Markov Chains and Random Walks

Stochastic Process

- A stochastic process is a sequence of random variables $\{X_t\}_{t=0}^{n}$.
- Usually the subindex t refers to time steps and if $t \in \mathbb{N}$, the stochastic process is said to be discrete.
- The random variable X_t is called the state at time t.
- If $n < \infty$ the process is said to be finite, otherwise it is said infinite.
- A stochastic process is used as a model to study the probability of events associated to a random phenomena.

Markov Chain

One simple case of stochastic events is the Markov Chains, where the property of the MC model is that the model does not remember the history of past events,

Markov Chains are defined on a finite set of states (S), where at time t, X_t could be any state in S, together with by the matrix of transition probability for going from each state in S to any other state in S, including the case that the state X_t remains the same at $t+1$.

In a Markov Chain, at any given time t, the state X_t is determined only by X_{t-1}.

That is the key property of Markov Chains and stochastic processes with that property are said to be Markovian or to have the Memoryless property.

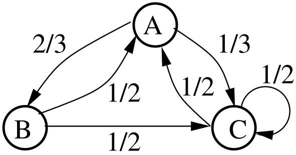
Markov-Chains: An important tool for CS

- One of the simplest forms of stochastic dynamics.
- Allows to model stochastic temporal dependencies
- Applications in many areas
  - Surfing the web
  - Design of randomizes algorithms
  - Random walks
  - Machine Learning (Markov Decision Processes)
  - Computer Vision (Markov Random Fields)
  - etc. etc.

Formal definition of Markov Chains

A finite, time-discrete Markov Chain, with finite state $S = \{1, 2, \ldots, k\}$ is a stochastic process $\{X_t\}$ s.t. for all $i, j \in S$, and for all $t \geq 0$,

$$\mathbf{Pr}\left[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i\right] = \mathbf{Pr}\left[X_{t+1} = j \mid X_t = i\right].$$



Transition probability matrix

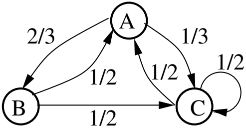
For $v, u \in S$, let $p^t_{u,v}$ be the probability of going from $u \rightsquigarrow v$ in exactly t steps i.e. $p^t_{u,v} = \mathbf{Pr}\left[X_{s+t} = v \mid X_s = u\right]$.

Formally for $s \geq 0$ and $t > 1$, $p^t_{u,v} = \mathbf{Pr}\left[X_{s+t} = v \mid X_s = u\right]$.

A MC besides being represented by its digraph, it also can be represented by its $|S| \times |S|$ transition matrix P, where for $u, v \in S$, the entrance (u, v) i P: $P_{u,v}$.

A times, we may use i $P^t_{u,v}$ to indicate entrance (u, v)in the matrix P, i.e $p^t_{u,v} = P^t_{u,v} = \mathbf{Pr}\left[X_{s+t} = v \mid X_s = u\right]$.

Transition matrix: Example

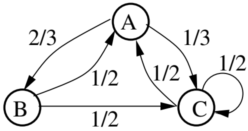


$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{array} = P$$

Notice the entrance (u, v) in P denotes the probability of going from $u \to v$ in one step.

Notice, in a MC the transition matrix is stochastic, so sum of transitions out of any state must be 1.

The power of the transition matrix



$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{array} = P$$

In ex. $\mathbf{Pr}\left[X_1 = C | X_0 = A\right] = P^1_{A,C} = 1/3.$
$\mathbf{Pr}\left[X_2 = C | X_0 = A\right] = P^1_{AB} P^1_{BC} + P^1_{AC} P^1_{CC} = 1/3 + 1/6 = P^2_{A,C}$

In general, assume a MC with k states and transition matrix P, let $u, v \in S$:

- What is the $\mathbf{Pr}\left[X_1 = u | X_0 = v\right]$, i.e. $= P_{v,u}$?
- What is the $\mathbf{Pr}\left[X_2 = u | X_0 = v\right] = P^2_{v,u}$?

The power of the transition matrix

Use Law Total Probability+ Markov propiety:

$$\begin{aligned}\mathbf{Pr}\left[X_2 = u | X_0 = v\right] &= \sum_{w=1}^{m} \mathbf{Pr}\left[X_1 = w | X_0 = v\right] \mathbf{Pr}\left[X_2 = u | X_1 = w\right] \\ &= \sum_{w=1}^{m} P_{v,w} P_{w,u} = P^2_{v,u}.\end{aligned}$$

In general $\mathbf{Pr}\left[X_t = w | X_0 = v\right] = P^t_{v,u}$ and $\mathbf{Pr}\left[X_{k+t} = w | X_k = v\right] = P^t_{v,u}$.

The argument can be generalized to

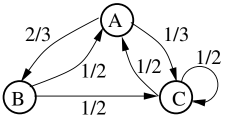
Given the transition matrix P of a MC, then for any $t > 1$,

$$P^t = P \cdot P^{t-1}.$$

Notice the entrance (u, v) in P^t denotes the probability of going from $u \to v$ in t steps.

Distributions at time t

Instead of fixing each initial state, we define X_0 to be a random variable, assigning to X_0 an initial distribution π_0, which is a row vector indicating at $t = 0$ the probability of starting in the corresponding state.
For example, in the MC:



we may consider,

$$\begin{array}{ccc} A & B & C \\ (0 & 0.3 & 0.6) \end{array} = \pi_0$$

Distributions at time t

Starting with an initial distribution π_0, what is the distribution of state v at time t?

$$\begin{aligned}
\pi_t[v] =& \mathbf{Pr}\left[X_t = v\right] \\
=& \sum_{u \in S} \mathbf{Pr}\left[X_0 = u\right] \mathbf{Pr}\left[X_t = v | X_0 = u\right] \\
=& \sum_{u \in S} \pi_0[u] P^t_{v,u}.
\end{aligned}$$

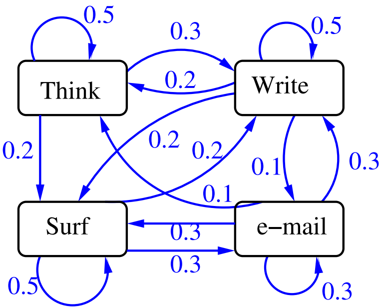
i.e. $\pi_t[y]$ is the probability at step t the system is in state y.

Therefore, $\pi_t = \pi_0 P^t$ and $\pi_{s+t} = \pi_s P^t$.

Example MC: Writing a research paper

Recall that Markov Chains are given either by a weighted digraph, where the edge weights are the transition probabilities, or by the $|S| \times |S|$ transition probability matrix P,

Example: Writing a paper $S = \{r, w, e, s\}$



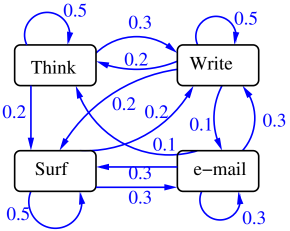
$$\begin{array}{c c} & \begin{array}{cccc} r & w & e & s \end{array} \\ \begin{array}{c} r \\ w \\ e \\ s \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} \end{array}$$

More on the Markovian property

Notice the memoryless property does not mean that X_{t+1} is independent from $X_0, X_1, \ldots, X_{t-1}$.

(For instance notice that intuitively we have:
$\mathbf{Pr}\left[\text{Thinking at } t+1\right] < \mathbf{Pr}\left[\text{Thinking at } t \mid \text{Thinking at } t-1\right]$).

Therefore, the dependencies of X_t on $X_0, \ldots, X_{t-1}$, are all captured by X_{t-1}.



Example of writing a paper

$\mathbf{Pr}\left[X_2 = s | X_0 = r\right]$ is probability that $t = 2$ we are in state s, if we start at $t = 0$ in state r.

