Markov Chains

- Discrete Time Markov Chains
- Examples of Markov Chains
- Multi-step State Transitions

Andrey Markov
A discrete time stochastic process associates a random variable $X_t$ with a sequence of “time” locations: $t = 0, 1, 2, \ldots$

We will assume that $X_t$ is a discrete random variable with a finite number of possible outcomes. (Can be generalized.)

Sometimes we call $X_t$ the state of the process at time $t$.

These random variables have a joint distribution:

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$$

In most real applications, time points are not independent.
What is Discrete Time?

- A regular sampling of real times in the world: One variable every second, hour, day, year, or ...
- The steps taken by some computational process: One variable for every “iteration” of some algorithm.
- Any other data with “sequential” structure. In computational biology, we may model genetic sequences (DNA, proteins, ...)

\[ X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \cdots \]

Wikipedia
Suppose $X_0 = 1$ is fixed, and for times $t = 1, \ldots, n$:

\[
P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)
\]

- Suppose $X_0=1$ is fixed, and for times $t = 1, \ldots, n$:
  \[
  X_t \in \{1, 2, \ldots, m\}
  \]

- We assign probability to each possible discrete sequence:
  
  There are $m^n$ possible sequences of length $n$.

  - Expensive to enumerate these probabilities in a table!
  - Often not needed.
Markov assumption: The probability of the next state depends only on the current state, and not the sequence of steps taken to reach the current state:

\[ P(X_{t+1} = j \mid X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i_t) \]

We define a Markov chain via a state transition matrix:

\[
p_{ij} = P(X_{t+1} = j \mid X_t = i) \]

\[
X_t \in \{1, \ldots, m\} \quad \sum_{j=1}^{m} p_{ij} = 1
\]
Joint Distribution of Markov Sequences

- **Joint Distribution:** The Markov assumption implies that

\[
P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_0)
\]

\[= P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1})
\]

(any process)

(Markov process)

- **Initial state:** From some (possibly degenerate) distribution

\[P(X_0)\]

- **State transition matrix:**

\[p_{ij} = P(X_{t+1} = j \mid X_t = i)\]

\[X_t \in \{1, \ldots, m\}, \quad \sum_{j=1}^{m} p_{ij} = 1\]

\[
\begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1m} \\
p_{21} & p_{22} & \cdots & p_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \cdots & p_{mm}
\end{bmatrix}
\]
State Transition Diagrams

**State transition diagram:**
- A *directed graph* with one node for each of \( m \) possible states
- Draw an edge from node \( i \) to node \( j \) if \( p_{ij} > 0 \)
- A sample from a Markov process is then a *record of nodes visited in a random walk in this graph*

**State transition matrix:**

\[
p_{ij} = P(X_{t+1} = j \mid X_t = i)
\]

\( X_t \in \{1, \ldots, m\} \quad \sum_{j=1}^{m} p_{ij} = 1 \)

\[
\begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1m} \\
p_{21} & p_{22} & \cdots & p_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \cdots & p_{mm}
\end{bmatrix}
\]
State Transition Diagrams

State transition diagram:

- A *directed graph* with one node for each of $m$ possible states
- Draw an edge from node $i$ to node $j$ if $p_{ij} > 0$
- A sample from a Markov process is then a *record of nodes visited in a random walk in this graph*

**Example:** *Bull and Bear Markets (Wikipedia)*

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$
Markov Property and Independence

\[ P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1}) \]

- For most choices of the state transition matrix, the states at different times are *not* independent. *This is useful!*

\[ p_{X_s X_t}(x_s, x_t) \neq p_{X_s}(x_s)p_{X_t}(x_t) \]

- But conditioned on the value of the present state, the past and future of a Markov process are independent:

\[ Y_t = \{X_0, X_1, \ldots, X_{t-1}\} \quad Z_t = \{X_{t+1}, X_{t+2}, \ldots, X_n\} \]

\[ p_{Y_t Z_t}(y_t, z_t \mid X_t = x_t) = p_{Y_t}(y_t \mid X_t = x_t)p_{Z_t}(z_t \mid X_t = x_t) \]
CS145: Lecture 17 Outline

- Discrete Time Markov Chains
- Examples of Markov Chains
- Multi-step State Transitions
Let $X_i$ be the number of days (including today) some machine has been broken, or $X_i=0$ if the machine is currently working.

