Oishi Banerjee and Cooper Raterink Summer 2020

CS109 Midterm Quiz

Take-Home Quiz information

Each quiz will be a 24-hour open-book, open-note, open-calculator exam. Quiz material will approximate about 2-3 hours of active work (*before* typesetting), though it may be more depending on how you prepare your submission. Unlike the problem sets, the take-home quizzes are ***strictly individual work***. Even course staff assistance will be limited to clarifying questions of the kind that might be allowed on a traditional, in-person exam. If you have questions during the exam, please ask them via our discussion forum. We will not have any office hours for answering quiz questions during the quiz.

For each problem, briefly explain/justify how you obtained your answer at a level such that a future CS109 student would be able to understand how to solve the problem. Please derive your mathematical expressions analytically, not via computerized simulation. For example, please use mathematical principles to solve counting problems, rather than writing a brute-force program to generate all possible outcomes. (That said, you may use code to *check* your analytically-derived answers.)

It is fine for your answers to be well-defined mathematical expressions including finite summations (but not integrals), products, factorials, exponentials, and combinations, unless the question *specifically* asks for a numeric quantity. Where numeric answers are required, the use of fractions is fine. You may use a calculator (e.g. WolframAlpha) to evaluate your mathematical expressions.

Honor Code Guidelines for Take-Home Quizzes

The take-home exams are open-book (open lecture notes, handouts, textbooks, course lecture videos, and internet searches for conceptual information, e.g., Wikipedia). Consultation of other humans in any form or medium (e.g., communicating with classmates, asking questions on forum websites such as StackOverflow) is prohibited. All work done with the assistance of any external material in any way (other than provided CS109 course materials) must include citation (e.g., "Referred to Wikipedia page on DeMorgan's Law for Question 2."). Copying solutions is unacceptable, even with citation. If by chance you encounter solutions to the problem, navigate away from that page before you feel tempted to copy. If you become aware of any honor code violations by any student in the class, your commitments under the Stanford Honor Code obligate you to inform course staff. *Please remember that there is no reason to violate your conscience to complete a take-home exam.*

Submission

You should upload your submission as a PDF to Gradescope. We provide a LaTeX template if you find it useful, but we will accept any legible submission. You can submit multiple times; we will only grade the last submission you submit before the deadline$^{\text{th}}$. Late submissions will not be accepted. ***Please double-check that you submit the right file.*** When uploading, PLEASE assign pages to each question.

I acknowledge and accept the letter and spirit of the Honor Code:

Name (typed or written): ____

1 Genetic Possibilities [18 points]

You're a cutting-edge geneticist creating animal hybrids in an entirely safe and ethical manner! You can make hybrids of these seven animals: turtles, ducks, platypi, bears, lions, bison and dragons. (*Note:* a turtle-duck is the same as a duck-turtle. In other words, the order in which you choose your species doesn't affect the end result.)

a. (2 points) If all hybrids are possible, how many different types of 3-animal hybrids could you make?

b. (4 points) If you randomly generate a 2-animal hybrid (where all possible 2-animal hybrids are equally likely), what are the chances you generate a hybrid that's part-bison and/or part-bear?

c. (6 points) You create 10 unique hybrid animals, but now you aren't sure what to feed them. You have 5 identical portions of fish and 15 identical portions of chips. How many distinct meal plans can you create for your animals? *Note:* You must give out all the food, distributing it in complete portions, but you don't need to feed every animal. Two meal plans are distinct if any animal receives a different amount of fish and/or chips in the two plans.

d. (6 points) You stop making hybrids and create a Multinomial Animal Generator. When you click Go, it randomly chooses a species and makes an animal of that species. Every time you run the machine, each type of small animal (turtle, duck, platypus) has a 20% chance of being chosen. Each type of large animal (bear, lion, bison, dragon) has a 10% chance of being chosen. If you run the machine 15 times, what is the probability you end up with exactly 3 animals from one species and exactly 2 animals of each of the other species?