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} \times \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} = \begin{array}{c} \begin{pmatrix} 0.31 & 0.34 & 0.09 & 0.26 \\ 0.21 & 0.38 & 0.14 & 0.27 \\ 0.14 & 0.33 & 0.21 & 0.32 \\ 0.07 & 0.29 & 0.26 & 0.38 \end{pmatrix} \end{array} \begin{array}{c} r \\ w \\ e \\ s \end{array}$$

$\mathbf{Pr}\left[X_1 = s | X_0 = r\right] = 0.07.$

Distribution on states

Recall π_t is the prob. distribution at X_t , over S.

For our example of writing a paper, if $t = 0$ (after waking up):

$$\begin{array}{cccccc} & & r & w & e & s \\ \pi_0 = & (& 0.2 & 0 & 0.3 & 0.5) \end{array}$$

$$\begin{pmatrix} 0.2 & 0 & 0.3 & 0.5 \end{pmatrix} \times \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.13 & 0.25 & 0.24 & 0.38 \end{pmatrix} = \pi_1$$

Therefore, we have $\pi_t = \pi_0 \times P^t$ and $\pi_{k+t} = \pi_k \times P^t$

Notice $\pi_t = (\pi_t[r], \pi_t[w], \pi_t[e], \pi_t[s])$

Stationary distributions: Writing a paper

- Suppose in the writing a paper example, the t is measured in minutes.
- To see how the Markov chain will evolve after 20 minutes i.e. $\mathbf{Pr}\left[X_{20} = s | X_0 = r\right]$ we must compute P^{20}, and to see if 5' later $\mathbf{Pr}\left[X_{25} = s | X_{20} = s\right]$.
- Matrices P^{20} and P^{25} may be almost identical.
- This indicates that in the long run, the starting state doesn't really matter,
- which implies that after a sufficiently long t: $\pi_t = \pi_{t+k}$, doesn't change when you do further steps.
- That is, for sufficient large t, the vector distribution converges to a π, $\pi_{t+1} = \pi_t P$, $\Rightarrow \pi = \pi P$. The stationary distribution

Stationary distributions: Writing a paper

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}^{20} = \begin{pmatrix} 0.1707318219 & 0.1707318219 & 0.18157173275 & 0.3116530652 \\ 0.1707317681 & 0.3360433708 & 0.18157177167 & 0.3116530893 \\ 0.1707316811 & 0.3360433559 & 0.18157183465 & 0.3116531282 \\ 0.1707315941 & 0.3360433410 & 0.18157189762 & 0.3116531671 \end{pmatrix}$$

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}^{25} = \begin{pmatrix} 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \end{pmatrix}$$

Therefore $\pi = (0.170732, 0.336043, 0.181572, 0.311653)$ is the stationary distribution of the writing a paper's problem.

Stationary distributions

A probability vector π is called a stationary distribution over S for P if it satisfies the stationary equations

$$\pi = \pi P.$$

Notice that if a MC has an stationary distribution π that means than after running a certain time the MC, then π the PMF for every r.v. X_i, $0 \leq i \leq n$.

How to find the stationary distribution

Given a finite MC with finite set of states $k = |S|$, let P be the $k \times k$ matrix of transition probabilities.

The stationary distribution $\pi = (\pi[1], \dots, \pi[k])$ over S, where $\pi_i = \pi[s_i]$. is defined by

$$(\pi[1], \dots, \pi[k]) = (\pi[1], \dots, \pi[k])P.$$

Therefore we have a system of k unknowns with k equations plus an extra equation: $\sum_{i=1}^{k} \pi[i] = 1$.

Stationary distributions: Example

In the writing a paper problem, we can transform $\pi = \pi P$ into 5 equations to get the value of π:

$$(\pi[t], \pi[w], \pi[e], \pi[s]) = (\pi[t], \pi[w], \pi[e], \pi[s]) \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}$$

$$\begin{aligned}
\pi[r] &= .5\pi[r] + .2\pi[w] + .1\pi[e], \\
\pi[w] &= .3\pi[r] + .5\pi[w] + .3\pi[e] + .2\pi[s], \\
\pi[e] &= .1\pi[w] + .3\pi[e] + .3\pi[s], \\
\pi[s] &= .2\pi[r] + .2\pi[w] + .3\pi[e] + .3\pi[s], \\
1 &= \pi[r] + \pi[w] + \pi[e] + \pi[s],
\end{aligned}$$

which yields, $\pi = (0.170732, 0.336043, 0.181572, 0.311653)$.

Stationary distributions

- Notice that $\{\pi[1], \ldots \pi[n]\}$ means π is an eigenvector of P with eigenvalue=1.
- A Markov Chain with k states and transition matrix P, it has a set of $k+1$ stationary equations with k unknowns $\{\pi[1], \ldots \pi[n]\}$, which are given by $\pi = \pi P$ together with $\sum_{u=1}^{k} \pi[u] = 1$:

$$\pi[u] = \sum_{v=1}^{k} \pi[v] P_{vu}, \ \forall 1 \leq v \leq k$$

- Linear algebra tells us that such a system either has a unique solution, or infinitely many solutions.
- We want a unique stationary distribution, so we will give conditions for MC that have a unique π.
- However, for MC with a huge number of states, it is a problem to get the stationary distribution by solving stationary equations.

An Example: The 2-SAT problem

Given a Boolean formula ϕ, constructed from:

- A set X of n Boolean variables,
- A set with m clauses $C_1, \dots C_m$, where each clause is the disjunction of exactly 2 literals, (x_i or $\bar{x}_i$),
- $\phi =$ conjunction of the m clauses.

The 2-SAT problem is to find an assignment $A^* : X \to \{0, 1\}$, which satisfies ϕ,
i.e, to find an A^* s.t. $A^*(\phi) = 1$.

Notice that if $|X| = n$, then $m \leq \binom{n}{2} \sim n^2$.

In general k-SAT$\in$ NP-complete, for $k \geq 3$. 2-SAT$\in$ P.

Randomized algorithm for 2-SAT problem

Given 2-SAT input ϕ on $X = \{x_i\}_{i=1}^n$ and $\{C_j\}_{j=1}^m$
Make $\forall 1 \leq i \leq n \; A(x_i) = 1$
while ϕ contains unsatisfied clause and the $\#$ steps is $\leq 2cn^2$
do
 pick and unsatisfied clause C_j
 choose u.a.r. one of the 3 literals and flip their value
 if ϕ is satisfied now **OUTPUT** the new truth assignment A^*
end while
OUTPUT ϕ is unsatisfiable

Notice if $\phi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_2) \wedge (x_1 \vee \bar{x}_2)$
does not has a $A^* \models \phi$.
Implement the algorithm on
$\phi = (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_3) \wedge (\bar{x}_1 \vee x_2) \wedge (\bar{x}_4 \vee \bar{x}_3) \wedge (x_4 \vee \bar{x}_1)$ which it has a
$A^* \models \phi$ ($A^*(x_1) = A^*(x_2) = A^*(x_4) = 1$ and $A^*(x_3) = 0$)

Analysis for 2-SAT algorithm

Given $\phi, |X| = n, \{C_j\}_{i=1}^{m}$, assume $\exists A^* \models \phi$.

- Let A_i the assignment at the i-th iteration.
- Let $X_i = |\{x_j \in X \mid A_i(x_j) = A^*(x_j)\}$.
- Notice $0 \leq X_i \leq n$. Moreover, when $X_i = n$, we found A^*.
- Analysis: Starting from $X_i < n$, how long to get $X_i = n$?
- Note that $\mathbf{Pr}[X_{i+1} = 1 \mid X_i = 0] = 1$.

As A^* satisfies ϕ and A_i no, that means there is a clause C_j that A^* satisfies but A_i not. So A^* and A_i disagree in the value of least one literal in the clause. Therefore,

For $1 \leq k \leq n-1$, $\mathbf{Pr}[X_{i+1} = k+1 \mid X_i = k] \geq 1/2$ and $\mathbf{Pr}[X_{i+1} = k-1 \mid X_i = k] \leq 1/2$.

