Quiz: Probability and Statistics (30 Marks)

Instruction:

- Please state reasons wherever applicable.
- Use precise mathematical arguments, no speeches.

Each question: 5 marks

1. Consider a discrete random variable X. Prove that for $a, b \in \mathbb{R}$ and $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$ where $F_X(\cdot)$ denotes the CDF of X.

2. A geometric random variable X with parameter p has PMF given by $p_X(k) = (1-p)^{k-1}p$. Derive the expression for its mean and variance.

3. In a game show, you have to choose one of three doors. Behind one door is a car and behind the other two doors is a goat. Once you choose a door, the presenter will open one of the other doors that has a goat. He offers you an opportunity to change your choice to the third unopened and unchosen door. Will you do so? Why? (Give mathematical/probabilistic proof justifying your arguments using Bayes rules/conditioning etc.)

4. Using probability axioms, prove the following
   - $P(\cap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$.
   - Probability that exactly one of the two events A and B occurs is $P(A) + P(B) - 2P(A \cap B)$

5. Give the definition of a sigma algebra. Prove that it is closed under countable intersections.

6. Airlines find that any passenger who books a seat, often fails to turn up with probability $\frac{1}{10}$. Indigo always sells 10 tickets for their 9 seat aeroplane. AirIndia sells 20 tickets for their 18 seat aeroplane. Which airline is more overbooked. Justify your answer. (Hint: Invoke Binomial $B(n, p)$ random variable)

Quiz 1 : Probability and Statistics

September 3, 2023

Question 1

Let X be a random variable with cdf $F_X(x)$. The CDF of a random variable X is defined as
$F_X(x_1) = P(\omega \in \Omega : X(\omega) \leq x_1) = \Sigma_{x<x_1} p_x(x)$ where p_x is the PMF.
For $a < b$, we can consider the following events:
- $C = X \leq a$
- $D = a < X \leq b$
- $E = X \leq b$

Then C and D are mutually exclusive and their union is the event E.
By the third axiom of probability, we know that
P(E) = P(D) + P(C)
$P(X \leq b) = P(a < X \leq b) + P(X \leq a)$
$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$
$P(a < X \leq b) = F_X(b) - F_X(a)$
Marking Scheme:
1 mark for cdf definition
2 marks for writing disjoint sets and invoking probability axiom
2 marks for logic

Question 2

Mean:

$$
\begin{aligned}
E[x] &= \sum_{x=1}^{\infty} x \cdot p(x) && \text{(0.5 mark)} \\
&= \sum_{x=1}^{\infty} x \cdot p \cdot (1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} (x) \cdot (1-p)^{x-1} && \text{(0.5 mark)} \\
&= p \left(1 \cdot (1-p)^0 + 2 \cdot (1-p) + 3 \cdot (1-p)^2 + \ldots\right) \\
&= p \left(1 + 2(1-p) + 3(1-p)^2 + \ldots\right) \\
&= p \left(\frac{1}{(1-(1-p))^2}\right) && \text{(1 mark)} \\
&= p \left(\frac{1}{p^2}\right) \\
&= \frac{1}{p} && \text{(0.5 mark)}
\end{aligned}
$$

Variance:

$$\begin{aligned}
\text{Var}(x) &= E(X^2) - [E(X)]^2 && \text{(0.5 mark)}\\
&= E(X(X-1)+X) - \frac{1}{p^2}\\
&= E(X(X-1)) + E(X) - \frac{1}{p^2} && \text{(0.5 mark)}\\
&= \sum_{x=1}^{\infty} x(x-1)p(1-p)^{x-1} + \frac{1}{p} - \frac{1}{p^2}\\
&= \left[2p(1-p) + 6p(1-p)^2 + 12p(1-p)^3 + \dots\right] + \frac{1}{p} - \frac{1}{p^2}\\
&= 2p(1-p)\left[1 + 3(1-p) + 6(1-p)^2 + \dots\right] + \frac{1}{p} - \frac{1}{p^2}\\
&= \frac{2p(1-p)}{p^3} + \frac{1}{p} - \frac{1}{p^2} && \text{(1 mark)}\\
&= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2}\\
&= \frac{1-p}{p^2} && \text{(0.5 mark)}
\end{aligned}$$

Question 3

Suppose you choose Door 1. Let us say presenter chooses door 3. Lets C_i denote the event that door i conceals a car is, and G denote the event that a goat is shown at door 3. (1/2 mark for mentioning bayes' law or conditional probability and 1/2 mark for proper assumptions of events) Using Conditional probability we get,

$$P(C_2|G) = \frac{P(C_2 \cap G)}{P(G)}$$

which we can further write as,

$$P(C_2|G) = \frac{P(C_2 \cap G|C_1) \cdot P(C_1) + P(C_2 \cap G|\neg C_1) \cdot P(\neg C_1)}{P(G)}$$

Now, let's find the required probabilities:
1. $P(C_2 \cap G|C_1) = 0$ because car cant be behind both the doors 1 and 2.

2. $P(C_2 \cap G|\neg C_1) = 1$ the probability that car is behind door 2 and goat is behind door 3 given that car is not behind door 1 can happen with probability 1.
3. $P(G|C_2) = 1$ probability that behind door 3 there is a goat, given that behind door 1 there is a car, these two events are independent, and the probability that behind door 3 there is a goat is equal to 1, as presenter always opens a door with a goat behind it.
4. $P(C_2) = P(C_1) = \frac{1}{3}$ because initially, the probability of the car being behind any door is equal ($\frac{1}{3}$ for each door).

