Markov Chains

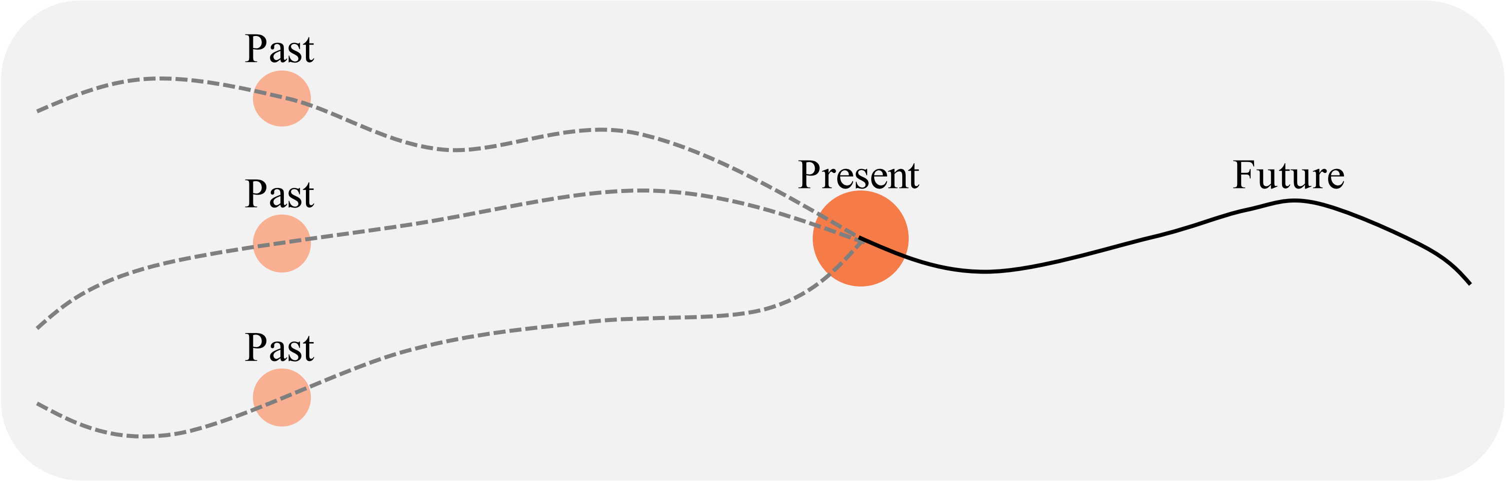
Outline

- **Stochastic Processes and Markov Property**
- **Markov Chains**
- **Chapman-Kolmogorov Equations**
- **Classification of States**
- **Invariant Measures, Time Averages, Limiting Probabilities**

Stochastic Processes and Markov Property

- Stochastic Process
  - Discrete-time: $\{X_n : n \geq 0\}$, integer number n indexed random variables
  - Continuous-time: $\{X(t) : t \geq 0\}$, real number t indexed random variables
  - Discrete state-space if each X_n or $X(t)$ has a countable range
  - Continuous state-space if each X_n or $X(t)$ has an uncountable range
  - Ex: Markov chains have discrete-time and discrete state-space

- Markov Property: Future conditioned on the present is in independent of the past



Markov Chains

- Markov Chain: Discrete time, discrete state space Markovian stochastic process.
  - Often described by its transition matrix P

- Ex: Moods {**C**ooperative, **J**udgmental, **O**ppositional} of a person as Markov chain

- Ex: A random walk process has state space of integers $\ldots, -2, -1, 0, 1, 2, \ldots$. For a fixed probability $0 \leq \mathrm{p} \leq 1$, the process either moves forward or backward:
  - $\mathrm{P}(X_{n+1} = i + 1 | X_n = i) = 1 - \mathrm{P}(X_{n+1} = i - 1 | X_n = i)$
  - The transition matrix has infinite dimensions and is sparse

<table>
<tr><th></th><th>…</th><th>-2</th><th>-1</th><th>0</th><th>1</th><th>2</th><th>…</th></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>-2</td><td>…</td><td>0</td><td>p</td><td colspan="4">0</td></tr>
<tr><td>-1</td><td>0</td><td>$1-p$</td><td>0</td><td>p</td><td colspan="3">0</td></tr>
<tr><td>0</td><td colspan="2">0</td><td>$1-p$</td><td>0</td><td>p</td><td colspan="2">0</td></tr>
<tr><td>1</td><td colspan="3">0</td><td>$1-p$</td><td>0</td><td>p</td><td>0</td></tr>
<tr><td>2</td><td colspan="4">0</td><td>$1-p$</td><td>0</td><td>…</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
</table>

