Markov Chains:

Theory and (Health) Applications

PM522a: Introduction to the Theory of Statistics

November 26, 2019

Abigail Horn

Postdoctoral Fellow, Preventive Medicine, USC

Abigail.Horn@usc.edu

www.AbigailHorn.com

References and Acknowledgements

- Kemeny JG, Snell JL. (1976) *Finite Markov chains*. Princeton, NJ: van Nostrand.
- Tsitsiklis, J., Bertsekas, D. Introduction to Probability (2nd Edition). Belmont, MA: Athena Scientific. (Course notes for MIT 6.041/6.431).
- Alexandre Jacquillat, Course notes from: 94.867: Decision Analytics for Business and Policy, Heinz College of Business, Carnegie Mellon University
- Zelda Zabinsky, Course notes from: IND E 411: Stochastic Models and Decision Analysis, Industrial Engineering Department, University of Washington

Outline

- Motivation
- Markov chains lecture
  - Definition and Properties
  - Classification of States and Chains
  - Ergodic chains
    - Steady-state Dynamics and Distributions
    - First passage time distributions and expected values
  - Absorbing chains
    - Absorbing probabilities:
    - Absorbing probabilities: Matrix equations
    - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

Stochastic Processes: Introduction

- Sometimes we need to consider uncertainty about a sequence of future events:
  - Uncertain weather in Napa valley every week over the grape season
  - Uncertain daily evolution of stock prices
  - Uncertain disease progression status
- We need probability models for systems that evolve over time or events in a probabilistic manner – stochastic processes

Stochastic Processes

- Suppose we take a series of observations of a random variable, X_0, X_1, X_2,...
- A stochastic process is an indexed collection of random variables $\{X_t\}$, where t is the index from a given set T. (The index t often denotes time.)

  Examples:

  - Roll a die 10 times, X_i = number on die on ith roll, i=1,2,...,10. Note that X_i takes integer values from 1 to 6. The stochastic process $\{X_t\} = \{X_1, X_2, ...\}$ denotes the sequence of rolls.
  - Sales of an item, X_t = number of items sold on day t, t=1,2,...Then the stochastic process $\{X_t\} = \{X_0, X_1, X_2, ...\}$ provides a mathematical representation of how the sales evolve starting today

Markov Chains: Introduction

- Stochastic process: A *sequence* of random variables
- Markov chains are special stochastic processes having:
    - A discrete sample space,
    - discrete time increments,
    - and a "memoryless" property, indicating that how the process will evolve in the future depends only on the present state of the process
- Markov *processes* are stochastic processes having:
    - A discrete sample space,
    - *continuous* time epochs
    - and the "memoryless" property

Today we will focus on Markov chains

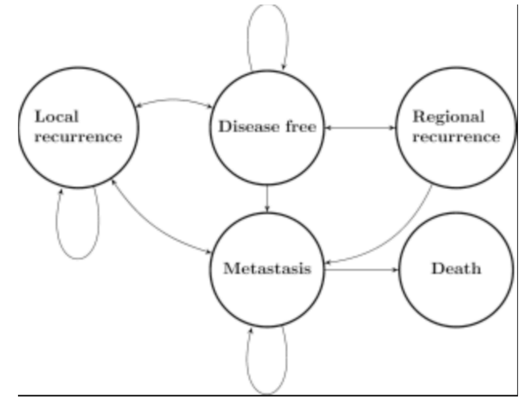
Why study Markov chains?

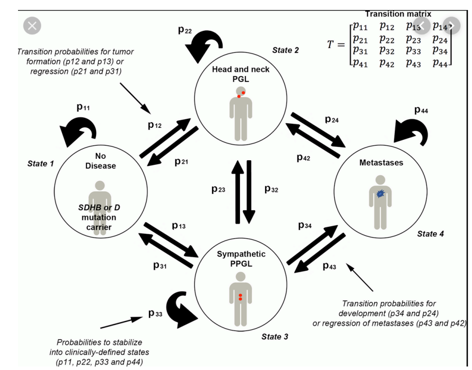
Relevance to many health applications, e.g.:

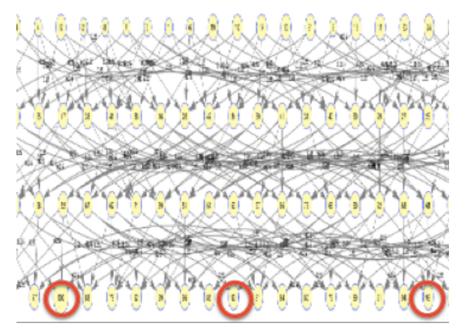
- Medical decision making
- Cost effectiveness studies
- Disease modeling
- Infectious disease spreading

Markov chains underlie very important methods, e.g.:

- Markov chain Monte Carlo methods
- Hidden Markov Models (time series etc.)







Markov Chains: Sample Questions to Ask

- What is the probability that the chain is in state i after n steps?
- What percent of the time is the chain in state i?
- What is the expected time it will take for the process to reach state i starting from state j?
- What is the expected time until the process reaches state i for the first time? or returns to state i?
- What is the probability that the chain will terminate at state i ?

Outline

- Motivation
- Markov chains lecture
  - Definition and Properties
  - Classification of States and Chains
  - Ergodic chains
    - Steady-state Dynamics and Distributions
    - First passage time distributions and expected values
  - Absorbing chains
    - Absorbing probabilities:
    - Absorbing probabilities: Matrix equations
    - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

States and State Space

- **State**: Description of the current "situation" of the system
  (Can be qualitative or quantitative)
- States are discrete $\boxed{\{X_n, n = 0, 1, ...\}}$
- The **state space** can be finite or infinite
- The state characterization must be mutually exclusive (no intersection) and exhaustive (include all possibilities)
- Examples:
  - Disease status
    - Susceptible, Infected, Recovered
  - Each week, the condition of a machine is determined by measuring the amount of electrical current it uses
    - High, Medium, Low, Failed
  - Weather monitoring
    - State: temperature (0 deg, 1 deg, …, 100 deg)

Markov Chain: Definition

- **Markov Chain**: A stochastic process that is
  1. Stationary: The transitions do not depend on the time step
  2. Memoryless: The transitions depend on the current state of the system, but not on past states (the *Markov Property*)

$$\boxed{P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i)}$$

- **Transition probabilities**

$$\boxed{P(X_{n+1} = j | X_n = i) = P_{ij}}$$

- Matrix representation: Storage of the transition probabilities

$$\boxed{P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots \\ P_{21} & P_{22} & P_{23} & \dots \\ \vdots & \vdots & \vdots & \ddots \\ & & & \dots \end{bmatrix}} \qquad \boxed{\sum_{j \in \mathcal{S}} P_{ij} = 1, \forall i \in \mathcal{S}}$$

Weather Example: Transition Probabilities

- Suppose the probability that tomorrow is dry is 0.8 if today is dry, but is 0.6 if it rains today.
- We write:

  P(dry tomorrow | dry today) = 0.8 = $P(X_1=0 \mid X_0=0)$

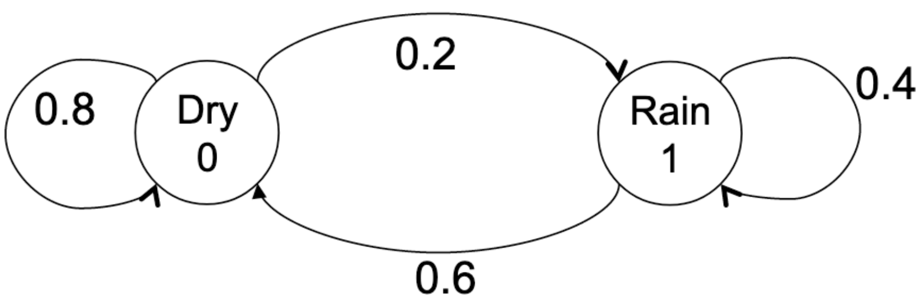
  P(dry tomorrow | rainy today) = 0.6 = $P(X_1=0 \mid X_0=1)$
- Or, for any day t, we write:

