CS5330: Optional problems on Markov Chains

April 10, 2020

Here are solution sketches to the optional problems on Markov chains. If anything is unclear, please talk to me or your TA.

1. For any transition matrix P, let $Q = (I + P)/2$. Argue that the Markov chain with transition matrix Q is aperiodic.

   For any i and j, if $P^t_{i,j} > 0$, then $Q^{t'}_{i,j} > 0$ for any $t' \geq t$. This is because the Markov chain using Q can spend $t' - t$ steps looping at i before taking the path of length t from i to j.

2. Consider the 2-state Markov chain that stays at the current state with probability p and moves to the other state with probability $1 - p$. Write down the transition matrix P, and find a simple expression for P^t.

   The transition matrix $P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$. The state after t steps is the same as the initial state exactly if there were an even number of flips among the t steps. Therefore, $P^t_{0,0} = P^t_{1,1} = \sum_{\text{even } i} \binom{t}{i} p^{t-1}(1-p)^i = \frac{1}{2}(1 + (2p-1)^t)$. (The last equality follows because $1 = \sum_{i=0}^t \binom{t}{i} p^{t-i}(1-p)^i$ and $(2p-1)^t = (p - (1-p))^t = \sum_{i=0}^t \binom{t}{i}(-1)^i p^{t-i}(1-p)^i$.)
   Hence, $P^t = \begin{bmatrix} (1 + (2p-1)^t)/2 & (1 - (2p-1)^t)/2 \\ (1 - (2p-1)^t)/2 & (1 + (2p-1)^t)/2 \end{bmatrix}$.

3. (a) For any matrix A, show that A and $A^\top$ have the same set of eigenvalues. (**Hint**: Use the fact that the eigenvalues of A are the roots of the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$, and also that the determinants of a matrix and of its transpose are the same. Show that the polynomials p_A and $p_{A^\top}$ are identical.)

   Note that $p_{A^\top}(\lambda) = \det(A^\top - \lambda I) = \det((A - \lambda I)^\top) = \det(A - \lambda I) = p_A(\lambda)$. So, A and $A^\top$ have the same set of eigenvalues.

(b) P is a *stochastic matrix* if it has non-negative entries and each of its rows sums to 1. Show that there exists a vector π such that $\pi P = \pi$.

If P is stochastic, then $uP^\top = u$ where u is the all-ones vector. Hence, 1 is an eigenvalue of $P^\top$, and therefore P also by part (a).

4. A *doubly stochastic* matrix is a stochastic matrix (see previous question) in which additionally all the columns sum to 1. Show that the uniform distribution is a stationary distribution for any Markov chain having a doubly stochastic transition matrix.

If the columns of P sum to 1, then $uP = u$ where u is the vector $\left[\frac{1}{n}, \dots, \frac{1}{n}\right]$.

5. Let $h_{\max}$ be the maximum hitting time between any pair of vertices in an n-vertex graph G. Show that the time for a random walk to visit every vertex is $O(h_{\max} \log n)$ with high probability. Conclude that the cover time is $O(h_{\max} \log n)$.
(**Hint**: Break a random walk of length $2k \cdot h_{\max}$ into k segments of length $2h_{\max}$. For any fixed vertex i, argue that i is visited in each segment with probability at least $1/2$. Set k so that each vertex is visited in some segment with high probability. To bound the cover time, use that $\mathbb{E}[X] = \sum_{k\geq 0} \Pr[X \geq k]$ for non-negative random variables X.)

Fix a vertex i. By Markov's inequality, a random walk of length $2h_{\max}$ visits i with probability at least $1/2$, no matter what the stating state is. So, the probability that i is not visited in a walk of length $4h_{\max}\log n$ is at most $(1/2)^{2\log n} = 1/n^2$. By the union bound, with probability $1 - 1/n$, each vertex is visited in a walk of length $4h_{\max}\log n$.

Let T denote the number of steps before every vertex is visited by a random walk. By the above argument, $\Pr[T \geq 2kh_{\max}] \leq n2^{-k}$. Therefore,

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{t=0}^{5h_{\max}\log n} \Pr[T \geq t] + \sum_{t > 5h_{\max}\log n} \Pr[T \geq t] \\
&\leq 5h_{\max}\log n + \sum_{i=1}^{\infty} \sum_{t=5ih_{\max}\log n}^{5(i+1)h_{\max}\log n - 1} \Pr[T \geq t] \\
&\leq 5h_{\max}\log n + \sum_{i=1}^{\infty} 5h_{\max}\log n \cdot \frac{n}{2^{5ih_{\max}\log n}} \\
&\leq 5h_{\max}\log n + \sum_{i=1}^{\infty} O(n^3 \log n) \cdot \frac{1}{n^{5i-1}}
\end{aligned}$$

where in the last line, we used the bound derived in class that $h_{\max} = O(n^3)$. It is easily seen that the last sum is bounded by $o(1)$. Hence, $\mathbb{E}[T] \leq 6h_{\max}\log n$.

