Discrete Time, Finite, Markov Chain

- A *stochastic process* $\mathbf{X} = \{X(t) : t \in T\}$ is a collection of random variables.
- $X(t) =$ the *state* of the process at time t.
- $\mathbf{X}$ is a *Discrete (finite) space* if for all t, X_t assumes values from a countably infinite (finite) set.
- If T is a countably infinite set we say that X is a *discrete time* process.

Definition

A discrete time stochastic process $X_0, X_1, X_2, \ldots$ is a *Markov chain* if

$$\begin{aligned}\Pr(X_t = a_t | X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_0 = a_0) \\ = \Pr(X_t = a_t | X_{t-1} = a_{t-1}) = P_{a_{t-1}, a_t}.\end{aligned}$$

Transition probability: $P_{i,j} = \Pr(X_t = j \mid X_{t-1} = i)$
Transition matrix:

$$\mathbf{P} = \begin{pmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}.$$

Probability distribution for a given time t:

$$\bar{p}(t) = (p_0(t), p_1(t), p_2(t), \ldots)$$

$$p_i(t) = \sum_{j \geq 0} p_j(t-1) P_{j,i},$$

$$\bar{p}(t) = \bar{p}(t-1)\mathbf{P}.$$

For any $n \geq 0$ we define the n-step transition probability

$$P^n_{i,j} = \Pr(X_{t+n} = j \mid X_t = i)$$

Conditioning on the first transition from i we have

$$P_{i,j}^{n} = \sum_{k \geq 0} P_{i,k} P_{k,j}^{n-1}. \tag{1}$$

Let $\mathbf{P^{(n)}}$ be the matrix whose entries are the n-step transition probabilities, so that the entry in the ith row and jth column is $P_{i,j}^{n}$. Then we have

$$\mathbf{P^{(n)}} = \mathbf{P} \cdot \mathbf{P^{(n-1)}},$$

and by induction on n

$$\mathbf{P^{(n)}} = \mathbf{P^{n}}.$$

Thus, for any $t \geq 0$ and $n \geq 1$,

$$\bar{p}(t+n) = \bar{p}(t)\mathbf{P^{n}}.$$

Example

Consider a system with a total of m balls in two containers.
We start with all balls in the first container.
At each step we choose a ball uniformly at random from all the balls and with probability $1/2$ move it to the other container.
Let X_i denote the number of balls in the first container at time i.
$X_0, X_1, X_2, ...$ defines a Markov chain with the following transition matrix:

$$p_{i,j} = \begin{cases} \frac{m-i}{2m} & j = i+1 \\ \frac{i}{2m} & j = i-1 \\ \frac{1}{2} & j = i \\ 0 & |i-j| > 1 \end{cases}$$

Randomized 2-SAT Algorithm

Given a formula with up to two variables per clause, find a Boolean assignment that satisfies all clauses.

Algorithm:

1. Start with an arbitrary assignment.
2. **Repeat** till all clauses are satisfied:
    1. Pick an unsatisfied clause.
    2. If the clause has one variable change the value of that variable.
    3. If the clause has two variable choose one uniformly at random and change its value.

What the is the expected run-time of this algorithm?
W.l.o.g. assume that all clause have two variables.
Assume that the formula has a satisfying assignment. Pick one such assignment.
Let X_i be the number of variables with the correct assignment according to that assignment after iteration i of the algorithm.
Let n be the number of variable.

$$Pr(X_i = 1 \mid X_{i-1} = 0) = 1$$

For $1 \leq t \leq n-1$,

$$Prob(X_i = t+1 \mid X_{i-1} = t) \geq 1/2$$

$$Prob(X_i = t-1 \mid X_{i-1} = t) \leq 1/2$$

Assume

$$Pr(X_i = 1 \mid X_{i-1} = 0) = 1$$

for $1 \leq t \leq n-1$,

$$Prob(X_i = t+1 \mid X_{i-1} = t) = 1/2$$

$$Prob(X_i = t-1 \mid X_{i-1} = t) = 1/2$$

Let D_t be the expected number of steps to termination when we have t correct variable assignments.
$D_0 = 1 + D_1$.
$D_t = 1 + \frac{1}{2}D_{t+1} + \frac{1}{2}D_{t-1}$
We "guess"

$$D_t = n^2 - t^2$$

.