  $P(X_t+1=0 \mid X_t=0) = 0.8$ and $P(X_t+1=0 \mid X_t=1) = 0.6$
- Suppose we are given the states of weather on days 0,1,2,3. That is, suppose we know that X0=0 X1=0 X2=1, X3=0 (dry, dry, rainy, dry).
- What is the probability that X4=0?
  - Mathematically, what is P(X4=0 | X3=0, X2=1, X1=0, X0=0) ?
  - We have P(X4=0 | X3=0) = 0.8, and, in writing this number we did not care about the values of X2 X1 X0
  - This observation is true for any values of X3 X2 X1 X0 and in fact for any t.
- Intuitively, given today's weather and the weather in the past, the conditional probability of tomorrow's weather is independent of weather in the past and depends only on today's weather

Weather Example: Transition Probabilities

- The weather chain:
  - p00 =P(Xt+1 = 0| Xt = 0) = 0.8
  - p10 =P(Xt+1 = 0| Xt =1) = 0.6
  - p01 =P(Xt+1 = 1| Xt = 0) = 1-P(Xt+1 = 0| Xt=0) = 0.2
  - p11 =P(Xt+1 = 1| Xt =1) = 1-P(Xt+1 = 0| Xt =1) = 0.4

Weather



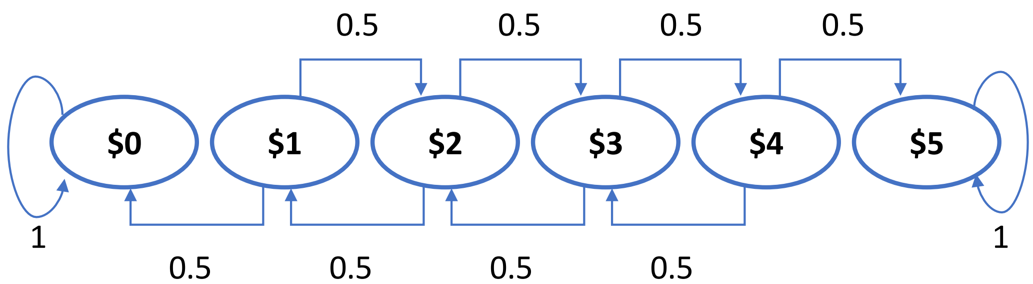
$$
\begin{array}{lll}
 & & \begin{array}{cc} 0 & 1 \end{array} \\
\textit{Dry} & 0 & \multirow{2}{*}{$\begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$} \\
\textit{Rain} & 1 &
\end{array}
$$

Gambler's Ruin Example: Properties

- Consider a gambling game with probability p=0.5 of winning on any turn, you start with $1, stop when you go broke or reach $5
- What are the random variables of interest, Xt?
  - X_t =$fortune on turn t
- What are the possible values (states) of the random variables?
  - {0,1,2,3,4,5}
- What is the index t?
  - turn of the game
- Does the Gambler's Ruin stochastic process satisfy the Markovian property?
  - Yes, intuitively, given your current gambling fortune and all past gambling fortunes, the conditional probability of your gambling fortune after one more gamble is independent of your past gambling fortunes and depends only on your current gambling fortune. More formally, P(X5=0 | X4=1, X3=2, X2=1, X1=2, X0=1) = P(X5=0 | X4=1) = 0.5.
- Is the Gambler's Ruin stochastic process stationary?
  - Yes, intuitively, the probability of winning is the same for all turns of the game. More formally, P(Xt+1=0 | Xt=1) = 0.5 for all t.

Gambler’s Ruin Example: Transition Probabilities

- State transition diagram:



- One-step transition matrix P:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Multi-step Transition Probabilities

- So far, we have only focused on one-step transition probabilities p_{ij}
  - But what if we are interested in the answer to the question, for example, if it is sunny today, what is the probability that it will be sunny day after tomorrow?
  - These are called multi-step (or n-step) transition probabilities.

- In particular, we want to find $P(X_t + n = j | X_t = i)$ which is denoted by $p_{ij}^{(n)}$

2-step Transition Probabilities: Weather Example

- Intuition: to go from state 0 to 0 in two steps we can either:
  - Go from 0 to 0 in one step and then go from 0 to 0 in one step OR
  - Go from 0 to 1 in one step and then go from 1 to 0 in one step
  - Therefore, $p^{(2)}_{00} = P(X_2 = 0 \mid X_0 = 0) = p_{00}p_{00}+p_{01}p_{10}$
- And more generally, $p_{00}^{(2)} = \sum_{k=0}^{1} p_{0k}p_{k0}$
- Now use the above intuition to write down the other 2-step transition probabilities
  - $p^{(2)}_{01}$, $p^{(2)}_{10}$, $p^{(2)}_{11}$

2-step Transition Probabilities: Weather Example

$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} \textit{Dry} \;\; 0 \\ \textit{Rain} \;\; 1 \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}$$

- These four two-step transition probabilities can be arranged in a matrix $P^{(2)}$ called the two-step transition matrix

$$P^{(2)} = \begin{pmatrix} p^{(2)}_{00} & p^{(2)}_{01} \\ p^{(2)}_{10} & p^{(2)}_{10} \end{pmatrix} = \begin{pmatrix} p_{00}p_{00} + p_{01}p_{01} & p_{00}p_{01} + p_{01}p_{11} \\ p_{10}p_{00} + p_{11}p_{10} & p_{10}p_{01} + p_{11}p_{11} \end{pmatrix}$$

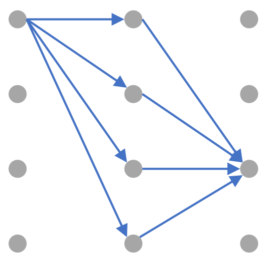
- Interpretation: $p_{01}^{(2)}$ is the probability that the weather the day after tomorrow will be rainy if the weather today is dry.

- An interesting observation: the two-step transition matrix is the square of the one-step transition matrix!
  - That is, $P^{(2)} = P^2$
  - Why? Recall matrix product

$$P^{(2)} = \begin{pmatrix} 0.8 & 0.4 \\ 0.6 & 0.2 \end{pmatrix}^2 = \begin{pmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{pmatrix}$$

2-step Transition Probabilities

- For a general Markov chain with states 0,1,...,S, to make a two-step transition from i to j, we go to some state k in one step from i and then go from k to j in one step.



$$\begin{aligned} P(X_{n+2}=j|X_n=i) &= \textstyle\sum_{k\in\mathcal{S}} P(X_{n+2}=j|X_{n+1}=k)P(X_{n+1}=k|X_n=i) \\ &= \textstyle\sum_{k\in\mathcal{S}} P_{ik}P_{kj} \\ &= \left(P^2\right)_{ij} \end{aligned}$$

- Therefore, the 2-step transition probability matrix is:

$$P^{(2)} = \begin{pmatrix} p_{00}^{(2)} & p_{01}^{(2)} & \cdots & p_{0S}^{(2)} \\ p_{01}^{(2)} & p_{11}^{(2)} & \cdots & p_{1S}^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ p_{S0}^{(2)} & p_{S1}^{(2)} & \cdots & p_{SS}^{(2)} \end{pmatrix}$$

