Final Exam

Elliott Pryor

11/25/2020

Problem 1

Statement

Choose one of the distribution videos of your peers to watch and critique. The only restriction is that this should be a group different from yours and from the project group you peer reviewed. You can find the videos here https://www.youtube.com/ playlist?list=PLMd58R18w9hcUaMGJGXRV0GFqOZ27Kt5w and https://spark.adobe.com/ video/GcMjER5KZAZLp. Answer the following questions.

a. For each distribution discussed: State the distribution that was used in the problem. Was the distributions used correctly with respect to their support? That is, was the random variable defined in a way that it could be reasonably modeled with the distribution? Briefly explain why or why not. *Make sure you answer these questions for both distributions!*

b. Choose the problem you found the most interesting out of the problems presented by the group. What made this particular problem more interesting than the other(s)?

c. Briefly reflect on what you learned with respect to how probability distributions (pmf/pdfs) are used for modeling random events in the real world as a result of participating in the distributions project. Specifically, can a probability model perfectly represent a random event that arises in the real world? Why is communicating model assumptions (i.e., assumptions that are required if a particular probability distribution is used) so important when applying probability models to real world problems? *There is no right answer here, please provide an honest and thoughtful reflection in ≈ 5-10 sentences.*

Solution

I watched the video on Geometric and Exponential Distributions.

a. The first problem was a geometric distribution. $X \sim Geom(p = 0.2)$. The problem was defined in a way that makes sense, counting the number of ice slabs a robot could cross. It could crash on the 1st, or never crash so geometric makes sense. I really liked the exponential distribution problem. They modeled an insider trading scenario where a company is likely to default on a loan on an exponential distribution ($X \sim Exp(\beta = 14)$). I know absolutely nothing about trading or money, but the premise seems reasonable. Although I would think that the company would be more likely to default later in the period instead of earlier. Since exponential is large close to zero, most of the probability is acquired at the beginning, and as seen in part b the company is likely to default earlier. I would guess they would default later once the company was closer to deadline and realize it was impossible to make it. But this is nit-picky. I think given that it follows an exponential distribution the problem was good.

b. As I mentioned in part a. I really liked the insider trading problem. I thought it was an entirely new problem that we had never seen anything like and was just interesting. I liked the truncated exponential distribution as it was new and clever distribution to apply. Also the questions were non-trivial, I liked the dedication to do IBP. It was just a hard problem that was applied in a way that we have never covered before and was just interesting.

c. No a probability model cannot perfectly model a real world scenario. Partially that is given in the definition of a model. You construct a model to get close to a real world scenario without dealing with the entire thing. But also, the models covered in this class are fairly simplistic. There are many assumptions made about independence and how trials behave. For instance it might not be that p is constant for geometric or binomial for every trial. Also there are lots of extraneous factors that can change outcomes of experiment and affect the distributions. It is important to communicate the assumptions so others know what factors we choose to ignore. This helps prevent misinformation/misinterpretation of the models and reminds us that it is only a model.

Problem 2

Statement

Let X be a random variable with *pdf* $f_X(x) = kX^2 I_{(0,2)}(x)$ where k is a constant

a. Find k

b. Let $Y = \frac{1}{X}$

   i. Draw a well-labeled graph showing the transformation from $\mathcal{X}$ to $\mathcal{Y}$ and state $\mathcal{Y}$

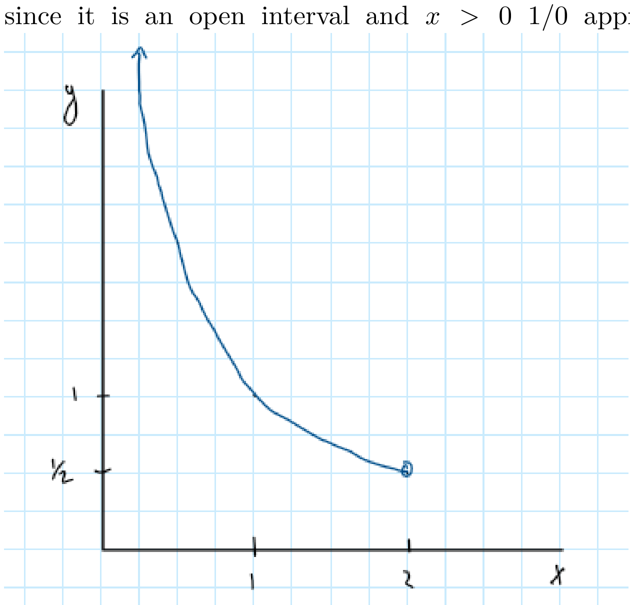
   ii. Find pdf $f_Y(y)$

   iii. **Extra Credit** What is the name of the distribution assumed for X.