Analysis for 2-SAT

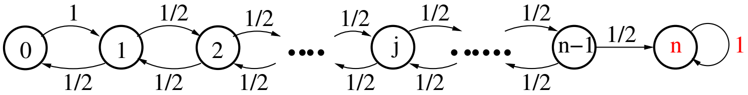
Therefore, the process $X_0, X_1, \ldots$ is not necessarily a MC,
As the probability X_i increases depends of wether A_i and A^* disagree in 1 or 2 variables in the unsatisfied C, which depend of the clauses considered in the past, so transition probabilities could be $1/2$ or 1 (not always the same)

Analysis for 2-SAT

Define a MC $Y_0, Y_1, Y_2, \ldots$ which is a pessimistic version of process $X_0, X_1, \ldots$, in the sense that Y_i measures exactly the same quantity than X_i but the probability of change will be exactly $1/2$:

- $Y_0 = X_0$ and $\mathbf{Pr}\left[Y_{i+1} = 1 \mid Y_i = 0\right] = 1$;
- $1 \leq k \leq n-1$, $\mathbf{Pr}\left[Y_{i+1} = k+1 \mid Y_i = k\right] = 1/2$;
- $\mathbf{Pr}\left[Y_{i+1} = k-1 \mid Y_i = k\right] = 1/2$.

Note that the time to reach n from $j \geq 0$:
in $\{Y_i\}_{i=0}^n$ is $\geq$ that in $\{X_i\}_{i=0}^n$.



MC for 2–SAT

Upper Bound on the time to arrive state n

Lemma: If a 2-CNF ϕ on n Boolean variables has satisfying assignment A^* The 2-SAT algorithm finds one in expected time $\leq n^2$.

Proof

Let h_j be the expected time for the process Y to go from state j to state n.

It suffices to prove that, when Y starts in state j to arrive to n is $\leq 2cn^2$.

Notice $h_n = 0$; $h_0 = h_1 + 1$;

We want a general recurrence on h_j, for $1 \leq j < n$

Define a rv Z_j counting the steps to go from states $j \to n$ in Y.

Then, with probability $1/2$ $Z_j = Z_{j-1} + 1$ and with probability $1/2$ $Z_j = Z_{j+1} + 1$. So $h_j = \mathbf{E}[Z_j]$.

$$\mathbf{E}[Z_j] = \mathbf{E}\left[\frac{Z_{j-1}+1}{2} + \frac{Z_{j+1}+1}{2}\right] = \frac{\mathbf{E}[Z_{j-1}]+1}{2} + \frac{\mathbf{E}[Z_{j+1}]+1}{2}.$$

Upper Bound on the time to arrive state n

From the previous bound we get $h_j = \frac{h_{j-1}}{2} + \frac{h_{j+1}}{2} + 1$.
So we get the $n+1$ equations,

$$h_n = 0$$
$$h_0 = h_1 + 1$$
$$h_j = \frac{h_{j-1}}{2} + \frac{h_{j+1}}{2} + 1 (\text{for } 0 \leq j \leq n-1)$$

Using induction, we can prove that the solution is

$$h_j = h_{j+1} + 2j + 1.$$

Therefore,
$h_0 = h_1 + 1 = h_2 + 1 + 3 = \cdots = \underbrace{h_n}_{=0} + \sum_{i=0}^{n-1}(2i+1) = n^2$.

$\square$

Error probability for 2-SAT algorithm

Theorem The 2-SAT algorithm give the correct answer NO if ϕ is not satisfiable. Otherwise, with probability $\geq 1 - \frac{1}{2^c}$ the algorithm return a satisfying assignment.

Proof

Let ϕ be satisfiable (otherwise the theorem holds).
Break the $2cn^2$ iterations into c blocks of $2n^2$ iterations.
For each block i, define a r.v. $Z =$ number of iterations from the start of the i-block until a solution is found.
Using Markov's inequality:

$$\mathbf{Pr}\left[Z > 2n^2\right] \leq \frac{n^2}{2n^2} = \frac{1}{2}.$$

Thwerefore, the probability that the algorithm fails to find a satisfying assignment after c segments (no block includes a solution) is at most $\frac{1}{2^c}$. $\square$

Properties of Markov chains: Recurrent

We would like to know which properties a Markov chain should have to assure the existence of a unique stationary distribution, i.e. that $\lim_{t\to\infty} P_t \to$ a stable matrix.

A state is defined to be recurrent if any time that we leave the state, we will return to it with probability 1.

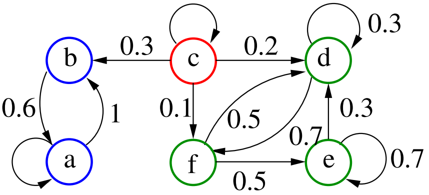
Formally, if at time t_0 the MC is in state s, s is recurrent if $\exists t$ such that $P_{s,s}^{t_0+t} = 1$. Otherwise the state is said to be transient.

A MC is said to be recurrent if every state is recurrent.

Intuitively, transience attempts to capture how "connected" a state is to the entirety of the Markov chain. If there is a possibility of leaving the state and never returning, then the state is not very connected at all, so it is known as transient.

More on Recurrent and Transient MC

Alternatively, given a MC $\{X_t\}$ with state set S, a $u \in S$ is transient if for $t > 0$,
$\mathbf{Pr}\left[X_t = u \text{ for infinitely many } t \,|\, X_0 = u\right] = 0.$
A $v \in S$ is recurrent if for $t > 0$,
$\mathbf{Pr}\left[X_t = u \text{ for infinitely many } t \,|\, X_0 = u\right] = 1.$



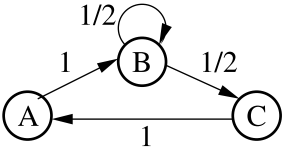
TRANSIENT: c

RECURRENT: a,b,d,e,f

For transient state, the number of times the chain visits s when starting at s is given by a geometric random variable in $G(p)$, where $p = \sum_{t \geq 1} P^t_{s,s}$.

Properties of Markov chains: Positive recurrent state

A recurrent state u has the property that the MC is i expected to return to u an infinite number of times.
However, when restricting to finite time the MC may not return to u in a finite number of steps, which contradicts the intuition for recurrence.
We need a further finer classification of recurrence states:
If $X_t = u$ define $\tau_u = \min\{\hat{t} \mid X_{t+\hat{t}} = u\}$, as the first return time to u.
Define a recurrent state u to be positive recurrent if $\mathbf{E}[\tau_u | X_0 = u] < \infty$. Otherwise u is said to be null recurrent state.



A MC with all states positive recurent.

Properties of Markov chains: Periodicity

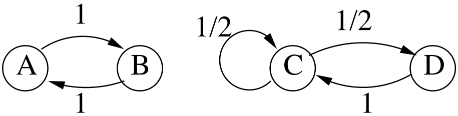
Define the period of $s_j \in S$ as $d(s_j) = \gcd\{t \in \mathbb{Z}^+ \mid P^n_{s_j,s_j} > 0\}$.
So from s_j the chain can return to s_j in periods of $d(s_j)$.

Define s_j to be periodic if $d(s_j) > 1$, and s_j to be aperiodic if $d(s_j) = 1$.

A Markov chain P is periodic if every state is periodic, otherwise it is aperiodic.

Periodicity: 1st. example

A state u in a MC has period$=t$ if only comes back to itself every t steps i.e. $P^i_{u,u} = 0, \forall i = t, 2t, 3t, \ldots$. Otherwise, the state is said to be aperiodic.



A,B periodic with period=2 C and D aperiodics

Notice for the left side Markov chain:

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, P^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ldots$$

$\Rightarrow \lim_{t\to\infty} P^t$ does not exist.

Periodicity: 1st. example

However, this specific Markov chain has a unique stationary distribution $\pi = (1/2, 1/2)$

$$\text{Using balence eq. } (\pi[A], \pi[B]) = (\pi[A], \pi[B]) \times \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{aligned} \pi[A] &= 0\pi[A] + 1\pi[B] \\ \pi[B] &= 1\pi[A] + 0\pi[B] \\ 1 &= \pi[A] + \pi[B] \end{aligned}$$

we get $\pi[A] = 1/2$ and $\pi[B] = 1/2$.

