Prob 140 Fall 2019 MIDTERM EXAM **A. Adhikari**

NAME (FIRST LAST): ________________________________ **SID:** ________________

BUILDING (circle one): VLSB/Dwinelle **SEAT NUMBER:** ________________

TIME AND CONDITIONS: 75 minutes; closed book/notes/internet; no calculator/computer

QUESTIONS AND ANSWERS

• There are 6 questions.

• **Give brief explanations or show calculations in each part of each question** unless the question says this is not required. You may use, without proof, any result proved or used in lecture, the textbook, homework, and labs, unless the question asks for a proof.

• You may answer any part of any question. If the answer to one part depends on another that you couldn't do, you can still provide an answer such as "The answer to part (a), divided by 2."

• **Please simplify all infinite sums.** Please leave all finite sums unsimplified, whether they are numerical or algebraic, unless the question requires a simplification.

GRADING

• The exam is worth 50 points. Each of the first 5 questions is worth 8 points. Question 6 is worth 10 points. Points for parts are indicated in the questions.

• We will give partial credit, but only for substantial progress towards a correct answer. We get to decide what "substantial progress" means.

• Please commit yourself to a single answer for each part of each question. If you give multiple answers (such as both True and False), please don't expect credit, even if the right answer is among those that you gave.

FORMAT

• There is a space for your name and SID number above each question. Please fill this in. It will ensure that we can identify your work during the scanning process.

• There is space for your answer below each question. **Please do not write outside the black boundary**; the scanner and Gradescope won't read it.

• If you need scratch paper please use the backs of the pages of the exam. But be aware that **they will neither be scanned nor graded**.

HONOR CODE

Data Science and the entire academic enterprise are based on one quality – integrity. We are all part of a community that doesn't fabricate evidence, doesn't fudge data, doesn't present other people's work as our own, doesn't lie and cheat. You trust that we will treat you fairly and with respect. We trust that you will treat us and your fellow students fairly and with respect. **Please abide by UC Berkeley's Honor Code:**

"As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others."

Your signature: ____

Name and SID: ______________________________

1. There are 100 participants in a randomized controlled experiment. Participants are assigned to the treatment and control groups in a way that ensures that there are at least 20 partcipants in each group. The randomization process is as follows.

- **Stage 1:** A fair coin is tossed for each participant. If the coin lands heads the participant is assigned to the treatment group, and if it lands tails the participant is assigned to control. If after all 100 tosses there are at least 20 participants in each group, the randomization is complete. If not, the process continues to Stage 2.

- **Stage 2 (if needed):** If the randomization is not complete at Stage 1, the process begins afresh. Participants are assigned to treatment and control based on new fair coin tosses as described above. If after all 100 tosses there are at least 20 participants in each group, the randomization is complete. If not, the process continues to Stage 3.

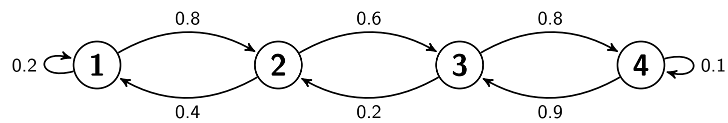
- **Subsequent Stages (if needed):** The process continues in this way, one stage at a time, until the randomization is completed.

a) [**5 points**] What is the chance that the randomization is completed in Stage 1?

b) [**3 points**] Let R be the number of stages required for the randomization to be completed. Find $E(R)$.

Name and SID: ____

2. Let $X_0, X_1, X_2, \ldots$ be a Markov Chain with time-homogeneous one-step transition probabilities displayed the diagram below. Suppose $P(X_0 = 1) = 1$.



a) [2 points] Is there a probability distribution π that satisfies the detailed balance equations for this chain? Why or why not?

b) [2 points] Fill in the blank with a number and show your calculations:

In the long run, the expected proportion of time the chain spends in State 2 is _______ times the expected proportion of time the chain spends in State 1.

c) [4 points] In which of the four states do you think X_{500} is most likely to be? Justify your answer with calculations.

Name and SID: ____________________

3. At a grocery store, the number of people in line at the checkout counter at 10 a.m. depends on the day of the week.

• On each weekday (Monday, Tuesday, Wednesday, Thursday, and Friday) the number of people in line at 10 a.m. has the Poisson (3) distribution, independently of all other days.

• On each weekend day (Saturday, Sunday) the number of people in line at 10 a.m. has the Poisson (6) distribution, independently of all other days.

Each day, each person in line brings their own grocery bag with chance 0.9, independently of all other people and days.

Suppose you pick one of the seven days next week at random, independently of the people in the store. Let G be the number of people who are in line at 10 a.m. that day and bring their own grocery bag.

a) [**4 points**] Find $E(G)$.

b) [**4 points**] Find $P(\text{you pick Thursday} \mid G = 5)$.

Name and SID: ____________________

4. A bowl of Halloween candy contains 40 Fun Size Snickers bars. There are 10 bars of each of four different kinds of Snickers: Original, Peanut Butter, Almond, and Crisper. Bars are drawn out one at a time at random without replacement. Let D be the number of draws till all four kinds have appeared.

a) [**4 points**] For each $n = 0, 1, 2, \ldots$, find $P(D > n)$.

b) [**2 points**] Find $E(D)$.

c) [**2 points**] Find the chance that the kind of bar that appears on Draw D is Original.

Name and SID: ______________________________

5. Let X have the Poisson (μ) distribution.

a) [3 points] Find $E\left(\frac{1}{X+1}\right)$.

b) [5 points] Find $E\left(\frac{1}{X+2}\right)$.

[Hint: First simplify the difference $\frac{1}{n+1} - \frac{1}{n+2}$ for a non-negative integer n.]

Name and SID: ______________________________

6. Let $X_1, X_2, \ldots, X_n$ be a random permutation of $1, 2, 3, \ldots, n$. That means the sequence $X_1, X_2, \ldots, X_n$ is equally likely to be any of the $n!$ permutations of $1, 2, 3, \ldots, n$. Here are two features of random permutations. You encountered the first one in your code-breaking lab.

- Transpositions: There is a *transposition* at a pair of indices $i \neq j$ if $X_i = j$ and $X_j = i$.
- Records: There is a *record* at an index $i > 1$ if $X_i > X_k$ for all $k = 1, 2, \ldots, i-1$. To take care of the edge case, the definition also says that there is always a record at index $i = 1$. For example, in the permutation 253461978 of the integers 1 through 9, the records are at indices 1, 2, 5, and 7.

a) [**4 points**] Find the expected number of transpositions in the random permutation $X_1, X_2, \ldots, X_n$.

b) [**4 points**] Fix an integer $i > 1$. Find the chance that the random permutation $X_1, X_2, \ldots, X_n$ has a record at index i.

c) [**2 points**] Find the expected number of records in the random permutation $X_1, X_2, \ldots, X_n$.