Introduction to Markov Chains

A Motivating Problem: st-Connectivity

- **Problem:** Given a graph G and vertices s, t in G, is there a walk from s to t?
- **Deterministic Algorithms:** Depth-First Search (DFS), Breadth-First Search (BFS).
  - Time Complexity: $O(|V| + |E|)$
  - Space Complexity: $O(|V|)$
- **Question:** Can we solve this using less space?

Random Walk Algorithm

```
import random
def random_walk_connectivity(G, s, t, T):
    u = s
    for i in range(T):
        if u == t:
            return True
        if random.getrandbits(1): # True with probability 1/2
            u = random.choice(G.neighbors(u))
    return False
```

- Are there situations where we can prove this is a **good** algorithm?
- How big should T be?

Random Walk Algorithm

- What are the time and space complexities in the worst case?

Math prerequisites

Conditional probability I

- We write $\Pr(A \mid B)$ for the probability of an event A given that another event B has already occurred.
- $\Pr(A \mid B)$ is pronounced "the probability of A given B"
- **Formula:** provided $\Pr(B) > 0$, we define

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- $\Pr(A \cap B)$ is the **joint probability** that both A and B occur.
- **Bayes rule:** $\Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$

Conditional probability II

- **Example 1:** Rolling a fair 6-sided die.
  - Let A be the event "rolling a 2": $\Pr(A) = 1/6$.
  - Let B be the event "rolling an even number": $\Pr(B) = 3/6 = 1/2$.
  - $\Pr(A \mid B) = \frac{\Pr(\text{rolling a 2 and an even number})}{\Pr(\text{rolling an even number})} = \frac{1/6}{1/2} = 1/3$.
- **Example 2:** Drawing cards from a standard deck.
  - What is the probability of drawing an Ace (A), given that the card drawn is a Heart (B)?
  - $\Pr(A \mid B) = \frac{\Pr(\text{Ace of Hearts})}{\Pr(\text{Heart})} = \frac{1/52}{13/52} = \frac{1}{13}$.

Stochastic processes

- A sequence of random variables indexed by time is a **stochastic process**.
- We often use discrete time $t \geq 0$ and so have a sequence $(X_0, X_1, X_2, \ldots)$.
- Let Ω be the **state space** where the X_t live
- **Example 3:** Repeatedly flipping a coin.
  - X_t is the outcome of the t-th coin flip.
- **Example 4:** A random walk on a graph.
  - X_t is the vertex the walker is on after t steps.

Markov Chains and Convergence

Markov Chains

- A random walk is a special case of a **Markov chain**
- Defining characteristics:
  - you take random steps, and
  - the distribution of the step you take depends only on the present state, not the history of how you got there.
- **Definition:** A stochastic process $(X_0, X_1, \ldots)$ is a **Markov chain** if:

$$\Pr\left(X_{t+1} = x_{t+1} \mid X_0 = x_0, \ldots, X_t = x_t\right) = \Pr\left(X_{t+1} = x_{t+1} \mid X_t = x_t\right)$$

- The distribution of the next step X_{t+1} depends only on the present state x_t, not the history.

Notation

- What is the probability we move to state $y \in \Omega$?
- It is allowed to depend on the present state x, but nothing else
- So we can consider function $P : \Omega \times \Omega \to [0, 1]$ where

$$P(x, y) = \Pr(X_{t+1} = y \mid X_t = x)$$

- It is conventional to put this data into a matrix and write $P_{xy} = P(x, y)$ or $P_{ij} = P(i, j)$.
- Get ready for linear algebra!

Transition Matrices

- The matrix P is a **right stochastic matrix**:
  - $P_{ij} \geq 0$ for all i, j.
  - Each row sums to 1: for every i, $\sum_j P_{ij} = 1$.
- **Example 3 again: flipping a coin**
  - State space $\Omega = \{H, T\}$
  - $P_{ij} = 1/2$ no matter what i and j are
- **Example 4 again: random walk on a graph G**
  - State space $\Omega = V(G)$
  - $P_{ij} = 1/d_i$ if j is a neighbor of i, 0 otherwise.