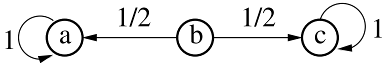
If one MC has at least one state s with self-transition $P_{s,s} > 0$ then the chain is aperiodic.

How to check if a MC is aperiodic

Given an irreducible MC with a finite number of states,

1. If there is at least one self-transition $P_{i,i}$ in the chain, then the chain is aperiodic.
2. If you can return from i to i in t steps and in k steps, where $\gcd(t, k) = 1$, then state i is aperiodic.
3. The chain is aperiodic if and only if there exists a positive integer k s.t. all entrances in matrix P^k are > 0 (for all pair of states (i, j) then $P^k_{i,j} > 0$).

Properties of Markov chains: Reducibility and irreducibility



This MC is sensitive to initial state.
In this MC, $\forall t$, $\lim_{t\to\infty} P^t$ exists,

$$P^t = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix},$$

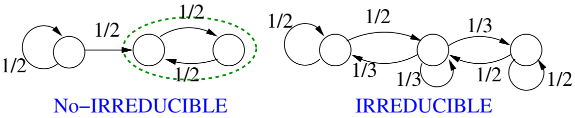
Solving the stationary equations

$$(\pi[1], \pi[2], pi[3]) = (\pi[1], \pi[2], pi[3]) \times \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix},$$

it turns out that we have infinite many stationary distributions $\pi = (p, 0, 1-p)$.

Properties of Markov chains: Irreducibility

A finite Markov chain P is irreducible if its graph representation $\vec{W}$ is strongly connected.
In irreducible $\vec{W}$, the system can't be trapped in small subsets of S
.



For finite Markov chains, an irreducible Markov chain is also denoted as ergodic.

Some relations among the previous classes of MC

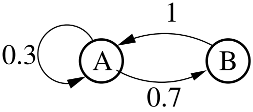
- If P is irreducible and contains a self-loop, then P is also aperiodic.
- If in a finite MC P all its states are irreducible then all the states are positive recurrent.
- If P is irreducible and finite all its states are positive recurrent, then the Markov chain has a unique stationary distribution.

Regular Markov Chain

A matrix A is defined to be regular if there is an integer $n > 0$ such that A^n contains only positive entries.

A Markov chain is a regular Markov Chain if its transition probability matrix P is regular.

Consider the following example:



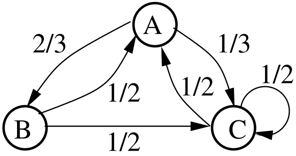
$$P = \begin{pmatrix} 0.3 & 0.7 \\ 1 & 0 \end{pmatrix} \qquad P^2 = \begin{pmatrix} 0.79 & 0.21 \\ 0.3 & 0.07 \end{pmatrix}$$

Properties of Regular MC

A finite state Markov Chain is regular if $\exists t < \infty$ such that for all states i, j, $P^t_{i,j} > 0$.

Notice that if a finite state MC is irreducible that means that for every pair of states i, j there is a t' s.t. $P^{t'}_{i,j} > 0$. If the MC is also aperiodic there is a value k s.t. for all pair of states (i, j), $P^k_{i,j} > 0$, which is exactly the definition of being regular. Therefore

Theorem A finite state Markov chain is irreducible and aperiodic if and only if it is regular.



Markov Chains: An issue about names

- For finite state Markov chains, many people denotes a that is aperiodic, irreducible, and positive recurrent as ergodic, as for instance in MU.
- However in this slides we use regular for finite MC that are aperiodic, irreducible, and positive recurrent, and reserve the name regular for irreducible MC.
- The mathematical reason for do so is nicely explained in the link: `https://math.stackexchange.com/questions/152491/is-ergodic-markov-chain-both-irreducible-and-aperiodic`
- However for infinite MC regularity is not easy to define.

Ex.: Gambler's Ruin

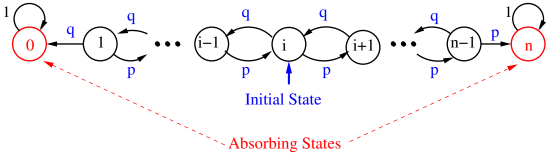
Model used to evaluate insurance risks.

- You place bets of 1€. With probability p, you gain 1 €, and with probability $q = 1 - p$ you loose your 1 € bet.
- You start with an initial amount of m €.
- You keep playing until you loose all your money or you arrive to have n €.
- Define bias factor $\alpha = q/p$; If $\alpha = 1$ then $p = q = 1/2$, so it is fair game. If $\alpha > 1$ you are more likely to loose than win; if $\alpha < 1$, the game is bias against you.
- The goal is finding the probability of winning i.e. starting in state m reaching state n.

Notice in this chain, once we enter in state 0 or in state n, we can't leave the state. Those states are called absorbing states.

Gambler's Ruin

The chain can be given either by a $(n+1)\times(n+1)$ transition matrix P, where for $0 \le i \le n$: $P_{i,(i+1)} = p$ and $P_{i,(i-1)} = q$, $P_{0,0} = P_{n,n} = 1$.



$$
\begin{array}{c c}
 & \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \cdots & n-1 & n \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ n-2 \\ n-1 \\ n \end{array} &
\begin{pmatrix}
1 & q & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & q & 0 & \cdots & 0 & 0 \\
0 & p & 0 & q & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & q & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & p & 1
\end{pmatrix}
\end{array}
$$

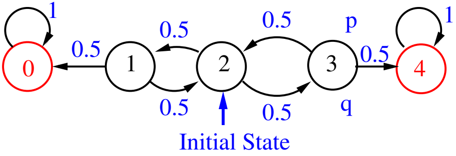
Gambler's Ruin

In a MC an absorbing state i is one for which $p_{i,i} = 1$.

The chain has two absorbing states, when the system arrives to one of them it never exit, it is absorbed.
Some of the questions to be asked about such a chain are:

- What is the probability that the process will eventually reach an absorbing state? absorption probability.
- On the average, how long will it take for the process to be absorbed? expected absorption probability.

Gambler's Ruin Example with 4 states



$$\begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\pi^0 = (0, 0, 1, 0, 0)$$
$$\pi^1 = (0, 1/2, 0, 1/2, 0)$$
$$\pi^2 = (1/4, 0, 2/4, 0, 1/4)$$
$$\vdots$$

Notice in this case the states 1,2, and 3 are transient and 0, 4 are absorbing states

Gambler's Ruin

Let $P_{i,n}$ denote the probability that the gambler with i € arrives to n €, before going broke.

Note that $1 - P_{i,n}$ is the corresponding probably that the gambler ruins.

Let us compute $P_{i,n}$:
Notice $P_{0,n} = 0$, $P_{n,n} = 1$ and $P_{i,n} = pP_{i+1,n} + qP_{i-1,n}$.

As $P_{i,n} = pP_{i,n} + qP_{i,n}$ then $P_{i+1,n} - P_{i,n} = \frac{q}{p}(P_{i,n} - P_{i-1,n})$.

In particular $P_{2,n} - P_{2,n} = \frac{q}{p}P_{1,n}$ (as $P_{0,n} = 0$)
and $P_{3,n} - P_{2,n} = \frac{q}{p}(P_{3,n} - P_{2,n}) = (\frac{q}{p})^2 P_{1,n}$
so $P_{i+1,n} - P_{i,n} = (\frac{q}{p})^i P_{1,n}$..

