

Homework 5: Discrete Random Variables

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.

Remember that you must tag your written problems on Gradescope.

Submission: You must upload a **pdf** of your written solutions to Gradescope under “HW 5 [Written]”. (Instructions as to how to upload your solutions to gradescope are on the course web page.)

Due Date: This assignment is due Wednesday October 29 at 11:59 PM.

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).

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.

1. Dice Try! [15 points]

You are playing a game that uses a fair 10-sided die whose faces are numbered 1, 2, ..., 10. The value of a roll is the number showing on the top of the die when it comes to rest. **Give all answers as simplified fractions.**

- Let X be the value of one roll of the die. Compute $\mathbb{E}[X]$ and $\text{Var}(X)$. For this problem, you can either calculate from first principles (i.e. definitions and theorems) or describe which variable from the zoo is appropriate and what values you get as a result.
- Let Y be the sum of the values of 8 independent rolls of the die. Compute $\mathbb{E}[Y]$ and $\text{Var}(Y)$. If you use independence, state **precisely** where in your computation you are using it.
- Let Z be the **average** of the values of 8 independent rolls of the die. Compute $\mathbb{E}[Z]$ and $\text{Var}(Z)$. If you use independence, state **precisely** where in your computation you are using it.

2. Table tennis [10 points]

A game in table tennis ends when a player has:

- at least 11 points *and*
- at least 2 more points than their opponent.

So, for example, a game at 11-10 is not over (no one has a two point lead), but a game at 12-10 is over.

Suppose a set is tied 10-10, and each point is won by one player (independently) with probability p . What is the expected number of points played before one player or the other wins the game? (Hint: You could calculate this with an infinite sum, in which case you may use wolframalpha to find a closed form. But a clever use of a variable from the zoo will save you quite a bit of work, and make it so an infinite summation is unnecessary.)

3. Instagram [16 points]

A photo-sharing startup offers the following service. A client may upload any number N of photos and the server will compare each of the $\binom{N}{2}$ pairs of photos with their proprietary image matching algorithms to see if there is any person that is in both pictures. Testing shows that the matching algorithm is the slowest part of the service, taking about 50 milliseconds of CPU time per photo pair. Hence, estimating the number of photos uploaded by each client is a key part of sizing their data center. The people in charge say that their gut feeling is that $N = 10$. You (the chief technical officer) say, “but N is a random variable”. For each of these possible distributions for N : What will the **expected** time (in milliseconds) for CPU demand per client be? (Your answer may be a function of N , p and/or λ , as appropriate).

Hint: Many of these computations can be done cleverly by using linearity of expectation along with the formula for variance and values from the zoo!

- the “distribution” where N is the same fixed number with probability 1?
- the geometric distribution with parameter p ?
- the Poisson distribution with parameter λ ?
- $N = 10X + 15$, where X is a Bernoulli random variable with parameter p ?

In each case, include as part of your answer the expected value of N and the variance of N . Make sure your answer is **not** in the form of a summation for this problem.

4. Binomial From Nowhere [12 points]

Consider repeatedly rolling a fair 7-sided die, each roll being independent of the others. Define the random variable Y to be the number of rolls until (and including) the first roll of a 6, and define the random variable X to be the number of 3's rolled before the first 6 is rolled. Show that $\mathbb{P}(X = j|Y = i)$, as j ranges over its possible values, is the probability mass function of a binomially distributed random variable and determine its parameters n and p .

- For concreteness, to start with suppose that $Y = 2$. Find $\mathbb{P}(X = 0|Y = 2)$, $\mathbb{P}(X = 1|Y = 2)$, $\mathbb{P}(X = 2|Y = 2)$. [3 points]
- Now generalize a bit to find $\mathbb{P}(X = 0|Y = i)$, $\mathbb{P}(X = 1|Y = i)$, $\mathbb{P}(X = 2|Y = i)$ where you let i be unknown (i.e., treat it as a variable); for simplicity assume that $i \geq 2$. [3 points]
- Now we're ready to solve the question we originally posed. Give a formula for $\mathbb{P}(X = j|Y = i)$. As j changes, you get the PMF of a binomial—what are n and p ? [6 points]

5. Feedback [1 point]

Answer these questions on the separate Gradescope box for this question.

Please keep track of how much time you spend on this homework and answer the following questions. This can help us calibrate future assignments and future iterations of the course, and can help you identify which areas are most challenging for you.

- How many hours did you spend working on this assignment (excluding any extra credit questions, if applicable)? Report your estimate to the nearest hour.
- Which problem did you spend the most time on?
- Any other feedback for us?

6. Go study :)

This homework is intentionally a bit shorter than the last few have been. Use the time to start on your studying :D