Elements of Mathematical Statistics
Homework Assignment #2

Due 10 March 2020 by the start of class.

Probability Distributions. Consider a fair 6-sided die with, instead of the usual dots on each side, a number written on each side. The numbers are N, $N+1$, ..., $N+5$ where N is a positive integer. When you roll the die, let the random variable X be the value of the number that is facing up. (E.g. if $N = 1$, this is a regular 6-sided die.)

(a) What is the distribution of X? Write out its p.m.f.

(b) Suppose $N = 5$. Find $\Pr[X \leq 7]$.

(c) Suppose you don't know the value of N, but you know $\Pr[N = 3] = 0.4$ and $\Pr[N = 4] = 0.6$. What is $\Pr[N \geq 2]$. *Hint: recognize that $N = 3$ and $N = 4$ form a partition and use the Law of Total probability.*

Probability Mass Function. Let X be a discrete random variable with the following p.m.f.

$$f(x) = \begin{cases} \frac{c}{2^x}, & x \in \{1, 2, 3, 4\} \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the value of c that makes this a valid p.m.f.

(b) Find $\Pr[X < 3]$.

(c) Find $\mathrm{E}[X]$.

(d) Find $\mathrm{E}\left[2^X\right]$.

Linearity of Expectation. The SAT is a multiple choice exam, with 5 possible answers for each question. To discourage test takers from guessing, ETS (the makers of the SAT) use the following point system: if you get a question right, you get 1 point; and if you get a question wrong, you lose 0.25 points.

(a) If one has no idea which answers to a question are right or wrong and chooses one's answer at random, what is the probability that the correct answer will be chosen?

(b) We can model this using a Bernoulli random variable, where $X = 1$ if the correct answer is chosen and $X = 0$ if an incorrect answer is chosen. What is the expectation of this random variable, $\mathrm{E}[X]$ when the answer to the question is chosen at random?

(c) We can further define another random variable, Y, to be the number of points earned on a question. The random variable Y is a transformation of the random variable X, specifically

$$Y = X - 0.25(1 - X) = 1.25X - 0.25.$$

Use linearity of expectation to show that $\mathrm{E}[Y] = 0$ if the answer to a question is chosen at random.

(d) What if you don't know which answer is right, but you do know that some answers are wrong? Find $\mathrm{E}[Y]$ when one answer is known to be false, and the given answer is chosen randomly from the remaining four. Then find $\mathrm{E}[Y]$ when two answers are known to be false, and the given answer is chosen randomly from the remaining three.

Expected Values. Four buses are transporting 200 UNM students to Las Cruces for a football game. The buses carry, respectively, 30, 40, 60, and 70 students.

(a) A student is selected at random—call her Alice. Let X be the number of students on Alice's bus. Find $\mathrm{E}[X]$.

(b) A bus driver is selected at random—call him Bob. Let Y be the number of students on Bob's bus. Find $\mathrm{E}[Y]$.

(c) Which one of these two expected values is larger? Explain why it is larger than the other.

Baking Bad. Walter White and Jesse Pinkman run an operation where they cook blueberry muffins. Let X be the number of blueberries found in one of their muffins and assume that X follows a Poisson distribution with parameter λ, that is $X \sim \mathrm{Pois}(\lambda)$.

(a) Determine the expected value and variance of X.

(b) Assume $\lambda = 9$. One of Walter and Jesse's customers gets very angry when he gets a muffin with 0 blueberries. What is the probability that this happens?

(c) Blueberries are in short supply! Walter and Jesse want to reduce the average number of blueberries per muffin (λ) to save costs. What is the smallest value of λ they can pick that still ensures no more than 1% of muffins have no blueberries. *Hint: your work on the previous question will be helpful here.*

(d) Assume now that the customer gets angry if he gets a muffin with 1 or fewer blueberries. What is the smallest value of λ that will ensure no more than 1% of muffins have 1 or fewer blueberries? *Hint: this is a harder problem to solve. Use any method you like, including guess-and-check or using R, to give a correct answer to 1 decimal place.*

(e) Assume $\lambda = 9$. Let the random variable Y be the number of blueberries found in a dozen (12) of Walter and Jesse's muffins. What is the distribution, mean, and variance of Y?

Build a Test Problem.

For this problem, try hard to *be creative*. Use an example from your life, work, or make something up. Your examples can be humorous or not.

Give an example of a Binomial random variable U and a Geometric random variable V. If possible, relate both of these random variables to independent Bernoulli trials of the same "experiment"—and remember that an experiment in this context doesn't need to be scientific, it just needs to be something for which you can identify successes and failures, like "my dog acts excited when I get back from school".

Ask between 2 and 4 test-type questions and give the answers to them (you do not need to show your work). If you need help coming up with questions you could ask, look at the questions you've previously seen on this homework assignment and see if they can give you any ideas.

Challenge Problem. **Rare Disease Testing.**

Refer to the Bayes Theorem example we did in class with drug testing. Recall that when the prevalence of a condition (the probability a randomly chosen person exhibits the condition) is very small, the probability of having the condition remains small even if someone tests positive for it.

Alf is worried that he may have cat scratch fever. Let the event D be $D = \{\text{Alf has the disease}\}$. Cat scratch fever is rare among Alf's species, so $\Pr[D] = 0.00001$. Alf's doctor has a test for cat scratch fever with sensitivity 0.99 and specificity 0.95. Better yet, the doctor's test can be taken multiple times and the results of each testing will be independent! Let the random variable X be the number of times Alf tests positive on n tests. Then, since the tests are independent, we have the following probability distributions for X based on whether Alf does or does not have the disease:

$$X \mid D \sim \text{Bin}(n, 0.99)$$
$$X \mid D^c \sim \text{Bin}(n, 0.05)$$

Now assume that Alf tests positive for k tests ($k \in \{0, 1, ..., n\}$). We can use Bayes Theorem as follows,

$$\Pr[D \mid X = k] = \frac{\Pr[X = k \mid D]\Pr[D]}{\Pr[X = k]},$$

where $\Pr[X = k] = \Pr[X = k \mid D]\Pr[D] + \Pr[X = k \mid D^c]\Pr[D^c]$ by the Law of Total Probability.

(a) Compute $\Pr[X = k \mid D]$ and $\Pr[X = k \mid D^c]$ using the appropriate Binomail pmf.

(b) Use Bayes Theorem to give an explicit expression for $\Pr[D \mid X = k]$ in terms of n and k.

(c) If Alf takes $n = 4$ tests and tests positive $k = 3$ times, what is the probability that he has the disease?

(d) How many consecutive tests does Alf need to test positive for, before the probability he has the disease is at least 0.99?