Problem Set 5

Due: Wednesday, November 9, by 11:59pm

Instructions

Solutions format, collaboration policy, and late policy. See PSet 1 for further details. The same requirements and policies still apply. Also follow the typesetting instructions from the prior PSets.

Solutions submission. You must submit your solution via Gradescope. In particular:

- Submit a *single* PDF file containing the solution to all tasks in the homework. Each numbered task should be solved on its own page (or pages). Follow the prompt on Gradescope to link tasks to your pages.
- Do not write your name on the individual pages – Gradescope will handle that.
- We encourage you to typeset your solution. The homepage provides links to resources to help you doing so using LaTeX. If you do use another tool (e.g., Microsoft Word), we request that you use a proper equation editor to display math (MS Word has one). For example, you should be able to write $\sum_{i=1}^{n} x^i$ instead of $\mathrm{x\hat{}1 + x\hat{}2 + ... + x\hat{}n}$. You can also provide a handwritten solution, as long as it is on a single PDF file that satisfies the above submission format requirements. It is your responsibility to make sure handwritten solutions are readable – we will *not* grade unreadable write-ups.

Task 1 – Random Grades [15 pts]

Every week, 25,000 students flip a 10,000-sided fair dice, numbered 1 to 10,000, to see if they can get their GPA changed to a 4.0. If they roll a 1, they win (they get their GPA changed). You may assume each student's roll is independent. Let X be the number of students who win.

a) For any given week, give the appropriate probability distribution (including parameter(s)), and find the expected number of students who win.

b) For any given week, find the exact probability that at least 2 students win. Give your answer to 5 decimal places.

c) For any given week, estimate the probability that at least 2 students win, using the Poisson approximation. Give your answer to 5 decimal places.

Task 2 – Instant Image [16 pts]

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 100 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. You (the chief technical officer) say, "N is a random variable, so we have to estimate the time using probability". What will the **expected** time (in milliseconds) for CPU demand per client be (as a function of p or λ or c) be if N has a distribution given by the following:
Make sure that each of your answers is **not** in the form of a summation for this problem. In each case include the expectation and variance of N as part of your answer.

Hint: Be careful about exactly when rules involving expectations apply.

a) the "distribution" where N equals some fixed positive integer c with probability 1?

b) the Poisson distribution with parameter λ?

c) the geometric distribution with parameter p?

d) $N = 10X + 7$, where X is a Bernoulli random variable with parameter p?

Task 3 – Binomial From Nowhere [12 pts]

Consider repeatedly rolling a fair 6-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 1's rolled before the first 6 is rolled. Show that $\mathbb{P}(X = j \mid Y = i)$, 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.

Task 4 – Sample Sampling Algorithm [18 pts]

Consider the following algorithm for generating a random sample S of size n from the set of integers $\{1, 2, \ldots, N\}$, where $0 < n < N$.

```
Sample(N, n):
  S ← ∅                          // S is a set of distinct integers, initially an empty set
  while |S| < n do
    x ← RollDie(N)               // x is the outcome of rolling a fair N-sided die
    S ← S ∪ {x}                  // if x is already in S it doesn't change
  return S
```

Let I be the number of die rolls until S is returned. Also, let I_i be the random variable which describes the number of rolls it takes from the time the set S has $i-1$ values to the first time a new value is added after that (i.e., the set S has i values).

a) What type of random variable from our zoo is I_i and what is/are the relevant parameter(s) for that random variable?

b) What is I in terms of the random variables I_i? Calculate $\mathbb{E}[I]$, expressing the result as a summation that depends on both N and n.

c) What is $\mathrm{Var}(I)$? You can leave your answer in summation form.

Task 5 – PDF [15 pts]

For this exercise, give exact answers as simplified fractions. Define function f_X by

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

a) Show that f_X has the properties required of a probability density function.

b) Compute the expectation of a random variable X with f_X as its PDF.

c) Compute the variance of a random variable X with f_X as its PDF.

Task 6 – Dart [8 pts]

You throw a dart at a circular target of radius r. Let X be the distance of your dart's hit from the center of the target. Your aim is such that $X \sim$ Exponential$(3/r)$. (Note that it is possible for the dart to completely miss the target.)

a) As a function of r, determine the value m such that Pr$(X < m) =$ Pr$(X > m)$. Then, for $r = 9$, give the value of m to 3 decimal places.

b) What is the probability that you miss the target completely? Give your answer to 3 decimal places.

Task 7 – Flea [16 pts]

A flea of negligible size is trapped in a large, spherical, inflated beach ball with radius r. (Recall that such a ball has volume $\frac{4}{3}\pi r^3$.) At this moment, it is equally likely to be at any point within the ball. Let X be the distance of the flea from the center of the ball. For X, find ...

a) the cumulative distribution function F_X.

b) the probability density function f_X.

c) the expected value $\mathbb{E}[X]$.

d) the variance $\mathrm{Var}(X)$