Matrix multiplication

- Suppose the present state is i. What is the distribution of the next state?
- This can be found by matrix multiplication. Let $\mu \in [0,1]^\Omega$ have a one in position i and zeros elsewhere.
- Then the distribution on Ω representing the state after one step is μP.
- **Note on transposes:** For Markov chains, we represent probability distributions as **row vectors** μ and multiply by matrices on the right: μP.
- This is likely the **dual** of everything you've ever done in linear algebra so far, where vectors were columns and multiplied on the left by matrices.

Example multiplications

$$\Omega = \{1, 2, 3\} \text{ and } P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Stationary Distributions

- A vector $\pi \in [0,1]^{\Omega}$ is a **stationary distribution** of P if it is a probability distribution (nonnegative entries which sum to 1) and:

$$\pi P = \pi$$

- That is, π is a **left eigenvector** of P with eigenvalue 1.
- If the present step is distributed according to π, all subsequent steps are also distributed according to π.
- **Goal:** Under what conditions does the distribution on the state converge to π as $t \to \infty$, regardless of the starting state?

Linear Algebra

- To find the eigenvalues of a matrix P, we first solve the **characteristic equation** $\det(\lambda I - P) = 0$ for λ.
- For stationary distributions we know the eigenvalue should be 1 so we can skip this step.
- Then to find the **left** eigenvector we solve $\pi P = \pi$ for π.
- This can be annoying, so it's a good idea to **guess** a solution if P looks "nice"

Examples I

$\Omega = \{1, 2\}$ and $P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$

Examples II

$$\Omega = \{a, b\} \text{ and } P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\Omega = \{x, y\} \text{ and } P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Examples III

$$\Omega = \{1, 2, 3\} \text{ and } P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Obstructions to Convergence

1. **Reducibility:** Can we reach every state from every state?
   - An **irreducible** chain has $k > 0$ such that $P^k_{ij} > 0$ for all i, j.
2. **Periodicity:** Can we only loop back to a state in multiples of $k > 1$ steps?
   - An **aperiodic** chain has period 1 for all states.

- A finite chain that is both irreducible and aperiodic is called **ergodic**.

Examples IV

Fundamental Theorem of Markov Chains

Let P be a finite, irreducible, aperiodic (ergodic) Markov chain. Then:

1. P has a **unique** stationary distribution π.
2. For **any** starting distribution μ, the distribution at time t converges to π:

$$\mu P^t \to \pi \text{ as } t \to \infty.$$

Reversible Markov Chains

- Finding stationary distributions directly can be hard.
- A Markov chain is **reversible** if there exists a distribution π satisfying the **detailed balance** condition:

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \text{for all } i, j$$

- **Fact:** If π satisfies detailed balance, it is a stationary distribution.
- *Intuition:* A reversible chain run backwards in time looks identical to the forward chain (when started in π).
- Many natural Markov chains are ergodic and reversible, and then it's much easier to find the stationary distribution

Lazy random walk I

- The **lazy random walk** on a simple graph G is defined as follows.
- At vertex i:
  - With probability $1/2$, stay at i.
  - With probability $1/2$, move to a uniform random neighbor of i.
- Let A be the adjacency matrix: $A_{ij} = 1$ if i and j are neighbors, 0 otherwise.
- $P_{ii} = 1/2$, $P_{ij} = 1/(2d_i)$ if j is a neighbor of i, and $P_{ij} = 0$ otherwise.

Lazy random walk II

- If G is d-regular, the transition matrix is $\frac{1}{2}I + \frac{1}{2d}A$.
- We can check detailed balance for π the uniform distribution.

Markov chains and algorithms

- Some computational problems are nicely solved by being able to choose an object at random from some probability distribution.
- For st-connectivity, suppose that we could sample a vertex of the graph **which is in the same component as s** uniformly at random.
- Some standard math (the coupon collector problem) says that after $n \log n$ samples we will have seen every vertex in the component of s at least once.
- This isn't long to wait!
- A random walk may give us this power, depending on the graph.
- The efficiency of the algorithm depends on how fast the chain converges to its stationary distribution.