Gambler's Ruin

On the other hand

$P_{i+1,n} = \sum_{k=0}^{i}(P_{k+1,n} - P_{k,n}) = \sum_{k=1}^{i}(P_{k+1,n} - P_{k,n}) + P_{1,n}$

$\Rightarrow P_{i+1,n} - P_{1,n} = \sum_{k=1}^{i}(P_{k+1,n} - P_{k,n}) = \sum_{k=1}^{i}(\frac{q}{p})^k P_{1,n}$

$\Rightarrow P_{i+1,n} = P_{1,n} + P_{1,n}\sum_{k=1}^{i}(\frac{q}{p})^k = \sum_{k=0}^{i}(\frac{q}{p})^k$

Using the geometric series equation $\sum_{j=0}^{i} x^j = \frac{1-x^{i+1}}{1-x}$.

$$P_{i+1,n} = \begin{cases} P_{1,n}\frac{1-(q/p)^{i+1}}{1-(q/p)}, & \text{if } p \neq q; \\ P_{1,n} \cdot (i+1) & \text{if } p = q = 1/2. \end{cases}$$

Gambler's Ruin

Choosing $i = n - 1$ and as $P_{n,n} = 1$, then

$$P_{1,n} = \begin{cases} \frac{1-(q/p)}{1-(q/p)^n}, & \text{if } p \neq q; \\ 1/n & \text{if } p = q = 1/2. \end{cases}$$

$$\text{Therefore }, P_{i,n} = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^n}, & \text{if } p \neq q; \\ i/n & \text{if } p = q = 1/2. \end{cases}$$

$\square$

Becoming rich or getting ruined

Using the deduced eq. for $P_{i,n}$:

- If $p > 1/2$ then $\frac{q}{p} < 1$ and $\lim_{n\to\infty} P_{i,n} = 1 - (q/p)^n > 0$. In this case, there is a positive probability that the gambler will become rich.
- If $p \leq 1/2$ then $\frac{q}{p} \geq 1$ and $\lim_{n\to\infty} P_{i,n} = 0$. So with probability 1 the gambler will get ruined.

For ex. if Pepet starts with 2 €, and $p = 0.6$, what is the probability that Pepet gets $n = 5$ €?
$P_{2,5} = \frac{1-(2/3)^2}{1-(2/3)^5} = 0.64$.

What is the probability that he will become infinitely rich?
$(n \to \infty)$ $P_{2,\infty} = 1 - (2/3)^2 = 0.56$.

Markov chains with absorbing states

- The Gambler ruin's Markov chain is an example of a Markov chain with one or more absorbing states, where the process stops.
- An absorbing state u has $p_{u,u} = 1$. In many application the absorbing MC has two states: 0 and n.
- Those MC are not irreducible (states 0 and n do not exit)
- Those MC play an important role in many "practical" stochastic processes: Biological, economical, and others.
- The limit probability distribution π of an absorbing MC has the absorbing probabilities for the absorbing state, and 0 for the other states. In the ex. of the Gambler's ruin if $p \neq q$, $\pi = \left(1 - (\frac{1-(q/p)^i}{1-(q/p)^n}), 0, \ldots, 0, \frac{1-(q/p)^i}{1-(q/p)^n}\right)$.
- The two important in those absorbing MC are:
  1. The absorption probability.
  2. The absorption time.

Fundamental Theorem of Markov Chains

Any finite, irreducible and aperiodic Markov chain P (i.e. regular) has the following properties:

1. The chain has a **unique** stationary distribution
$$\pi = (\pi[0], \pi[1], \pi[2], \ldots, \pi[n]).$$
2. $\lim_{t\to\infty} P^t$ exists and its row are copies of the stationary distribution π.

Recall that any finite state MC has a stationary distribution, but it may not be unique.
If we have a a periodic state i, $\pi[i]$ is not necessarily the limit probability of being in state i, but the frequency of being in state i.

Number of steps: Expected first passage

Given a regular Markov chain with a set S of states , $|S| = r$, and a unique stationary distribution π,

We want to compute the expected first recurrence time for $u \in S$, $h_{u,u}$,
i.e. the expected number of steps we need so that starting from u we return for first time to return to u.
Intuitively, in the long run we expect the MC to be in state u a fraction of $\pi[u]$, so $h_{u,u} \sim 1/\pi[u]$

The expected first passage from u to v is denoted $h_{u,v}$. and denoted mean first passage time.

In the particular case of random walks (2-SAT, 3-SAT, Gambler Ruin, etc.) $h_{u,v}$ will be denoted as the hitting time.

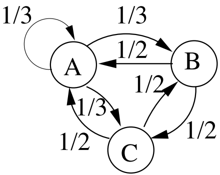
Computing $h_{u,u}$ using π

Theorem. In a finite, regular Markov chain with $|S| = r$ and a unique stationary distribution π, for $u \in S$

$$h_{u,u} = \frac{1}{\pi[u]}.$$

This technique is important and it is called first step analysis: it consist in breaking down the possibilities resulting from the first step in the MC.

Proof For $u, v \in S$, $h_{uv} = \mathbf{E}\left[\#\text{ steps } u \to v\right] = \sum_{w=1}^{k} P_{u,w} \mathbf{E}\left[\#\text{ steps } u \to v | 1st.\text{ step } u \to w\right]$



$$P = \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \end{array}$$

$\pi = (3/7, 2/7, 2/7)$

Proof of the Theorem

Two cases for w:

$(w = v)$ Then the expected time $u \to v$ is 1.

$(w \neq v)$ We take 1 step $v \to w$. By the Markovian propriety, we have to concentrate in state w:

$$\begin{aligned}
h_{uv} &= P_{u,v} + \sum_{w \neq v} P_{u,w}(1 + \underbrace{\mathbf{E}\left[\text{time from } w \to v\right]}_{h_{w,v}}) \\
&= P_{u,v} + \sum_{w \neq v} P_{u,w}(1 + h_{wv}) \\
&= P_{u,v} - P_{u,v}(1 + h_{vv}) + \sum_{w=1}^{r} P_{u,w}(1 + h_{wv}) \\
&= \underbrace{-P_{u,v} h_{vv}}_{\diamond} + \underbrace{\sum_{w=1}^{k} P_{u,w}(1 + h_{wv})}_{*} \qquad (1)
\end{aligned}$$

Proof of the Theorem: Term $(*)$

Let J be the $k \times k$ matrix of 1's, then $1 + h_{wv} = (J + H)[w, v]$, where $H = (h_{v,u})$:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} h_{AA} & h_{AB} & h_{AC} \\ h_{BA} & h_{BB} & h_{BC} \\ h_{CA} & h_{CB} & h_{CC} \end{pmatrix} = \begin{pmatrix} 1 + h_{AA} & 1 + h_{AB} & 1 + h_{AC} \\ 1 + h_{BA} & 1 + h_{BB} & 1 + h_{BC} \\ 1 + h_{CA} & 1 + h_{CB} & 1 + h_{CC} \end{pmatrix}$$

$$\sum_{w=1}^{k} h_{u,w}(1+h_{wv}) = \sum_{w=1}^{k} H[u,w](J+H)[w,v] = (H\times(J+H))[w,v].$$

So the sum is just the entrance (w, v) in the matrix $H \times (J + H)$.

Proof of the Theorem: Term ($\Diamond$)

Introduce $r \times r$ diagonal matrix D, where $D_{v,v} = h_{v,v}$,
so $P_{u,v} h_{v,v} = (P \times D)[u, v]$.

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \times \begin{pmatrix} h_{AA} & 0 & 0 \\ 0 & h_{BB} & 0 \\ 0 & 0 & h_{CC} \end{pmatrix} = \begin{pmatrix} h_{AA}/3 & h_{BB}/3 & h_{CC}/3 \\ h_{AA}/2 & 0 & h_{CC}/2 \\ h_{AA}/2 & h_{BB}/2 & 0 \end{pmatrix}$$

Ending the proof

Substituting $\diamond$ and $*$ in equation (1):
$h_{u,v} = -(P \times D)[u, v] + (H \times (J + M))[n, v]$.

As it is true $\forall(u, v)$, we have
$H = -PD + P(J + H) = -PD + PJ + PH$

Multiply both sides by the stationary distribution:
$\pi H = -\pi PD + \pi PJ + \pi PH$.
But by the stationary equation: $\pi P = \pi$
$\Rightarrow \pi H = -\pi D + \pi J + \pi H \Rightarrow \pi D = \pi J$

Notice πJ is just the k-dimensional vector $(1, 1, \ldots, 1)$
$(\sum_i \pi[i] = 1)$

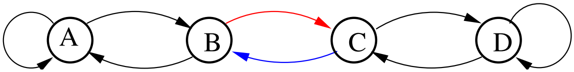
$\pi D = (\pi[1]h_{11}, \pi[2]h_{22}, \ldots, \pi[k]h_{kk}) \Rightarrow \pi[u] = \frac{1}{h_{u,u}}$ $\square$

Reversible Markov Chain

- For regular MC $\lim_{t\to\infty} P^t$ has all rows the same: the stationary π.
- If $|S|$ small, we can compute π by solving the stationary equations: $\pi = \pi P$.
- There is a nice property for MC, which makes more easy to compute the stationary distribution π of those MC: The reversibility.

