Homework 4: Random Variables

Version 3: We corrected the example in Problem 2; X is negative when there are more $3-6$'s than there are $1-2$'s. We corrected line 2 in the example for problem 6 (a student can only get a not-already-given-out test)

For each problem, remember you must briefly explain/justify how you obtained your answer, as correct answers without an explanation will not receive full credit. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide.

In general, your goal in an explanation is to write enough that a student from class who has attended lecture, but not read the problem yet, could understand your approach, verify your reasoning, and believe your answer is correct. While we do not usually need to see arithmetic, you must include enough work that in principle one could rederive your answer with only a scientific calculator.

Unless a problem states otherwise, you should leave your answer in terms of factorials, combinations, etc., for instance 26^7 or $26!/7!$ or $26 \cdot \binom{26}{7}$ are all good forms for final answers.

Instructions as to how to upload your solutions to gradescope are on the course web page.

Remember that you must tag your written problems on Gradescope.

Submission: You must upload a **pdf** of your written solutions to Gradescope under "HW 4 [Written]". (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.)

Due Date: This assignment is due at 11:59 PM Wednesday April 28 (Seattle time, i.e. GMT-7).
You will submit the written problems as a PDF to gradescope. Please put each numbered problem on its own page of the pdf (this will make selecting pages easier when you submit), and ensure that your pdfs are oriented correctly (e.g. not upside-down or sideways). The coding problem will also be submitted to gradescope.

Collaboration: Please read the full collaboration policy. If you work with others (and you should!), you must still write up your solution independently and name all of your collaborators somewhere on your assignment.

Problems 4 and 5 will be much easier after learning linearity of expectation on Friday.

1. CDF to PMF [7 points]

Let X be a discrete random variable that takes integer values from 1 to 10 (inclusive), and has the following cumulative distribution function:

$$F_X(n) = \begin{cases} 0 & \text{if } n < 1 \\ \frac{(n+1)(n+2)}{132} & \text{if } 1 \leq n \leq 10 \\ 1 & \text{if } n > 10 \end{cases}$$

Find the probability mass function (PMF) for X.

2. Keep It Rolling [10 points]

You roll a die independently 100 times. Let X represent the difference between the number of times the result is 1 or 2, and the number of times the result is 3 through 6.

(a) What is the support of X? [4 points]

(b) What is the probability that $X = 4$? [6 points]

Note that we are looking at difference without absolute value. So if you rolled a 4, a 2, then a 1, the difference would be 1.
If you rolled a 4, a 5, then a 1, the difference would be -1.

3. Turnabout is Fair Play [8 points]

You want to open a casino with a new card game. In the game, a dealer will shuffle a (full) deck of (standard) cards so they are fully-shuffled (i.e. every ordering is equally likely). The dealer then flips over the top three cards.

To play, a player pays $\$1$.

- If all three cards are diamonds, the player gets a payout of $\$10$ (not including the dollar they already paid)
- If two cards are diamonds, the player gets a payout of $\$5$
- If none of the cards are diamonds, the player gets no payout.

You want all the games in your casino to be **fair games** – the expected profit (payout minus cost) of your game should be $\$0$. What should the payout be if exactly one of the cards is a diamond? Give your answer as a simplified fraction (don't worry if your number couldn't be paid out with standard U.S. currency)

4. Socks [14 points]

You have 10 pairs of socks. 3 pairs are bright stripes, 5 pairs are checkered designs, and 2 pairs are solid colors. Both socks in a pair are identical, but each of the pairs are distinct from each other.

A pair of socks is "properly matched" if they are from the same pair. A pair of socks is "fashionably mismatched" if they are from different pairs, but from the same group (for example, two bright stripe socks could be fashionably mismatched; but a solid color sock and a checkered sock could not be).

You (uniformly) randomly pair the socks with each other. In each of these problems clearly define any random variables you need.

(a) What is the expected number of properly matched socks? [7 points]

(b) What is the expected number of fashionably mismatched socks? [7 points]

5. Better Drinks [11 points]

You just bought a set of n flavored syrups[1] to mix into coffee. Since you need the caffeine, but hate the taste of coffee, you mix in exactly four different syrup flavors into each of your drinks.

Your coffee will taste good only if the four flavors you mix in are pairwise compatible – that is, if for every pair in the four flavors, those two are compatible. Each pair of flavors is compatible (independently) with probability p. What is the expected number of 4-flavor-combinations that will taste good?

(a) Carefully define variables for this problem. [4 points]

(b) Calculate the expectation. [7 points]

6. Returning Exams [20 points]

Suppose a teacher is returning n tests back to her n students. They show up one at a time to take their test back. However, rather than handing them their own test, the teacher does the following: The first person gets a uniformly random test. After that each person gets their own test, if the teacher hasn't already handed it back to someone who arrived earlier; otherwise, they get a uniformly random one of the remaining tests. For example, if there are 3 people A, B, C, then the possible outcomes are

- A gets their own test (probability $1/3$); B must get their own test; C must get their own test. [Overall probability $1/3$]

[1] vanilla, french vanilla, peppermint, caramel, hazelnut, pumpkin spice,...

- A gets B's test (probability $1/3$); B gets a random test, it could be C's (probability $1/2$); C gets A's test. [Overall probability $1/6$]
- A gets B's test (probability $1/3$); B gets a random test, it could be A's (probability $1/2$); C gets their own test. [Overall probability $1/6$]
- A gets C's test (probability $1/3$); B must get their own test; C must get A's test. [Overall probability $1/3$]

We are interested in determining the probability that the n^{th} person to go up to the professor gets their own test returned. When there are n people, call the probability person n gets their own test returned r_n. The example above says that $r_3 = 1/3 + 1/6 = 1/2$.

We strongly recommend (but won't require) that you write some code and repeatedly run the experiment described in this problem for a few values of n to get estimates of r_n. This won't prove that the answer is what your code shows, but it will give you a sense of whether your answers are reasonable before you get too deep into complicated formulas.

(a) Suppose the first person got person n's test, what is the probability that Person n gets their own test? (this part is not supposed to be the hard part) [3 points]

(b) Suppose the first person get their own test back, what is the probability that Person n gets their own test? (still not the hard part) [3 points]

(c) Now let's think about the hard case. Suppose that the first person gets person i's test (where i is not 1 or n).

- Explain why people $2, ..., i-1$ get their own tests back. [1 point]
- Give the probability that person n gets their own test back conditioned on the first person getting person i's test. You should use r as defined in the introduction to this problem in your expression (don't try to get an exact number yet – this expression will be complicated) [5 points]

(d) Use the law of total probability (and parts a-c) to give an expression for r_n in terms of other value(s) of r and constants. [4 points]

(e) Give a non-recursive and fully simplified expression for r_n ($n \geq 3$). Justify this expression by assuming it is correct for $r_1, r_2, \ldots, r_{n-1}$ and simplifying your expression from the last step (you're essentially doing an inductive step of an inductive proof checking your formula here). [4 points]