CSE5401 Lec. 6

Markov chains

1. Definition of Markov chain
2. Characterizing unique stationary distributions
3. Hitting times

In the analysis of randomized algorithms, we frequently use Markov chains

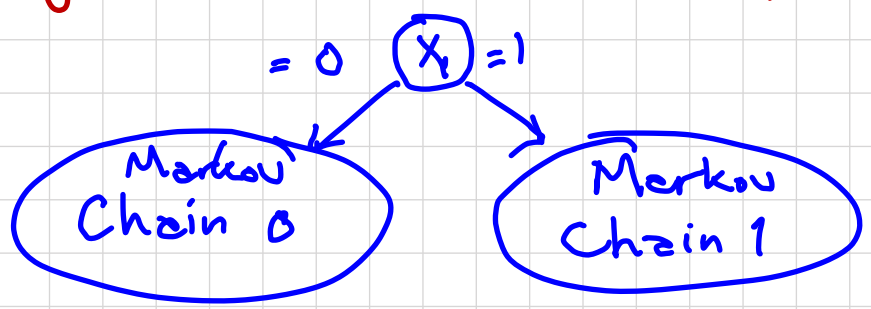
Def^n: A Markov chain is a sequence of random variables $\{X_t\}_{t=1}^{\infty}$ with the Markov property: for all sequences $x_1, \dots, x_t$,

$$Pr[X_t = x_t \mid X_1 = x_1, \dots, X_{t-1} = x_{t-1}] = Pr[X_t = x_t \mid X_{t-1} = x_{t-1}].$$

In other words: the "state" of the t^{th} random variable completely determines the distribution of the $t+1^{th}$. A random walk is an example. If every X_t takes values from a finite set (which will be the case for all of our applications) we can identify the values with integers $\{1, \dots, n\}$, and then the distribution is naturally represented by a matrix $P_{ij} = Pr[X_t = j \mid X_{t-1} = i]$, and a distribution on the values corresponds to a vector $\vec{x}^{(t)}$ with $x_i^{(t)} = Pr[X_t = i] \in [0,1]$, and $\sum_{i=1}^{n} x_i^{(t)} = 1$ $\left(\|\vec{x}^{(t)}\|_1 = 1\right)$, as with our random walks. If for all t, t' $Pr[X_t = j \mid X_{t-1} = i] = Pr[X_{t'} = j \mid X_{t'-1} = i]$ for all i, j, then we say that the Markov chain is homogeneous. In this case, we can think of the Markov chain as a random walk on a directed graph, with probabilities given by the entries P_{ij}.

We are interested in the long-run behavior of a Markov chain. In many cases, the distribution on states at some large time index will converge to a fixed distribution independent of what x_1 is. It is instructive to consider when this may fail to occur. What are some examples of Markov chains where for all t, the distribution at time t depends on (gives information about) x_1?

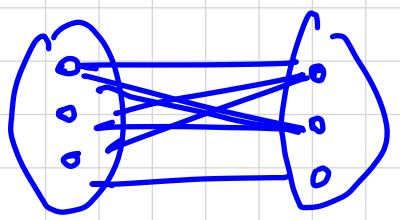
Example 1:



Not strongly connected.

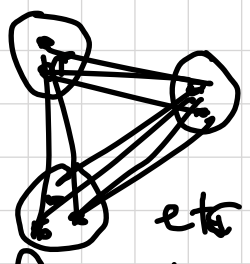
X_1 is determined by whether X_t is in the 1st comp. or the 2nd.

Example 2: Bipartite



X_1 determined to be on left or right by X_t

Ex. 2 can be generalized to cycles of any length



So we see that the chain's graph must be "aperiodic":

Def^n: We say that a Markov chain is aperiodic if for all states x $\gcd\{t : p^t(x,x) > 0\} = 1$.

The fundamental theorem of Markov chains is that these are the only two barriers. Otherwise, the distribution converges

Theorem If P is the transition matrix of a strongly connected homogeneous finite Markov chain, then there exists π^* such that for all distributions π_0, $\lim_{t\to\infty} P^t \pi_0 = \pi^*$ iff it is aperiodic.

We won't prove this in full generality, but we will prove a special case that is sufficient for us: when P is symmetric. In this case, we know that there exist n orthogonal eigen-

vectors $\vec{u}_1, \ldots, \vec{u}_n$ with real eigenvalues $\lambda_1, \ldots, \lambda_n$ for P.