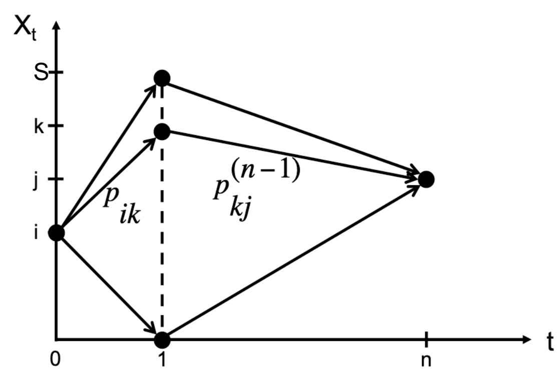
Multi-step Transitions: Chapman-Kolmogorov equations

- For a general Markov chain with states 0,1,...,M, the n-step transition from i to j means the process goes from i to j in n time steps
- Let m be a non-negative integer not bigger than n. The Chapman-Kolmogorov equation is:

$$p_{ij}^{(n)} = \sum_{k=0}^{S} p_{ik}^{(m)} p_{kj}^{(n-m)}$$

- Interpretation: if the process goes from state i to state j in n steps then it must go from state i to some state k in m (less than n) steps, and then go from k to j in the remaining n-m steps.
- Consider the case when n=1:

$$p_{ij}^{(n)} = \sum_{k=0}^{S} p_{ik}^{(1)} p_{kj}^{(n-1)}$$



Multi-step Transitions: Matrix representation

- The $p_{ij}^{(n)}$ are the elements of the n-step transition matrix, $P^{(n)}$
- In matrix notation,

$$P^{(n)} = P^{(m)} P^{(m-n)}$$

- This implies that the n-step transition matrix is the n^th power of the one-step transition, and so on:

$$\begin{aligned} P^{(n)} &= P \cdot P^{(n-1)} \\ &= P \cdot P \cdot P^{(n-2)} \\ &\vdots \\ &= P \cdot P \cdot P \cdots P = P^{n} \end{aligned}$$

Outline

- Motivation
- Markov chains lecture
  - Definition and Properties
  - Classification of States and Chains
  - Ergodic chains
    - Steady-state Dynamics and Distributions
    - First passage time distributions and expected values
  - Absorbing chains
    - Absorbing probabilities:
    - Absorbing probabilities: Matrix equations
    - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

Classification of States

- **Absorbing state**: Once you get in, you can never leave

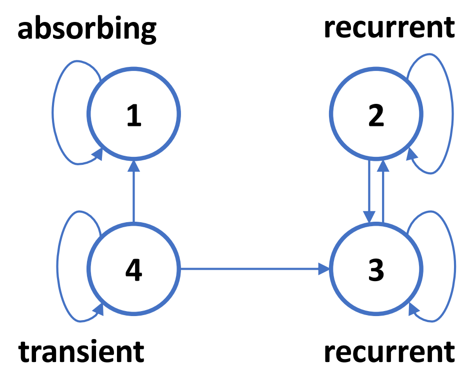
$$\boxed{P_{ii} = 1}$$

- **Transient state**: Once you leave, you may not come back

$$\boxed{P(T^i < \infty) < 1}$$

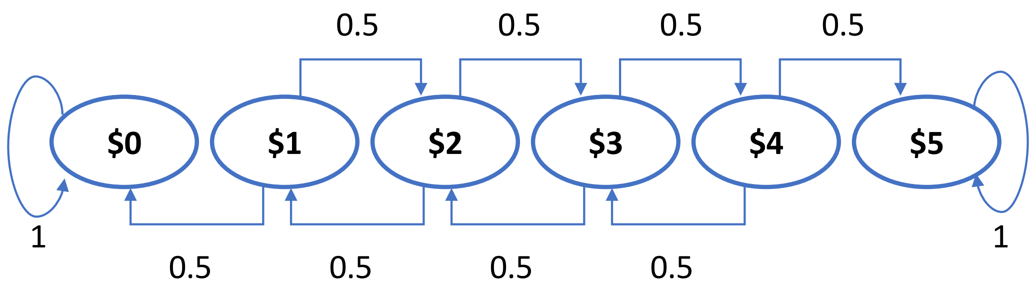
- **Recurrent state**: You will come back eventually

$$\boxed{P(T^i < \infty) = 1}$$



Classification of States: Gambler's Ruin Problem

- State transition diagram:



  - Transient states?
  - Recurrent states?
  - Absorbing states?

Irreducibility

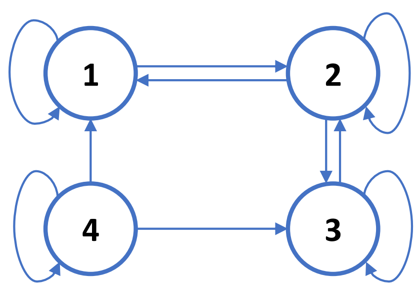
- Two states i and j **communicate** if there exists a path from i to j and a path from j to i with non-zero probabilities

$$\boxed{\text{there exist } x_1,\ldots,x_p \text{ such that } P_{ix_1}P_{x_1x_2}\ldots P_{x_{p-1}x_p}P_{x_pj} > 0}$$

- A chain is **irreducible** if all its states communicate

$$\boxed{\text{there exist } y_1,\ldots,y_q \text{ such that } P_{jy_1}P_{y_1y_2}\ldots P_{y_{q-1}y_q}P_{y_qi} > 0}$$

- If a chain is irreducible, either all its states are recurrent ("most of the time"), or all its states are transient

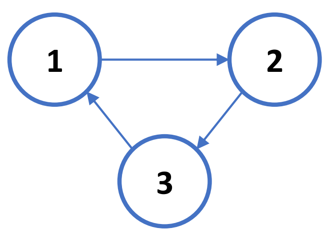


Periodicity

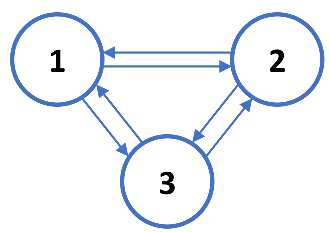
- A state i is **periodic** if it is visited only in a number of steps which is a multiple of an integer d > 1:

$$\boxed{gcd\left\{n \geq 1; P_{ii}^{n} > 0\right\} > 1}$$

- The state i is **aperiodic** otherwise
- A state with a self-loop (i.e. $p_{ii}>0$) is always aperiodic



periodic

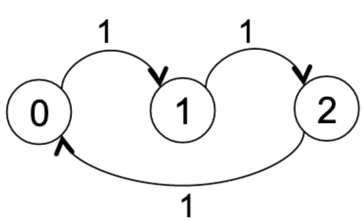


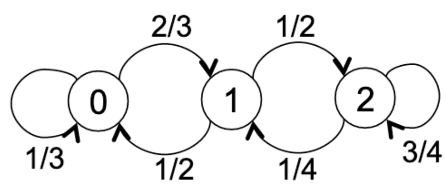
aperiodic

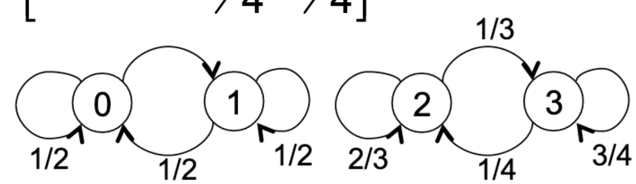
Irreducibility and Periodicity: Example

- Which of the following chains are Irreducible? Periodic?

