Set-up equation to describe scenario

$$q_R = 1/2(1 + q_R) + 1/6(1 + q_{RA}) + 1/6(1 + q_A) + 1/6$$

- Now need expressions for q_{RA} and q_A

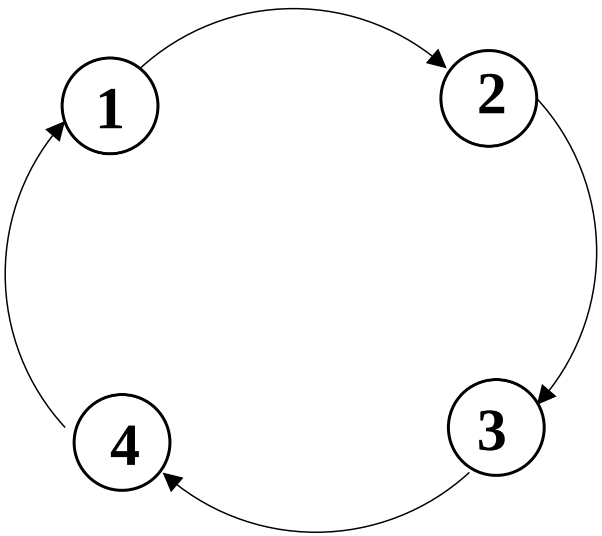
$$\begin{aligned} q_{RA} &= 1/2(1 + q_{RA}) + 1/6(1 + q_R) + 1/6(1 + q_A) + 1/6 \\ q_A &= 1/2(1 + q_A) + 1/6(1 + q_R) + 1/6(1 + q_{RA}) + 1/6 \end{aligned}$$

- Series of linear equations - simply solve to get:

$$q_R = 6, \quad q_A = 6, \quad , q_{RA} = 6, \quad q_G = 0$$

Periodicity

- Consider the standard, working, traffic lights



$$\begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

- If the traffic lights start at `red` then the process is

$$\texttt{red} \rightarrow \texttt{red/amber} \rightarrow \texttt{green} \rightarrow \texttt{amber} \rightarrow \texttt{red}$$
$$\text{state } 1 \rightarrow \text{state } 2 \rightarrow \text{state } 3 \rightarrow \text{state } 4 \rightarrow \text{state } 1$$

  – this has a period of 4

Periodicity (Definition)

The period of a state k of a Markov chain is the greatest common divisor of the set

$$\{i \geq 0 : P^i_{k,k} > 0\}$$

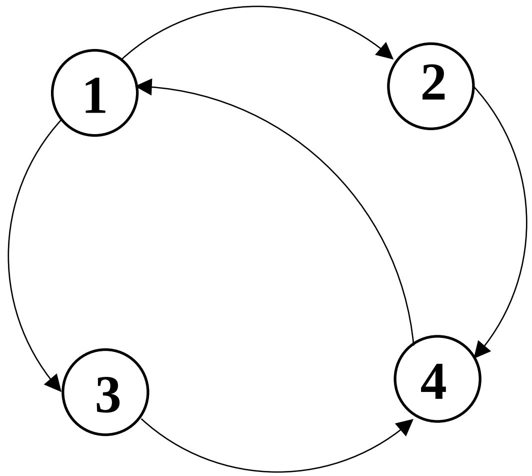
A state k is aperiodic if it has period 1.

- A Markov chain is aperiodic if all states are aperiodic
  - a state k is aperiodic if for some time i

$$P^i_{k,k} > 0 \text{ and } P^{i+1}_{k,k} > 0$$

  - if a chain is irreducible and aperiodic then it is regular ergodic
    it will have a limiting distribution

Periodicity - Example 1 from UW



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

- Possible transitions are

$$1 \to 2 \to 4 \to 1$$
$$1 \to 3 \to 4 \to 1$$

- Looking at the eigenvalues of $\mathbf{P}$

$$\lambda^4 - \lambda = 0; \quad \lambda = 0, 1, e^{2\pi i/3}, e^{4\pi i/3}$$

  – three eigenvalues satisfy $|\lambda| = 1$

Periodicity - Example 1 from UW

- The transition matrices for 2,3 and 4 steps are:

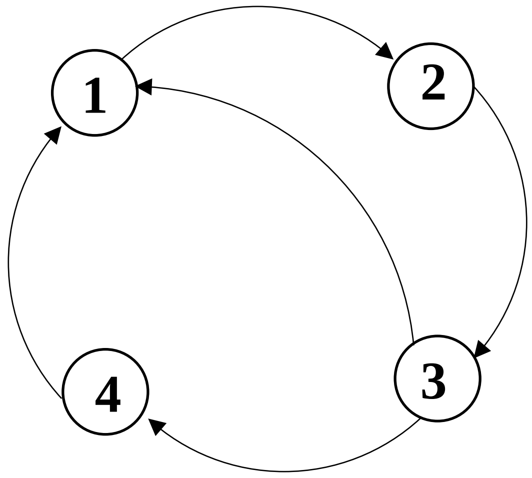
$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}, \mathbf{P}^3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{P}^4 = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