Moreover, these can be found by the Rayleigh quotients:

$$\vec{u}_1 = \arg\max_{\vec{u}} \frac{\vec{u}^T P \vec{u}}{\|u\|_2^2}, \quad \vec{u}_2 = \arg\max_{\vec{u} \cdot \vec{u}_1 = 0} \frac{\vec{u}^T P \vec{u}}{\|u\|_2^2}, \ldots, \quad \vec{u}_n = \arg\min_{\vec{u}} \frac{\vec{u}^T P \vec{u}}{\|u\|_2^2}$$

where the eigenvalues are the resp. objective values.

Symmetry may seem limiting: e.g., even in an undirected graph, $P_{ij} = \frac{1}{\deg(i)}$ so symmetry requires $\deg(i) = \deg(j)$ for all i, j in the same connected component. But, we can address this by introducing an appropriate probability of remaining at state i ("adding self-loops"): let $M = 1 + \max_i \deg(i)$, and given a graph with adjacency matrix G, suppose vertex i stays put w.p. $1 - \frac{\deg(i)}{M}$ and o.w. (w.p. $\frac{\deg(i)}{M}$) transitions to j s.t. $G_{ij} = 1$ w.p. $\frac{1}{\deg(i)}$. (i.e., adding $M - \deg(i)$ self-loops). Then the Markov chain has transition matrix $P = \mathrm{diag}\left(1 - \frac{\deg(i)}{M}\right) + \frac{1}{M} G$ which is symmetric

To prove the convergence theorem, we'll use the following helpful identity:

Lemma For any symmetric transition matrix P of a homogeneous finite Markov chain, for all $\vec{x}$ $\|\vec{x}\|_2^2 - \vec{x}^T P \vec{x} = \frac{1}{2} \sum_{i,j} P_{ij} (x_i - x_j)^2$

Proof: Simply expand the RHS: $\sum_{i,j} P_{ij} (x_i - x_j)^2 = \sum_{i,j} P_{ij} x_i^2 - 2 P_{ij} x_i x_j + P_{ij} x_j^2$

Since P is a transition matrix, $\sum_j P_{ij} = 1$; since it is symmetric $\sum_i P_{ij} = 1$ also. Thus, we have $\sum_i x_i^2 - \vec{x}^T P \vec{x} + \sum_j x_j^2$, from which the claim follows directly. ▨

This lemma gives us one more important linear-algebraic characterization of a basic graph property:

Theorem For any symmetric transition matrix P of a homo-

geneous finite Markov chain, the number of eigenvectors with eigenvalue 1 equals the number of connected components

Proof First, let's suppose the graph has k connected components $C_1, \dots, C_k \subseteq \{1, \dots, n\}$. What are k orthogonal vectors with eigenvalue 1? Let $\vec{v}^{(i)} = \begin{cases} 1/|C_i| & \text{on } j \in C_i \\ 0 & \text{o.w.} \end{cases}$ $\vec{v}^{(i)} + \vec{v}^{(j)}$ are orthogonal because $C_i \cap C_j$ are disjoint. Because they are constant on each connected component, $\frac{1}{2}\sum_{i,j} P_{ij}(v_i^{(k)} - v_j^{(k)}) = 0$, so $\|\vec{v}^{(k)}\|_2 = \vec{v}^{(k)} P \vec{v}^{(k)}$, i.e., they have eigenvalue 1 by the lemma. So there are at least k such eigenvectors. To see there are at most k, consider any $\vec{v}$ s.t. $\vec{v}^T P \vec{v} = \|\vec{v}\|_2^2$ (eigenvalue 1). by the lemma, $\sum_{i,j} P_{ij}(v_i - v_j)^2 = 0$. But $P_{ij} \geq 0$, so this implies $v_i = v_j$ whenever $P_{ij} > 0$. Therefore, there exist $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ s.t. $\vec{v} = \sum_{i=1}^{k} \alpha_k \vec{v}^{(i)}$, and we see that nonzero $\vec{v}$ cannot be orthogonal to $\vec{v}^{(1)}, \dots, \vec{v}^{(k)}$. Thus, $\lambda^{(k+1)} \neq 1$. ▨