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad P = \begin{bmatrix} \tfrac{1}{3} & \tfrac{2}{3} & 0 \\ \tfrac{1}{2} & 0 & \tfrac{1}{2} \\ 0 & \tfrac{1}{4} & \tfrac{3}{4} \end{bmatrix} \qquad P = \begin{bmatrix} \tfrac{1}{2} & \tfrac{1}{2} & 0 & 0 \\ \tfrac{1}{2} & \tfrac{1}{2} & 0 & 0 \\ 0 & 0 & \tfrac{2}{3} & \tfrac{1}{3} \\ 0 & 0 & \tfrac{1}{4} & \tfrac{3}{4} \end{bmatrix}$$







Irreducible

Period = 3

Irreducible

Aperiodic

Reducible, 2 classes Each

class is aperiodic

Ergodicity

- A Markov Chain is **ergodic** if it is
  - Irreducible
  - Recurrent (i.e., all its states are recurrent)
  - Aperiodic (i.e., all its states are aperiodic)

- Why do we care?
  - If a Markov chain is irreducible and recurrent, then its long-term behavior does not depend on initial conditions
  - If a Markov chain is also aperiodic, then its long-term behavior reaches “steady state,” i.e., does not change from stage to stage

Outline

- Motivation
- Markov chains lecture
    - Definition and Properties
    - Classification of States and Chains
    - Ergodic chains
        - Steady-state Dynamics and Distributions
        - First passage time distributions and expected values
    - Absorbing chains
        - Absorbing probabilities:
        - Absorbing probabilities: Matrix equations
        - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

Unconditional Probability in state j at time n

- The transition probability $p_{ij}^{(n)}$ is a conditional probability, $P(X_n=j \mid X_0=i)$
- How do we "un-condition" the probabilities?
- That is, how do we find the (unconditional) probability of being in state j at time n, $P(X_n=j)$?
- Uncondition on the probabilities $P(X_0=i)$ defining the initial state distribution:

$$\begin{aligned} P(X_n = j) &= \sum_{i=0}^{S} P(X_n = j | X_0 = i) P(X_0 = i) \\ &= p_{ij}^{(n)} P(X_0 = i) \end{aligned}$$

Unconditional Probabilities: Weather Example

- If initial conditions are unknown, we might assume it's equally likely to be in any initial state:

$$P(X_0 = 0) = \frac{1}{2} = P(X_0 = 1)$$

- Then, what is the probability that it is Dry (state 0) in two days?

- P(X = 0 in 2 days) = P(in state 0 at time 2) =

$$P^{(2)} = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} 0 & 1 \\ \multicolumn{2}{c}{\left(\begin{array}{cc} 0.76 & 0.24 \\ 0.72 & 0.28 \end{array}\right)} \end{array}\begin{array}{cc} \\ \textit{Dry} & 0 \\ \textit{Rain} & 1 \end{array}$$

$$\begin{aligned} P(X_2 = 0) &= \sum_{i=0}^{2} p_{i2}^{(2)} P(X_0 = i) \\ &= p_{00}^{(2)} P(X_0 = 0) + p_{10}^{(2)} P(X_0 = 1) \\ &= (0.76)(0.5) + (0.72)(0.5) = 0.74 \end{aligned}$$

Steady-State Probabilities

- As n gets large, what happens? What is the probability of being in any state?
- Consider the 5-step transition probability for the weather example:

$$P^{(5)} = \begin{pmatrix} 0.8 & 0.4 \\ 0.6 & 0.2 \end{pmatrix}^5 = \begin{pmatrix} 0.7501 & 0.2499 \\ 0.7498 & 0.2502 \end{pmatrix}$$

- In the long-run (e.g., after 5 or more days), the probability of being in state j converges

  → No longer depends on initial state

- These probabilities are called the steady-state probabilities

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j \qquad \longleftrightarrow \qquad \boxed{\lim_{n\to+\infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ & & & \cdots \end{bmatrix}}$$

Observations about this limit

- This limit exists for any "irreducible ergodic" Markov chain
- The behavior of this important limit depends on properties of states i and j and the Markov chain as a whole.
    - If j is transient, then $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ for all i
        - Intuitively, the probability that the Markov chain is in a transient state after a large number of transitions tends to zero.
    - If the chain is periodic, the limit will not exist

Stationary Distribution

- How can we find the probabilities π_j without calculating $P^{(n)}$ for very large n?
- The following are the steady-state equations:

$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}$	Once you reach this distribution, you stay there
$\pi^T = P\pi^T$	Above, in matrix form
$\sum_{j=0}^{M} \pi_j = 1$	The π_j form a probability distribution
$\pi_j \geq 0$	The π_j are defined (for ergodic chains)

- Solve a system of linear equations!

Solving for Stationary Distribution

$$\pi^T P = \pi^T$$

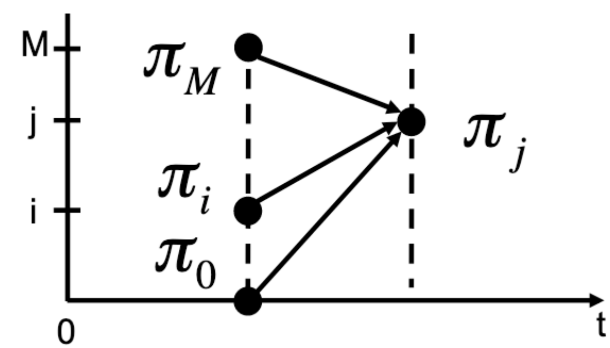
$$\sum_{i=0}^{M} \pi_i = 1$$

$$\begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_M \end{bmatrix} \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0M} \\ p_{10} & p_{11} & \cdots & p_{1M} \\ \vdots & \vdots & & p_{(M-1)M} \\ p_{M0} & p_{M1} & \cdots & p_{MM} \end{bmatrix} = \begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_M \end{bmatrix}$$

$$\begin{aligned} \pi_0 p_{00} + \pi_1 p_{10} + \cdots + \pi_M p_{M0} &= \pi_0 \\ \pi_0 p_{01} + \pi_1 p_{11} + \cdots + \pi_M p_{M1} &= \pi_1 \\ \vdots &= \vdots \\ \pi_0 p_{0M} + \pi_1 p_{1M} + \cdots + \pi_M p_{MM} &= \pi_M \\ \pi_0 + \pi_1 + \cdots + \pi_M &= 1 \end{aligned}$$

General idea:

Go from steady state to steady state



Other (Intuitive) Properties

- Another interpretation is that π_j is the fraction of time the process is in state j (in the long-run)

  Meaning, the stationary distribution corresponds to the long-run frequency of each state's occupancy:

$$\boxed{\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_n = j) = \pi(j), \forall j \in \mathcal{S}}$$

- The longer it takes to "return" to any state, the less time you spend in that state

$$\boxed{\pi_j = \frac{1}{E(T^j)}, \forall j \in \mathcal{S}}$$

First passage time:

$$\boxed{T^i = \min\{n \geq 1; X_n = i | X_0 = i\}}$$

Typical Applications of Steady-State Probabilities / Ergodic Chains

- **Expected recurrence time**: the expected number of steps between consecutive visits to a particular (recurrent) state
  - To answer questions such as, What is the expected number of sunny days between rainy days?
- **First passage time**: the time at which the Markov chain visits a particular state for the first time
  - Typical quantities: First passage time probability, Expected first passage time
- Number of visits to a state at a certain time
- Long-run expected average cost per unit time (if a cost is incurred or gain rewarded every time a Markov chain visits a specific state)

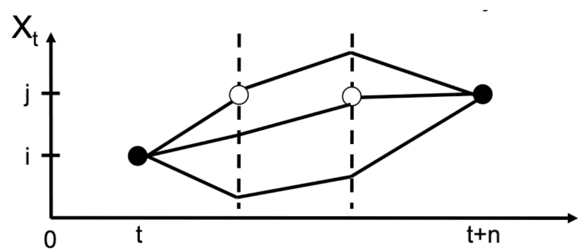
Expected Recurrence Time

- The expected recurrence time, denoted μ_{jj} , is the expected number of transitions between two consecutive visits to state j.
- Due to the frequency interpretation of the probabilities, the steady state probabilities, π_j, are related to the expected recurrence times, μ_{jj}, as

$$\mu_{jj} = \frac{1}{\pi_j} \qquad j = 0, 1, ..., S$$

First Passage Times

- Definition: **The first passage time from state i to state j is the number of transitions made by the process in going from state i to state j for the first time**
  - When i = j, this first passage time is the recurrence time for state i
- Let $f_{ij}^{(n)}$ = probability that the first passage time from state i to state j is equal to n
- What is the difference between $f_{ij}^{(n)}$ and $p_{ij}^{(n)}$?



