CSE 312: Foundations of Computing II — Winter 2021

Problem Set 8 (due Wednesday, March 10 at 11:59pm)

Directions:

Answers: *For each problem, remember you must briefly explain/justify how you obtained your answer, as correct answers without an explanation will receive* ***no credit****. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. It is fine for your answers to include summations, products, factorials, exponentials, or combinations; you dont need to calculate those all out to get a single numeric answer, for instance* 26^7 *or* $26!/7!$ *or* $26 \cdot \binom{26}{7}$.

Your solutions need to be concise and clear. We will take off points for lack of clarity or for excess verbosity. Please see section worksheet solutions (posted on the course website) to gauge the level of detail we are expecting.

Please clearly indicate your final answer, in such a way as to distinguish it from the rest of your explanation.

Groups: *This homework is to be completed in groups of 1 or 2. Specific guidelines about collaboration are available on the syllabus, but every group will be submitting their own submission. Please cite any collaboration at the top of your submission. Instructions are available here as to how to add groupmates to your submission.*

Before you start your homework, write down the list of people you collaborated with. Remember, you can only collaborate outside your group by discussing the problems at a high level. Provide names and email addresses.

Submission: *You must upload a* ***pdf*** *of your written solutions to Gradescope under "PSet 8 [Written]". Problem 6 is a coding problem, so under "Pset 8 [Coding]" you will upload a .py file called cse312_pset8_markov_chains.py. There is* ***no extra credit*** *on this pset. (Instructions as to how to upload your solutions to Gradescope are on the course web page.) The use of latex is highly recommended.*

Note that if you want to hand-write your solutions, you'll need to scan them. We will take off points for hand-written solutions that are difficult to read due to poor handwriting and neatness.)

Please cite any collaboration at the top of your submission (beyond your group members, who should already be listed).

1. Markov's Inequality (10 points)

Give an example of a discrete random variable for which Markov's Inequality is tight in the following sense.

(a) [5 Points] To begin, describe the pmf for a nonnegative random variable X with mean 10, such that

$$Pr(X \geq 5 \cdot E(X)) = \frac{1}{5}.$$

(b) [5 Points] Next, describe the pmf for a nonnegative random variable X such that given any constant $c > 1$

$$\Pr(X \geq c \cdot E(X)) = \frac{1}{c}.$$

Your construction should be parameterized by c.

Hint: You can decide on a positive mean for this RV as well, and find a PMF to match.

2. MLE-1 (10 points)

Suppose that $x_1, \ldots, x_n$ are i.i.d. realizations from a probability density function

$$f_X(x) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the Maximum Likelihood Estimate for θ. (You don't need to check second order conditions.)
You cannot use Wolfram Alpha or any other graphing calculator to solve this problem. You must show your work.

3. MLE-2 (10 points)

Individuals in a certain country are voting in an election between 3 candidates: A,B and C. Suppose that independently each person votes for candidate A with probability θ_1, for candidate B with probability θ_2 and for candidate C with probability $1 - \theta_1 - \theta_2$. (Thus, $0 \leq \theta_1 + \theta_2 \leq 1$.) The parameters θ_1, θ_2 are unknown. Suppose that $x_1, \ldots, x_n$ are n independent, identically distributed samples from this distribution. (Let $n_A =$ number of x_is equal to A, let $n_B =$ number of x_is equal to B, and let $n_C =$ number of x_is equal to C.) What are the maximum likelihood estimates for θ_1 and θ_2 in terms of n_A, n_B, and n_C? (You don't need to check second order conditions.)
You cannot use Wolfram Alpha or any other graphing calculator to solve this problem. You must show your work.

4. Markov Chains (15 points)

This topic is covered in class on Wednesday 3/3. A **discrete-time stochastic process (DTSP)** is a sequence of random variables $X_0, X_1, X_2, \ldots$, where X_t is the value at time t. For example, the temperature in Seattle or stock price of TESLA each day, or which node you are at after each time step on a random walk on a graph.

