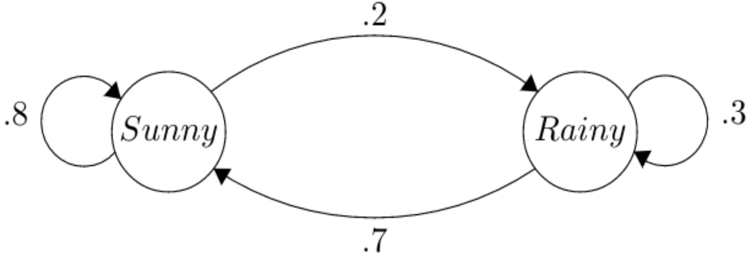
Markov Chains



Introduction

- Markov chains are a tool for studying *stochastic processes* that evolve over time.
- A sequence (X_t) of random variables has the *Markov property* if for all t, the probability distribution for X_{t+1} depends only on X_t, but not on $X_{t-1}, \ldots, X_0$. Also called the Memorylessness Property.
- Markov chains are often modeled using *directed* graphs. The states are represented as nodes, and an edge from state i to state j is weighted with probability $p_{i,j}$.
- Markov chains can be written in matrix form (using the adjacency matrix).
- The state distribution at time t is $q_t = q_0 \cdot P^t$.

Random Walk

- Let $G = (V, E)$ be a directed graph, and let $\omega : E \to [0, 1]$ be a weight function so that $\sum_{v:(u,v)\in E} \omega(u, v) = 1$ for all nodes u.
- Let $u \in V$ be the *starting node*.
- A *weighted random walk on G starting at u* is the following discrete Markov chain in discrete time.
  - Beginning with $X_0 = u$, in every step t, the node X_{t+1} is chosen according to the weights $\omega(X_t, v)$, where v are the neighbors of X_t.
- The *sojourn time* T_i of state i is the time the process stays in state i.
- The *hitting time* $T_{i,j}$ is the random variable counting the number of steps until visiting j the first time when starting from state i. Given by the formula: $h_{i,j} = 1 + \sum_{k \neq j} (p_{i,k} \cdot h_{k,j})$

Stationary Distribution & Ergodicity

- A distribution π over the states is called *stationary distribution* of the Markov chain with transition matrix P if $\pi = \pi \cdot P$.
- A Markov chain is *irreducible* if all states are reachable from all other states. That is, if for all $i, j \in S$ there is some $t \in \mathbb{N}$, such that $p_{i,j}^{(t)} > 0$.
- A state i has period k if any return to state i must occur in multiples of k time steps. A state with period $k = 1$ is called aperiodic. If all states of a Markov chain are aperiodic, the entire chain is aperiodic.
- If a finite Markov chain is irreducible and aperiodic, then it is called *ergodic*.
- If a Markov chain is ergodic it holds that $\lim_{t \to \infty} q_t = \pi$ where π is the unique stationary distribution of the chain.

Google's PageRank Algorithm

- Google's PageRank algorithm is based on a Markov chain obtained from a variant of a random walk of the "Web Graph".
- Google's idea is to model a *random surfer* who follows hyperlinks in the web graph, i.e., performs a simple random walk. After sufficiently many steps, the websites can be ranked by how many times they were visited.
- Let W be a random surfer matrix, and let $\alpha \in (0, 1)$ be a constant. Denote further by R the matrix in which all entries are $1/n$. The following matrix M is called the *Google Matrix*: $M = \alpha \cdot W + (1 - \alpha) \cdot R$
- In every step, with probability $1 - \alpha$, the random surfer "gets bored" by the current website and surfs to a new random site.

Simple Random Walks

- Let G be a graph with m edges. The stationary distribution π of any simple random walk on G is $\pi_u = \frac{\deg(u)}{2m}$
- For simple random walks, hitting time $h_{u,u}$ is $\frac{2m}{\deg(u)}$
- The *cover time* $cov(v)$ is the expected number of steps until all nodes in G were visited at least once, starting at v.