  - $p_{ij}^{(n)}$ includes paths that visit j
  - $f_{ij}^{(n)}$a does not include paths that visit j

Observations on First Passage Times

- First passage times are random variables and have probability distributions associated with them
- $f_{ij}^{(n)}$ = probability that the first passage time from state i to state j is equal to n
- These probability distributions can be computed using a simple idea: condition on where the Markov chain goes after the first transition:
  - For the first passage time from i to j to be n>1, the Markov chain has to transition from i to k (different from j) in n=1, and then the first passage time from k to j must be n-1.
- This concept can be used to derive recursive equations for $f_{ij}^{(n)}$

  (and is frequently used in finding relationships in Markov chains)

First Passage Times: Recursive Equations

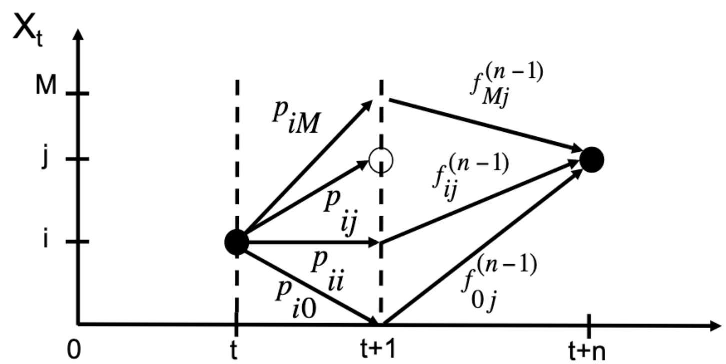
- The first passage times satisfy a recursive relationship:

$$f_{ij}^{(1)} = p_{ij}^{(1)} = p_{ij}$$

$$f_{ij}^{(2)} = \sum_{k \neq j} p_{ik} f_{kj}^{(1)}$$

$$\vdots$$

$$f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}$$



Probability of Ever Reaching j from i

- If the chain starts out in state i, what is the probability that it visits state j at some future time?
- This probability is denoted f_{ij} , and

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$$

- If f_{ij}=1, then the chain starting at i definitely reaches j at some future time, in which case $f^{(n)}_{ij}$ is a genuine probability distribution for the first passage time.
- On the other hand, if f_{ij}<1, the chain starting at i may never reach j. In fact, the probability that this happens is 1- f_{ij}.

Expected First Passage Time

- The expected first passage time from i to j is:

$$\mu_{ij} = \infty \qquad \text{if} \quad f_{ij} < 1$$

$$\mu_{ij} = E\left[f_{ij}^{(n)}\right] = \sum_{n=1}^{\infty} n f_{ij}^{(n)} \qquad \text{if} \ f_{ij} = 1$$

- There is another way to compute the expected first passage times μ_{ij} , which is to use the same idea as for finding recursive relations for first passage time probabilities: condition on where the chain goes after one transition

$$\mu_{ij} = \underbrace{1 p_{ij}}_{\text{Goes in 1 step}} + \sum_{k \neq j} \underbrace{p_{ik}}_{\text{Goes in } (1+\mu_{kj}) \text{ steps}} (1 + \mu_{kj}) = \sum_{k} p_{ik} + \sum_{k \neq j} p_{ik}\mu_{kj} = 1 + \sum_{k \neq j} p_{ik}\mu_{kj}$$

- System of equations:

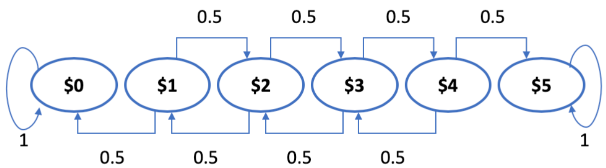
$$\boxed{\mu_{ij} = 1 + \sum_{\substack{k=0 \\ k \neq j}}^{M} p_{ik}\mu_{kj}}$$

Outline

- Motivation
- Markov chains lecture
    - Definition and Properties
    - Classification of States and Chains
    - Ergodic chains
        - Steady-state Dynamics and Distributions
        - First passage time distributions and expected values
    - Absorbing chains
        - Absorbing probabilities:
        - Absorbing probabilities: Matrix equations
        - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

Absorbing Markov Chains

Gambler's Ruin absorbing chain



- Quantities of interest:
  - Probability of absorption at k starting from i, f_{ik}
  - Expected time to absorption at k
  - Expected number of times process is in state j before being absorbed

- Absorbing states
  - Absorbing state: once the chain visits state k, it remains there forever, $p_{kk} = 1$
  - We are interested in f_{ik} when i is a transient state and k is an absorbing state

Linear Equations for Absorption Probabilities

- As before, we condition on the first transition of the Markov chain to get:

$$f_{ik} = \sum_{j=0}^{M} p_{ij} f_{jk} \qquad \textbf{for } \boldsymbol{i} = 0,1,\ldots,\boldsymbol{M}$$

with f_{kk}=1 for absorbing state k, and f_{jk}=0 when j is a recurrent state

Matrix Equations for Absorption Probabilities

- First, it is convenient to consider the canonical form of matrix P in aggregated version by uniting all transient states and all absorbing states:
  - Let Q represent transient states, for *t* transient states
  - Let R represent absorbing states, for *r* absorbing states
  - Then we have

$$P = \begin{pmatrix} Q & R \\ \mathbf{0} & I_r \end{pmatrix}$$

  Where

  - Q is a t-by-t matrix
  - R is a nonzero t-by-r matrix,
  - $\mathbf{0}$ is an r-by-t zero matrix, and
  - I_r is the r-by-r identity matrix.

- Q describes the probability of transitioning from some transient state to another
- R describes the probability of transitioning from some transient state to some absorbing state

Matrix Equations for Absorption Probabilities

- With Q and R, the equation for absorbing probabilities,

$$f_{ik} = \sum_{j=0}^{M} p_{ij} f_{jk} \qquad \text{for } i = 0,1,\ldots,M$$

can be decomposed into transient to absorbing transitions, and transient to transient followed by transient to absorbing transitions, as:

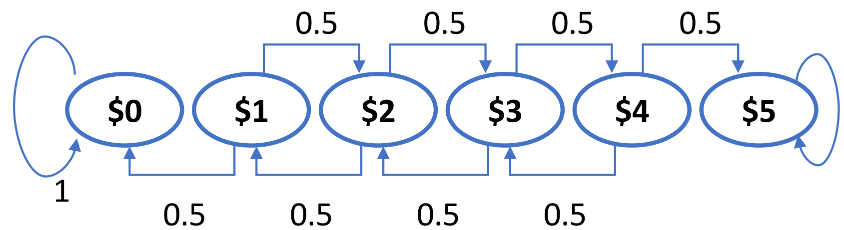
$$f_{ik} = p_{ij} + \sum_{j \in Q} p_{ij} f_{jk}$$

- Which can be rewritten in matrix form, and solved for the absorbing probability matrix A as:

$$A = R + QA$$
$$A = (I - Q)^{-1} R$$

Matrix Equations for Absorption Probabilities: Gambler's Ruin

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



- Reorganizing the matrix into canonical form:

$$P = \left[\begin{array}{cccc|cc} 0 & .5 & 0 & 0 & 0 & .5 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Q

R

$$A = (I - Q)^{-1} R$$

$$A = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{pmatrix}$$

Applying Absorbing Chain Theory to Ergodic Chains

- To answer certain questions, and especially if we are interested in more detailed behavior of the process going to j, we can apply absorbing chain theory to ergodic chains
    - For example, if we want to calculate the mean number of times the process will be in each of the other states before reaching j for the first time
    - To do this, we make j into an absorbing state $\longrightarrow$ absorbing process with single absorbing state
    - Then, the behavior of this process before absorption is exactly the same as the behavior of the original process before the first passage time to j
- Provides an alternative way to find the first passage time probabilities: Use the matrix relation derived for absorbing chains
    - Advantages: Matrix-based, non-recursive

Applying Absorbing Chain Theory to Ergodic Chains:

First Passage Probabilities – Matrix Equation

To find first passage to state o:

- Adopt the “single absorbing state” notion discussed above, creating a new chain where transient states are all states but state o has been removed. Then we want to find the probability that the contamination travels through transient states for time steps n-1 and only gets “absorbed” at o at exactly time step n. This can be found as follows:

$$F_{so}(n) = [(\mathbf{I} - \mathbf{P}(o|o))^{n-1}\mathbf{p}(o)]_s$$

  where P(o|o) is the transition matrix P with row and column o removed, and p(o) is the o^{th} column of P with element o removed.

