CS 70 Summer 2020 — Discrete Mathematics and Probability Theory

Amin Ghafari, Yining Liu, Khalil Sarwari

Final

- You may consult three handwritten double-sided sheets of notes. Apart from that, you may not look at books, notes, etc. Calculators, phones, computers, and other electronic devices are prohibited unless they are part of the recording submission.
- There are 9 questions on this exam, worth a total of 100 points.
- No clarification will be provided on the exam questions.
- Note that the questions vary in difficulty
- You may, without proof, use theorems and facts that were proven in the lecture, notes, discussions, and/or in homeworks unless explicitly mentioned otherwise.
- **You have 150 minutes to work on the exam. You will then have 45 minutes for scanning and uploading your answers. Late submissions will be penalized.**
- **You may use up to** $\min(x, 15)$ **minutes of the scanning time to continue working on your exam if you have lost** x **minutes due minor technical issues during the exam.**

1 (19 Points) Short Answers I

Unless otherwise stated, you must show all your work in order to get full credit.

(a) (4 points) Prove the following statement using **induction**: given $n \in \mathbb{N}$, if S is a set of cardinality n, then the power set of S has cardinality 2^n.

(b) (3 points) For each of the following sets, **state** whether it is finite, countably infinite, or uncountable. **No need to justify/show work**.

(i) The set of prime numbers

(ii) The set of infinite sequences of integers

(iii) The set of computer programs

(c) (2 points) Show that if $k > 0$ is an integer, then $k+1$ is coprime to $k^2 + 2k$.

(d) (2 points) Compute the following: $\left(\sum_{i=1}^{10} i^{16}\right) \mod 17$.

(e) (3 points) Solve the following system of congruences for x (i.e. solve for the unique solution modulo 60).

$$\begin{aligned} x &\equiv 2 \pmod{3}, \\ x &\equiv 3 \pmod{4}, \\ x &\equiv 4 \pmod{5}. \end{aligned}$$

(f) (3 points) Let p, q and r be polynomials of degree at most 2 over $GF(23)$ such that

$$\begin{array}{lll} p(1) = 1 & p(2) = 5 & p(3) = 4 \\ q(1) = 3 & q(2) = 7 & q(3) = 15 \\ r(1) = 1 & r(2) = 3 & r(3) = 1 \end{array}$$

Let $f(x) = p(x) + 2q(x) - 5r(x)$. Find $f(x)$. **Simplify your answer for full credit.**

Hint: You do not need to calculate the polynomials p, q, r.

(g) (2 points) Alice wants to send Bob a message of length n while guarding against k_e erasure errors and k_g general errors. How many total packets does she need to send?

2 (15 Points) Short Answers II

Unless otherwise stated, you must show all your work in order to get full credit.

(a) (2 points) We roll a 6-sided fair die 4 times. How many possible strictly ascending sequences of numbers are there?

(b) (3 points) We choose n numbers from $\{1, 2, \ldots, 9\}$ uniformly at random with replacement and define x as the product of these numbers. What is the probability that x is divisible by 35? (n is a positive integer.)

(c) (3 points) Give a combinatorial proof for the following equation. (An algebraic proof receives 0 points.)

$$\binom{n}{2} = \sum_{i=1}^{n-1} i$$

(d) (3 points) How many subsets of $\{1, 2, ..., 2n\}$ are there that do not contain any elements x and y satisfying the equation $x + y = 2n + 1$? For example, $x = 1$ and $y = 2n$ cannot be in the subset at the same time. (n is a positive integer.)

(e) (2 points) We have n fair coins and m biased coins in a bag. The biased coins land on heads with probability p. Without looking, we pick one uniformly at random and flip it. It lands heads. What is the probability that it is a fair coin, given that it landed heads? Leave your answer in terms of n, m and p.