A machine that is working today will be broken tomorrow with probability $p$. Otherwise, with probability $1-p$, it keeps working.

On any given day, a broken machine is repaired with probability $r$. Otherwise, with probability $1-r$, it remains broken.

After being broken for $m$ days, it is always replaced with a working machine.

$$
\begin{pmatrix}
1-p & p & 0 & 0 & \cdots & 0 \\
r & 0 & 1-r & 0 & \cdots & 0 \\
r & 0 & 0 & 1-r & \cdots & 0 \\
& \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & 0 & \cdots & 0
\end{pmatrix}
$$
Let $X_i$ be the number of customers in a line (queue) during (short) time period $i$, where exactly one event happens:

- With probability $s$, a customer is served and leaves the queue, unless already there are no customers ($X_i=0$).
- With probability $r$, a new customer joins the queue, unless the queue is already at its maximum capacity $m$.
- Otherwise, the number of customers remains the same ($X_{i+1}=X_i$).

This is a discrete Markov process with $m+1$ states.
Are characters independent?

- Assume letters are independent and equally common:
  \[ P(X_t) = \frac{1}{27} \text{ for } X_t \in \{a, b, c, \ldots, z, \_\} \]

- Assume letters are independent and follow frequencies of real text:
  saade ve mw hc n ctt da k cethctocusoscelalwo gx fgrsnoh, tvcttaf aetnlbilo fc lhd okleutsndycoshtbogo cct ib nhcaoopcnfni ngent
A first-order Markov model encodes probability of each letter, given previous letter.

A second-order (bigram) Markov model encodes probability of each letter, given previous two letters (state is letter pairs).

A third order (trigram) Markov model encodes probability of each letter, given previous three letters (state is letter triples).
Order-1: t I amy, vin. id wht omanly heay atuss n macon aresethe hired boutwhe t, tl, ad torurest t plur I wit hengamind tarer-plarody thishand.

Order-2: Ther I the heingoid of-pleat, blur it dwere wing waske hat trooss. Yout lar on wassing, an sit." "Yould," "I that vide was nots ther.

Order-3: I has them the saw the secorrow. And wintails on my my ent, thinks, fore voyager lanated the been elsed helder was of him a very free bottlemarkable,

Order-4: His heard." "Exactly he very glad trouble, and by Hopkins! That it on of the who difficentralia. He rushed likely?" "Blood night that.

Claude Shannon’s Markov chain simulator (1948): To construct [an order 1 model] for example, one opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter recorded, etc. It would be interesting if further approximations could be constructed, but the labor involved becomes enormous at the next stage.
Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

\[ r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i) \]

where \( r_{ij}(1) = p_{ij} \)

State transition matrix:

\[ p_{ij} = P(X_{t+1} = j \mid X_t = i) \]

\( X_t \in \{1, \ldots, m\} \)

\[ \sum_{j=1}^{m} p_{ij} = 1 \]
\[ \pi_{ti} = P(X_t = i) \quad \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \]

After \( n \) time steps:

\[ \pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n \]

State Transition Matrix:

\[
P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}
\]

\[ p_{ij} = P(X_{t+1} = j \mid X_t = i) \]
Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

\[ r_{ij}(n) = P(X_n = j \mid X_0 = i) \]

where \( r_{ij}(1) = p_{ij} \)

- First consider the special case where \( n=2 \):

\[
P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{m} P(X_2 = j, X_1 = k \mid X_0 = i)
\]

\[
P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{m} P(X_2 = j \mid X_1 = k) P(X_1 = k \mid X_0 = i)
\]

\[
r_{ij}(2) = \sum_{k=1}^{m} p_{ik} p_{kj} = \sum_{k=1}^{m} r_{ik}(1) p_{kj}
\]
Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

\[ r_{ij}(n) = P(X_n = j \mid X_0 = i) \quad \text{where } r_{ij}(1) = p_{ij} \]