Chapman-Kolmogorov Equations

- Probability of going from state x to state y in n steps

$$p_{x,y}^{<n>} = \mathrm{P}(X_{k+n} = y | X_k = x)$$

- To go from x to y in $n + m$ steps, go through state z in the nth step

$$p_{x,y}^{<n+m>} = \sum_{z \in \mathcal{X}} p_{x,z}^{<n>} p_{z,y}^{<m>}$$

- Using transition matrices

$$P^{n+m} = P^n P^m$$

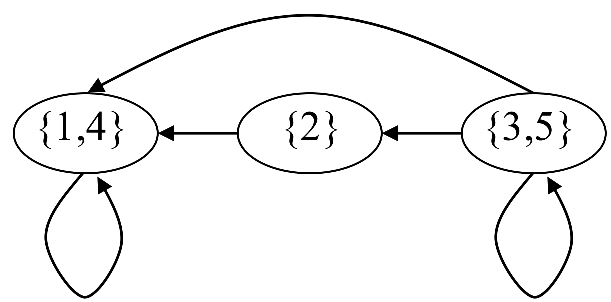
Classification of States: Communication

- State y is accessible from state x if $p_{x,y}^{<n>} > 0$ for some n.
- Contrapositive: If state y is not accessible from x, then $p_{x,y}^{<n>} = 0$ for all n.

  P(Reaching y ever | Starting in x)=$\sum_{n=0}^{\infty} p_{x,y}^{<n>}$=0
- States (x, y) communicate if y is accessible from x and x is accessible from y
- Ex: Communication is a relation on $(\mathcal{X} \times \mathcal{X})$. This relation is reflexive, symmetric and transitive. Hence, it is an equivalence relation.
- The communication relation splits $\mathcal{X}$ into equivalence classes: Each class includes the set of states that communicate with each other.
- Ex: The transition matrix below on the left creates classes {1,4}, {2}, {3,5}. We can define an aggregate state Markov chain whose states are these classes as below in the middle. The new chain is likely to end up in {1,4} below on the right.

	1	2	3	4	5
1				+	
2	+			+	
3	+	+			+
4	+				
5		+	+	+	

	1,4	2	3,5
1,4	+		
2	+		
3,5	+	+	+



Classification of States: Periodicity

- Ex: The transition matrix below on the left creates classes {1,2,4} and {3,5}. These classes are not accessible from each other, so the chain decomposes into two chains, with transition matrices on the right.

	1	2	3	4	5
1		+		+	
2	+				
3					+
4		+			
5			+		

	1	2	4
1		+	+
2	+		
4		+	

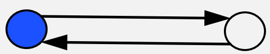
	3	5
3		+
5	+	

- An irreducible Markov chain has only one class of states. A reducible Markov chains as two examples above illustrate either eventually moves into a class or can be decomposed. In view of these, limiting probability of a state in an irreducible chain is considered. Irreducibility does not guarantee the presence of limiting probabilities.
- Ex: A Markov chain with two states $\boldsymbol{\mathcal{X}} = \{x, y\}$ such that $p_{x,y} = p_{y,x} = 1$. Starting in state x, we can ask for $p_{x,x}^{<n>}$. This probability has a simple but periodic structure: It is 1 when n is even; 0 otherwise. The limit of $p_{x,x}^{<n>}$ does not exist as n approached infinity.
- To talk about limiting probabilities, we need to rule out periodicity. Period $d(x)$ of state x is the greatest common divisor (gcd) of all the integers in $\{n \geq 1 \colon p_{x,x}^{<n>} > 0\}$.

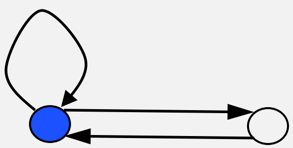
$$d(x) = gcd\{n \geq 1 \colon p_{x,x}^{<n>} > 0\}.$$

