PSET 6: Answer Key
Discrete random variables

1 Background info

- Name
- How long did this problem set take you?
- How difficult was this problem set? very easy 1 2 3 4 5 very challenging

2 Calculate probabilities in a sample space S

Events A and B are contained within a sample space S. Given that $\Pr(A) = 0.65$, $\Pr(B) = 0.3$ and $\Pr(A \cap B) = 0.1$, find:[1]

a. $\Pr(A \cup B)$

Solution

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = 0.65 + 0.30 - 0.10 = 0.85.$$

b. $\Pr(A \cap B^c)$

Solution

$$\Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B) = 0.65 - 0.10 = 0.55.$$

c. $\Pr[(A \cap B^c) \cup (B \cap A^c)]$

Solution

$$\Pr[(A \cap B^c) \cup (B \cap A^c)] = \Pr(A) + \Pr(B) - 2\Pr(A \cap B) = 0.65 + 0.30 - 2(0.10) = 0.75.$$

$$\Pr(A \cup B) - \Pr(A \cap B) = 0.85 - 0.10 = 0.75$$

3 Random variables

We have been discussing random variables. Provide the following explanations in your own words:

a. What is a random variable?

Solution.
A random variable X is a random process or variable with a numerical outcome; it is a measurable function that assigns a numerical value to each outcome in a sample space. It translates randomness into numbers so we can compute probabilities and expectations.

b. What is the difference between upper case and lower case x? How (if at all) do they matter?

Solution. X denotes the random variable itself. x denotes a particular realized value of X. They matter because probability statements use the variable X and evaluate at realized points

[1] Inspired by Grimmer HW 8.4

c. If I'm thinking about something like $p_X(x_0)$, what do all the parts/pieces mean?

Solution. For a discrete X, $p_X(x_0) = \Pr(X = x_0)$ is the probability mass function (pmf) evaluated at x_0: the probability that X equals x_0.

4 Survey Says

A survey has 54% respondents 50 or older and 46% respondents under 50. Within the survey, on a particular question, 9.5% of the 50-plus population agrees strongly while 2.7% of under 50 respondents agree strongly.

1. What is the probability someone selected at random is 50 or older?

   Solution. Given: $\Pr(50+) = 0.54$, $\Pr(<50) = 0.46$, and $\Pr(\text{SA}|50+) = 0.095$, and $\Pr(\text{SA}|<50) = 0.027$, where SA = "strongly agrees."

   We have $\Pr(50+) = \boxed{0.54}$.

2. The selected individual strongly agrees with the survey question. Now what is the likelihood that person is 50 or older? Explain your reasoning and SHOW ALL YOUR WORK

   Solution.

   Computing $\Pr(50+|\text{SA})$:

$$\begin{aligned}
\Pr(\text{SA}) &= \Pr(50+)\Pr(\text{SA}|50+) + \Pr(<50)\Pr(\text{SA}|<50) && \text{Law of total probability} \\
&= 0.54 \cdot 0.095 + 0.46 \cdot 0.027 = 0.06372. \\
\Pr(50+|\text{SA}) &= \frac{\Pr(\text{SA}|50+)\Pr(50+)}{\Pr(\text{SA})} && \text{Beyes' theorem} \\
&= \frac{0.54 \cdot 0.095}{0.06372} \approx \boxed{0.8051}.
\end{aligned}$$

3. Are the two answers above the same or different? Explain.

   Solution. Different. Conditioning on "strongly agrees" changes the prior because older respondents are shown to be more likely to strongly agree, so the posterior rises from 0.54 to ≈ 0.805.

4. (for fun, no points) What is the survey question?

   Solution.

5 PMF vs CMF

Consider the following function: $f(x) = \frac{1}{6}$. Find the pmf and cmf of the function and provide them in a table below.