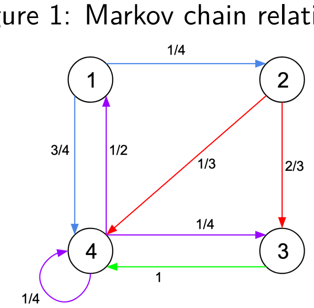
Formally, a discrete time finite **Markov Chain** is a DTSP, with the additional following three properties:

I. ...has a finite **state space** $\mathcal{S} = \{s_1, \ldots, s_n\}$ which it bounces between, so each $X_t \in \mathcal{S}$.

II. ...satisfies the **Markov property**. A DTSP satisfies the Markov property if the future is (conditionally) independent of the past given the present. Mathematically, it means, $P(X_{t+1} = x_{t+1} | X_0 = x_0, X_1 = x_1, \ldots, X_{t-1} = x_{t-1}, X_t = x_t) = P(X_{t+1} = x_{t+1} | X_t = x_t)$.

III. ...has **transition probabilities** that do not vary over time. Meaning, if we are at some state s_i, we transition to another state s_j with probability *independent* of the current time. Due to this property and the previous, the transitions are governed by n^2 probabilities: the probability of transitioning from one of n current states to one of n next states. These are stored in a square $n \times n$ **transition probability matrix (TPM)** P, where $P_{ij} = P(X_{t+1} = s_j | X_t = s_i)$ is the probability of transitioning from state s_i to state s_j for any/every value of t.

Figure 1: Markov chain relationship



In this problem we will explore the Markov chain shown in Figure 1: It is perhaps useful to think of this Markov chain as a probabilistic finite automaton.

(a) [6 Points] The Markov chain associated with Figure 1 has the TPM shown in Figure 2. This means that

Figure 2: Transition Probability Matrix P

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

if, for example, we are in state 4 at some time t (i.e. $X_t = 4$), then at the next time step we will be in state 1 with probability 1/2, state 3 with probability 1/4 or state 4 with probability 1/4. Use the law of total probability to compute $P(X_2 = 4 \mid X_0 = 1)$. Show your work.

(b) [3 Points] Suppose we weren't sure where we started. That is, let $q = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ be such that $P(X_0 = i) = q_i$, where q_i is the i^{th} element of the vector q (i.e., we start at one of the 4 states uniformly at random). Think of this vector q as our belief distribution of where we are at time $t = 0$. First, compute qP, the matrix-product of q and P, the transition probability matrix. (Notice that q is a row vector that is multiplying P on the left.) What do the entries in qP represent? (Notice that qP is the following 4-dimensional row vector):

$$qP = \left(\sum_{i=1}^{4} q_i P_{i1}, \quad \sum_{i=1}^{4} q_i P_{i2}, \quad \sum_{i=1}^{4} q_i P_{i3}, \quad \sum_{i=1}^{4} q_i P_{i4} \right)$$

Give your answer to the first question as **4 simplified fractions**.

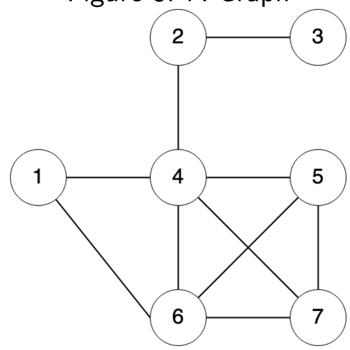
(c) [6 Points] The **stationary distribution** of a Markov chain is the $|\mathcal{S}|$-dimensional row vector π such that the matrix equation $\pi P = \pi$ holds (and π is a valid probability mass function, i.e., $\sum_{i \in \mathcal{S}} \pi_i = 1$). For our example, $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ is 4-dimensional, and contains 4 probabilities which sum to 1. The intuition/interpretation of π is that it gives the probabilities of being in each state in the "long run". That is, for t large enough, $\pi_i = P(X_t = i) = P(X_{t+1} = i) = P(X_{t+2} = i) = \ldots$. (See the next problem for more on this.) Using your answer to the previous part, explain in 1-2 sentences why $\pi P = \pi$ is called the stationary distribution, and solve for it.

5. Random Walk Analysis (15 points)

[**Written**] In Figure 1, there is an undirected graph with 7 nodes.