Reversible Markov Chain

- Intuitively assume the MC below has the appropriate probabilities to have stationary distribution π. Then. for sufficiently large t, $p_{B,C}^{t+1} = \pi[B]P_{B,C}^{t}$ (red), and $p_{C,B}^{t+1} = \pi[C]P_{C,B}^{t}$ (blue).
- So in stationary distribution, the rate $B \rightarrow C$ = rate $C \rightarrow A$, and this holds for every pair of adjacent states.
  i.e. For such MC, $\forall u, v \in S\, \pi[u]P_{u,v} = \pi[v]P_{v,u}$.



If a MC P has a stationary distribution π, this means π is the joint PMF for $X_0, X_1, \ldots, X_n$. Assume that we run backwards the process: as well $X_n, X_{n-1}, \ldots, X_0$, an π is also the joint PMF of this time-reversal process. Then we say that the MC is reversible.

Reversible Markov Chain

Given a Markov Chains P, with a finite state S and a unique stationary distribution π, we say that the Markov Chain is time reversible if for all pair $u, v \in S$, it satisfies the balance equations:

$$\pi[u]P_{u,v} = \pi[v]P_{v,u}.$$

The name reversible is due to the fact that we can run the MC in the reverse and we have the same values.

The next theorem shows that if the balance equation holds for some distribution $\hat{\pi}$ then it must be a stationary distribution

Reversible Markov Chain

Theorem Let P be Markov Chain with estates S. If π is a probability vector satisfying the balance equations $\pi[u]P_{u,v} = \pi[v]P_{v,u}, \forall u, v \in S$, then π is a stationary distribution.

Proof: Check the stationary distribution holds, i.e. $\pi = \pi P$

$$
\begin{aligned}
(\pi P)[v] = &\sum_{u \in S} \pi[u]P_{u,v} = \sum_{u \in S} \pi[v]P_{v,u} \\
=&\pi[v] \sum_{u \in S} P_{v,u} = \pi[v]. \qquad \square
\end{aligned}
$$

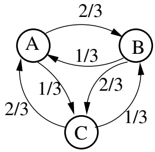
Given a finite-state MC, which is reversible, to find a stationary distribution: Solve the balance equations together with the equation $\sum_{v \in S} \pi[v]$.

What happens if the Markov Chain is not reversible

Not all Markov Chains are time reversible.

If there is no solution to the time reversibility equations , the way to find πc is to use the stationary equations, which always yield a solution (provided the state-space in not too large).

The following MC is not reversible:



To prove it, find the stationary distribution $\pi = (1/3, 1/3, 1/3)$ and notice that $\pi[B]P_{B,C} = \frac{2}{9} \neq \frac{1}{6} = \pi[C]P_{C,B}$.

Testing if a MC is reversible: Kolmogorov's loop criterion

It is desirable to verify reversibility before finding the stationary vector π.
Recall A MC is reversible if for every finite sequence of states $i_0, i_1, i_2, \dots i_k$ we have $p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{k-1},i_k} p_{i_k,i_0} = p_{i_0,i_k} \cdots p_{i_1,i_0}$

Kolmogorov's loop criterion: A Markov transition matrix P is reversible iff for every loop of distinct states, the forward loop probability product equals the backward loop probability product.

But for large number of states n, the number of loops could be exponential.

Kolmogorov's loop criterion

- A two-state MC is always reversible as $p_{1,2}p_{2,1} = p_{2,1}p_{1,2}$.
-
- If P is symmetric (bistochastic), then. $p_{i,j} = p_{j,i}, \forall i, j \in S$, Kolmogorov's criterion is satisfied and P is reversible.
- If the zeros in a regular MC P are not symmetric, then the chain is not reversible:
- There is a nice algorithm based in matrix operations to check if a MC P is reversible. Brill, Cheung, Hlynka, Jiang: *Reversibility checking for Markov chains*, Comm. on Stochastic Analysis, 12: 2, 129–135, (2018)

Symmetric matrix

As said P is symmetric ($\forall i, j \in S$ then $P_{i,j} = P_{j,i}$) the MC is reversible.

Lemma If P is symmetric then it has a unique stationary distribution π which is the uniform distribution, i.e. $\forall i \in S, \pi[i] = 1/n$, where $n = |S|$.

Proof A regular MC with symmetric transition matrix is also reversible.

Then $\forall i, j \in S, \pi[i]P_{i,j} = \pi[j]P_{j,i} \Rightarrow \pi[i] = \pi[j]$. If we have n states each with the same stationary distribution then $\pi = (1/n, \ldots, 1/n)$ $\square$

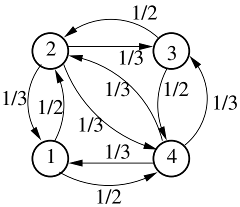
As a symmetric transition matrix P is also double stochastic we also can state that: If we have that the transition matrix P of a Markov chain is stochastic then the MC has unique stationary distribution π which is the uniform distribution.

Example: Random walk on a graph

Given G by its adjacency matrix, a walker moves to a randomly from vertex i to a neighbor with probability $1/d(i)$.

$$G = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Interpret random walks on G as a Markov chain and give the transition matrix P.



$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/4 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

Example: Random walk on a graph

Is P reversible?

Notice all non-diagonal 0s in P are symmetric.
There are 3 loops: (a) $(1 \to 2 \to 4 \to 1)$ (b) $(2 \to 3 \to 4 \to 2)$, (c) $(1 \to 2 \to 3 \to 4 \to 1)$.
For the (a), (b) and (c) Markov's loop criteria works, so yes.

Using the previous fact determine the stationary distribution of the MC.

As it is reversible, using the balance equations:

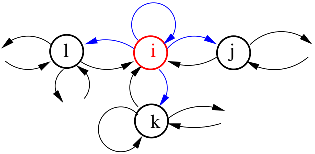
$$\begin{aligned}
\frac{\pi[1]}{2} &= \frac{\pi[2]}{3}; \\
\frac{\pi[2]}{3} &= \frac{\pi[3]}{2}; \\
\frac{\pi[2]}{3} &= \frac{\pi[4]}{3}; \\
1 &= \pi[1] + \pi[2] + \pi[3] + \pi[4];
\end{aligned}$$

We get
$\pi = \left(\frac{1}{5}, \frac{3}{10}, \frac{1}{5}, \frac{3}{10}\right)$

Computing $h_{i,v}$

We proved that for finite, regular Markov chains if π is the unique stationary distribution, then for any $u \in S$, $h_{u,u} = \frac{1}{\pi[u]}$.

Let us see for any $u, v \in S$, how to compute $h_{v,u}$:



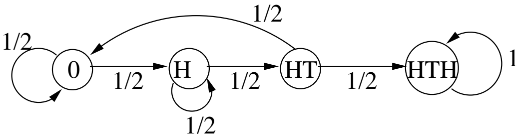
$$h_{i,n} = 1 + p_{i,i} h_{i,n} + p_{i,j} h_{j,n} + p_{i,k} h_{k,n} + p_{i,l} h_{l,n}.$$

Computing $h_{i,v}$: Example

A fair coin is tossed repeatedly and independently. Find the expected number of tosses till the pattern HTH appears.

Computing $h_{i,v}$: Example

A fair coin is tossed repeatedly and independently. Find the expected number of tosses till the pattern HTH appears.



