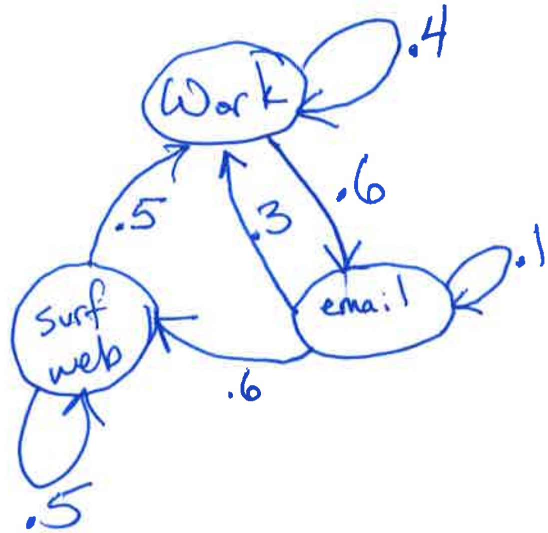
# Markov chains!

Example: (copied from lecture notes of Ryan O'Donnell at CMU)

A day at the office:




N states $\{1, 2, \ldots, N\}$

Directed graph possibly with self-loops

Each edge has a weight corresponding to a probability. So they are non-negative.

For every vertex/state, the sum of the edge weights (i.e., probabilities) of outgoing edges is $= 1$.

We start at a state & at each time step we follow an outgoing edge based on the probabilities of the outgoing edges.

Can model by a $N \times N$ matrix $P$

where $P(i,j)$ = weight of edge $i \to j$

= $\Pr$(going from $i$ to $j$)

For earlier example:

let 1 = Work, 2 = email, 3 = surf web,

$$P = \begin{bmatrix} .4 & .6 & 0 \\ .3 & .1 & .6 \\ .5 & 0 & .5 \end{bmatrix}$$

Time $t = 0, 1, 2, \ldots$

Let $X_t$ = state at time $t$

$X_0$ = initial state

for $t \geq 1$, $X_t$ is a random variable.

$$\Pr(X_1 = 2 \mid X_0 = 1) = .6$$

$$\Pr(X_1 = 3 \mid X_0 = 3) = .5$$

In general, $\Pr(X_1 = j \mid X_0 = i) = P(i,j)$

Moreover, for all $t \geq 1$,

$$\Pr(X_t = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i_{t-1})$$
$$= \Pr(X_t = j \mid X_{t-1} = i_{t-1})$$
$$= P(i_{t-1}, j)$$

Thus the process is "memoryless."
The probability of going to state $j$ at time $t$ only depends on the current state $X_{t-1}$.
The previous states don't matter.

Known as "Markov Property."

What is $\Pr(X_2 = 2 \mid X_0 = 1)$?
Try all possibilities for state at time 1.

$$\Pr(X_2 = 2 \mid X_0 = 1)$$
$$= \Pr(X_2 = 2 \mid X_1 = 2)\Pr(X_1 = 2 \mid X_0 = 1)$$
$$+ \Pr(X_2 = 2 \mid X_1 = 3)\Pr(X_1 = 3 \mid X_0 = 1)$$
$$+ \Pr(X_2 = 2 \mid X_1 = 1)\Pr(X_1 = 1 \mid X_0 = 1)$$
$$= (.1)(.6) + 0 + (.6)(.4)$$
$$= .3$$

For general Markov chains,

$$\Pr(X_{t+2}=j \mid X_t=i)$$
$$=\sum_{k=1}^{N}\Pr(X_{t+2}=j \mid X_{t+1}=k)\Pr(X_{t+1}=k \mid X_t=i)$$
$$=\sum_k P(k,j)P(i,k)$$
$$=P^2(i,j)$$

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

Suppose $X_0$ is randomly distributed according to some distribution $\mu_0$.

What is the distribution of $X_1$?

$$\Pr(X_1=j)=\sum_{i=1}^{N}\Pr(X_0=i)\Pr(X_1=j \mid X_0=i)$$
$$=\sum_i \mu(i)P(i,j) \qquad [\ \mu\ ]\left[\ \overset{j}{P}\ \right]$$
$$=(\mu P)(j)$$
$$=\mu_1(j)$$

If $X_0 \sim \mu_0$ then $X_t \sim \mu_t$ where

$$\mu_t = \mu_0 P^t$$

Take the earlier simple example on 3 states.

$$P = \begin{bmatrix} .4 & .6 & 0 \\ .3 & .1 & .6 \\ .5 & 0 & .5 \end{bmatrix} \qquad P^2 = \begin{bmatrix} .34 & .3 & .36 \\ .45 & .19 & .36 \\ .45 & .3 & .25 \end{bmatrix}$$

$$P^7 = \begin{bmatrix} .405413 & .269831 & .324756 \\ .405546 & .270497 & .323957 \\ .40528 & .27063 & .32409 \end{bmatrix}$$

$$P^{15} = \begin{bmatrix} .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \end{bmatrix}$$

This means that regardless of where you start at time 0, for big enough $t$,

$$Pr(X_t = 1) \approx .405405$$

$$Pr(X_t = 2) \approx .27027$$

$$Pr(X_t = 3) \approx .324324$$

There is a $\pi = (.405405, .27027, .324324)$ and for large $t$,

$$\mu_t \approx \pi$$

more precisely, $\lim_{t \to \infty} \mu_t = \pi$

This $\pi$ is called the stationary distribution, or the invariant distribution,

because $\pi = \pi P$

if $X_t \sim \pi$ then $X_{t+1} \sim \pi$

So once we're in distribution $\pi$ then we stay in $\pi$.