6. Let $h_{\min}$ be the minimum hitting time between any pair of distinct vertices in an n-vertex graph G. The goal of this problem is for you to show that

$$C(G) \geq \Omega(h_{\min} \cdot \log n)$$

where $C(G)$ is the cover time of the graph.

(a) Consider a random walk $X_0, X_1, \ldots$ where the initial state $X_0 = x$ is arbitrary. Choose a random permutation $\pi : [n] \to [n]$. For a state i, let T_i be the first time that all the states $\pi(1), \pi(2), \ldots, \pi(i)$ have been visited. Show that:

$$\mathbb{E}[T_1] \geq \left(1 - \frac{1}{n}\right) h_{\min}$$

With probability $1/n$, $\pi(1) = x$, and in this case, $T_1 = 0$. In all other cases, $\mathbb{E}[T_1 \mid \pi(1) \neq x] \geq h_{\min}$. Hence, $\mathbb{E}[T_1] \geq \frac{1}{n} \cdot 0 + \left(1 - \frac{1}{n}\right) \cdot h_{\min}$.

(b) Observe that the probability that $\pi(i)$ is visited after states $\pi(1), \ldots, \pi(i-1)$ is $\frac{1}{i}$. Using this, argue that $T_i - T_{i-1} = 0$ with probability $1 - \frac{1}{i}$.

Suppose for some $1 \leq k \leq i$, $\pi(k)$ is the last visited vertex among $\{\pi(1), \ldots, \pi(i)\}$. Then, by definition, $T_k = T_{k+1} = \cdots = T_i$. Hence, unless $k = i$, $T_{i-1} = T_i$.

(c) Conditioned on $\pi(i)$ being visited after $\pi(1), \ldots, \pi(i-1)$, show that $\mathbb{E}[T_i - T_{i-1}] \geq h_{\min}$.

Let $\pi(k)$ be the last visited vertex among $\{\pi(1), \ldots, \pi(i-1)\}$. Then, $T_{i-1} = T_k$. Furthermore, if $\pi(i)$ is visited after $\pi(k)$, then $T_i - T_{i-1} = T_i - T_k$ is the length of the walk between $\pi(k)$ and $\pi(i)$. No matter what k is, the expected length of this walk is at least $h_{\min}$.

(d) Conclude that:

$$\mathbb{E}[T_n] \geq \left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right) \cdot h_{\min}$$

From the previous parts, $\mathbb{E}[T_i - T_{i-1}] \geq \frac{1}{i} h_{\min}$. Hence, $\mathbb{E}[T_n] = \mathbb{E}[T_1] + \sum_{i=2}^{n} \mathbb{E}[T_i - T_{i-1}] \geq \left(1 - \frac{1}{n}\right) h_{\min} + \sum_{i=2}^{n} \frac{1}{i} h_{\min} = \left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right) h_{\min}$. This is called the *Matthews' bound*.

7. Show that if μ and ν are two distributions on $[n]$ with probability mass functions $f : [n] \to \mathbb{R}$ and $g : [n] \to \mathbb{R}$ respectively,

$$\|\mu - \nu\| = \frac{1}{2} \sum_{i=1}^{n} |f(i) - g(i)|$$

By definition, $\|\mu - \nu\| = \max_{S\subseteq[n]} \sum_{i\in S}(f(i) - g(i))$. It's clear that the maximizing set $S = \{i : f(i) > g(i)\}$. So, $\|\mu - \nu\| = \sum_{i:f(i)>g(i)}(f(i) - g(i))$. Note that $\sum_{i:f(i)>g(i)} f(i) - g(i) = \sum_{i:f(i)\le g(i)} g(i) - f(i)$, and so, $\sum_i |f(i) - g(i)| = 2\sum_{i:f(i)>g(i)} f(i) - g(i)$, which implies our claim.

8. Show that the Markov chain for k-coloring graphs of maximum degree Δ discussed in class is irreducible, if $k \ge \Delta + 2$. Moreover, prove that the stationary distribution of the Markov chain is the uniform distribution on k-colorings.