- Computed recursively via the Chapman-Kolmogorov equation:

\[ r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj} , \quad \text{for } n > 1, \text{ and all } i, j, \]

\[ P(X_n = j \mid X_0 = i) = \sum_{k=1}^{m} P(X_{n-1} = k \mid X_0 = i) \cdot P(X_n = j \mid X_{n-1} = k, X_0 = i) \]
Multi-step State Transitions

Given the current state, we would like to predict what state we will be in at multiple steps into the future:

\[ r_{ij}(n) = P(X_n = j \mid X_0 = i) \]

where \( r_{ij}(1) = p_{ij} \)

Computed recursively via the Chapman-Kolmogorov equation:

\[ r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}, \quad \text{for } n > 1, \quad \text{and all } i, j, \]

With random initial state:

\[ P(X_n = j) = \sum_{i=1}^{m} P(X_0 = i)r_{ij}(n) \]

**Marginal distribution of state after \( n \) steps.**
Reminder: Matrix Multiplication

\[ y = Ax = \begin{bmatrix} a_1^T & a_2^T & \cdots & a_m^T \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix} \]

\[ x^T y \in \mathbb{R} = [x_1 \ x_2 \ \cdots \ x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i. \]

\[ y^T = x^T A = x^T \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} x^T a_1 \\ x^T a_2 \\ \vdots \\ x^T a_n \end{bmatrix} \]

\[ C = AB = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix} \]

Zico Kolter, 2012
State Transitions & Matrix Multiplication

\[ \pi_{ti} = P(X_t = i) \]

\[ \pi_{1j} = \sum_{i=1}^{m} p_{ij} \pi_{0i} \]

\[ \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \]

Textbook convention:

\[ \pi_1^T = \pi_0^T P \]

Each row of \( P \) sums to one.

Alternative convention:

\[ \pi_1 = P^T \pi_0 \]

Each column of \( P^T \) sums to one.
Multi-Step State Transitions

\[ \pi_{ti} = P(X_t = i) \quad p_{ij} = P(X_{t+1} = j \mid X_t = i) \]

\[ \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \]

\[ P = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1m} \\
p_{21} & p_{22} & \cdots & p_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \cdots & p_{mm}
\end{bmatrix} \]

State Distribution after \( n \) time steps:

\[ \pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n \]

\[ \pi_n = P^T \pi_{n-1} = P^T P^T \pi_{n-2} = (P^n)^T \pi_0 \]

\( P^n \) multiplies the square matrix \( P \) by itself \( n \) times.

This is not equivalent to raising the entries of \( P \) to the power \( n \).
Example: Up-to-Date or Behind?

Let us introduce states 1 and 2, and identify them with being up-to-date and behind, respectively. Then, the transition probabilities are

\[ p_{ij} = \begin{cases} 0.8 & \text{if } j = 1 \text{ and } i = 2 \\ 0.6 & \text{if } j = 1 \text{ and } i = 1 \\ 0.2 & \text{if } j = 2 \text{ and } i = 2 \\ 0.4 & \text{if } j = 2 \text{ and } i = 1 \\ 0.0 & \text{otherwise} \end{cases} \]

\[ r_{ij}(n) = P(X_n = j \mid X_0 = i) = P^n \]

### Derivation of the Chapman-Kolmogorov equation

The probability of transitioning from state \( i \) to state \( j \) in \( n \) steps can be expressed as

\[ r_{ij}(n) = \sum_{k=1}^{m} p_{ik} r_{kj}(n-1) \]

For the cases of Examples 6.1 and 6.2, respectively. There are some interesting observations about the limiting behavior of the transition probabilities as a function of the number of transitions.

### Figures

- Figure 6.1: Transition probability graph in Example 6.1.
- Figure 6.2: Transition probability graph for the cases of Examples 6.1 and 6.2.
- Figure 6.3: Time-1 step transition probabilities for the "up-to-date/behind" Example 6.1.
- Figure 6.4: Time-1 step transition probabilities for the "up-to-date/behind" Example 6.1.
Example: Spiders and the Fly

The transition probability graph and the transition probability matrix for the "spiders-and-fly" Example 6.2.

