CS 237: *Probability in Computing*
Boston University

Professors: Alina Ene, Tiago Januario
November 22, 2024

Final Exam Announcement

Times and Places: Tuesday, December 17^{th}, 3pm-5pm:

HAR105 - Regular time room (2-hour exam)

IECB12 - Extended time room based on each student's accommodation

Crib sheet: The exam is *closed-book* and *closed-devices*. You may bring one double-sided *hand-written* crib sheet on an $8\frac{1}{2} \times 11$ sheet of paper. (Writing on a computer and then printing is not allowed. Printed crib sheets can be collected from students at any point during the final exam) Preparing a crib sheet can be a useful study aid, so take time in selecting material for it.

Exam procedure: When you enter the room, we will ask you to leave your bag and coat at the podium and direct you to one of the seats. **We ask that you don't seat next to your friends / collaborators.** You will be allowed to take only your crib sheet, pens, pencils and erasers with you. We will provide scrap paper. **Please turn off your cell phone for the duration of the exam and leave it in your bag.** If you have questions at any point during the exam, raise your hand, and a staff member will assist you. If you are done with the exam before the exam period is over, please hand in your exam to an instructor, pick up your stuff and quietly leave the room.

Material covered: You are responsible for material covered in all the lectures, assigned reading for the lectures, all homework assignments, and all discussions. The course home page has a record of the material covered in lectures. Many useful materials, including discussion handouts and homework solutions are posted on Piazza under Resources.

Practicing for the exam: Review lecture slides (including Top Hat questions), homework problems, and discussion handouts. Make sure you have tried all solved exercises for the topics we covered in our textbook (P). **At least one of the problems mentioned in the two previous sentences will appear on the exam.**

Practice problems in this handout are not intended to provide a comprehensive review of the topics covered on the final. This handout is not representative of the length of the exam.

Skills:

- You should be comfortable with the concepts from probability we studied and be able to use them to solve problems of the level you have seen in homework assignments.
- You should be able to recognize the named distributions we discussed (e.g., Bernoullli, Binomial, Geometric, Negative Binomial, Uniform, Exponential, and Poisson) and be able to figure out the corresponding parameters.
- You should be able to calculate the probability of an event and find the PMF/PDF, CDF, the expectation, and the variance of a random variable.
- You should be able to justify every step using the laws of probability we studied.

- You should be able to use Markov and Chebyshev inequalities.
- Review the applications we studied and be ready to analyze variants of these applications.
- Review material covered in the mathematical prerequisites, including combinatorics, arithmetic and geometric sums, derivatives, and integrals.

Final Exam Study Questions

Students are encouraged to post their solutions to these practice problems to Piazza. Please use a separate Piazza post for each problem.

1. (**Birthday Party**) Ali throws a birthday party and invites eleven friends to join him to get bubble tea. The twelve of them head to Chatime, and each of them chooses a flavor at random independent of each other. They all pick a flavor from the following three: taro milk tea, matcha sunrise and mango green tea.
   (a) What is the probability that no one gets the taro milk tea?
   (b) What is the probability that exactly one person gets the matcha sunrise?
   (c) What is the probability that everyone gets the same tea?

2. (**XOR-ing bits**) The exclusive or (XOR, denoted as $\oplus$) of two bits is equal to 1 if exactly one of the bits is 1, and it is equal to 0 otherwise. We independently toss a fair coin 3 times to obtain 3 bits b_1, b_2, b_3 (we map a heads to 1 and a tails to 0). Let $Y_1 = b_1 \oplus b_2$, $Y_2 = b_2 \oplus b_3$, and $Y_3 = b_1 \oplus b_3$.
   (a) Find the PDF of each random variable Y_1, Y_2, Y_3.
   (b) Are Y_1, Y_2, Y_3 pairwise independent? Justify your answer.
   (c) Are Y_1, Y_2, Y_3 mutually independent? Justify your answer.

3. (**Brunch**) You go to Tatte for brunch with your significant other. They get a Croissant Sandwich with probability 1/15, a Breakfast Sandwich with probability 2/15, an Avocado Tartine with probability 2/5 a Multigrain Avo Smash Tartine with probability 1/5 and a Winter Shakshuka with probability 1/5. The pricing of the products is given below:
   - Breakfast / Croissant Sandwhiches : \$9.50,
   - Avocado Tartine / Multigrain Avo Smash Tartine : \$10,
   - Winter Shakshuka: \$13.