Solution. Uniform PMF: $p_X(x) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$, and 0 otherwise. CMF: $F_X(x) = \Pr(X \le x)$ equals

$$F_X(x) = \begin{cases} 0, & x < 1 \\ \frac{k}{6}, & k - 1 < x \le k,\ k = 1, 2, \dots, 6 \\ 1, & x \ge 6. \end{cases}$$

x	$p_X(x)$	$F_X(x) = \Pr(X \leq x)$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	1

6 Getting a traffic ticket

You drive to work 5 days a week for a full year (50 weeks), and with probability $p = 0.04$ you get a traffic ticket on any given day, independent of other days. Let X be the total number of tickets you get in the year.[2]

a. What is the probability that the number of tickets you get is exactly equal to the expected value of X?

Solution.

Drive 5 days/week for 50 weeks (250 independent days) at probability $p = 0.04$ of a ticket. Let the random variable $X \sim \text{Bin}(n = 250, p = 0.04)$. Then $\mathbb{E}[X] = np = 10$.

$\Pr(X = \mathbb{E}[X]) = \Pr(X = 10)$ (exact binomial):

$$\Pr(X = \mathbb{E}[X]) = \binom{250}{10}(0.04)^{10}(0.96)^{240} \approx \boxed{0.1276881}.$$

b. Calculate approximately the probability in (a) using a Poisson approximation.

Solution.

Set up the Poisson approximation with $\lambda = np = 10$:

$$\Pr(X = 10) \approx e^{-10}\frac{10^{10}}{10!} \approx \boxed{0.1251100}.$$

7 Obtaining requests for information

X is a discrete random variable. It takes the value of the number of days required for a governmental agency to respond to a request for information. X is distributed according to the following PMF:[3]

$$f(x) = e^{-6}\frac{6^x}{x!} \text{ for } X \in \{0, 1, 2...\}$$

[2]Inspired by BT 2.41

[3]Inspired by Grimmer HW10.2

a. Given this information, what is the probability of a response from the agency in 5 days or less?

Solution.

$X \sim \text{Poisson}(\lambda = 6)$ with pmf $f(x) = e^{-6}\frac{6^x}{x!}$, $x = 0, 1, 2, \ldots$.

$$\Pr(X \leq 5) = \sum_{x=0}^{5} e^{-6}\frac{6^x}{x!} \approx \boxed{0.4456796}.$$

b. What is the probability the agency response takes more than 10 but less than 13 days?

Solution.

$\Pr(10 < X < 13) = \Pr(X = 11) + \Pr(X = 12)$:

$$e^{-6}\frac{6^{11}}{11!} + e^{-6}\frac{6^{12}}{12!} \approx \boxed{0.03379344}.$$

c. What is the probability the agency response takes more than 5 days?

Solution.

$\Pr(X > 5) = 1 - \Pr(X \leq 5) \approx 1 - 0.4456796 = \boxed{0.5543204}$.

d. Suppose using X you generate a new variable, **Responsive**. **Responsive** equals 1 if an agency responds in 5 days or less and 0 otherwise. What is the expected value of **Responsive**?

Solution.

Define responsive as $= \mathbf{1}\{X \leq 5\}$. Then from part (a):

$$\mathbb{E}[\textbf{Responsive}] = \Pr(X \leq 5) \approx \boxed{0.4456796}.$$

e. What is the variance of **Responsive**?

Solution.

For a binary random variable, Var(**Responsive**) $= p(1-p)$, where $p = \Pr(X \leq 5)$:

$$\text{Var} \approx 0.4456796 \times 0.5543204 \approx \boxed{0.247}.$$

8 Modeling electoral outcomes

Suppose we've developed a model predicting the outcome of the upcoming midterm elections in a state with 4 Congressional districts. In each district there are two candidates, a Republican and a Democrat. We have reason to believe the following PMF describes the distribution of potential election results where $K \in \{0, 1, 2, 3, 4\}$ and is the number of seats won by Republican candidates in the upcoming election.

$$\Pr(K = k|\theta) = \binom{4}{k} \theta^k (1-\theta)^{4-k}$$

Based on polling information, we think the appropriate value for θ is 0.423.[4]

a. What's the expected number of seats Republicans will win in the upcoming election?