$$D_n = 0.$$

$$D_t = 1 + \frac{1}{2}(n^2 - (t-1)^2) + \frac{1}{2}(n^2 - (t+1)^2) =$$

$$1 + \frac{1}{2}(2n^2 + 2t^2 + 2) = n^2 - t^2$$

$$D_0 = 1 + D_{n-1} = 1 + n^2 - 1 = n^2.$$

Theorem

Assuming that the formula has a satisfying assignment the expected run-time to find that assignment is $O(n^2)$.

Theorem

There is a one-sides error randomized algorithm for the 2-SAT problem that terminates in $O(n^2 \log n)$ time, with high probability returns an assignment when the formula is satisfiable, and always returns "UNSATISFIABLE" when no assignment exists.

Proof.

The probability that the algorithm does not find an assignment when exists in $2n^2$ steps is bounded by $\frac{1}{2}$. □

Classification of States

Definition

State i is *accessible* from state j if for some integer $n \geq 0$, $P^n_{i,j} > 0$. If two states i and j are accessible from each other we say that they *communicate*, and we write $i \leftrightarrow j$.

In the graph representation $i \leftrightarrow j$ if and only if there are directed paths connecting i to j and j to i.

The communicating relation defines an equivalence relation. That is, the relation is

1. Reflexive: for any state i, $i \leftrightarrow i$;
2. Symmetric: if $i \leftrightarrow j$ then $j \leftrightarrow i$; and
3. Transitive: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

Definition

A Markov chain is *irreducible* if all states belong to one communicating class.

Lemma

A finite Markov chain is irreducible if and only if its graph representation is a strongly connected graph.

$r_{i,j}^t =$ the probability that starting at state i the first transition to state j occurred at time t, that is,

$$r_{i,j}^t = \Pr(X_t = j \text{ and for } 1 \leq s \leq t-1,\ X_s \neq j \mid X_0 = i).$$

Definition

A state is *recurrent* if $\sum_{t \geq 1} r_{i,i}^t = 1$, and it is *transient* if $\sum_{t \geq 1} r_{i,i}^t < 1$. A Markov chain is recurrent if every state in the chain is recurrent.

The expected time to return to state i when starting at state j:

$$h_{j,i} = \sum_{t \geq 1} t \cdot r_{j,i}^t$$

Definition

A recurrent state i is *positive recurrent* if $h_{i,i} < \infty$. Otherwise, it is *null recurrent*.

Example - null recurrent states

States are the positive numbers.

$$P_{i,j} = \begin{cases} \frac{i}{i+1} & j = i+1 \\ 1 - \frac{i}{i+1} & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

The probability of not having returned to state 1 within the first t steps is

$$\prod_{j=1}^{t} \frac{j}{j+1} = \frac{1}{t+1}.$$

The probability of never returning to state 1 from state 1 is 0, and state 1 is recurrent.

$$r_{1,1}^{t} = \frac{1}{t(t+1)}.$$

$$h_{1,1} = \sum_{t=1}^{\infty} t \cdot r_{1,1}^{t} = \sum_{t=1}^{\infty} \frac{1}{t+1} = \infty$$

State 1 is null recurrent.

Lemma

In a finite Markov chain,

1. *At least one state is recurrent;*
2. *All recurrent states are positive recurrent.*

Definition

A state j in a discrete time Markov chain is *periodic* if there exists an integer $\Delta > 1$ such that $\Pr(X_{t+s} = j \mid X_t = j) = 0$ unless s is divisible by Δ. A discrete time Markov chain is *periodic* if any state in the chain is periodic. A state or chain that is not periodic is *aperiodic*.

Definition

An aperiodic, positive recurrent state is an *ergodic* state. A Markov chain is *ergodic* if all its states are ergodic.

Corollary

Any finite, irreducible, and aperiodic Markov chain is an ergodic chain.