Rename the states {H, HT, HTH} as $\{1, 2, 3\}$.

$$h_{0,3} = 1 + \frac{1}{2}h_{0,3} + \frac{1}{2}h_{1,3},$$
$$h_{1,3} = 1 + \frac{1}{2}h_{1,3} + \frac{1}{2}h_{2,3},$$
$$h_{2,3} = 1 + \frac{1}{2}h_{0,3},$$

which yields $h_{0,3} = 10$.

Random walks

- An algorithmic paradigm.
- Given a finite, connected graph $G = (V, E)$ with $|E| = m$, $|V| = n$, a random walk on G is a MC defined by the sequence of moves of a particle between the vertices of G.
- A random walk on G, probability starts from a given $v \in V$, and if v has $d(v)$ outgoing neighbors, then the probability that the walk moves to u is $1/d(v)$, where $\mathcal{N}(u) =$ set of neighbors of u and $|\mathcal{N}(u)| = d(u)$.

The generic algorithm:

Given $G = (V, E)$, $v \in V$
for repeat for T steps **do**
  Chose u.a.r. (with probability $= 1/d(v)$ a $u \in \mathcal{N}(v)$)
  $v = u$
end for

Random walks: Definitions

Given a connected graph $G = (V, E)$ define:

1. The hitting time $h_{v,u}$ from v to u, that is the expected number of steps for the random walk to go from v to u (for first time).
2. The cover time $C_{v,u}$ from v as the expected number of steps that a walk will take in starting from v visiting all vertices in G.
3. The cover time of G, C_G as $\max_{v \in V} C_v$.

Random walks and Markov Chains

Theorem: A random walk on an undirected G is aperiodic iff G is not bipartite.

Proof G is bipartite iff it does not have cycles with odd number of edges.

In an undirected G there is always a path of length 2 from $v \to v$.

If G is bipartite the RW is periodic with period 2.

If G is not bipartite it has an odd cycle, and traversing that cycle we have an odd-length path $v \to v \Rightarrow$ the Markov chain is aperiodic. $\square$

We assume the given undirected G is not bipartite and it is connected,

then the MC defined by RW on G is irreversible and aperiodic, so by the Fundamental Theorem the random walk converges to a stationary distribution π.

Random walks and Markov Chains

The next theorem shows the stationary distribution π only depends of sequence degree in G.

Theorem: A random walk on G converges to a stationary distribution $\pi = (\pi[u])_{u \in V}$, where $\pi[u] = d(u)/2m$.

Proof. First we prove π is a true distribution: For $G = (V, E)$:

$\sum_{u \in V} d(u) = 2|E| \Rightarrow \sum_{u \in V} \pi[u] = \sum_{u \in V} d(u)/2|E| = 1.$

Let P be the trasition matrix of the MC, then $\forall u \in V$,
as $\pi = \pi P \Rightarrow \pi[u] = \sum_{v \in \mathcal{N}(u)} \frac{d(v)}{2|E|} \frac{1}{d(v)} = \frac{d(u)}{2|E|}$. $\square$

As we already know that if P is regular, $h_{u,u} = 1/\pi[u]$, then
Corollary: For $u \in V$, $h_{u,u} = 2|E|/d(u)$.

Random walks and Markov Chains

Lemma: Given $G = (V, E)$, if $(u, v) \in E$ then $h_{v,u} < 2|E|$.

Proof: Let $u \in V$, from the previous corollary $h_{u,u} \frac{2|E|}{d(u)}$,
on the other hand we also know $h_{u,u} = \sum_{w \in \mathcal{N}(u)} (1 + h_{w,u})$.
Therefore, $\frac{2|E|}{d(u)} = \frac{1}{d(u)} \sum_{w \in \mathcal{N}(u)} (1 + h_{w,u})$

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

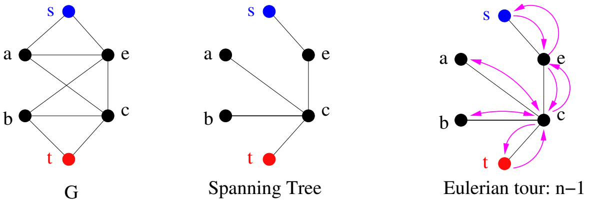
So $h_{u,u} < 2|E|$ $\square$

Random walks and Markov Chains

Corollary: Given $G = (V, E)$, its cover time $C_G \leq 4|E||V|$.

Proof: Given G with $|V| = n, |E| = m$, find a spanning tree T_G with $n-1$ edges of G, then traverse T_G using a cyclic Eulerian tour, and the number of steps to traverse it is an upper bound to C_T.

That can be done in $O(m+n)$ using BFS.



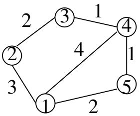
Let $v_0, v_1, \ldots v_{2n-2} = v_0$ the resulting sequence in the tour.
The expected time of going through the tour is $\leq C_G$, so

$$\sum_{i=0}^{2n-3} h_{v_i,v_{i+1}} < (2n-2)2m < 4nm.$$

Random walk on a weighted connected graph

Let $G(V, E)$ be finite connected graph with $|V| = n$, and positive weights $w_{i,j} = w_{j,i}$ for every $(i, j) \in E$ for $i, j \in V$.
Define a MC random walk on the nodes V by making

$$P_{i,j} = \frac{w_{i,j}}{\sum_k w_{i,k}}.$$



$$P = \begin{pmatrix} 0 & 3/9 & 0 & 4/9 & 2/9 \\ 3/5 & 0 & 2/5 & 0 & 0 \\ 0 & 2/3 & 0 & 1/3 & 0 \\ 3/9 & 0 & 1/6 & 0 & 1/6 \\ 2/3 & 0 & 0 & 1/3 & 0 \end{pmatrix}$$

- This chain is irreducible (by connectivity of G)

Random walk on a weighted connected graph

- To see this chain is reversible,

Solve the balance equations for $\forall i, j \in S$, using the fact that $w_{i,j} = w_{j,i}$:

$$\pi[i]P_{i,j} = \pi[j]P_{j,i} \Rightarrow \frac{\pi[i]w_{i,j}}{\sum_k w_{i,k}} = \frac{\pi[j]w_{j,i}}{\sum_k w_{j,k}} \Rightarrow \frac{\pi[i]}{\sum_k w_{i,k}} = \frac{\pi[j]}{\sum_k w_{j,k}}$$

For a constant $c > 0$, and for all $i \in S$, let $\frac{\pi[i]}{\sum_k w_{i,k}} = c$.
Then $\pi[i] = c \sum_k w_{i,k}$ and as $\sum_{i \in S} \pi[i] = 1$, then

$$c = (\sum_{i \in S} \sum_{k \in \mathcal{N}(i)} w_{i,k})^{-1} \Rightarrow \pi[i] = \frac{\sum_k w_{i,k}}{\sum_{i \in S} \sum_{k \in \mathcal{N}(i)} w_{i,k}}.$$

Notice in the case all weights are 1, $P_{i,j} = 1/d(i)$.

Algorithm to check $s - t$ connectivity in undirected G

Given a $G = (V, E)$, with $|V| = n, |E| = m$, and $s, t \in V$ we want to find a path from $s \to t$.

Deterministically we can do it in $O(n + m)$ (DFS) or Dijstra $O(m + n \lg n)$, however they need $\Omega(n)$ space.

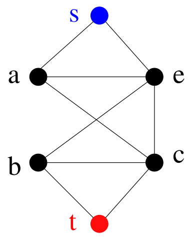
We produce a randomized algorithm, based in RW, that uses $O(n^3)$ steps and $O(\lg n)$ bits of space. At each step only needs to remember the last position, i.e. $\Theta(1)$ time.

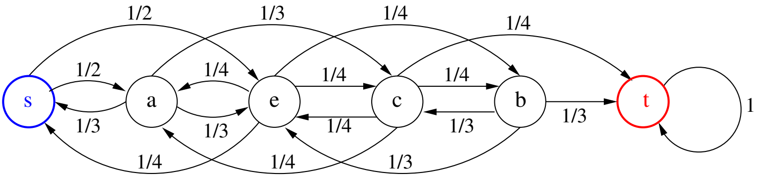
Moreover, no need of large or complicated data structure.