(f) (2 points) Let $X \sim \text{Poisson}\left(\frac{1}{2}\right)$. Calculate $\mathbb{E}[X!]$. (Note: we really mean $X!$, as in "the factorial of X".) Recall that for $|r| < 1, \sum_{i=0}^{\infty} ar^i = \frac{a}{1-r}$

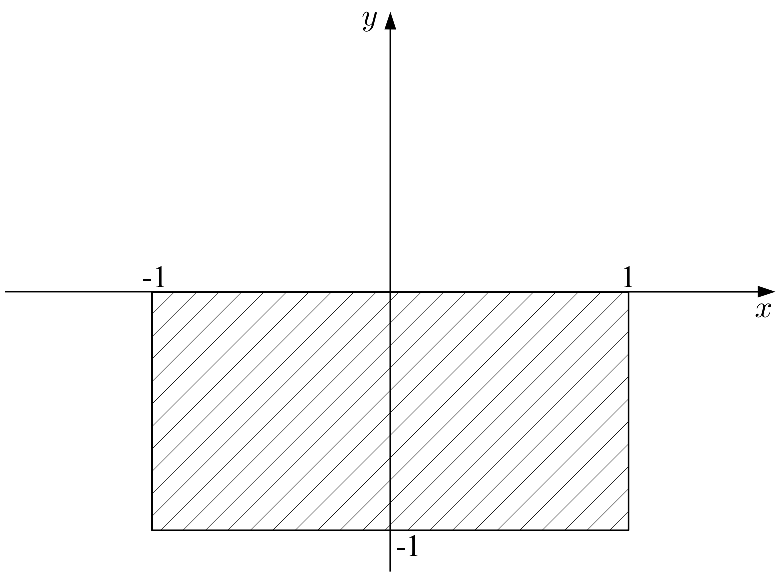
3 (6 Points) Joints

Let X and Y be two continuous random variables. The joint density of (X,Y) is uniform on the shaded region below, and 0 outside the shaded region. Mathematically, the figure consists of a rectangle.

(a) (1 point) What is the joint density $f_{X,Y}$ on the shaded region?

(b) (2 points) Set up, but do not evaluate the integrals for the values of $f_X(x)$ and $f_Y(y)$ on the shaded region.

(c) (3 points) Are X and Y independent? **Justify your answer.**



4 (5 Points) Cubes

Suppose you have a $3 \times 3 \times 3$ inch cube of solid brown wood. You paint all 6 faces white and chop it up into $1 \times 1 \times 1$ inch subcubes. There are 27 subcubes in total. You then toss all of these in a bag, and with your eyes closed, you take one out and roll it. Opening your eyes, you notice that the 5 faces that are showing are brown. What's the probability that the face you can't see (i.e. 6^{th} face) is also brown? **Show all your work for full credit.**

5 (12 Points) If It's Any Indication

Let G be an undirected graph on n vertices where each of possible edge of the graph is included with probability p, independent from every other edge. This means that for each edge e of the $\binom{n}{2}$ total edges, that edge e is a part of G with probability p, and is missing from the graph with probability $1-p$. A vertex is called *isolated* if it is adjacent to no vertices of the graph.

For the following parts, your answer should be an expression in terms of n and p. Show all your work for full credit.

(a) (2 points) Find the expected degree of a vertex.

(b) (2 points) Find the variance in the degree of a vertex.

(c) (4 points) Find the expected number of isolated vertices.

(d) (4 points) Find the variance of the number of isolated vertices.

6 (10 Points) Markov Chains

Kevin and Shahzar are playing a game with 2 bins. On each turn, Shahzar picks a bin uniformly at random. If the bin already has a ball in it, Shahzar does nothing. If the bin doesn't have a ball, Shahzar throws a ball into that bin**. On the same turn, Kevin picks a bin uniformly at random, independently of Shahzar, and empties it (removes any balls that are in the bin).

** Note that this means the maximum number of balls in any bin is 1.

(a) (4 points) For a particular bin, construct a two-state Markov chain for the number of balls it contains at the end of a turn. **Clearly indicate the states and transition probabilities of the chain.**

(b) (2 points) Does the Markov chain converge to a unique invariant distribution? **Justify your answer.**

(c) (4 points) Find the invariant distribution(s) of the Markov chain. **Show all your work for full credit.**

7 (11 points) Distributed Distribution

Robel took his Organic Chemistry exam and wanted to see the distribution of scores. Since the web portal site was down, Robel was unable to access the true distribution. Robel decided to survey students, and get an estimate of the distribution for himself.

Let n be the number of students who fill out his survey. For each student i who fills out his survey, the score they report X_i follows $X_i \sim \text{Exp}\left(\frac{1}{\mu}\right)$, where μ is the true mean score on the exam. All X_i are independent and identically distributed (i.i.d.), with mean μ and variance μ^2. Note that this means the reported scores can be anywhere in $[0, \infty)$.