Example: The Gambler's Ruin

- Consider a sequence of independent, two players, fair gambling games.
- In each round a player wins a dollar with probability $1/2$ or loses a dollar with probability $1/2$.
- $W^t =$ the number of dollars won by player 1 up to (including) step t.
- If player 1 has lost money, this number is negative.
- $W^0 = 0$. For any t, $E[W^t] = 0$.
- Player 1 must ends the game if he loses ℓ_1 dollars ($W^t = -\ell_1$); player 2 must terminates when she loses ℓ_2 dollars ($W^t = \ell_2$).
- Let q be the probability that the game ends with player 1 wining ℓ_2 dollar.
- If $\ell_2 = \ell_1$, then by symmetry $q = 1/2$. What is q when $\ell_2 \neq \ell_1$?

$-\ell_1$ and ℓ_2 are recurrent states. All other states are transient. Let

P_i^t be the probability that after t steps the chain is at state i.

For $-\ell_1 < i < \ell_2$, $\lim_{t\to\infty} P_i^t = 0$.

$$\lim_{t\to\infty} P_{\ell_2}^t = q.$$

$$\lim_{t\to\infty} P_{\ell_1}^t = 1 - q.$$

$$E[W^t] = \sum_{i=-\ell_1}^{\ell_2} iP_i^t = 0$$

$$\lim_{t\to\infty} \mathbf{E}[W^t] = \ell_2 q - \ell_1(1-q) = 0.$$

$$q = \frac{\ell_1}{\ell_1+\ell_2}.$$

Stationary Distributions

$$\bar{p}(t+1) = \bar{p}(t)\mathbf{P}$$

Definition

A *stationary distribution* (also called an *equilibrium distribution*) of a Markov chain is a probability distribution $\bar{\pi}$ such that

$$\bar{\pi} = \bar{\pi}\mathbf{P}.$$

Theorem

Any finite, irreducible, and aperiodic (ergodic) Markov chain has the following properties:

1. *The chain has a unique stationary distribution* $\bar{\pi} = (\pi_0, \pi_1, \ldots, \pi_n)$;
2. *For all* j *and* i, *the limit* $\lim_{t\to\infty} P_{j,i}^t$ *exists and it is independent of* j;
3. $\pi_i = \lim_{t\to\infty} P_{j,i}^t = \frac{1}{h_{i,i}}$.

- For any distribution vector $\bar{p}$

$$\pi = \lim_{t \to \infty} \bar{p}\mathbf{P}^t.$$

- $$\frac{1}{\pi_i} = h_{i,i} = \sum_{t=1}^{\infty} t \cdot r_{i,i}^t$$

Proof

We use:

Lemma

For any irreducible, ergodic Markov chain, and for any state i, *the limit* $\lim_{t\to\infty} P_{i,i}^t$ *exists, and*

$$\lim_{t\to\infty} P_{i,i}^t = \frac{1}{h_{i,i}}.$$

Using the fact that $\lim_{t\to\infty} P^t_{i,i}$ exists, we now show that for any j and i $\lim_{t\to\infty} P^t_{j,i} = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}$.

For $j \neq i$ we have $P^t_{j,i} = \sum_{k=1}^{t} \cdot r^k_{j,i} P^{t-k}_{i,i}$.

For $t \geq t_1$, $\sum_{k=1}^{t_1} r^k_{j,i} P^{t-k}_{i,i} \leq \sum_{k=1}^{t} r^k_{j,i} P^{t-k}_{i,i} = P^t_{j,i}$.