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 1.0 & 0 & 0 & 0 \\
2 & 0.3 & 0.4 & 0.3 & 0 \\
3 & 0 & 0.3 & 0.4 & 0.3 \\
4 & 0 & 0 & 0 & 1.0 \\
\end{array}
\]

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 1.0 & 0 & 0 & 0 \\
2 & 0.5 & 0.25 & 0.24 & 0.09 \\
3 & 0.09 & 0.24 & 0.25 & 0.42 \\
4 & 0 & 0 & 0 & 1.0 \\
\end{array}
\]

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 1.0 & 0 & 0 & 0 \\
2 & 0.5 & 0.17 & 0.17 & 0.16 \\
3 & 0.16 & 0.17 & 0.17 & 0.50 \\
4 & 0 & 0 & 0 & 1.0 \\
\end{array}
\]

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 1.0 & 0 & 0 & 0 \\
2 & 0.55 & 0.12 & 0.12 & 0.21 \\
3 & 0.21 & 0.12 & 0.12 & 0.55 \\
4 & 0 & 0 & 0 & 1.0 \\
\end{array}
\]

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 1.0 & 0 & 0 & 0 \\
2 & 2/3 & 0 & 0 & 1/3 \\
3 & 1/3 & 0 & 0 & 2/3 \\
4 & 0 & 0 & 0 & 1.0 \\
\end{array}
\]

\[
r_{ij}(n) = P(X_n = j \mid X_0 = i) = P^n
\]
Markov Convergence Questions

As $n \to \infty$, does $r_{ij}(n)$ converge to something?

$\begin{align*}
0.5 &amp; \quad 0.5 \\
1 &amp; \quad 1 &amp; \quad 1
\end{align*}$

$r_{11}(n) = \quad r_{22}(n) = \quad r_{33}(n) =$
Markov Convergence Questions

As \( n \to \infty \), does \( r_{ij}(n) \) depend on the initial state \( i \)?

![Transition probability graph](attachment:transition_graph.png)

<table>
<thead>
<tr>
<th>UpD</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

\[ r_{ij}(1) \]

\[ r_{ij}(\infty) \]

![Transition probability matrices](attachment:transition_matrices.png)
Classification of Markov Chain States

- **Accessible:** State $j$ is accessible from state $i$ if $r_{ij}(n)$ is positive for some $n$. The set of states accessible from state $i$ then equals

$$A(i) = \{ j \mid r_{ij}(n) > 0 \text{ for at least one time } n \}$$

- **Recurrent:** If a recurrent state is visited once, it will be visited an infinite number of times. State $i$ is recurrent if and only if

$$\text{for all } j \in A(i), i \in A(j).$$

- **Transient:** Any state that is not recurrent is transient, and with probability one will only be visited a finite number of times

- **Absorbing:** A state $i$ is absorbing if the process never leaves that state: $p_{ii} = 1$
Classification of Markov Chain States

- **Accessible:** State \( j \) is accessible from state \( i \) if \( r_{ij}(n) \) is positive for some \( n \). The set of states accessible from state \( i \) then equals

\[
A(i) = \{ j \mid r_{ij}(n) > 0 \text{ for at least one time } n \}
\]

- **Recurrent:** If a recurrent state is visited once, it will be visited an infinite number of times. State \( i \) is recurrent if and only if

\[
\text{for all } j \in A(i), i \in A(j).
\]

- **Recurrent Class:** A collection of recurrent states for which all pairs are accessible from each other, or communicate

A pair of states \((i,j)\) communicate if and only if the state transition diagram has directed paths from \( i \) to \( j \), and from \( j \) to \( i \).
Classification of Markov Chain States

Accessible: State $j$ is accessible from state $i$ if $r_{ij}(n)$ is positive for some $n$. The set of states accessible from state $i$ then equals

$$A(i) = \{j \mid r_{ij}(n) > 0 \text{ for at least one time } n\}$$

Recurrent: If a recurrent state is visited once, it will be visited an infinite number of times. State $i$ is recurrent if and only if

for all $j \in A(i), i \in A(j)$.