Markov Chain Examples with Different Periods

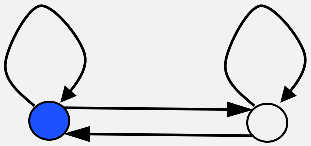
2 States



Period 2 = gcd{2,4, ... }



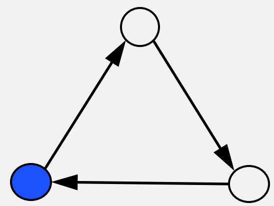
Period 1 = gcd{1,2, ... }



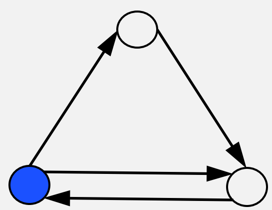
Period 1 = gcd{1,2, ... }

All possible transitions with 2 communicating states ⇒The same period

3 States

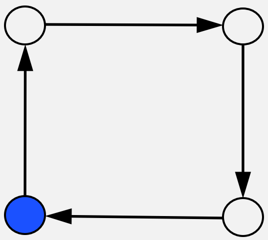


Period 3 = gcd{3,6, ... }

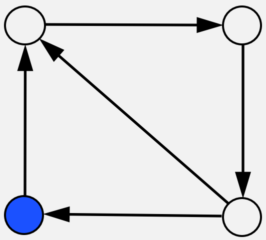


Period 1 = gcd{2,3, ... }

4 States

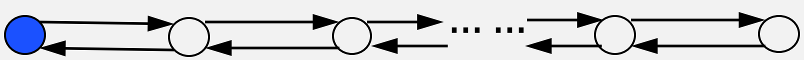


Period 4 = gcd{4,8, ... }



Period 1 = gcd{4,7,.. }

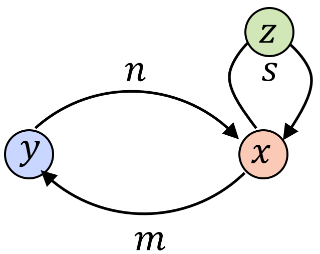
Many States



Period 2 = gcd{2,4,6 ... }

Period is a Class Property

- Period of any two states in the same class are the same.
  - For classes with two states only, see the last page
  - Consider classes with at least three states
  - Consider x, y such that $p_{x,y}^{<m>} > 0$ and $p_{y,x}^{<n>} > 0$ for some m and n.
    - Such m, n exist because x, y are in the same class
    - » Period of state x, $d(x) = \gcd\{s \geq 1 : p_{x,x}^{<s>} > 0\}$
    - » By definition of m, n and for any s with $p_{x,x}^{<s>} > 0$.
      - $p_{y,y}^{<n+m>} \geq p_{y,x}^{<n>} p_{x,y}^{<m>} > 0$ and $p_{y,y}^{<n+s+m>} \geq p_{y,x}^{<n>} p_{x,x}^{<s>} p_{x,y}^{<m>} > 0$
      - Such $s \geq 1$ exists because x communicates with another (third) state z in its class
    - » $d(y)$ divides both $n + m$ and $n + s + m$
    - » $d(y)$ divides every s with $p_{x,x}^{<s>} > 0$
      - $d(y)$ divides gcd of such s
    - » Hence, $d(y)$ divides $d(x)$.
  - Repeat by changing the roles
    - » $x \leftrightarrow y \Rightarrow d(y)$ divides $d(x)$.
  - Periods $d(x)$ and $d(y)$ divide each other ⇒ they must be equal.



Classification of States: Recurrence

- A state is called recurrent if the chain returns to the state in finite steps with probability 1.
  - The first time state visits state y after starting at state x is a random variable $\tau_{x,y}$:
$$\tau_{x,y} = \min\{n \geq 1: X_n = y \text{ and } X_0 = x\}$$
  - This variable is also called the hitting time
  - Recurrent state x iff $\mathrm{P}(\tau_{x,x} < \infty) = 1$; Otherwise, transient state.

- A recurrent state has only finite value of hitting time.
- A positive recurrent state has $\mathrm{E}(\tau_{x,x}) < \infty$. Positive recurrence $\Rightarrow$ recurrence.
  - Ex: Heavy tail hitting time distributions, e.g., Pareto, can have infinite expected value.