Since the chain is irreducible $\sum_{t=1}^{\infty} r^t_{j,i} = 1$ For any $\epsilon > 0$ there exists (a finite) $t_1 = t_1(\epsilon)$ such that $\sum_{t=1}^{t_1} r^t_{j,i} \geq 1 - \epsilon$.

$$\begin{aligned}
\lim_{t\to\infty} P^t_{j,i} &\geq \lim_{t\to\infty} \sum_{k=1}^{t_1} r^k_{j,i} P^{t-k}_{i,i} = \sum_{k=1}^{t_1} r^k_{j,i} \lim_{t\to\infty} P^t_{i,i} \\
&= \lim_{t\to\infty} P^t_{i,i} \sum_{k=1}^{t_1} r^k_{j,i} \geq (1-\epsilon) \lim_{t\to\infty} P^t_{i,i}.
\end{aligned}$$

Similarly,

$$P_{j,i}^{t} = \sum_{k=1}^{t} r_{j,i}^{k} P_{i,i}^{t-k} \leq \sum_{k=1}^{t_1} r_{j,i}^{k} P_{i,i}^{t-k} + \epsilon,$$

$$\begin{aligned}
\lim_{t\to\infty} P_{j,i}^{t} &\leq \lim_{t\to\infty} \Big(\sum_{k=1}^{t_1} r_{j,i}^{k} P_{i,i}^{t-k} + \epsilon\Big) \\
&= \sum_{k=1}^{t_1} r_{j,i}^{k} \lim_{t\to\infty} P_{i,i}^{t-k} + \epsilon \\
&\leq \lim_{t\to\infty} P_{i,i}^{t} + \epsilon.
\end{aligned}$$

For any pair i and j

$$\lim_{t\to\infty} P_{j,i}^{t} = \lim_{t\to\infty} P_{i,i}^{t} = \frac{1}{h_{i,i}}.$$

Let

$$\pi_i = \lim_{t\to\infty} P^t_{j,i} = \frac{1}{h_{i,i}}.$$

We show that $\bar{\pi} = (\pi_0, \pi_1, \dots)$ forms a stationary distribution. For every $t \geq 0$, $P^t_{i,i} \geq 0$, and thus $\pi_i \geq 0$. For any $t \geq 0$, $\sum_{i=0}^n P^t_{j,i} = 1$, and thus

$$\lim_{t\to\infty} \sum_{i=0}^{n} P^t_{j,i} = \sum_{i=0}^{n} \lim_{t\to\infty} P^t_{j,i} = \sum_{i=0}^{n} \pi_i = 1,$$

and π is a proper distribution. Now,

$$P^{t+1}_{j,i} = \sum_{k=0}^{n} P^t_{j,k} P_{k,i}.$$

Letting $t \to \infty$ we have

$$\pi_i = \sum_{k=0}^{n} \pi_k P_{k,i},$$

proving that $\bar{\pi}$ is a stationary distribution.

Suppose that there was another stationary distribution $\bar{\phi}$.

$$\phi_i = \sum_{k=0}^{n} \phi_k P_{k,i}^t,$$

and taking the limit as $t \to \infty$ we have

$$\phi_i = \sum_{k=0}^{n} \phi_k \pi_i = \pi_i \sum_{k=0}^{n} \phi_k.$$

Since $\sum_{k=0}^{n} \phi_k = 1$, we have $\phi_i = \pi_i$ for all i, or $\bar{\phi} = \bar{\pi}$.

Computing the Stationary Distribution

1. Solve the system of linear equations $\bar{\pi}\mathbf{P} = \bar{\pi}$.

2. Solving equilibrium equations.

Theorem

Let S be a set of states of a finite, irreducible, aperiodic Markov chain. In the stationary distribution, the probability that the chain leaves the set S equals the probability that it enters S.

Proof.

For any state i:

$$\sum_{j=0}^{n-1} \pi_j P_{j,i} = \pi_i = \pi_i \sum_{j=0}^{n-1} P_{i,j}$$

$$\sum_{j \neq i} \pi_j P_{j,i} = \sum_{j \neq i} \pi_i P_{i,j}.$$

Theorem

Consider a finite, irreducible, and ergodic Markov chain on n states with transition matrix $\mathbf{P}$. If there are non-negative numbers $\bar{\pi} = (\pi_0, \ldots, \pi_{n-1})$ such that $\sum_{i=0}^{n-1} \pi_i = 1$, and for any pair of states i, j,

$$\pi_i P_{i,j} = \pi_j P_{j,i},$$

then $\bar{\pi}$ is the stationary distribution corresponding to $\mathbf{P}$.