   Recall that each move in the Markov chain is to pick a random vertex v from the graph, a random color $c \in [k]$, and to color v with c if permitted and to otherwise leave the coloring unchanged.

We can check that the uniform distribution is stationary by verifying the time-reversibility conditions. Let P be the transition matrix for the Markov chain. If f and g are two distinct k-colorings such that that $P_{f,g} > 0$, then $P_{f,g} = P_{g,f} = \frac{1}{nk}$. If $f = g$, time-reversibility is trivial.

We now verify irreducibility. Consider two proper k-colorings f and g, and fix an ordering of the vertices. Attempt to recolor the vertices in this order to match f and g. Suppose you get stuck, in the sense that you are trying to recolor a vertex v from $f(v)$ to $g(v)$, but there is a vertex w later in the order which is a neighbor of v, and $f(w) = g(v)$. So, we cannot recolor v from $f(v)$ to $g(v)$. But in this case, consider an intermediate coloring in which w is recolored to a color different from the current colors of its neighbors as well as $g(v)$. This is possible if $k > \Delta + 1$. After all such neighbors w are recolored, we can then recolor v from $f(v)$ to $g(v)$ as desired, and continue.

9. Consider the following random walk on the hypercube $\{0,1\}^n$: with probability $1/(n+1)$, stay at current vertex; otherwise, with probability $1/(n+1)$ for each of the n neighbors, go to one of the neighbors. Note that the self-loop probability is $1/(n+1)$.

   An alternative way to view the walk is that for current state x, a random $i \in \{0,1,\dots,n\}$ is picked uniformly at random. If $i = 0$, x doesn't change; otherwise, x_i is flipped.

   Consider the following coupling (X_t, Y_t).

   - Suppose X_t and Y_t differ at only one coordinate i_0. Then, if X_t picks $i = 0$, Y_t picks i_0; if X_t picks i_0, then Y_t picks $i = 0$; else, both X_t and Y_t pick the same i.
   - Suppose X_t and Y_t differ at the subset of coordinates $S \subseteq [n]$, where $|S| > 1$. Fix a bijection $\pi : S \to S$ such that $\pi(i) \ne i$ for all $i \in S$. Then, if X_t picks $i = 0$, then Y_t also picks $i = 0$; if X_t picks $i \notin S$, then Y_t also picks i; if X_t picks $i \in S$, then Y_t picks $\pi(i)$.

   Observe that the distance between X_t and Y_t never increases. Analyze separately the expected time needed for the distance to decrease to 1 and then the expected time for the distance to go from 1 to 0. Use this to give a bound on the expected coupling time and, hence, the mixing time for this Markov chain.

Let Δ_t be the number of coordinates X_t and Y_t differ. We first claim that if $\Delta_t > 1$, $\mathbb{E}[\Delta_{t+1} \mid \Delta_t] = \Delta_t\left(1 - \frac{2}{n+1}\right)$. This is true because with probability $\frac{1}{n+1}$ (when both pick $i = 0$) plus

$\frac{n-\Delta_t}{n+1}$ (when both pick $i \notin S$), $\Delta_{t+1} = \Delta_t$, and with probability $\frac{\Delta_t}{n+1}$, $\Delta_{t+1} = \Delta_t - 2$. Hence, $\mathbb{E}[\Delta_{t+1} \mid \Delta_t] = \frac{1+n-\Delta_t}{n+1}\Delta_t + \frac{\Delta_t}{n+1}(\Delta_t - 2) = \Delta_t \cdot (1 - \frac{2}{n+1})$. Hence, $\Pr[\Delta_t > 1] \le \mathbb{E}[\Delta_t] \le n\left(1 - \frac{2}{n+1}\right)^t$. Let T_1 be the first time Δ_t becomes 1. The above implies that $\mathbb{E}[T_1] = O(n \log n)$.

Now if $\Delta_t = 1$, Δ_{t+1} becomes 0 with probability $2/(n+1)$ (when X_{t+1} is either 0 or the coordinate where the two chains differ). So, T_0, the first time Δ_t becomes 0, is T_1 plus a geometric random variable with parameter $2/(n+1)$. Hence, $\mathbb{E}[T_0] = O(n \log n) + (n+1)/2 = O(n \log n)$. So, $t_{\mathrm{mix}}(1/3) = O(n \log n)$ and hence, $t_{\mathrm{mix}}(\varepsilon) = O(n \log n \log 1/\varepsilon)$.