- This expression represents the probability of the contamination bouncing around in the transient states (i.e. any state but state o) for n-1 time steps, then at the n^{th} time step going for the first time to o, from any connected node.
- So in absorbing chain notation, we can see the P(o|o) matrix as the Q matrix and the p(o) column vector as a one-state R matrix.

Summary

Definitions and properties

- State space
- Transition matrix
- Classes

Ergodic chains

- Stationary distribution
- First passage probabilities
- Expected first passage times

Absorbing chains

- Absorbing probabilities

Outline

- Motivation
- Markov chains lecture
  - Definition and Properties
  - Classification of States and Chains
  - Ergodic chains
    - Steady-state Dynamics and Distributions
    - First passage time distributions and expected values
  - Absorbing chains
    - Absorbing probabilities:
    - Absorbing probabilities: Matrix equations
    - Applying absorbing chain theory to ergodic chains
- Challenge problem: Pass-the-gift
- Example from research

Pass-the-gift problem: Setup

Suppose you are at a dinner party. The host wants to give out a door prize that is wrapped in a box. Everyone (including the host) sits around a circular table and each person is given a fair coin. Initially the host is holding the box. He/she flips his coin. If it is heads, the box is passed to the right; if it is tails, it is passed to the left. The process is repeated by whichever guest is holding the box. (Heads, they pass right; tails, they pass left.) The game ends when the last person to receive the box finally gets it for the first time. That person gets to keep the box as the winner of the game.

Pass-the-gift problem: Questions

1) What is your probability of winning the game?
   - *How would you calculate this?*
   - *Does it depend on where you sit?*

2) Given you win, what is the expected length of the game?
   - *How would you calculate this?*
   - *Does this depend on where you sit?*

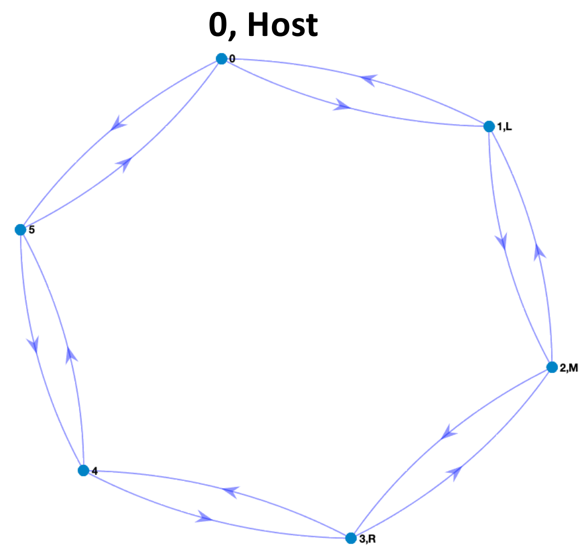
Pass-the-gift problem:

1) Probability of Winning

- How would you calculate this?

  Hint: Apply absorbing chain theory to regular chain:

  - *Make absorbing states out of Winning or Losing*
  - *Calculate probability of Winning as being absorbed into Winning state*



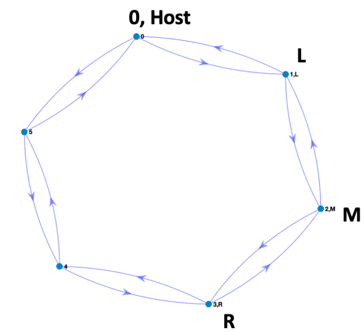
Pass-the-gift problem:

1) Probability of Winning

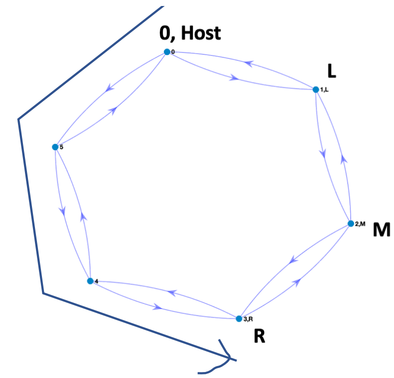
How do we think about winning and losing?

- Label my position M, the person to my right R, and to my left L
- What's necessary for me to win?
  - Because I'm sandwiched in between R and L, to get to me the gift has to first go either to R or to L to get to me (if I'm next to the host, the host is R or L).
  - So for me to win, either:
    - the gift first travels from the host to R, without visiting me or L, and then travels all around the circle from R to L without cycling back to reach me,
    - OR,
    - The gift first travels from the host to L, without visiting me or R, and travels all around the circle from L to R without reaching me.

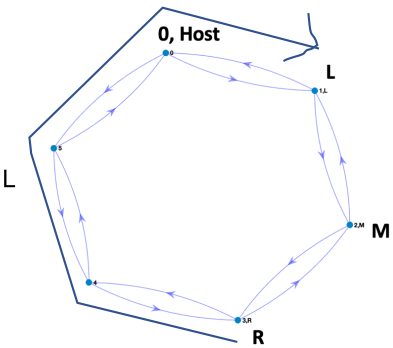
Labeling M, R, L



Probability of traveling first from host to R without visiting M or L



Probability of traveling from R to L without visiting me



Pass-the-gift problem:

1) Probability of Winning

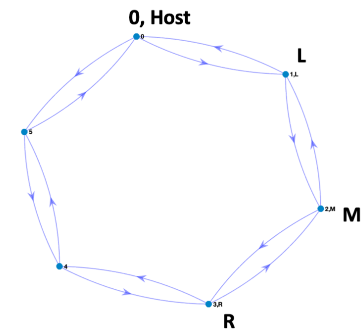
Let’s define these probabilities:

- Let p_RL = probability the gift travels from R to L without reaching me
  - By symmetry, p_RL = p_LR , so let’s call this p.
  - Also by symmetry, p does not depend on my location, since it is the probability going all the way around from wherever I am
- Let q_R_i= probability gift travels first to R without visiting L or me, if I am in position i
- Let q_L_i = probability gift travels first to L without visiting R or me

Then we have:

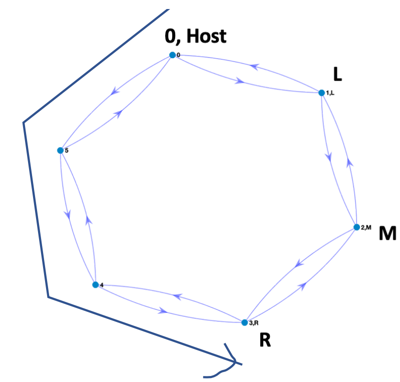
$$\text{Prob(I win, if I'm in position i)} = p(q\_R_i + q\_L_i)$$