Recurrent Class: A collection of recurrent states for which all pairs are accessible from each other, or communicate

Irreducible: A Markov chain is irreducible if it has only one recurrent class (plus possibly some transient states)
Examples of Markov Chain Decompositions

Sec. 6.2 Classification of States

1. Single class of recurrent states

2. Single class of recurrent states (1 and 2) and one transient state (3)

3. Two classes of recurrent states (class of state 1 and class of states 4 and 5) and two transient states (2 and 3)

Figure 6.7: Examples of Markov chain decompositions into recurrent classes and transient states.

To the presence or absence of a certain periodic pattern in the times that a state is visited. In particular, a recurrent class is said to be periodic if its states can be grouped in $d > 1$ disjoint subsets $S_1, ..., S_d$ so that all transitions from one subset lead to the next subset; see Fig. 6.8. More precisely, if $i \in S_k$ and $p_{ij} > 0$, then

- $\{ j \in S_{k+1}, \text{ if } k = 1, ..., d - 1 \}$
- $j \in S_1, \text{ if } k = d$.

A recurrent class that is not periodic, is said to be aperiodic.

Thus, in a periodic recurrent class, we move through the sequence of subsets in order, and after $d$ steps, we end up in the same subset. As an example, the recurrent class in the second chain of Fig. 6.7 (states 1 and 2) is periodic, and the same is true of the class consisting of states 4 and 5 in the third chain of Fig. 6.7. All other classes in the chains of this figure are aperiodic.

Absorbing State

Not Irreducible (Reducible)
Periodic States

- **Periodic**: A state $j$ in a Markov chain is periodic if there exists an integer $d > 1$ such that $P(X_{t+s}=j \mid X_t=j) = 0$ unless $s$ is divisible by $d$.

- **Aperiodic**: A Markov chain in which no recurrent states are periodic.
CS145: Lecture 18 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains
We Sec. 6.1 Discrete-Time Markov Chains

At the probability matrix readily apparent.

Visualize the entire model in a way that can make some of its major properties "self-transition"). By recording the numerical values of graph it is also helpful to lay out the model in the so-called Specification of Markov Models possible for the next state to be the same as the current one. Even though the state does not change, we still view this as a state transition of a special type (a "transition probability matrix of earlier states.

All of the elements of a Markov chain model can be encoded in a

• (a) the set of states
• (b) the set of possible transitions, namely, those pairs (s, s')
• (c) the numerical values of those transitions

A Markov chain model is specified by identifying variables 

\[ \pi_{ti} = P(X_t = i) \]

\[ \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \]

After n time steps:

\[ \pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P^2 = \pi_0^T P^n \]

**State Transition Matrix:**

\[ P = \begin{bmatrix}
    p_{11} & p_{12} & \cdots & p_{1m} \\
    p_{21} & p_{22} & \cdots & p_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    p_{m1} & p_{m2} & \cdots & p_{mm}
\end{bmatrix} \]

\[ p_{ij} = P(X_{t+1} = j \mid X_t = i) \]

**State Transition Diagram:**
Steady-State (Equilibrium) Distribution

\[ \pi_{ti} = P(X_t = i) \quad \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \]

After \( n \) time steps:

\[ \pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n \]

If state distribution converges:

\[ ||\pi_n - \pi_{n-1}|| \to 0 \]

The equilibrium distribution is the unique solution to \( m+1 \) linear equations:

\[ \pi_j^* = \sum_{i=1}^{m} \pi_i^* p_{ij}, \quad j = 1, \ldots, m. \]

\[ \sum_{i=1}^{m} \pi_i^* = 1. \]

\[ \pi^* = P^T \pi^* \]

**Alternative Interpretation:** Eigenvector of state transition matrix whose eigenvalue is 1.
Conditions for Steady-State Distribution

A unique steady-state distribution exists if and only if:
- The Markov chain is \textit{aperiodic} (not periodic).
- The Markov chain is \textit{irreducible} (single recurrent class).