(a) [3 Points] Suppose we perform a random walk on the graph from figure 1, starting at node 3. This means that if we are at a particular node i in the graph at some time step, then in the following time step, we will transition to each of its neighbors with equal probability. So if it has one neighbor, we transition there with probability 1. If it has 3 neighbors, we transition to each of those with probability 1/3, and so on. This defines a Markov chain whose states are the nodes in the graph and $X_t = i$ if after t steps the random walk is at vertex number i. Fill out the 7×7 transition probability matrix P in Figure 4 with simplified fractions; we've filled out the first row for you (available in the LaTeX template linked alongside this pset). The row represents the current state and the column represents the next state. So the first row represents our transition probabilities *from* state 1 to all 7 states, and so on.

Figure 3: A Graph



(b) [4 Points] Suppose that $Pr(X_t = j | X_0 = 3) = q_j^t$, and let the vector $q^t = (q_1^t, q_2^t, q_3^t, q_4^t, q_5^t, q_6^t, q_7^t)$. Thus, we have, for example,

$$q^0 = (0, 0, 1, 0, 0, 0, 0)$$
$$q^1 = (0, 1, 0, 0, 0, 0, 0)$$

and

$$q^2 = (0, 0, 1/2, 1/2, 0, 0, 0).$$

Recall that we can compute the vectors q^t using vector matrix products as follows:

$$q^{t+1} = q^t P.$$

Again, note that on the right hand side, we are multiplying a row vector by a matrix.

Your coding assignment will be to compute these vectors. This will allow you to numerically observe the convergence to the stationary distribution.

At each time $t \geq 1$, in your code, also compute

$$\Delta_t = \sum_{i=1}^{7} (q_i^t - q_i^{t-1})^2.$$

Note that Δ_t is measuring how far apart the distribution over states is at time $t-1$ from the distribution over states at time t.

For each $\epsilon \in \{0.5, 0.1, 0.01, 0.001, 0.000001\}$, we will report the value of the first time t such that

$$\Delta_t \leq \epsilon.$$

As your response to this question, include the relevant times, the distributions at those times, and the $\epsilon = 0.000001$ graph produced by the wrapper code.

Complete the coding assignment in Problem 6 to answer this question and part d.

(c) [6 Points] Consider a random walk on an undirected graph with n vertices and m edges, where the degree of vertex i is d_i. (The degree of a vertex is the number of neighbors it has in the graph. So in the graph above, for example, $d_1 = 2$ and $d_4 = 5$.) Prove that

$$\pi_i = \frac{d_i}{2m}$$

is the stationary distribution of the random walk. That is, show that for $\pi_i = \frac{d_i}{2m}$ it holds that $\sum_{i=1}^{n} \pi_i = 1$ and that $\pi = \pi P$, where $\pi = (\pi_1 \ldots, \pi_n)$ and P is the transition probability matrix of the Markov chain.

This shows, for example, that for the random walk on the graph shown in Figure 3.

$$\pi = \left(\frac{2}{20}, \frac{2}{20}, \frac{1}{20}, \frac{5}{20}, \frac{3}{20}, \frac{4}{20}, \frac{3}{20}\right).$$

Make sure your proof that $\pi_i = \frac{d_i}{2m}$ works for an arbitrary undirected graph , and not just for our specific Markov Chain.

(d) [2 Points] Finally, please compute again, for each of the times you found in part (b), the value of

$$\Delta_t^\pi,$$

where

$$\Delta_t^\pi = \sum_{i=1}^{7}(q_i^t - \pi_i)^2.$$

Provide a 2 sentence discussion/interpretation of what you've found.

Figure 4: Transition Probability Matrix P of Random Walk

$$P = \begin{bmatrix} 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \\ TODO & TODO & TODO & TODO & TODO & TODO & TODO \end{bmatrix}$$

6. Random Walk (5 points)

[**Coding**] Implement the functions __init__() walk(), and delta() in cse312_pset8_markov_chains.py. The rest of the code is provided as wrapper code to evaluate when you pass the ϵ thresholds for Δ and to generate helpful plots. You do not have to worry about the main function or the plot function.

Details as to the implementation of these functions is given as part of the Markov Chain introduction in Problem 5. The starter code has hints and guidelines in the comments, so be sure to read it carefully.

There will be no autograding for this assignment, either on Gradescope or Ed. **You should still turn in your code on Gradescope under "Pset 8 [Coding]", however it will not be automatically graded.** Instead, we will manually check that you did in fact write your own code as a pair, and that the graphs and data provided as part of Problem 5 are accurate.