Labeling M, R, L



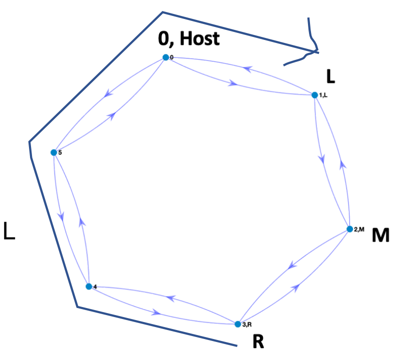
q_R_i:

Probability of traveling first from host to R without visiting M or L



p:

Probability of traveling from R to L without visiting me

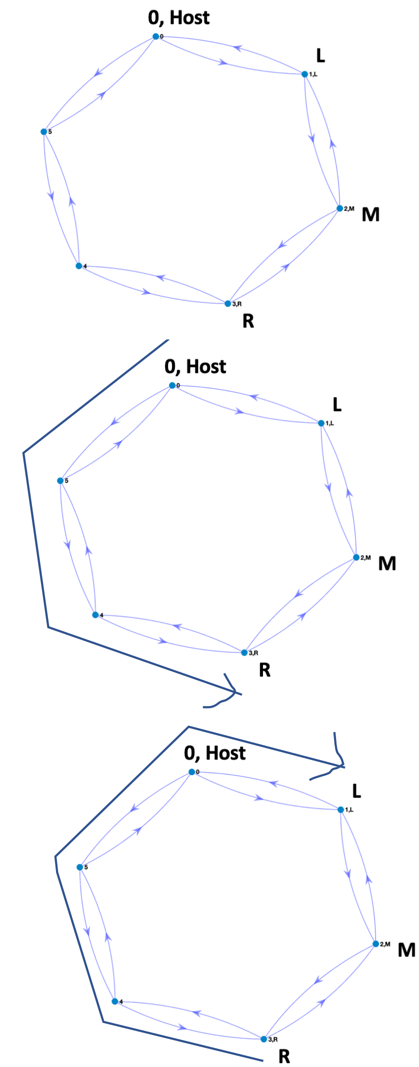


Pass-the-gift problem:

1) Probability of Winning

How to calculate p, q_R_i , q_L_i ?

We can calculate each of these probabilities by transforming our circular chain into an absorbing Markov chain!

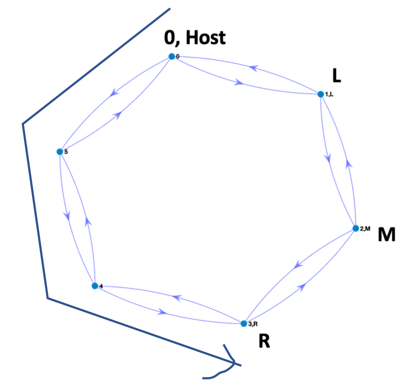


Pass-the-gift problem:

1) Probability of Winning

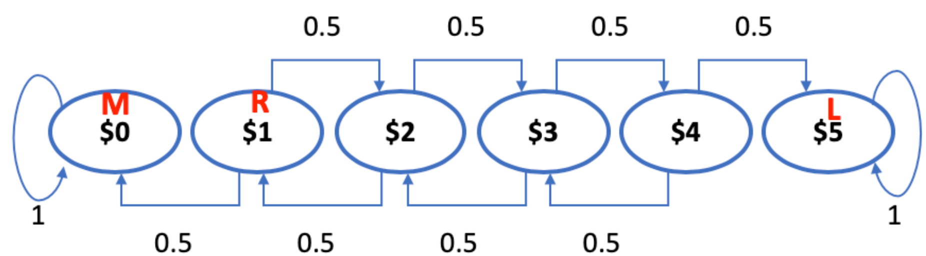
1a) Absorbing Markov chain to calculate p —

(Recall p is the probability the gift travels from R to L without visiting me, or vice versa)



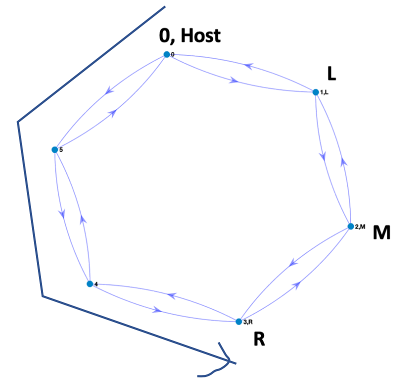
- *Create two absorbing states:*
  - *Win — it reaches everyone before me*
  - *Loose — it reaches me before it has reached everyone else*

- This is equivalent to the Gambler's Ruin Markov chain, labeling my location as state 0, $R \longleftrightarrow \$1$, and $L \longleftrightarrow \$5$ (and the other states in between):

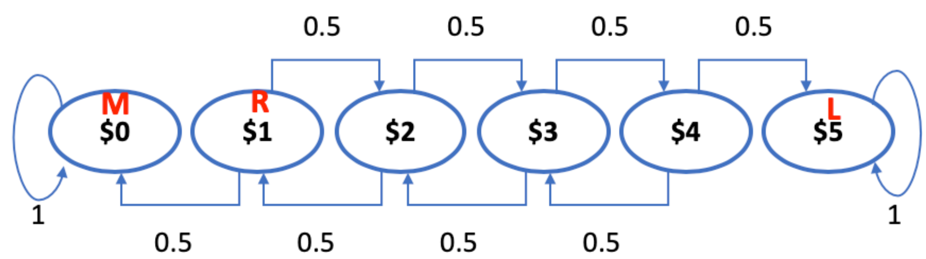


Pass-the-gift problem:

1) Probability of Winning



1a) Absorbing Markov chain to calculate p —



- The probability p of the gift traveling from R to L before passing by me is the probability of being absorbed at L ($5). Any way of loosing — that is, any way in which I am reached first — is equivalent to being absorbed at M ($0).

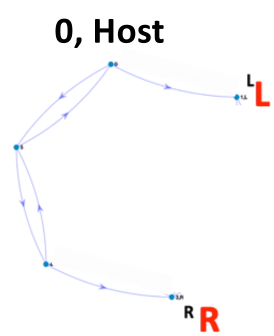
- Earlier we calculated these probabilities for the Gambler's Ruin problem with a $0 to $5 spread being the possible states, and found that the probability of being absorbed at 5 starting from position 1, $A_{1,5} = 0.2$

Pass-the-gift problem:

1) Probability of Winning

1b) Absorbing Markov chain to calculate q_R_i and q_L_i

- We are interested in the probability that the gift's path takes it to R before L and me, or takes it to L before R and me
- Since we are only interested in comparing these 2 events — we can create two absorbing states: one at R and one at L
- So we have a chain that goes from R to L, with 5 states total — my location (state M) knocked out because we have conditioned on the process going first to R or L before me

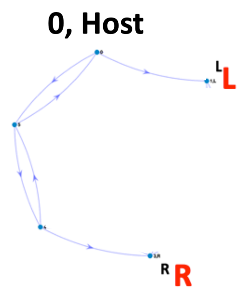


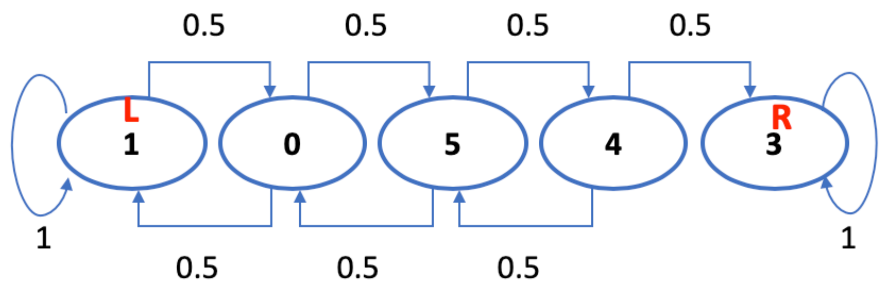
Pass-the-gift problem:

1) Probability of Winning

1b) Absorbing Markov chain to calculate *q_RL$_i$ and q_LR$_i$*

- Let's say we're in position 2; then we have the absorbing chain:



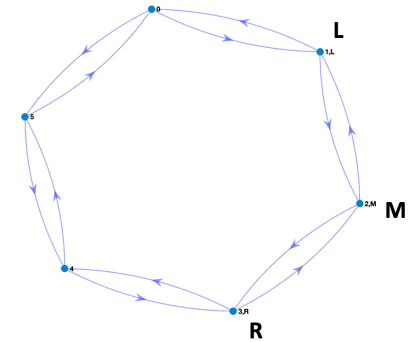


- So we are interested in the probability that the process, which starts with the host at 0, goes to L first: $A_{0,4}$, or goes to R first: $A_{0,2}$
- We can calculate both of these absorbing probabilities. But what else do we notice about **q_RL$_i$ and q_LR$_i$**?