5. $P(G)$ This total probability can be written as:

$$P(G) = P(G|C_1) \cdot P(C_1) + P(G|\neg C_1) \cdot P(\neg C_1)$$

$$P(G) = 1 \cdot \left(\frac{1}{3}\right) + (1) \cdot \left(\frac{2}{3}\right)$$

Reason : Using 3rd point and 4th point we have made the above substitions

$$P(G) = \frac{1}{3} + \frac{2}{3}$$

$$P(G) = 1$$

(2 marks for finding the probabilities)
So:

$$P(C_2|G) = \frac{1 \cdot \frac{2}{3}}{1} = \frac{2}{3}$$

(1/2 mark)

$$P(C_3|G) = 0$$

By axioms of probability,

$$P(C_1|G) = 1 - \frac{2}{3} - 0 = \frac{1}{3}.$$

(1/2 mark)
Therefore,

$$P(C_2|G) > P(C_1|G)$$

(1/2 mark)
So we should switch the doors because the probabilty of finding the car in other door his higher than door we chose first. (1/2 mark)

Note : Any other method with proper mathematical proof will be graded in a similar fashion for the approach.

Question 4

For Part 1

We have

$P(\bigcap_{i=1}^{n} A_i) = (P(\bigcup_{i=1}^{n} A_i^c)^c)$ [de Morgan's law]

$= 1 - P(\bigcup_{i=1}^{n} A_i^c)$

We have

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\Rightarrow P(A \cup B) \leq P(A) + P(B)$

By using induction on the above inequality for n events, we have

$P(\bigcup_{i=1}^{n} A_i^c) \leq \sum_{i=1}^{n} P(A_i^c)$

So, by using the equation, we get

$\Rightarrow 1 - P(\bigcap_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i^c)$

$\Rightarrow P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$

MARKING SCHEME:
+0.5 for using de Morgan's law
+2.0 for stating and proving above inequality
+0.5 if stating Boole's inequality (the inequality above) without proof

For Part 2

The given problem implies that either (i) A happens and B does not or (ii) A does not happen and B happens. So, we can write the desired probability as :-

$P(A \text{ or } B) = P(A \cap B^c) + P(A^c \cap B)$ (1 mark)

By law of total probability applied on event A, we have

$P(A) = P(A \cap B^c) + P(A \cap B) \ldots (1)$ (0.5 marks)

By law of total probability applied on even B, we have

$P(B) = P(B \cap A^c) + P(B \cap A) \ldots (2)$ (0.5 marks)

Adding equations (1) and (2), we get
$\Rightarrow P(A) + P(B) = 2P(A \cap B) + P(A^c \cap B) + P(A \cap B^c)$
$\Rightarrow P(A^c \cap B) + P(A \cap B^c) = P(A) + P(B) - 2P(A \cap B)$ (0.5 marks)
MARKING SCHEME :
+1 mark for writing correct expression for the problem
+0.5 mark for applying law of total probability on A
+0.5 mark for applying law of total probability on B
+0.5 mark for solving the two equations

Question 5

Definition : Event space or sigma-algebra $\mathcal{F}$ is a collection of measurable sets (1/2 marks for Definition) that satisfy:

1. $\emptyset \in \mathcal{F}$ - (1/2 marks)
2. $A \in \mathcal{F} \implies A^C \in \mathcal{F}$ (closed under complementation) - (1/2 marks)
3. $A_k \in \mathcal{F} \implies \bigcup_{k \in I} A_k \in \mathcal{F} k \in I$ (closed under countable union) - (1/2 marks)

Now, we will prove that sigma-algebra $\mathcal{F}$ is closed under countable intersections:

$$
\begin{aligned}
A_k \in \mathcal{F} &\implies A_k^C \in \mathcal{F} && \text{(closed under complementation) - 1/2 marks} \\
&\implies \left(\bigcup_{k \in I} A_k^C\right) \in \mathcal{F} && \text{(closed under union) - 1/2 marks} \\
&\implies \left(\bigcup_{k \in I} A_k^C\right)^C \in \mathcal{F} && \text{(closed under complementation) - 1/2 marks} \\
&\implies \left(\bigcap_{k \in I} (A_k^C)^C\right) \in \mathcal{F} && \text{(de Morgan's Law) - 1 marks} \\
&\implies \bigcap_{k \in I} A_k \in \mathcal{F} \\
&\implies \mathcal{F} \textit{ is closed under intersection.} \text{ - 1/2 marks}
\end{aligned}
$$

Question 6

The plane which has the higher probability of more people showing up than the number of seats is more overbooked. (1 Mark)
Probability that k people show up for a flight with l seats is

$$\binom{k}{l}\left(\frac{9}{10}\right)^{k}\left(\frac{1}{10}\right)^{l-k}(1Mark)$$

Here $\binom{k}{l}$ stands for $\frac{k!}{(l!)(k-l)!}$ and $k \geq l$
Let N people come to the indigo flight and M people for Air india The probability that more people show up for the indigo flight is than number of seats is-

$$P(N > 9) = P(N = 10)$$

$$= \binom{10}{10}\left(\frac{9}{10}\right)^{10}\left(\frac{1}{10}\right)^{0}(1Mark)$$

$$= \left(\frac{9}{10}\right)^{10}$$

Probability that more people show up for Air india flight than number of seats is-

$$P(M > 18) = P(M = 19) + P(M = 20)$$

$$= \binom{20}{19}\left(\frac{9}{10}\right)^{19}\left(\frac{1}{10}\right)^{1} + \binom{20}{20}\left(\frac{9}{10}\right)^{20}\left(\frac{1}{10}\right)^{0}(1.5Marks)$$

Thus $P(M > 18) > P(N > 9)$. Hence, Air India is more over booked (0.5 Marks)