   Let X be the price of your s.o.'s order.
   (a) Find the PDF of X,
   (b) Find the CDF of X,
   (c) Find $\mathbb{E}(X)$,
   (d) Find $\text{Var}(X)$.

4. (**Dice**) In all of the parts, assume that each die is a fair 6-sided die with the numbers $1, 2, 3, 4, 5, 6$ written on its sides and that all tosses are independent.

(a) We roll two dice. Find the expected value of the sum of the two numbers obtained.

(b) We roll two dice. Find the variance of the sum of the two numbers obtained.

(c) We repeatedly roll two dice and take the sum of the two numbers obtained. What is the probability that we get a sum of 4 before a sum of 7?

(d) We repeatedly roll two dice and take the sum of the two numbers obtained. What is the expected number of rolls until we get a sum of 4 or a sum of 7? (For example, if we get a sum of 4 on the first roll, the number of rolls is 1.)

5. (**Grade**) It's finals season and 100 students are taking the CS 111 final this year. For each student i let X_i be the random variable defined by:

$$X_i = \begin{cases} 1 & \text{if student } i \text{ gets an A in the exam} \\ 0 & \text{otherwise} \end{cases}$$

Let $X = \sum_i X_i$. Note that the value of X equals the number of students who get an A on the exam. We model each X_i as a Bernoulli random variable with parameter $p = 1/3$, and X as a binomial distribution with parameters $p = 1/3$ and $n = 100$.

(a) Find $\mathbb{E}(X_i)$ for any i.

(b) Find $\mathbb{E}(X)$.

(c) Now we consider a similar set up for course projects, where students are allowed to collaborate, and the probability of each student getting an A is $p = 1/2$. Argue that the binomial distribution is no longer a good model for the value of X.

(d) Prove that, despite part (c), $\mathbb{E}(X)$ still equals np (with $p = 1/2$).

6. (**Lottery**) Ludmila buys a lottery ticket every day. She alternates between buying a Massachusetts lottery ticket and a USA Lottery ticket. For instance, if she buys an MA ticket on Monday, she will buy a USA ticket on Tuesday, and again an MA ticket on Wednesday and so on. Ludmila will stop playing the lottery as soon as she wins a prize. Every MA lottery ticket is a winning ticket with probability p_{MA} independently of any other ticket, and every USA lottery ticket is a winning ticket with probability p_{USA} independently of any other ticket. Let X be the number of tickets Ludmila buys. Suppose that on the first day she buys an MA lottery ticket.

(a) Find $\Pr(X = 1)$.

(b) Find $\Pr(X = 2)$.

(c) Note that this distribution satisfies: $\mathbb{E}(X \mid X > 2) = \mathbb{E}(X) + 2$. Use this fact and the law of total expectation to prove that:

$$\mathbb{E}(X) = \frac{2 - p_{MA}}{p_{MA} + p_{USA} - P_{MA} p_{USA}}$$

7. (**Waiting Time**) Alice goes to the post office to mail a letter, and she is equally likely to find either 0 or 1 customers ahead of her. The service time for the customer ahead, if present, is exponentially distributed with parameter λ. Find the CDF of Alice's waiting time.

8. (**Randomly Chosen Variables**) You are making random variables for your computer simulation, and you want to randomly create each variable's distribution.

   (a) You roll a fair 4-sided die to decide the distribution of random variable X. The numbers you can roll and their resulting distributions are as follows:

   1: $Geometric(\frac{1}{11})$
   2: $Poisson(\frac{22}{7})$
   3: $Exponential(\frac{5}{16})$
   4: $Binomial(15, \frac{1}{4})$

   Based on this setup, find $\mathbb{E}(X)$.

   (b) You now spin a spinner that can land on all numbers in $(0, 10]$ uniformly at random. If the spinner lands on t, you create a new random variable Y such that $Y \sim Exp(\frac{1}{2t^2})$. Find $\mathbb{E}(Y)$.