Solution

a. We know that integrating over the entire support of X, should yeild an answer of 1. So, $1 = \int_0^2 kx^2 dx = k\int_0^2 x^2 dx = k\left(1/3x^3\Big|_0^2\right) = k(8/3)$. So $k = 3/8$.

b. i. We know that $x \in (0,2)$ so $0 < x < 2$. Then with $y = 1/x$ we have $1/0 > y > 1/2$ since it is an open interval and $x > 0$ $1/0$ approaches ∞. So we have $y \in (1/2, \infty)$.



   ii. Clearly y is strictly decreasing on $x = (0,2)$. So we can use the method of univariate transformations for monotone functions from our notes. We have that $Y = g(X)$ then $f_Y(y) = -f_x(g^{-1}(y))\left(\frac{d}{dy}(g^{-1}(y))\right)$. In our case we have $g = 1/X$ and thus $g^{-1}(y) = 1/Y$. Then $\frac{d}{dy}(g^{-1}(y) = -1/Y^2$. So plugging all this in we get $f_Y(y) = -(8/3)(1/y)^2(-1/y^2) = (8/3)(1/y^4)$. So $f_Y(y) = (8/3)(1/y^4)I_{(1/2,\infty)}(y)$

iii. **Extra Credit** TODO

Problem 3

Statement

A submarine is missing and is presumed to have equal probability of being sunk in one of three Regions off of the Gulf Coast. If the submarine is actually down in Region i, let the probability of a successful search in that region be equal to $1 - \alpha_i$ for i = 1, 2, 3. For example, the probability a search of Region 1 will recover the sub, given that it sunk in Region 1 is $1 - \alpha_1$. What is the conditional probability the submarine is sunk in Region 1, given that the search of Region 1 was unsuccessful? Be sure to clearly define events and probabilities!

Solution

Let X be the random variable indicating if the search was successful. We know that $X \sim Binom(1 - \alpha_1)$. We say 0 indicates unsuccessful search and 1 indicates successful search. Let Y be the random variable indicating what region the ship was sunk in.

We are searching for $P(Y = 1|X = 0)$. This is a basic application of Bayes' Rule

$$P(Y = 1|X = 0) = \frac{P(X = 0|Y = 1)P(Y = 1)}{P(X = 0|Y = 1)P(Y = 1) + P(X = 0|Y = 2)P(Y = 2) + P(X = 0|Y = 3)P(Y = 3)}$$

Since it is equally likely to be in any of the regions we know that $P(Y = 1) = P(Y = 2) = P(Y = 3) = 1/3$. For $P(X = 0|Y = i)$ where $i = 2, 3$ we know this must be 1 as there cannot be a successful search if the ship isn't in the region. So we are left with:

$$P(Y = 1|X = 0) = \frac{1/3 \cdot P(X = 0|Y = 1)}{1/3\,(P(X = 0|Y = 1) + 2)} = \frac{P(X = 0|Y = 1)}{P(X = 0|Y = 1) + 2}$$

We then know the probability of an unsuccessful search in the region where the sink is: $1 - P(success) = 1 - (1 - \alpha_1) = \alpha_1$. So we get

$$P(Y = 1|X = 0) = \frac{\alpha_1}{\alpha_1 + 2}$$

Problem 4

Statement

Let X and Y be random variables, where X is the number of workplace injuries occurring in a factory on any given day and Y is the level of activity in the factory. The joint distribution of X and Y is given by

$$f_{XY}(x, y) = \begin{cases} \frac{e^{-y}y^x}{6x!} & 0 < y < 6, \quad x = 0, 1, 2, ... \\ 0 & otherwise \end{cases}$$

a. Are X, Y statistically independent? Explain.

b. One can show that $Y \sim Uniform(0, 6)$ and $X|Y \sim Poisson(\lambda = y)$ if $y \in (0, 6)$ Use that information to find:

   i. E(X|Y = y)

   ii. Var(X|Y = y)

   iii. E(Y)

iv. Var(Y)

c. Show $E[XY] = 12$ show each step clearly

d. Show $Var(X) = 6$ show each step clearly.