  $A_{0,4} + A_{0,2} = 1$

  → q_R$_i$ + q_L$_i$ = 1 for all i !

We didn't really need a Markov chain for this!



Pass-the-gift problem:

1) Probability of Winning

Bringing everything together, we have:

Prob(I win, if I'm in position i) = $p^*(q\_R_i + q\_L_i)$

= p^*1

For the example with 6 states total, p = 0.2 (calculated in Gambler's Ruin)

This makes sense: since there are 5 ways to choose i, and p is the same for all i , p = 1 / 5 = 0.2

Pass-the-gift problem:

2) Expected length of the game given you win

- How would you calculate this?
    - *Hint: use conditional expectation and apply absorbing chain theory to the ergodic chain*

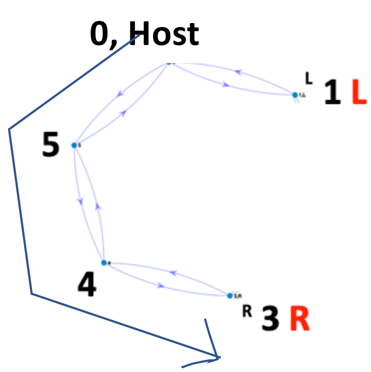
Pass-the-gift problem:

2) Expected length of the game given you win

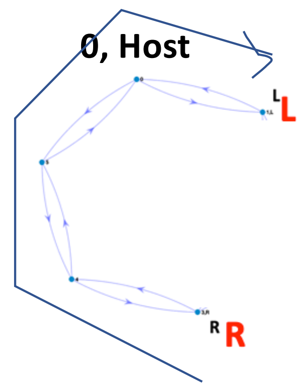
We can use the same logic as before for thinking about how I won:

- The process started by either going first to R or first to L (“segment 1”)
- If the process goes first to R, I win if it goes from R to L (“segment 2”, given R first), then L to me (“segment 3”, given R first).
- If the process goes first to L, I win if it goes from L to R, then R to me.

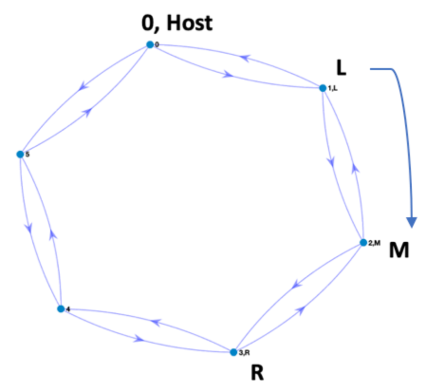
Segment 1



Segment 2



Segment 3



Pass-the-gift problem:

2) Expected length of the game given you win

Then by the total expectation theory, we have:

E[Length of game given I win] =

P(Host to R) * (E[segment 1 (Host to R)] + E[segment 2 (R to L)] + E[segment 3 (L to M)]) +

P(Host to L) * (E[segment 1 (Host to L)] + E[segment 2 (L to R)] + E[segment 3 (R to M])

Note that:

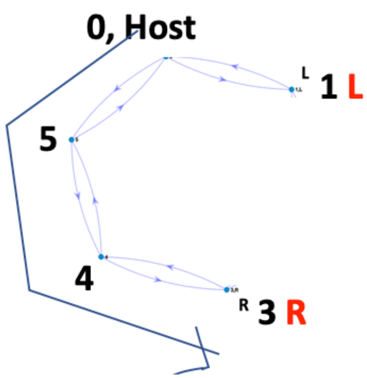
- P(Host to L) = 1 – P(Host to R)
- E[segment 2 (R to L)] = E[segment 2 (L to R)]
- E[segment 3 (L to M)]) = E[segment 3 (R to M])

Pass-the-gift problem:

2) Expected length of the game given you win

Segment 1: Host to R (or Host to L)

- In part (1b), we have already found the probability that the process goes first to R or L. This is $A_{0,R}$ or $A_{0,L}$ for the “knocked-out” chain with 2 absorbing states
- The expected length of Segment 1 is equal to the time to absorption $E[T_{0,R}]$ or $E[T_{0,L}]$

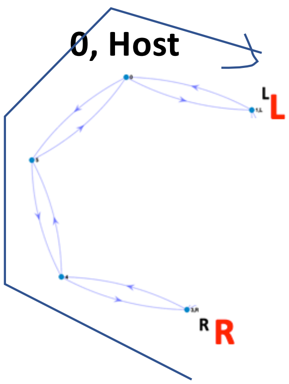


Pass-the-gift problem:

2) Expected length of the game given you win

Segment 2: R to L (equivalently, L to R)

- Given and the process went first to R, and I won, the process has to have gone from R to L before me – so we can think of this as a conditional space with a transient chain with my position knocked out.
- The expected time to reach L from R is then equal to the expected first passage time to R starting from L in this knocked-out chain, $E[F^*_{RL}]$
  - Denoting with * this first passage time, F^*_{ij}, because this is not the full chain

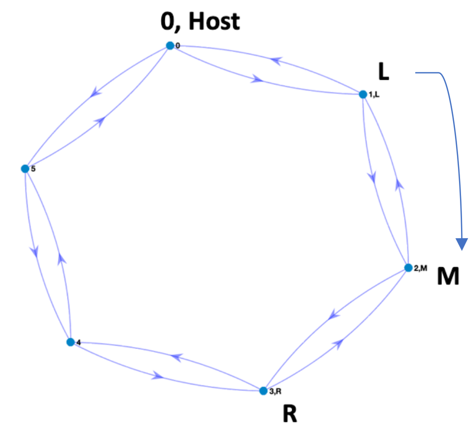


Pass-the-gift problem:

2) Expected length of the game given you win

Segment 3: L to me (equivalently, R to me)

- Now that all other states have been reached, and the process is at state L, all that needs to happen is for it to move from L to me. This is simply the expected first passage time from L to me in the full chain, $E[F_{LM}]$, since any of the previous states may be re-traversed



Pass-the-gift problem:

2) Expected length of the game given you win

Bring it all together using conditional expectation:

E[Length of game given I win] =

P(Host to R) * (E[segment 1 (Host to R)] + E[segment 2 (R to L)] + E[segment 3 (L to M)]) +

(1- P(Host to R)) * (E[segment 1 (Host to L)] + E[segment 2 (L to R)] + E[segment 3 (R to M])

= P(Host to R) * (E[$A_{0,R}$] + E[$F^*_{R,L}$] + E[$F_{L,M}$]) +

(1- P(Host to R)) * (E[$A_{0,L}$] + E[$F^*_{L,R}$] + E[$F_{R,M}$])

noting that E[$F^*_{R,L}$] = E[$F^*_{L,R}$] and E[$F_{L,M}$] = E[$F_{R,L}$]