# Markov Chains

Markov chains:

- Powerful tool for sampling from complicated distributions
- rely only on local moves to explore state space.
- Many use Markov chains to *model* events that *arise* in nature.
- We *create* Markov chains to *explore* and *sample from* problems.

2SAT:

- Fix some assignment $A$
- let $f(k)$ be expected time to get all $n$ variables to match $A$ if $n$ currently match.
- Then $f(n) = 0$, $f(0) = 1 + f(1)$, and $f(k) = 1 + \frac{1}{2}(f(k+1) + f(k-1))$.
- Rewrite: $f(0) - f(1) = 1$ and $f(k) - f(k+1) = 2 + f(k-1) - f(k)$
- So $f(k) - f(k+1) = 2k + 1$
- deduce $f(0) = 1 + 3 + \cdots + (2n-1) = n^2$
- so, find with probability 1/2 in $2n^2$ time.
- With high probability, find in $O(n^2 \log n)$ time.

More general formulation: *Markov chain*

- State space $S$
- markov chain begins in a *start state* $X_0$, moves from state to state, so output of chain is a sequence of states $X_0, X_1, \ldots = \{X_t\}_{t=0}^{\infty}$
- movement controlled by matrix of *transition probabilities* $p_{ij}$ = probability next state will be $j$ given current is $i$.
- thus, $\sum_j p_{ij} = 1$ for every $i \in S$
- implicit in definition is *memorylessness property:*

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

- Initial state $X_0$ can come from any probability distribution, or might be fixed (trivial prob. dist.)
- Dist for $X_0$ leads to dist over sequences $\{X_t\}$
- Suppose $X_t$ has distribution $q$ (vector, $q_i$ is prob. of state $i$). Then $X_{t+1}$ has dist $qP$. Why?

- Observe $\Pr[X_{t+r} = j \mid X_t = i] = P^r_{ij}$

Graph of MC:

- Vertex for every state
- Edge $(i, j)$ if $p_{ij} > 0$
- Edge weight $p_{ij}$
- weighted outdegree 1
- Possible state sequences are paths through the graph

Stationary distribution:

- a $\pi$ such that $\pi P = \pi$
- left eigenvector, eigenvalue 1
- steady state behavior of chain: if in stationary, stay there.
- note stationary distribution is a sample from state space, so if can get right stationary distribution, can sample
- lots of chains have them.
- to say which, need definitions.

Things to rule out:

- infinite directed line (no stationary)
- 2-cycle (no stationary)
- disconnected graph (multiple)

Irreducibility

- any state can each any other state
- i.e. path between any two states
- i.e. single strong component in graph

Persistence/Transience:

- $r^{(t)}_{ij}$ is probability first hit state $j$ at $t$, given start state $i$.
- $f_{ij}$ is probability *eventually* reach $j$ from $i$, so $\sum r^{(t)}_{ij}$
- expected time to reach is *hitting time* $h_{ij} = \sum t r^{(t)}_{ij}$

- If $f_{ij} < 1$ then $h_{ij} = \infty$ since might never reach. Converse not always true.
- If $f_{ii} < 1$, state is *transient.* Else *persistent.* If $h_{ii} = \infty$, *null persistent.*

Persistence in finite graphs:

- graph has strong components
- *final* strong component has no outgoing edges
- Nonfinal components:
  - once leave nonfinal component, cannot return
  - if nonfinal, nonzero probability of leaving in $n$ steps.
  - so guaranteed to leave eventually
  - so, vertices in nonfinal components are transient
- Final components
  - if final, will stay in that component
  - If two vertices in same strong component, have path between them
  - so nonzero probability of reaching in (say) $n$ steps.
  - so, vertices in final components are persistent
  - geometric distribution on time to reach, so expected time finite. Not null-persistent

  Conclusion:

  - In finite chain, no null-persistent states
  - In finite irreducible chain, all states non-null persistent (no transient states)

Periodicity:

- Periodicity of a state is max $T$ such that some state only has nonzero probability at times $a + Ti$ for integer $i$
- Chain *periodic* if some state has periodicity $> 1$
- In graph, all cycles containing state have length multiple of $T$
- Easy to eliminate: add self loops
- slows down chain, otherwise same

Ergodic:

- aperiodic and non-null persistent
- means might be in state at any time in (sufficiently far) future

Fundamental Theorem of Markov chains: Any irreducible, finite, aperiodic Markov chain satisfies:

- All states ergodic (reachable at any time in future)
- unique stationary distribution $\pi$, with all $\pi_i > 0$
- $f_{ii} = 1$ and $h_{ii} = 1/\pi_i$
- number of times visit $i$ in $t$ steps approaches $t\pi_i$ in limit of $t$.