Solution

e. No, they are not independent. It is not possible to factor the term y^x into the product of two functions of one variable. So by the factorization theorem they are not independent

f. i. $E(X|Y = y)$ is just y. This is easy to see if we let $Z = X|Y = y \sim Poisson(\lambda = y)$. So we are looking for $E(Z)$ which is expectation of a Poisson distribution, which is just λ. So in our case, $\lambda = y$. Thus $E(X|Y = y) = y$ for $y \in (0, 6)$ otherwise 0.

ii. $Var(X|Y = y)$ is also just y. Using the same Z as above, we are looking for variance of a Poisson distribution, which is $\lambda = y$. Thus $Var(X|Y = y) = y$ for $y \in (0, 6)$ otherwise 0.

iii. We know that $Y \sim Uniform(0, 6)$ so, $E(Y)$ is the midpoint of this interval. So $E(Y) = 3$.

iv. We know that $Y \sim Uniform(0, 6)$ so, $Var(Y)$ is $\frac{(6-0)^2}{12} = 36/12 = 3$.

g. We want to find $E(XY)$. We know from class that $E(XY) = E_y(YE_x(X|Y))$ by iterated expectation. So we know from b.i, that $E_x(X|Y)$ is just y. So now we are looking for $E_y(Y^2) = Var(Y) + (E(Y))^2$. Both of which we found in part b. So $E(XY) = E(Y^2) = 3 + 3^2 = 12$

h. We want to find $Var(X)$. By theorem 5.15 we know that $Var(X) = E_Y(Var(X|Y)) + Var_Y(E(X|Y))$. We know from part b that $Var(X|Y) = Y$ and $E(X|Y) = Y$. So we are looking for $E(Y) + Var(Y)$. Thus we have $Var(X) = E(Y) + Var(Y) = 3 + 3 = 6$

Problem 5

Statement

A coal-operated power plant has two smokestacks: one with a cleaning device known as a *scrubber* and another without a scrubber. An environmental engineer measures the amount of pollutant in each smokestack by weighing the particulate pollution in a random sample of a fixed volume of air. Let X be the amount of pollutant in a sample from the smokestack *without the scrubber*, and Y be the amount of pollutant in a sample of the same volume from the smokestack *with the scrubber*. The joint distribution can be written as,

$$f_{XY}(x, y) = \begin{cases} 1 & 0 \leq x \leq 2, \;\; 0 \leq y \leq 1, \;\; 2y \leq x \\ 0 & otherwise \end{cases}$$

Find the probability that the sample taken from the smokestack *with the scrubber* (Y) shows a reduction in the amount of pollutant by one-third or more (as compared to the sample taken from the smokestack *without the scrubber* (X)).

a. Make a clearly labeled graph to show the joint support of $f_{XY}(x, y)$ and the region of interest for finding the probability

b. Find the probability of interest **using integration**. *Make sure you have written a clear probability statement using proper notation along with its corresponding integral.*

c. **Extra Credit**. Explain how/why this particular problem can be solved without using integration, and find the probability using geometry.

Solution

a. Graph of the region of interest and the constraints. The yellow marks the joint support. The green marks region of interest within the support.

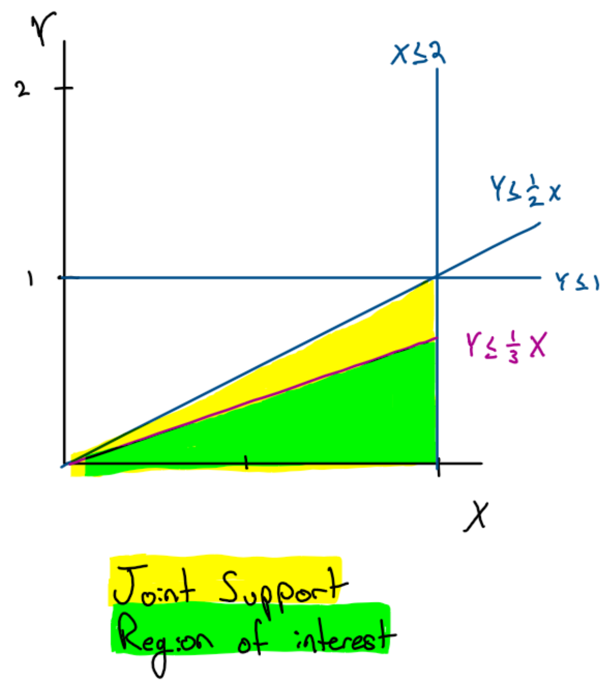


Figure 1: Region of integraton

b.

$$\begin{aligned}P(Y \leq 1/3X) &= \int_0^2 \int_0^{x/3} 1 dy dx \\ &= \int_0^2 y|_0^{x/3} \, dx \\ &= \int_0^2 x/3 dx \\ &= x^2/6|_0^2 \\ &= 4/6 = 2/3\end{aligned}$$

c. This can be done geometrically since the pmf is constant across the region of integration. So we are essentially finding the area of the triangle with base 2 height 2/3. So area is $1/2bh = 1/2 * 2 * 2/3 = 2/3$ as expected.