Properties of steady-state distributions:
- For any transient state $i$, $\pi_i^* = 0$
- For all state pairs $(i,j)$, \( \lim_{n \to \infty} r_{ij}(n) = \pi_j^* \)

Linear equations defining unique steady-state distribution:

\[
\pi_j^* = \sum_{i=1}^{m} \pi_i^* p_{ij}, \quad j = 1, \ldots, m.
\]
\[
\sum_{i=1}^{m} \pi_i^* = 1.
\]
Violations of Steady-State Conditions

Periodic (not aperiodic) Markov chains:

Multiple recurrent classes (not irreducible):

Both aperiodic and irreducible:

Transient state:

\[ \pi_3^* = 0 \]
Example: 2-State Stationary Distribution

\[
\begin{pmatrix}
0.5 & 0.5 \\
0.2 & 0.8
\end{pmatrix}
\begin{pmatrix}
(2/7, 5/7) \\
(0.5, 0.5) \\
(0.2, 0.8)
\end{pmatrix} = \begin{pmatrix}
2/7 \\
5/7
\end{pmatrix}
\]

\[\pi_1 = 2/7, \quad \pi_2 = 5/7\]

- Assume process starts at state 1. \( P(X_1 = 1) = 1 \)

- \( P(X_1 = 1, \text{ and } X_{100} = 1) = \approx \pi_1 = \frac{2}{7} \)

- \( P(X_{100} = 1 \text{ and } X_{101} = 2) \approx \pi_1 p_{12} = \frac{2}{7} \cdot \frac{1}{2} = \frac{1}{7} \)
A finite state space Markov chain has a stationary distribution if and only if it is aperiodic, and its corresponding directed graph has one strongly connected component.
Example: Card Shuffling

- Markov chain states: All \( n! \) orderings of \( n \) distinct cards
- Transitions: Pick one of the \( n \) cards uniformly at random, and move that card to the top of the deck.

**Stationary distribution:** Uniform distribution over permutations

**Proof.**

Consider a state \( C = c_1, \ldots, c_n \). The chain can move to this state from the \( n \) states \( C(j) = c_2, \ldots, c_1, \ldots, c_n \), where card \( c_1 \) is in place \( j \) of the deck. There are \( n \) such states, and each has probability \( 1/n \) to move to state \( C \).

\[
\pi_C = \sum_{j=1}^{n} \pi_{C(j)} \frac{1}{n}
\]

\[
\frac{1}{n!} = \sum_{j=1}^{n} \frac{1}{n} \frac{1}{n!}
\]
Visit Frequency Interpretation

\[ \pi_j = \sum_k \pi_k p_{kj} \]

- (Long run) frequency of being in \( j \): \( \pi_j \)
- Frequency of transitions \( k \rightarrow j \): \( \pi_k p_{kj} \)
- Frequency of transitions into \( j \): \( \sum_k \pi_k p_{kj} \)

![Diagram of Markov Chain](image)
Example: Google’s Pagerank

Consider the following Markov chain:

- One state for each of $m$ webpages
- At each time step, choose one of the outgoing links from a page with equal probability
- Rank of a page: Fraction of time that a “random surfer” spends on that page (equilibrium distribution)

Good Properties:

- Webpages are important if many other pages link to them
- Webpages are important if other important pages link to them
Consider the following Markov chain:

- One state for each of $m$ webpages
- At each time step, choose one of the outgoing links from a page with equal probability
- Rank of a page: Fraction of time that a “random surfer” spends on that page (equilibrium distribution)

Problems:

- There may be many absorbing states, so the desired equilibrium distribution does not exist!
- Results can be sensitive to single link changes, and thus “noisy”
Example: Google’s Pagerank

Consider the following Markov chain:

- At each time step, \textit{with probability} \( q \), choose one of the outgoing links from a page with equal probability.
- At each time step, \textit{with probability} \( 1 - q \), choose one of \( m \) pages uniformly.
- Rank of a page: Fraction of time that a “random surfer” spends on that page.

Good Properties:

- Webpages are important if many other pages link to them.
- Webpages are important if other important pages link to them.
- Steady-state distribution always exists.
Example: Google’s Pagerank

Because web graph is sparse, it is possible to compute the probabilities of reaching each page even if the number of pages is very large.

\[ \pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n \]
CS145: Lecture 19 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains
Consider a simplified class of Markov chains where all states are one of two types:

- **Transient:** Visited finite number of times.
- **Absorbing:** Once entered, process never leaves (self-transition probability equals 1).

Practical examples of such Markov chains:
- Games with *win* and *loss* states.
- Computer systems with *failure* states.