For the following parts, show all your work for full credit.

(a) (2 points) Robel was looking through the survey results as they came in and saw a score above 90. However, he does not remember the exact score. What is the expected value of the reported score he saw, given it was above 90? You may use μ in your answer.

Robel is now interested in estimating the true mean of the distribution. It then makes sense to consider $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, which can be interpreted as the sample mean of the reported scores from the survey.

(b) (2 points) **Determine $\mathbb{E}[\hat{\mu}_n]$, and $\text{Var}(\hat{\mu}_n)$.**

(c) (3 points) Using Chebyshev's inequality, find the tightest upper bound on the probability that the sample mean $\hat{\mu}_n$ is at least 10 points above the true mean μ.

(d) (4 points) Robel knows that 100 students filled out his survey. He later finds out from the web portal site that the true mean was actually 50. Using the Central Limit Theorem as a means of approximation, **What the probability that $\hat{\mu}_n$ is *not* within 1 point of the true mean $\mu = 50$? Leave your answer in terms of the standard normal CDF Φ.**

8 (11 points) Graphs, Polynomials, and Counting: The Ultimate CS70 Crossover!

Recall that a graph is n-colorable if you can color the vertices using n colors such that no adjacent vertices have the same color. Now instead of just validity, we can count also *how many* colorings (if any) exist. We define $P_G(x)$ as the number of ways of coloring a graph G with x colors. For example, if $P_G(n) = 0$ for a graph G and a positive integer n, then G is not n-colorable (there are zero ways to n-color G).

It turns out that this function is always a polynomial for a given graph G. Assume vertices are distinguishable. **For each of the following parts, show all your work for full credit.**

(a) (2 points) Let G be the triangle graph (a complete graph of 3 vertices). Calculate $P_G(5)$.

(b) (4 points) Let G be a complete graph of n vertices.

  (i) Find the polynomial $P_G(x)$.

  (ii) Use the polynomial you found in part (i) to show that the minimum number of colors needed to color the complete graph of n vertices is n.

(c) (3 points) Let G be a tree of n vertices. Find the polynomial $P_G(x)$.

(d) (2 points) Let G be a graph of n vertices that contains at least one edge. Prove that the sum of the coefficients in $P_G(x)$ is 0.

9 (11 points) Lineage Tracing

Consider the following model for tracking cell division. A petri dish begins with a single white cell. At the start of each time step $t \geq 1$, every cell in the petri dish will divide into two cells that inherit its color**. Then, each white cell in the petri dish will become green with probability p, independently. Then the timestep ends.

**This means at the end of any timestep t, there are 2^t cells in the dish.

For each of the following parts, show all your work for full credit.

(a) (2 points) What is the probability that a particular cell at the end of timestep n is green?

(b) (3 points) The **descendants** of a cell are the two cells that it divided into and all of their descendants. A cell is an **ancestor** of any of its descendants. Note that the first cell in the petri dish is the ancestor of all cells.

Suppose we observed that a cell is green at the end of time step n. Given this observation, what is the probability that the green mutation occurred in an ancestor of the observed cell at timestep t, where $t \leq n$?

(c) (3 points) What is the expected number of green cells at the end of time step n? Give your answer as an expression in terms of n and p.

(d) (3 points) Suppose $p > 1/2$. Prove that as n tends to ∞, the probability that every cell at time step n is green tends to 1.

Submission

- Keep the recording going;
- Scan answer booklet and cheatsheets into PDF;
- Submit to Gradescope by 11:15PM PDT. If Gradescope is being really slow, you may submit your exam PDF using a private Piazza post and attaching your pdf to the private post;
- Stop the recording;
- Upload recording to your Google Drive, Box, DropBox, etc;
- Submit the link to your uploaded recording using this form:
  https://forms.gle/qnY76ZQs9Rjx77qL6
  by 11:59PM 8/14 PDT latest;
- Submit your cheatsheets to Gradescope under "Final Cheat Sheets". If you did not use any cheatsheets, you must submit 6 blank pages. If you already submitted the cheatsheets and did not make any changes, then you do not need to resubmit.

If you have technical issues during the exam, you should report these issues when you submit your exam by making a private post on Piazza. You should attach your exam pdf to this post if you have it.

You are strong!