9. (**Getting Ready**) You are going to hit Tavern (formerly Tavern in the Square) with your two best friends: Janette and Ayush. You're ready to go but your friends are still getting ready. The time T_J (measured in minutes) it takes Janette to get ready follows an exponential distribution with parameter $\lambda_J = 1/3$. On the other hand, Ayush, who's a perfume enthusiast, takes time $T_A \sim \text{Exponential}(\lambda_A = 1/5)$. T_J and T_A are independent of each other.

   (a) You plan to leave as soon as both your friends are ready, what is the probability that it will take longer than 4 minutes to leave your place?

   (b) You remember that you heard from your friend Hao that both Ayush and Janette were talking about you behind your back, allegedly saying your new boots are "not as *turnt* as you think" (while actually they are *poppin' fresh*). At this point, you just want to make sure you don't show up at Tavern alone, and so you decide to leave with whichever "friend" is ready first. What is the probability you will wait longer than 3 minutes?

10. (**Normal Distribution**) Let $X \sim \text{Normal}(0, 1)$ and $Y \sim \text{Normal}(1, 4)$.

    (a) Find $\Pr(X \leq 1.5)$, $\Pr(X \leq -1)$, and $\Pr(-1 \leq Y \leq 1)$.

    (b) Find the PDF of the random variable $\frac{1}{2}(Y - 1)$.

    (c) Let $Z = 2X + Y$. Find $\Pr(0 \leq Z \leq 2)$.

11. (**Randomized Algorithm**) Random Company Inc. has designed a randomized algorithm that seems to have a strange behavior. If we run the algorithm more than once on the same data, it has different running times. The running time of the algorithm (in seconds) has expectation 10 and variance 25. We run the algorithm for 1 minute and, if the algorithm does not terminate by that time, we stop it and restart it from scratch. Each time we restart, we get an independent run of the algorithm.

    (a) Use Markov inequality to give an upper bound on the probability that we have to restart the algorithm (in one run).

    (b) Use Chebyshev's inequality to give an upper bound on the probability that we have to restart the algorithm (in one run).

(c) Give an upper bound the probability that we run the algorithm more than 10 times.

12. (**Random variable**) Let X be a nonnegative random variable with $\mathbb{E}(X^3) = k$. Prove that $\Pr(X \geq a) \leq \frac{k}{a^3}$ for any $a > 0$.

13. (**Inequalities**) Let $X_1, ..., X_n$ be independent Bernoulli random variables with parameters $p_1, \ldots, p_n$, respectively. Let $X = \sum_{i=1}^{n} X_i$.

    (a) Find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ in terms of $p_1, \ldots, p_n$.

    For each of the following parts, decide which of the inequalities from class (Markov, Chebyshev, or the union bound) to apply, and give the smallest upper bound on the probability that you can (in terms of $p_1, \ldots, p_n$).

    (b) Give an upper bound on $\Pr(X \geq 2)$.

    (c) Give an upper bound on $\Pr(X \geq 2 + \mathbb{E}(X))$.

14. (**The Candy Bar Problem**) A candy bar is made up of $2k+1$ unit-length blocks arranged in a line. You choose one of the $2k$ boundaries between two blocks uniformly at random and cut the bar at this boundary. Let the random variable X be the length of the *longer* of the two resulting pieces.

    (a) Compute $\mathbb{E}(X)$ in the case $k = 2$.

    (b) Compute $\mathbb{E}(X)$ as a function of k. Check your answer against the value you obtained in part (a).

    (c) Compute the variance $\mathrm{Var}(X)$ as a function of k. (You may use the fact that the sum of the squares of the first m positive integers is $\sum_{i=1}^{m} i^2 = \frac{1}{6}m(m+1)(2m+1)$.)

    (d) Use Chebyshev's inequality together with parts (b) and (c) to derive an upper bound on the probability that the longer piece has length at least $\frac{15k+4}{8}$.

15. (**Continuous Distributions**)

    Let X be a continuous random variable with the following PDF:

$$f(x) = \begin{cases} 1/3 & \text{if } -2 \leq x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

    (a) Find $\Pr(X = -1)$.

    (b) Sketch the graph of the PDF above.

    (c) Find the CDF of X.

    (d) Find $\Pr(1/2 \leq X \leq 1)$.

16. (**Algorithms involving Randomness**)

    (a) Consider the reservoir sampling algorithm from lecture, and recall that it stores a single item in memory. Suppose the algorithm has processed the first 10 items of the stream. Compute the probability that the item stored in memory is the first item.