Solution.

We are given $K \sim \text{Bin}(n = 4, \theta = 0.423)$ denote the number of Republican seats.

$$\mathbb{E}[K] = n\theta = 4(0.423) = \boxed{1.692}.$$

b. Given this PMF, what's the probability that no Republican legislators win in the upcoming election?

Solution.

$\Pr(K = 0) = (1-\theta)^4 = (0.577)^4 \approx \boxed{0.1108417}$.

c. What's the probability that Republican legislators win a majority of the seats in this state?

Solution.

Republican majority is defined as strictly more than half of 4, so we compute $\Pr(K \geq 3)$:

$$\Pr(K \geq 3) = \binom{4}{3} \theta^3 (1-\theta) + \theta^4 \approx \boxed{0.20670111}.$$

[4] Inspired by Grimmer HW10.3. Data from *genericpolling* (https://www.realclearpolling.com/polls/state-of-the-union/generic-congressional-vote)

d. A prominent political pundit declares they are certain that Republicans will win a majority of seats in the next election and offers the following bet. If Republicans win a majority of the seats, we must pay the pundit \$15.00. If Republican's fail to win a majority of states, we will win \$20.00. Based on our model, should we take this bet? **Hint: Think of the betting outcomes as a random variable. Find the expected value of this random variable.**

Solution.

Bet 1: Pay \$15 if Republicans win majority, and receive \$20 otherwise. Let $M = \mathbf{1}\{K \geq 3\}$ represent the binary random variable for Republican majority.

$$\begin{aligned}
\mathbb{E}[\text{payout}] &= (-15)\Pr(M=1) + 20\Pr(M=0) \\
&= (-15)\Pr(K \geq 3) + 20(1 - \Pr(K \geq 3)) \\
&= -15(0.20670111) + 20(0.79329889) \approx \boxed{\$12.77}.
\end{aligned}$$

Since our expected value is greater than 0, take the bet.

e. Suppose we are offered a second bet with a more complicated structure. In this case we'll receive \$100 if the Republicans win a majority, \$50 if neither party wins a majority and we'll have to pay \$200 if the Democrats win a majority. Should we take this bet?

Solution.

Bet 2: Receive \$100 if GOP majority ($K \geq 3$); \$50 if tie ($K = 2$); pay \$200 if Dem majority ($K \leq 1$). Computing the necessary probabilities:

$$\begin{aligned}
\Pr(K=2) &= \binom{4}{2}\theta^2(1-\theta)^2 \approx \boxed{0.35742392} \\
\Pr(K \leq 1) &= \Pr(K=0) + \Pr(K=1) \\
&= \Pr(K=0) + \binom{4}{1}\theta(1-\theta)^3 \\
&\approx \boxed{0.43587497}.
\end{aligned}$$

Expected value:

$$\begin{aligned}
\mathbb{E}[\$] &= 100\Pr(K \geq 3) + 50\Pr(K=2) - 200\Pr(K \leq 1) \\
&\approx 100(0.20670111) + 50(0.35742392) - 200(0.43587497) \\
&\approx \boxed{-\$48.63}.
\end{aligned}$$

Since expected value is negative, do not take this bet.

9 AI and Resources statement

Please list (in detail) all resources you used for this assignment. If you worked with people, list them here as well. It is not enough to say that you used a resource for help, you need to be specific on the link and *how* it was helpful. W/R/T gen AI tools (including GPT, etc.) you cannot use them to do work on your behalf – you cannot put in any of the questions, etc. You can ask for help on logic / sample problems. If you do use GPT or other AI tools, you need to provide a link to your chat transcript. Any suspected academic integrity violations will be immediately reported.