- The process is not regular ergodic
  - it does not settle down to a unique stationary distribution

$$\mathbf{x}^{(0)}\mathbf{P} = \mathbf{x}^{(0)}\mathbf{P}^4 = \mathbf{x}^{(0)}\mathbf{P}^7 = \ldots$$

  - states are periodic with period 3

Periodicity - Example 2 from UW



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

- Possible transitions are

$$1 \rightarrow 2 \rightarrow 3 \rightarrow 1$$
$$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$$

- Possible to return to state 1 after 3/4/6/7/8 ...
  – the period of state 1 is 1 - it is aperiodic

Periodicity - Example 2 from UW

- Looking at the eigenvalues of $\mathbf{P}$

$$\lambda^4 - \lambda/2 - 1/2 = 0; \quad \lambda = 1, -0.176 \pm 0.861i, -0.648$$

  – one eigenvalues satisfy $|\lambda| = 1$

- The transition matrices for 2,3,4,5,6 and 7 steps are:

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{P}^3 = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{P}^4 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/4 & 0 & 1/2 & 1/4 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}$$

$$\mathbf{P}^5 = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/4 & 0 & 1/2 & 1/4 \\ 1/2 & 1/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}, \mathbf{P}^6 = \begin{bmatrix} 1/4 & 0 & 1/2 & 1/4 \\ 1/2 & 1/4 & 0 & 1/4 \\ 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \mathbf{P}^7 = \begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/8 & 1/4 & 1/2 & 1/8 \\ 1/4 & 0 & 1/2 & 1/4 \end{bmatrix}$$

Periodicity - Example 2 from UW

- For this system looking at the leading diagonals $(P^i_{k,k} > 0)$
  - states 1,2,3 non-zero steps: 3,4
  - state 4 non-zero steps: 4,7

- Lowest common divisor is 1 for all states:
  - system is aperiodic
  - system also irreducible (all states communicate)
  - system is therefore regular ergodic

- Unique stationary (limiting) distribution

$$\mathbf{P}^{\infty} = \begin{bmatrix} 2/7 & 2/7 & 2/7 & 1/7 \\ 2/7 & 2/7 & 2/7 & 1/7 \\ 2/7 & 2/7 & 2/7 & 1/7 \\ 2/7 & 2/7 & 2/7 & 1/7 \end{bmatrix}$$

Periodicity and Eigenvalues (Reference)

- Consider a Markov Chain with transition matrix $\mathbf{P}$ that has a period 3
  - there will be three eigenvalues of $\mathbf{P}$ with magnitude 1
  - the values of these eigenvalues will be: $1, e^{2\pi i/3}, e^{4\pi i/3}$

  Example 1 from UW is a Markov Chain of this form:

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}; \quad \text{eigenvalues: } 1, e^{2\pi i/3}, e^{4\pi i/3}, 0$$

- More generally for a Markov Chain with period δ
  - there will be δ eigenvalues of $\mathbf{P}$ with magnitude 1
  - the values of these eigenvaues will be: $1, e^{2\pi i/\delta}, e^{4\pi i/\delta}, \ldots$

Ergodic Theorem (Reference) / Detailed Balance

If $X_o, X_1, \ldots$ is a regular ergodic Markov chain with stationary distribution $\boldsymbol{\pi}$, then for any function $f(x)$ as $N \to \infty$

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \to \sum_{k\in\mathcal{S}} \pi_k f(k)$$

where $\mathcal{S}$ is the state-space of the Markov chain.

A transition matrix $\mathbf{P}$ and a distribution $\boldsymbol{\pi}$ are in detailed balance if

$$\pi_j P_{j,k} = \pi_k P_{k,j}$$

where $\mathcal{S}$ is the state-space of the Markov chain.

Detailed Balance & Stationary Distribution

A distribution $\boldsymbol{\pi}$ satisfies detailed balance with transition matrix $\mathbf{P}$

$$\pi_j P_{j,k} = \pi_k P_{k,j}$$

What happens to this distribution after 1-step?

$$\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}\mathbf{P}$$

Now consider element k of the distribution

$$\tilde{\pi}_k = \sum_j \pi_j P_{j,k} = \sum_j \pi_k P_{k,j} = \pi_k \left(\sum_j P_{k,j} \right) = \pi_k$$

- $\boldsymbol{\pi}$ is a stationary distribution of $\mathbf{P}$

Summary

- Finite-state (discrete) Markov chains
  - current state completely determines probability of next state
- Range of attributes to characterise Markov Chains
  - limiting and stationary distributions
  - wait-times
  - periodicity and nature of states (communicate, transient etc)
- Next lecture examines continuous state-space systems