Justify all except uniqueness here.
Finite irreducible aperiodic implies ergodic (since finite irreducible implies non-null persistent)
Intuitions for quantities:

- $h_{ii}$ is expected return time
- So hit every $1/h_{ii}$ steps on average
- So $h_{ii} = 1/\pi_i$
- If in stationary dist, $t\pi_i$ visits follows from linearity of expectation

## Random walks on undirected graphs:

- general Markov chains are directed graphs. But undirected have some very nice properties.
- take a connected, non-bipartite undirected graph on $n$ vertices
- states are vertices.
- move to uniformly chosen neighber.
- So $p_{uv} = 1/d(u)$ for every neighbor $v$
- stationary distribution: $\pi_v = d(v)/2m$
- unqiqueness says this is only one
- deduce $h_{vv} = 2m/d(v)$

Definitions:

- Hitting time $h_{uv}$ is expected time to reach $u$ from $v$
- commute time is $h_{uv} + h_{vu}$
- $C_u(G)$ is expected time to visit all vertices of $G$, starting at $u$
- *cover time* is $\max_u C_u(G)$ (so in fact is max over any starting distribution).

- let's analyze max cover time

Examples:

- clique: commute time $n$, cover time $\Theta(n \log n)$
- line: commute time between ends is $\Theta(n^2)$
- lollipop: $h_{uv} = \Theta(n^3)$ while $h_{vu} = \Theta(n^2)$ (big difference!)
- also note: lollipop has edges added to line, but higher cover time: adding edges can increase cover time even though improves connectivity.

general graphs: adjacent vertices:

- lemma: for adjcaent $(u, v)$, $h_{uv} + h_{vu} \leq 2m$
- proof: new markov chain on *edge traversed following vertex MC*
  - transition matrix is *doubly stochastic:* column sums are 1 (exactly $d(v)$ edges can transit to edge $(v, w)$, each does so with probability $1/d(v)$)
  - In homework, show such matrices have *uniform* stationary distribution.
  - Deduce $\pi_e = 1/2m$. Thus $h_{ee} = 2m$.
- So consider suppose original chain on vertex $v$.
  - suppose arrived via $(u, v)$
  - expected to traverse $(u, v)$ again in $2m$ steps
  - at this point will have commuted $u$ to $v$ and back.
  - so conditioning on arrival method, commute time $2m$ (thanks to memorylessness)

General graph cover time:

- theorem: cover time $O(mn)$
- proof: find a spanning tree
- consider a dfs of tree-crosses each edge once in each direction, gives order $v_1, \ldots, v_{2n-1}$
- time for the vertices to be visited in this order is upper bounded by commute time
- but vertices adjacent, so commute times $O(m)$
- total time $O(mn)$
- tight for lollipop, loose for line.

Tighter analysis:

- analogue with electrical networks

- Assume unit edge resistance
- Kirchoff's law: current (rate of transitions) conservation
- Ohm's law
- Gives effective resistance $R_{uv}$ between two vertices.

- Theorem: $C_{uv} = 2mR_{uv}$
- (tightens previous theorem, since $R_{uv} \leq 1$)
- Proof:
  - Suppose put $d(x)$ amperes into every $x$, remove $2m$ from $v$
  - $\phi_{uv}$ voltage at $u$ with respect to $v$
  - Ohm: Current from $u$ to $w$ is $\phi_{uv} - \phi_{wv}$
  - Kirchoff: $d(u) = \sum_{w \in N(u)}$ currents $= \sum_{w \in N(u)} \phi_{uv} - \phi_{wv} = d(u)\phi_{uv} - \sum \phi_{wv}$
  - Also, $h_{uv} = \sum(1/d(u))(1 + h_{wv})$
  - same soln to both linear equations, so $\phi_{uv} = h_{uv}$
  - By same arg, $h_{vu}$ is voltage at $v$ wrt $u$, if insert $2m$ at $u$ and remove $d(x)$ from every $x$
  - add linear systems, find $h_{uv} + h_{vu}$ is voltage difference when insert $2m$ at $u$ and remove at $v$.
  - now apply ohm.