In earlier example, $\pi = \pi P$ means

$$\pi(1) = .4\pi(1) + .3\pi(2) + .5\pi(3)$$
$$\pi(2) = .6\pi(1) + .1\pi(2) + 0\pi(3)$$
$$\pi(3) = 0\pi(1) + .6\pi(2) + .5\pi(3)$$

3 variables, 3 equations. but to get a unique solution also need to use that:

$$1 = \pi(1) + \pi(2) + \pi(3).$$

Any $\pi$ satisfying $\pi P = \pi$ is called a stationary distribution.

Want to know when there is a stationary distribution $\pi$ that we eventually reach for all initial distributions $\mu_0$?

Need that there is a unique stationary distribution.
Otherwise if there are $\pi_1, \pi_2$ both are stationary.
Then for $X_0 \sim \pi_1$ we don't reach $\pi_2$
& for $X_0 \sim \pi_2$ we don't reach $\pi_1$.

One condition we need is that the weighted graph representing the Markov chain is strongly connected. Thus for all $i, j$ there exists a $t$ where

$$P^t(i,j) > 0$$

Path from $i$ to $j$ of length $t$.

Called irreducible condition.

We also need that it doesn't have any periodicities.

E.g., if it's bipartite, then if we start on the left side we know at even times we're on the left & at odd times we're on the right.

We need that for all $i$,

$$\gcd\{t : P^t(i,i) > 0\} = 1$$

(for the bipartite example the gcd = 2)

This is called aperiodic condition.

Ergodic MC = irreducible & aperiodic

equivalent condition:

There exists $t$ such that for all $i,j$

$$P^t(i,j) > 0$$

(Note this changes the order of quantification from irreducible:

irreducible: $\forall i,j \ \exists t \ \ P^t(i,j) > 0$

ergodic: $\exists t \ \forall i,j \ \ P^t(i,j) > 0$ )

# Fundamental theorem of Markov chains:

For ergodic Markov chain, there is a unique stationary distribution $\pi$ & for all $\mu_0$,

$$\lim_{t \to \infty} \mu_t = \pi$$

In other words, So for all $i$,

$$\lim_{t \to \infty} P^t(i,j) = \pi(j)$$

Thus, for ergodic MC, no matter our initial distribution, we eventually reach a unique stationary distribution $\pi$.

But what is $\pi$?

Note $\pi P = \pi$

this means $\pi$ is a (left) eigenvector of $P$ with eigenvalue 1.

Suppose the MC is symmetric.

So $P(i,j) = P(j,i)$ for all $i,j$.

Then we claim that $\pi$ is uniformly distributed over $\{1, \ldots, N\}$

(assuming $P$ is ergodic).

Why?

Let's check that for $\pi = \text{uniform}(N)$ that $\pi P = \pi$

Look at the $j^{th}$ entry of $\pi P$:

$$(\pi P)(j) = \sum_{i=1}^{N} \pi(i) P(i,j)$$

$$= \frac{1}{N} \sum_{i=1}^{N} P(i,j)$$

$$= \frac{1}{N} \sum_{i=1}^{N} P(j,i) \quad \text{since } P(i,j) = P(j,i)$$

$$= \frac{1}{N} \quad \text{since } \sum_{i=1}^{N} P(j,i) = 1$$

(sum of outgoing edges is $=1$)

$\boxtimes$

Reversible condition is a weighted version of symmetry.

A MC is reversible with respect to $\pi$

if for all $i,j$:

$$\pi(i)P(i,j) = \pi(j)P(j,i)$$

Such a $\pi$ is a stationary distribution.

Thus if $P$ is ergodic then this $\pi$ is the unique stationary distribution & we eventually reach it.

Why is $\pi$ a stationary distribution?

Some proof: Look at $j^{th}$ entry of $\pi P$:

$$(\pi P)(j) = \sum_{i=1}^{N} \pi(i)P(i,j)$$

$$= \sum_{i=1}^{N} \pi(j)P(j,i)$$

$$= \pi(j)\sum_{i=1}^{N} P(j,i)$$

$$= \pi(j) \quad \text{since } \sum_{i=1}^{N} P(j,i) = 1.$$

$\boxtimes$

When it's not reversible it's difficult to determine $\pi$, since $N$ is often HUGE.

## Example:

Want to generate a random matching of an input graph $G=(V,E)$.

Let $\Omega$ = set of all matchings of $G$.
of all size

Choose $X_0$ arbitrarily (e.g., $X_0 = \emptyset$)

From $X_t \in \Omega$,

- choose an edge $e$ uniformly at random from $E$
- Let $X' = \begin{cases} X_t \cup e & \text{if } e \notin X_t \\ X_t - e & \text{if } e \in X_t \end{cases}$
- If $X' \in M$ (i.e., if $X'$ is a matching) then set $X_{t+1} = \begin{cases} X' & \text{with prob. } 1/2 \\ X_t & \text{with prob. } 1/2 \end{cases}$
  else set $X_{t+1} = X_t$.

Note, for all $M \in \Omega$,

$P(M,M) \geq 1/2$, thus $P$ is aperiodic.

Also, for all $M \in \Omega$, we can get to $M' = \emptyset$ & then from $M' = \emptyset$ we can get to every $M'' \in \Omega$. Thus $P$ is irreducible.

Therefore $P$ is ergodic.

Note, for all $M, M' \in \Omega$,

$P(M,M') = P(M',M)$

So $P$ is symmetric.

Therefore, $\pi = \text{uniform}(\Omega)$ is the unique stationary distribution.

But what's a big enough $t$ so that:

$\mu_t \approx \pi$?