Proof.

$$\sum_{i=0}^{n-1} \pi_i P_{i,j} = \sum_{i=0}^{n-1} \pi_j P_{j,i} = \pi_j.$$

Thus $\bar{\pi}$ satisfies $\bar{\pi} = \bar{\pi}\mathbf{P}$, and $\sum_{i=0}^{n-1} \pi_i = 1$, and $\bar{\pi}$ must be the unique stationary distribution of the Markov chain. $\square$

Theorem

Any irreducible aperiodic Markov chain belongs to one of the following two categories:

1. *The chain is ergodic. For any pairs of states* i *and* j*, the limit* $\lim_{t\to\infty} P^t_{j,i}$ *exists and is independent of* j*. The chain has a unique stationary distribution* $\pi_i = \lim_{t\to\infty} P^t_{j,i} > 0$*.*
   or
2. *No state is positive recurrent. For all* i *and* j*,* $\lim_{t\to\infty} P^t_{j,i} = 0$*, and the chain has no stationary distribution.*

Example: A Simple Queue

Discrete time queue.

At each time step, exactly one of the following occurs:

- If the queue has fewer than n customers, then with probability λ a new customer joins the queue.
- If the queue is not empty, then with probability μ the head of the line is served and leaves the queue.
- With the remaining probability the queue is unchanged.

$X_t =$ the number of customers in the queue at time t.

$$\begin{aligned} P_{i,i+1} &= \lambda \text{ if } i < n \\ P_{i,i-1} &= \mu \text{ if } i > 0 \\ P_{i,i} &= \begin{cases} 1-\lambda & \text{if } i = 0 \\ 1-\lambda-\mu & \text{if } 1 \le i \le n-1 \\ 1-\mu & \text{if } i = n. \end{cases} \end{aligned}$$

The Markov chain is irreducible, finite, and aperiodic, so it has a unique stationary distribution $\bar{\pi}$.

We use $\bar{\pi} = \bar{\pi}\mathbf{P}$ to write

$$\begin{aligned}
\pi_0 &= (1-\lambda)\pi_0 + \mu\pi_1, \\
\pi_i &= \lambda\pi_{i-1} + (1-\lambda-\mu)\pi_i + \mu\pi_{i+1},\ 1 \le i \le n-1, \\
\pi_n &= \lambda\pi_{n-1} + (1-\mu)\pi_n.
\end{aligned}$$

$$\pi_i = \pi_0 \left(\frac{\lambda}{\mu}\right)^i$$

Adding the requirement $\sum_{i=0}^{n} \pi_i = 1$, we have

$$\sum_{i=0}^{n} \pi_i = \sum_{i=0}^{n} \pi_0 \left(\frac{\lambda}{\mu}\right)^i = 1,$$

$$\pi_0 = \frac{1}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i}.$$

For all $0 \leq i \leq n$,

$$\pi_i = \frac{\left(\frac{\lambda}{\mu}\right)^i}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i}. \tag{2}$$

Use cut sets to compute the stationary probability:
For any i, the transitions $i \rightarrow i+1$ and $i+1 \rightarrow i$ are a cut-set.

$$\lambda \pi_i = \mu \pi_{i+1}.$$

$$\pi_i = \pi_0 \left(\frac{\lambda}{\mu} \right)^i.$$

Removing the limit on n, the Markov chain is no longer finite. The Markov chain has a countably infinite state space. It has a stationary distribution if and only if the following set of linear equations has a solution with all $\pi_i > 0$:

$$
\begin{aligned}
\pi_0 &= (1-\lambda)\pi_0 + \mu\pi_1 \\
\pi_i &= \lambda\pi_{i-1} + (1-\lambda-\mu)\pi_i + \mu\pi_{i+1}, \, , i \geq 1.
\end{aligned}
$$

$$
\pi_i = \frac{\left(\frac{\lambda}{\mu}\right)^i}{\sum_{i=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^i} = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right)
$$

is a solution of the above system of equations.
All of the π_i are greater than 0 if and only if $\lambda < \mu$.
If $\lambda > \mu$, no stationary distribution, each state in the Markov chain is transient.
If $\lambda = \mu$ there is no stationary distribution, and the queue length will become arbitrarily long, but now the states are null recurrent.