The clock on the number of steps is due to the fact that a Markovian RW does not know where it has visited the whole graph (because the Markovian property).

Algorithm to check $s - t$ connectivity in undirected G

s-t **Connectivity** $G = (V, E), s, t$
Start a RW from s
If the RW reaches t in $\leq 4n^3$ steps,
  there is a path $s - t$
    Otherwise there is no path





$s - t$ connectivity algorithm

Theorem The $s - t$ connectivity algorithm returns the correct answer with probability $\geq 1/2$ and in $O(n^3)$ steps using $O(\lg n)$ bits of memory.

Using Markov's inequality:
$\mathbf{Pr}$ [RW has not visited all vertices after $2C_G$ steps] $\leq \frac{1}{2}$,
by the previous corollary $C_G \leq 4|E||V| = 4nm \leq 4n^3$.

Notice if we set the clock to $200nm$ the above theorem tells us that the failure probability is reduced to $1/200$.

Markov Chain Monte Carlo technique

The Monte Carlo methods are a collection of tools for estimating values through sampling and simulations.

The Markov Chain Monte Carlo technique (MCMC) is a particular technique to sample from a desired probability distribution.
MCMC for sampling
Input: A large, but finite, set S (matching, coloring, independent sets), a weight function $w : S \to \mathbb{R}^+$;
Objective: Sample $u \in S$, from a given probability distribution given by w,

$$\pi[u] \sim \frac{w(u)}{\sum_{v \in S} w(v)}$$

Technique: Construct and ad-hoc MC which converges to the distribution we want.

Technique

Given a state space S ($|S|$ may be very large) to form the MC, which is regular (or better symmetric):

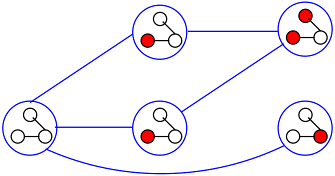
1. Connect the state space.
2. Define carefully the transition probabilities.
3. Starting at any state u do a random walk until arriving to stationary distribution π
   The simpler case is to aim for π be the uniform distribution.
4. Bound the maximal number streps we need to walk until arriving to π.

Example: Sample the set of independent vertices in G

Given a graph $G = (V, E)$ the $I \subseteq V$ is independent set if there is no edge between any two vertices in I.

Consider the Markov chain on all the set of independent subsets of V, generated by

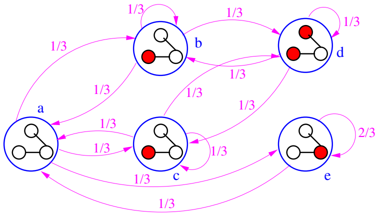
Given $G = (V, E)$
I_0 is an arbitrary independent set in G
To go from an independent set I_t to I_{t+1}
  choose u.a.r. $v \in V$
  if $v \in I_t$ then $I_{t+1} = I_t \setminus \{v\}$
  if $v \notin I_t$ and adding v still independent, $I_{t+1} = I_t \cup \{v\}$
  Otherwise $I_{t+1} = I_t$



Example: Sample the set of independent vertex in a graph

We have a $G = (V, E)$ and $n = |V|$, An we have a set S of state, each state an independent subset of V. So $|S| \sim 2^n$.
Notice in the MC every state $I \in S$ of differs from its neighbors $\mathcal{N}(I)$ in one vertex.
For $I \in S$, with probability $1/n$ choose $v \in V$:

- If $I_i \cup \{v\}$ is not independent, stay in x.
- If $\{v\}$ in I_i go to new state I_j without v.
- If $\{v\}$ is not in I_i and form an i.s. go to I_j.



$$P = \begin{array}{c|ccccc} & a & b & c & d & e \\ \hline a & 0 & 1/3 & 1/3 & 0 & 1/3 \\ b & 1/3 & 1/3 & 0 & 1/3 & 0 \\ c & 1/3 & 0 & 1/3 & 1/3 & 0 \\ d & 0 & 1/3 & 1/3 & 1/3 & 0 \\ e & 1/3 & 0 & 0 & 0 & 2/3 \end{array}$$

$$\pi = (1/5, 1/5, 1/5, 1/5, 1/5)$$

Sampling independent vertex in a G

Given $G = (V, E)$, $|V| = n$, and want to sample uniformly from all the N independent sets of vertices in G, including the set with 0 elements.

Make a random walk on a Markov chain on the finite but large state space $S = \{I_1, I_2, \ldots, I_N\}$, of all independent vertices in G.

Two states I_i, I_j are directly connected iff their size differs in one vertex, i.e. if their Hamming distance $|I_i \oplus I_j| = 1$.

Sampling independent vertex in a G

The transition matrix P:

$$P_{I_i,I_j} = \begin{cases} \frac{1}{n} & \text{if } |I_i \oplus I_j| = 1 \\ 1 - \frac{\mathcal{N}(I_i)}{n} & \text{if } |I_i| \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Notice, P is aperiodic (self-loops) and irreversible (connected) so it converges to a stationary distribution.

Moreover, as $P_{I_i,I_j} = P_{I_j,I_i}$ then P is symmetric and therefore it has a uniform stationary distribution $(1/N, 1/N, 1/N, \dots, 1/N)$.

How long do we have to go in the RW to get the stationary distribution?

Card Shuffling

Shuffling a n-card deck so we can get a uniform distribution on the set all all possible $n!$ permutations (if $n = 52 \Rightarrow 52! \sim 10^{77}$).
Ways to shuffle n-cards:

- Random Transpositions: Pick two cards uar, and switch them. Repeat.
- Card at Random to Top: Pick a card uar at put it on the top of the deck. Repeat.
- Riffle Shuffle: Split the deck into two parts according to $\text{Bin}(n, 1/2)$, and mix the two parts. Repeat.

Card Shuffling

How many repetitions do we need to "get close" to a uniform permutation?

- Random Transpositions: 100 repetitions.
- Card at Random to Top: 300 repetitions.
- Riffle Shuffle: 8 repetitions.

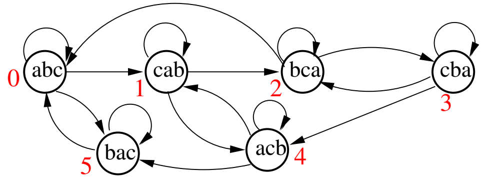
What do we mean by get close? to get with a fraction of the uniform permutation, (20%).

D. Bayers, P. Diaconis: Trailing the dovetail shuffle to its lair *Ann. Appl. Prob.*, (2):2, 294–315, 1992

Card Shuffling: Card at Random to Top

- All shuffling processes are Markov chain, with $n!$ states, and they differ in the transition probabilities.
- In the Card at Random to Top model the transition probability is $1/n$.
- Any of the $n!$ permutations can be reached from any permutation, so the chain is irreducible.
- With probability $1/n$ the state remains the same, so each state is aperiodic, and the Markov is aperiodic and so it has a unique stationary distribution.

Example: 3 cards and the MC for the shuffle model



MC for suffling
three cards abc

All transitions = 1/3

$$
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{pmatrix}
1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 1/3 & 1/3 & 0 & 0 & 1/3 \\
1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\
0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\
1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\
0 & 1/3 & 0 & 0 & 1/3 & 1/3
\end{pmatrix}
\end{array}
$$

The stationary distribution of the shuffling MC

- Recall the stationary distribution π satisfies the equation $\pi = \pi \cdot P$.
- If s is a state of the chain, denote as $\mathcal{N}(s)$ the set of states that can reach s in 1 step, then $n = |\mathcal{N}(s)|$.
- As the top card in s could have been in n different positions $n = |\mathcal{N}(s)|$: $\pi[s] = \frac{1}{n} \sum_{u \in \mathcal{N}(s)} \pi[u]$.
- Therefore as the uniform distribution satisfies these equations, then asymptotically it must coincide with the π.
- But we would like to shuffle the cards just a finite number of times.
- How many times should we shuffle until the distribution is close to uniform?