- Ex: Starting with $X_0 = x$, let N_x be the number times the chain is in x:
$$N_x = 1_{X_0=x} + 1_{X_1=x} + 1_{X_2=x} + \cdots$$
  - We have
$$\mathrm{E}(N_x|X_0 = x) = \mathrm{E}\left(\sum_{n=0}^{\infty} 1_{X_n=x}|X_0 = x\right) = \sum_{n=0}^{\infty} \mathrm{E}(1_{X_n=x}|X_0 = x) = \sum_{n=0}^{\infty} p_{x,x}^{<n>}$$

The last term is more operational as it is based on transition probabilities

Recurrence Related Derivations

- The expected value, of the number of times the chain is in x, $\mathrm{E}(N_x|X_0=x)=\sum_{n=0}^{\infty}p_{x,x}^{<n>}$ can also be written as

$$\mathrm{E}(N_x|X_0=x)=\frac{1}{1-\mathrm{P}(\tau_{x,x}<\infty)}$$

  - Note that to be in state x at time $n\geq 1$, the chain must come to state x for the first time in time k for $k=1\ldots n$. This probabilistic reasoning yields

$$p_{x,x}^{<n>}=\sum_{k=1}^{n}\mathrm{P}(\tau_{x,x}=k)p_{x,x}^{<n-k>}$$

  - On the other hand,

$$\sum_{n=0}^{\infty}p_{x,x}^{<n>}-1=\sum_{n=1}^{\infty}p_{x,x}^{<n>}=\sum_{n=1}^{\infty}\sum_{k=1}^{n}\mathrm{P}(\tau_{x,x}=k)p_{x,x}^{<n-k>}$$

$$=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\mathrm{P}(\tau_{x,x}=k)p_{x,x}^{<n-k>}=\sum_{k=0}^{\infty}\mathrm{P}(\tau_{x,x}=k)\sum_{n=k}^{\infty}p_{x,x}^{<n-k>}$$

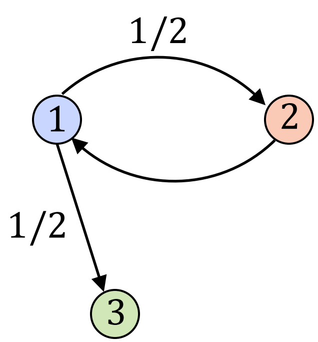
$$=\sum_{k=0}^{\infty}\mathrm{P}(\tau_{x,x}=k)\sum_{n=0}^{\infty}p_{x,x}^{<n>}=\mathrm{P}(\tau_{x,x}<\infty)\sum_{n=0}^{\infty}p_{x,x}^{<n>}$$

  - Hence, $\mathrm{E}(N_x|X_0=x)=\sum_{n=0}^{\infty}p_{x,x}^{<n>}=\frac{1}{1-\mathrm{P}(\tau_{x,x}<\infty)}$.

- If $\mathrm{P}(\tau_{x,x}<\infty)=1$, the state x is recurrent and $\mathrm{E}(N_x|X_0=x)=\sum_{n=0}^{\infty}p_{x,x}^{<n>}=\infty$.
- If $\mathrm{P}(\tau_{x,x}<\infty)<1$, the state x is transient and $\mathrm{E}(N_x|X_0=x)=\sum_{n=0}^{\infty}p_{x,x}^{<n>}<\infty$.

Infinite Hitting Time

- $\mathrm{P}(\tau_{x,x} < \infty) < 1 \Leftrightarrow \mathrm{P}(\tau_{x,x} = \infty) > 0$



Example:

- $\mathrm{P}(\tau_{1,1} = \infty) = \frac{1}{2}$ and $\mathrm{P}(\tau_{1,1} = 2) = \frac{1}{2}$
- N_1: Number of times to visit state 1
  - $N_1 = 1$ wp $\frac{1}{2}$, $N_1 = 2$ wp $\left(\frac{1}{2}\right)^2$
  - $N_1 = k$ wp $\left(\frac{1}{2}\right)^k$
- $\mathrm{E}(N_1) = 2 = \frac{1}{1-\frac{1}{2}} = \frac{1}{1-\mathrm{P}(\tau_{1,1}<\infty)}$
- $\sum_{k=0}^{\infty} \mathrm{P}(\tau_{1,1} = k)$?
  - $\lim_{n\to\infty} \sum_{k=0}^{n} \mathrm{P}(\tau_{1,1} = k) = 0 + \frac{1}{2} + 0 + 0 + \cdots = \frac{1}{2}$
  - $\mathrm{P}(\tau_{1,1} = \infty) + \lim_{n\to\infty} \sum_{k=0}^{n} \mathrm{P}(\tau_{1,1} = k) = \frac{1}{2} + \frac{1}{2} = 1$