Random Walks on Undirected Graph

Let $G = (V, E)$ be a finite, undirected, and connected graph.

Definition

A *random walk* on G is a Markov chain defined by the movement of a particle between vertices of G. In this process, the place of the particle at a given time step is the state of the system. If the particle is at vertex i, and i has $d(i)$ outgoing edges, then the probability that the particle follows the edge (i, j) and moves to a neighbor j is $1/d(i)$.

Lemma

A random walk on an undirected graph G is aperiodic if and only if G is not bipartite.

Proof.

If the graph is bipartite then the random walk is periodic, with a period $d = 2$.
If the graph is not bipartite, then it has an odd cycle, and by traversing that cycle we have an odd length path from any vertex to itself. $\square$

Theorem

A random walk on G converges to a stationary distribution π, where

$$\pi_v = \frac{d(v)}{2|E|}.$$

Proof.

Since $\sum_{v \in V} d(v) = 2|E|$,

$$\sum_{v \in V} \pi_v = \sum_{v \in V} \frac{d(v)}{2|E|} = 1,$$

and π_v is a proper distribution over $v \in V$.

Let $N(v)$ be the set of neighbors of v. The relation $\bar{\pi} = \bar{\pi}\mathbf{P}$ gives

$$\pi_v = \sum_{u \in N(v)} \frac{d(u)}{2|E|} \frac{1}{d(u)} = \frac{d(v)}{2|E|}$$

$h_{v,u}$ denotes the expected number of steps to reach u from v.

Corollary

For any vertex u in G,

$$h_{u,u} = \frac{2|E|}{d(u)}.$$

Lemma

If $(u, v) \in E$, *then* $h_{v,u} < 2|E|$.

Proof.

Let $N(u)$ be the set of neighbors of vertex u in G. We compute $h_{u,u}$ in two different ways.

$$\frac{2|E|}{d(u)} = h_{u,u} = \frac{1}{d(u)} \sum_{w \in N(u)} (1 + h_{w,u}).$$

Hence

$$2|E| = \sum_{w \in N(u)} (1 + h_{w,u}),$$

and we conclude that $h_{v,u} < 2|E|$. □

Definition

The *cover time* of a graph G is the maximum over all vertices of the expected time to visit all nodes of the graph starting the random walk from that vertex.

Lemma

The cover time of $G = (V, E)$ *is bounded above by* $4|V| \cdot |E|$.

Proof.

Choose a spanning tree on G, and an Eulerian cycle on the spanning tree.
Let $v_0, v_1, \ldots, v_{2|V|-2} = v_0$ be the sequence of vertices in the cycle.

$$\sum_{i=0}^{2|V|-3} h_{v_i, v_{i+1}} + h_{v_{2|V|-2}, v_1} < (2|V| - 2)2|E| < 4|V| \cdot |E|,$$

□

Application: An $s - t$ Connectivity Algorithm

Given an undirected graph $G = (V, E)$, and two vertices s and t in G.
Let $n = |V|$ and $m = |E|$.
We want to determine if there is a path connecting s and t.
Easily done in $O(m)$ time and $\Omega(n)$ space.
$s - t$ **Connectivity Algorithm**

- Start a random walk from s.
- If the walk reaches t within $4n^3$ steps, return that there is a path. Otherwise, return that there is no path.

Theorem

The algorithm returns the correct answer with probability $1/2$*, and it only errs by saying that there is no path from* s *to* t *when there is such a path.*

Proof.

If there is no path, the algorithm returns the correct answer.
If there is a path, the expected time to reach t from s, is bounded by $4nm < 2n^3$.
By Markov's inequality, the probability that a walk takes more than $4n^3$ steps to reach s from t is at most $1/2$. □

The algorithm must keep track of its current position, which takes $O(\log n)$ bits, and the number of steps taken in the random walk, which also takes only $O(\log n)$ bits (since we count up to only $4n^3$).