To characterize such Markov chains:
- From each *transient* state, what is the probability of each *absorbing* state?
- From each *transient* state, what is the expected time until *absorption*?
Computing Absorption Probabilities

\[ a_i = P(X_n \text{ eventually becomes equal to the absorbing state } s \mid X_0 = i) \]

**Theorem:**
Consider a Markov chain in which each state is either transient or absorbing. We fix a particular absorbing state \( s \). Then, the probabilities \( a_i \) of eventually reaching state \( s \), starting from \( i \), are the unique solution of the equations

\[
\begin{align*}
    a_s &= 1, \\
    a_i &= 0, \quad \text{for all absorbing } i \neq s, \\
    a_i &= \sum_{j=1}^{m} p_{ij} a_j, \quad \text{for all transient } i.
\end{align*}
\]

**Proof:**
\[
\begin{align*}
a_i &= P(A \mid X_0 = i) = \{ \text{state } s \text{ is eventually reached} \} \\
    &= \sum_{j=1}^{m} P(A \mid X_0 = i, X_1 = j) P(X_1 = j \mid X_0 = i) \quad \text{(total probability thm.)} \\
    &= \sum_{j=1}^{m} P(A \mid X_1 = j)p_{ij} \quad \text{(Markov property)} \\
    &= \sum_{j=1}^{m} a_j p_{ij}.
\end{align*}
\]
**Computing Absorption Probabilities**

\[ a_i = \mathbf{P}(X_n \text{ eventually becomes equal to the absorbing state } s \mid X_0 = i) \]

**Theorem:**
Consider a Markov chain in which each state is either transient or absorbing. We fix a particular absorbing state \( s \). Then, the probabilities \( a_i \) of eventually reaching state \( s \), starting from \( i \), are the unique solution of the equations

\[
\begin{align*}
    a_s &= 1, \\
    a_i &= 0, \quad \text{for all absorbing } i \neq s, \\
    a_i &= \sum_{j=1}^{m} p_{ij} a_j, \quad \text{for all transient } i.
\end{align*}
\]

From each initial state \( i \), what is the probability \( a_i \) of eventually reaching state 1?

\[
\begin{align*}
    a_1 &= 1, & a_2 &= \frac{2}{3}, & a_3 &= \frac{1}{3}, & a_4 &= 0
\end{align*}
\]
Example: The Gambler’s Ruin

- At each round, win $1 with probability $p$, lose $1$ with probability $(1-p)$.
- Gambler plays until wins a target of $m$, or loses all money.
- Starting with $i$, what is the probability that the gambler wins or loses?

Let $a_i$ be probability of loss, $1 - a_i$ be probability of win.

\[
1 - a_i = \begin{cases} 
\frac{1 - \rho^i}{1 - \rho^m} & \text{if } \rho \neq 1, \\
\frac{i}{m} & \text{if } \rho = 1.
\end{cases}
\]

\[
\rho = \frac{1 - p}{p}
\]

Detailed analysis in B&T Example 7.11.
Expected Time to Absorption

\[ \mu_i = \mathbb{E}[\text{number of transitions until absorption, starting from } i] \]

**Theorem:**

The expected times \( \mu_i \) to absorption, starting from state \( i \) are the unique solution of the equations

\[
\mu_i = 0, \quad \text{for all recurrent states } i,
\]

\[
\mu_i = 1 + \sum_{j=1}^{m} p_{ij} \mu_j, \quad \text{for all transient states } i.
\]

From each initial state \( i \), what is the expected number of steps to absorption?

\[
\mu_1 = 0 \quad \mu_2 = \mu_3 = \frac{10}{3} \quad \mu_4 = 0
\]

*Matches mean of a geometric distribution with success probability of 0.3.*
Absorption Probabilities and Expected Time to Absorption

Example 6.12. (Gambler’s Ruin)

The absorption probabilities \( p_i \) graph in which states 4 and 5 have been lumped into the absorbing state starting from one of the transient states. For the purposes of this problem, absorption in state 6 occurs with probability \( p_i \).

### First phase:
Find probability of reaching (being absorbed by) each recurrent class.

### Second phase:
Find equilibrium distribution of states within each of these recurrent classes.