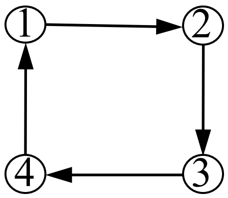
Invariant Measures

- Invariant measure ρ, possibly infinite dimensional, column vector with $\rho \geq 0$ satisfying

$$\rho^T = \rho^T P$$

  - Viewing transition matrix P as an operator, the invariant measure is the fixed point of the operator; successive applications of the operator does not move the invariant measure.
  - Invariant measure is not unique: ρ invariant $\Rightarrow 2\rho$ invariant
  - Towards uniqueness, normalize the invariant measure:
  - $\pi = \frac{\rho}{\rho^T \mathbf{1}}$ for $\rho^T \mathbf{1} < \infty$, where $\mathbf{1}$ is a column vector of ones.
  - Invariant probability measure π satisfies
    - » Invariance: $\pi^T = \pi^T P$
    - » Normalization: $\pi^T \mathbf{1} = 1$
    - » Nonnegativity: $\pi \geq 0$
- Ex: Consider a 4-state Markov Chain with

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

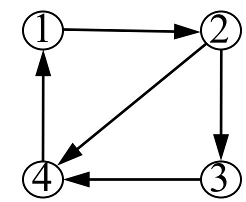


  - This chain has invariant measures $\left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]$, $[1, 1, 1, 1]$, $[2, 2, 2, 2]$ or $[a, a, a, a]$ for $a \geq 0$
  - Among these, the only invariant probability is $\left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]$

Invariant Measure and Time Averages

- Ex: Consider a 4-state Markov Chain with

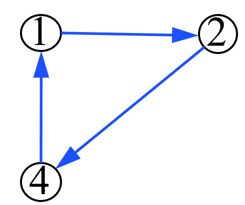
$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

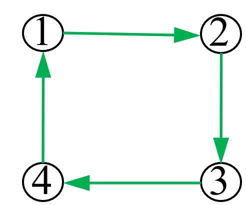


  – This chain has invariant measures $\left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right]$, [2, 2, 1, 2], [4, 4, 2, 4] or $[2a, 2a, a, 2a]$ for $a \geq 0$
  – Among these, the only invariant probability is $\left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right]$ as

$$\left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right]$$

  – Consider two cycles, triangle and square, defined as
  – Think of the Markov Chain as $\frac{1}{2}$triangle+ $\frac{1}{2}$square.
  – In the triangle, the chain takes 3 steps to come back.
  – In the square, it takes 4 steps.





  – In 7 steps, the chain returns to state 1 by visiting {1,2,4} twice and {3} once on average
  – $\mathrm{E}(\tau_{1,1}) = 3.5 = \frac{1}{2}3 + \frac{1}{2}4$ and $\mathrm{E}\left(\sum_{n=0}^{\tau_{1,1}} 1_{X_n=1} | X_0 = 1\right) = \mathrm{E}\left(\sum_{n=0}^{\tau_{1,1}} 1_{X_n=2} | X_0 = 1\right) = \mathrm{E}\left(\sum_{n=0}^{\tau_{1,1}} 1_{X_n=4} | X_0 = 1\right) = 1$, whereas $\mathrm{E}\left(\sum_{n=0}^{\tau_{1,1}} 1_{X_3=1} | X_0 = 1\right) = 0.5$.
  – An invariant measure turns out to be the expected number of visits to a particular state: $\left[1, 1, \frac{1}{2}, 1\right]$
  – The invariant probability is $\left[\frac{1}{3.5}, \frac{1}{3.5}, \frac{0.5}{3.5}, \frac{1}{3.5}\right] = \left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right]$

Invariant Measure, Time Average & Limiting Probability

- In the previous example, time averages are 1/3.5, 1/3.5, 1/7, 1/3.5 represent the percentage of time the chain stays in states 1, 2, 3, 4.
- In general, time average random variable is not over single cycle but over N steps for $N \to \infty$:

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N} 1_{X_n=x}}{N}$$

- Consistency Result: An irreducible and positive recurrent Markov chain X_n has
  - The unique invariant probability π, and
  - Time average converges to this invariant probability almost surely $\frac{\sum_{n=0}^{N} 1_{X_n=x}}{N} [\to as]\, \pi_x$
- The consistency result implies that we do not have to separately search for invariance probability and time averages; it suffices to find one of these. But the result is not operational.