2 Ticket Sales [22 points]

You manage the box office of a small but popular island theater. You are starting to sell tickets for tonight's performance. On average, you sell 1 ticket every 5 minutes.
(*Note:* For full credit, don't include any infinite summations in your answers. Finite summations are okay.)

a. (4 points) What's the probability that you sell no tickets in the next twenty minutes?

b. (6 points) You have 150 tickets to sell for a performance in 12 hours. What's the probability that you'll run out of tickets before tonight's performance? Please give a numerical answer. Assume selling the 150th ticket means you've run out of tickets.

c. (6 points) If the theater puts on 100 performances every year, what is the probability you'll run out of tickets at least 20 times this year? You may refer to your answer from part b as p. Please give an exact answer in terms of p (do not approximate).
Note: You may assume that each show's tickets sales are independent from the sales for other performances. Also assume that the probability that any single show will run out of tickets is equal to part b's answer.

d. (6 points) Give a numerical estimate of your answer from (c), using an appropriate random variable for your approximation.

3 Misleading Averages [16 points]

In a large city with questionable government ethics, there are three separate neighborhoods.

In this city, many residents live in the Lower Ring neighborhood; we define L as the event where a resident lives in the Lower Ring. A resident's in the Lower Ring has yearly income (in amount of gold earned per year) represented by the following PDF, where c is a mystery constant:

$$f(x) = \begin{cases} c * e^{10-0.25x} & \text{if } 40 < x < 60 \\ 0 & \text{else} \end{cases}$$

The remaining residents (L^C) live in two other neighborhoods. A third of these residents live in the Upper Ring, each making exactly 500 gold per year, while two-thirds live in the Middle Ring, each making exactly 200 gold per year.

According to city officials, everyone in the city is extremely prosperous and happy, because the average income of city residents is 100 gold coins per year. In other words E[I] = 100, where I is a random variable representing a resident's income. Your job is to find out more about the demographics of the city.

a. (6 points) What is the value of the constant c in the PDF given above?

b. (4 points) What is the average income of residents living in the Middle or Upper Ring? That is, what is $E[I|L^C]$?

c. (6 points) What fraction of the population lives in the Lower Ring? That is, what is P(L)? (*Hint:* When computing $E[I|L]$, use WolframAlpha to make things way easier. This link may help: https://www.wolframalpha.com/input/?i=integral+from+a+to+b+of+f%28x%29)

4 Pentapox Outbreak [26 points]

(*Hint:* This problem requires you to juggle multiple probabilities, some of which will only be useful on the later sub-problems. We recommend taking notes while reading through the problem statement and defining your events/random variables clearly.)

The world is plagued by a sudden outbreak of Pentapox! Some cases of Pentapox have no externally visible symptoms. Local scientists administer a simple two-step screening test to all outwardly asymptomatic patients. Based on prior studies, they estimate that *10% of all test subjects have Pentapox.*

- **Step 1: Blood Pressure Check.** Pentapox tends to raise patients' blood pressure, so in the first step of the test doctors check each test subject's blood pressure. Here's what we know:
  - Doctors say a patient has "high" blood pressure if their diastolic blood pressure is above 80.
  - Among subjects *without* Pentapox, diastolic blood pressure is normally distributed with a mean of 70 and a variance of 25.
  - Doctors estimate that 80% of subjects *with* Pentapox have high blood pressure.

a. (6 points) What is the probability a randomly chosen subject *without* Pentapox has high blood pressure? Please find a numerical answer.

- **Step 2: Breathing Check.** Pentapox also makes breathing more difficult, so in the second step doctors test each subject's lung function and rate it as either "healthy" or "poor." Here's what we know:
  - 25% of all subjects have poor lung function.
  - 89% of subjects with Pentapox have poor lung function.

b. (6 points) What is the probability that a subject with poor lung function has Pentapox?

c. (6 points) What is the probability that a subject without Pentapox has poor lung function?

- **Test Result.** The screening test gives a positive result for Pentapox if the subject has both poor lung function *and* high blood pressure. Otherwise the result is negative.[1]
  - *Note:* Having poor lung function is conditionally independent from having high blood pressure, given whether a patient has Pentapox or not.

d. (8 points) What is the probability that a subject who tests positive actually has Pentapox? You may refer to the answers of prior sub-problems as a, b, and c respectively.

[1]Science is difficult, especially during a public-health crisis! This screening test is not necessarily the best it could be. As a completely optional not-for-credit exercise, you can keep poking at the numbers we've given here to find a screening test you think is better.