Now we can prove the main theorem. We'll first note necessity: since a symmetric P implies that any edge gives a path from i back to i of length 2, P is either bipartite (has only paths of length 2) or has an odd cycle. In the latter case, since it is connected, it is aperiodic. (since $\gcd(2, k) = 1$ for any odd k.) Clearly if P is bipartite then π_0 supported on any one i never converges. (at even/odd steps, it switches membership from one set to the other w.p. 1.) So, if P is aperiodic, it is not bipartite, and hence it does not have -1 as an eigenvalue. Since we have assumedit

to be connected, there is only one eigenvector, $v^{(1)}$, with eigenvalue 1. Note that indeed, by the lemma, $\frac{\vec{v}^T P \vec{v}}{\|\vec{v}\|_2^2} = 1 - \frac{1}{2\|\vec{v}\|_2^2} \sum_{i,j} P_{ij}(v_i - v_j)^2 \leq 1$ since $\|\vec{v}\|_2^2, P_{ij}, (v_i - v_j)^2 \geq 0$. Furthermore, we similarly have $\frac{1}{2}\sum_{i,j} P_{ij}(v_i + v_j)^2 = \|\vec{v}\|_2^2 + \vec{v}^T P \vec{v}$ so $\frac{\vec{v}^T P \vec{v}}{\|\vec{v}\|_2^2} = -1 + \frac{1}{2\|\vec{v}\|_2} \sum_{i,j} P_{ij}(v_i + v_j)^2 \geq -1$, hence $\lambda_n \geq -1$. As we saw on the homework (essentially), if P is not bipartite, $\lambda_n \neq -1$. Thus, $\lambda = \max\{|\lambda_2|, |\lambda_n|\} < 1$. Thus, as in the homework, the projection of $P^t \pi^{(0)}$ on $\{\lambda_2, \ldots, \lambda_n\}$ shrinks by λ^t, and hence $\|\vec{v}^{(1)} - P^t \pi^{(0)}\|_2 \leq \lambda^t \|\pi^{(0)}\|_2 \to 0$ as $t \to \infty$. ▨

In particular, observe that the distribution converges to a uniform distribution on states. This will be important later. Another characterization of the limiting distribution is in terms of the # steps to return to a vertex:

Defn: The hitting time of j from i, h_{ij}, is the expected # of steps before the homogeneous finite Markov chain started at i reaches state j.

Theorem For an aperiodic, strongly connected, homogeneous finite Markov chain, $\pi^* = \lim_{t\to\infty} P^t \pi^{(0)}$ satisfies $\pi_i = \frac{1}{h_{ii}}$ for all i.

Proof: First observe that for all i,j, $h_{ij} = P_{ij} + \sum_{k\neq j} P_{ik}(1 + h_{kj})$.

$$= 1 + \sum_{k\neq j} P_{ik} h_{kj}$$

Let $D = \text{diag}(h_{ii})$ and N be given by $N_{ii} = 0$, $N_{ij} = h_{ij}$ for all $i \neq j$. (Thus, $(D+N)_{ij} = h_{ij}$ for all i,j) Our equations may be written: $D + N = E + PN$ (where $E_{ij} = 1$ for all i,j). Algebra yields $(I-P)N = E - D$, so $\pi^{*T}(I-P)N = \pi^{*T}(E-D)$. But $\pi^{*T}P = \pi^{*T}$, so $\underbrace{\pi^{*T}(I-P)}_{=0}N = \pi^{*T}(E-D)$.

Therefore, $(\pi_1^* h_{11}, \ldots, \pi_n^* h_{nn}) = (\underbrace{\sum_j \pi_1^*}_{1}, \ldots, \underbrace{\sum_i \pi_n^*}_{1})$ ▨

This allows us to bound the time it takes for a random walk to visit any vertex in a connected graph

Lemma If $(u,v) \in E$, for the random walk on G, $h_{u,v} + h_{v,u} \leq 2|E|$

Proof: Let's represent the random walk on G as a Markov chain on edges, where if the original random walk transitions from u to v on step t, then our edge chain is in state (u,v). This chain has $2|E|$ states, and its limiting distribution is uniform. Thus, from (u,v), the expected # steps to return to (u,v) is $2|E|$ by our lemma. This is a path $v \to u$ followed by (u,v) back to v, so indeed, $h_{v,u} + h_{u,v} \leq 2|E|$ ▨

Theorem For any initial state, the expected # of steps to visit every vertex of a connected graph is $\leq 2|E|(|V|-1)$.

Proof: Consider any spanning tree T of G. There is a path that traverses each edge in T twice (once in each direction). The expected time to visit the vertices of G in this order starting from the given v_0 $(v_0, v_1, \ldots, v_{2(|V|-1)})$ is

$$\sum_{i=0}^{2|V|-3} h_{r_i, v_{i+1}} = \sum_{(u,w) \in T} h_{u,w} + h_{w,u} \leq (|V|-1) \cdot 2|E| \text{ by the lemma.}$$ ▨

One application of this is a very space-efficient algorithm for deciding if there is a path between a given pair of vertices s & t: take a random walk from s. If we reach t in $4|E|(|V|-1)$ steps, a path exists; and conversely if one exists, the probability we fail to reach t is $< \frac{1}{2}$ by Markov's inequality. We can amplify this by independent repetition.