- Towards an operational method, let us introduce limiting probability

$$\pi_y = \lim_{n\to\infty} p_{x,y}^{<n>}$$

- Note the limiting probability is independent of the initial state x; possible only in an aperiodic chain
- Crude methodology: Keep multiplying the transition matrix by itself to obtain P^n until its rows converge to each other so that any one of the rows can be taken as the limiting probability.
- Issues with the crude methodology:
  - No assurance of convergence
  - No relation between limiting probability, time average and invariant measure

Main Result
Invariant Measure=Time Average=Limiting Probability

Main Result: For an irreducible Markov chain with a period of 1, if an invariant probability measure π exists, i.e., a solution to $\pi^T = \pi^T P$, $\pi^T \mathbf{1} = 1$, $\pi \geq 0$ then

- the Markov chain is positive recurrent,
- π is unique,
- π is also the limiting probability,
- for each state x, $\pi_x > 0$.

- Since irreducible & positive recurrent chains have time average [$\rightarrow$ *as*] invariant measure, π computed above is also the time average

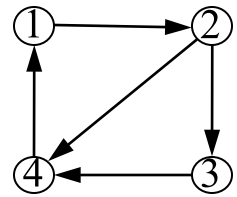
- All we have to check is 1) irreducible, 2) aperiodic 3) solution to $\pi^T = \pi^T P$, $\pi^T \mathbf{1} = 1$, $\pi \geq 0$.
- The solution to $\pi^T = \pi^T P$, $\pi^T \mathbf{1} = 1$, $\pi \geq 0$ is $\mathbf{1}^T(I - P + \mathbb{1})^{-1}$, where I is the identity matrix and $\mathbb{1}$ is the matrix of ones, both of these matrices have the same size as the transition matrix P.
  - To obtain this, $\pi^T = \pi^T P$ implies $\pi^T(I - P) = \mathbf{0}$.
  - Hence, $\pi^T(I - P + \mathbb{1}) = \mathbf{0}^T + \pi^T \mathbf{1} = 1^T$, where $\mathbf{0}$ is the column vector of only 0s.
  - When the Markov chain is irreducible $(I - P + \mathbb{1})$ can be shown to have the inverse $(I - P + \mathbb{1})^{-1}$, so

$$\pi^T = \mathbf{1}^T(I - P + \mathbb{1})^{-1}$$

Limiting Probability Example

- Ex: Consider a 4-state Markov Chain with

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



  - The chain is irreducible and aperiodic, main result applies
  - $I - P + \mathbb{1} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 2 & 1/2 & 1/2 \\ 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 2 \end{bmatrix}$, in R "IP1=rbind(c(2,0,1,1),c(1,2,1/2,1/2),c(1,1,2,0),c(0,1,1,2))".
  - $(I - P + \mathbb{1})^{-1} = \dfrac{\begin{bmatrix} 6.5 & 3 & -2 & -4 \\ -3.5 & 7 & 0 & 0 \\ -1.5 & -5 & 8 & 2 \\ 2.5 & -1 & -4 & 6 \end{bmatrix}}{14}$, in R "solve(IP1)"
  - $\mathbf{1}^T (I - P + \mathbb{1})^{-1} = \left[\frac{4}{14}, \frac{4}{14}, \frac{2}{14}, \frac{4}{14}\right] = \left[\frac{2}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}\right]$, in R "c(1,1,1,1) %*% solve(IP1)"
  - On the other hand, P^n rows convergence to $\left[\frac{4}{14}, \frac{4}{14}, \frac{2}{14}, \frac{4}{14}\right]$:

$$P^{15} = \frac{\begin{bmatrix} 3.9375 & 5.2500 & 1.7500 & 3.0625 \\ 3.0625 & 3.9375 & 2.6250 & 4.3750 \\ 3.5000 & 2.6250 & 2.6250 & 5.2500 \\ 5.2500 & 3.5000 & 1.3125 & 3.9375 \end{bmatrix}}{14}, \quad P^{30} = \frac{\begin{bmatrix} 3.84 & 4.05 & 2.10 & 4.01 \\ 4.02 & 3.84 & 2.02 & 4.12 \\ 4.18 & 3.86 & 1.91 & 4.05 \\ 4.05 & 4.18 & 1.93 & 3.84 \end{bmatrix}}{14} \text{ and } P^{60} = \frac{\begin{bmatrix} 4.00 & 4.00 & 2.00 & 4.00 \\ 4.00 & 4.00 & 2.00 & 4.00 \\ 4.00 & 4.00 & 2.00 & 4.00 \\ 4.00 & 4.00 & 2.00 & 4.00 \end{bmatrix}}{14}$$

Summary

- Stochastic Processes and Markov Property
- Markov Chains
- Chapman-Kolmogorov Equations
- Classification of States
- Invariant Measures, Time Averages, Limiting Probabilities