

Quiz 1: Probability and Statistics

August 25, 2022

1 5 Marks

1. The CDF of a random variable X is defined as $F(x) = \mathcal{P}(\omega \in \Omega : \mathcal{X}(\omega) \leq x) = \sum_{x \leq x_1} p_X(x)$ where p_X is the PMF. Prove that:
 $\mathcal{P}(a < X \leq b) = F_X(b) - F_X(a)$

Solution: Since the outcomes for random variables define events in the event space, which is why we are able to assign probabilities to such outcomes. Let X be a random variable with cdf $F_X(x)$. 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

$$\begin{aligned} P(E) &= P(D) + P(C) \\ P(X \leq b) &= P(a < X \leq b) + P(X \leq a) \\ P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ P(a < X \leq b) &= F_X(b) - F_X(a) \end{aligned}$$

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

Solution: For the geometric distribution, the range $R_x = 1, 2, 3, \dots$ and the PMF is given by $P_x(k) = (1-p)^{k-1}p$, for $k = 1, 2, \dots$, where $0 < p < 1$. Thus we can write

$$\begin{aligned} E[X] &= \sum_{x_k \in R_x} x_k P_x(x_k) \\ E[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\ E[X] &= p \sum_{k=1}^{\infty} k(1-p)^{k-1} \end{aligned}$$

We know that:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \text{ for } |x| < 1$$

Taking derivative of this equation with respect to x

$$\frac{d}{dx} \sum_{k=0}^{\infty} x^k = \frac{d}{dx} \frac{1}{1-x}$$

Thus we have

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

For the expectation we can write

$$E[X] = p \frac{1}{(1-(1-p))^2}$$

$$E[X] = p \frac{1}{p^2}$$

$$E[X] = \frac{1}{p}$$

For Variance

$$Var[X] = E[X^2] - E[X]^2$$

$$E[X^2] = \sum_{k=0}^{\infty} k^2 p_x(k)$$

$$E[X^2] = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1}$$

$$E[X^2] = p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1}$$

Let $(1-p) = q$

$$\sum_{x=1}^{\infty} x^2 q^{x-1} = \frac{1+q}{(1-q)^3}$$

Substituting in $E[X^2]$

$$E[X^2] = p * \frac{(1+(1-p))}{(1-(1-p))^3}$$

$$E[X^2] = p * \frac{2-p}{p^3}$$

$$E[X^2] = \frac{2-p}{p^2}$$

$$Var[X] = \frac{2-p}{p^2} - \frac{1}{p^2}$$

$$Var[X] = \frac{1-p}{p^2}$$

3. Two cards are chosen from a standard deck of 52 cards. Suppose that you win 2 Rs for each heart selected, and lose 1 Rs for each spade selected. Other suits (clubs or diamonds) bring neither win nor loss. Let X denote your winnings. Determine the probability mass function $p_X(x)$.

Solution: Considering all the possible cases for the Winnings

- Winnings > 4 Cases: None
 - $P(\text{With replacement}) = 0$
 - $P(\text{Without replacement}) = 0$
- Winnings $= 4$ Cases: Both hearts
 - $P(\text{With replacement}) = \frac{13}{52} * \frac{13}{52} = \frac{169}{2704} = \frac{1}{16}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{12}{51} = \frac{156}{2652} = \frac{1}{17}$
- Winnings $= 3$ Cases: None
 - $P(\text{With replacement}) = 0$
 - $P(\text{Without replacement}) = 0$
- Winnings $= 2$ Cases: One heart, other club or diamond
 - $P(\text{With replacement}) = 2 * (\frac{13}{52} * \frac{26}{52}) = \frac{676}{2704} = \frac{1}{4}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{26}{51} + \frac{26}{52} * \frac{13}{51} = \frac{676}{2652} = \frac{13}{51}$
- Winnings $= 1$ Cases: One heart, One spade
 - $P(\text{With replacement}) = 2 * (\frac{13}{52} * \frac{13}{52}) = \frac{338}{2704} = \frac{1}{8}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{13}{51} + \frac{13}{52} * \frac{13}{51} = \frac{338}{2652} = \frac{13}{102}$
- Winnings $= 0$ Cases: Both clubs, both diamonds, one club one diamond
 - $P(\text{With replacement}) = \frac{13}{52} * \frac{13}{52} + \frac{13}{52} * \frac{13}{52} + 2 * (\frac{13}{52} * \frac{13}{52}) = \frac{676}{2704} = \frac{1}{4}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{12}{51} + \frac{13}{52} * \frac{12}{51} + 2 * (\frac{13}{52} * \frac{13}{51}) = \frac{650}{2652} = \frac{25}{102}$
- Winnings $= -1$ Cases: One spade, Other club or diamond
 - $P(\text{With replacement}) = 2 * (\frac{13}{52} * \frac{26}{52}) = \frac{676}{2704} = \frac{1}{4}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{26}{51} + \frac{26}{52} * \frac{13}{51} = \frac{676}{2652} = \frac{13}{51}$
- Winnings $= -2$ Cases: Both spades
 - $P(\text{With replacement}) = \frac{13}{52} * \frac{13}{52} = \frac{169}{2704} = \frac{1}{16}$
 - $P(\text{Without replacement}) = \frac{13}{52} * \frac{12}{51} = \frac{156}{2652} = \frac{1}{17}$
- Winnings < -2 Cases: None
 - $P(\text{With replacement}) = 0$
 - $P(\text{Without replacement}) = 0$

4. For a random variable X with mean μ , its variance $Var(X)$ is defined as $E[(X - \mu)^2]$. Prove that $Var(aX + b) = a^2 Var(X)$ for arbitrary constants a and b

Solution:

$$Var(X) = E[(X - \mu)^2]$$

$$Var(X) = E[X^2 + (\mu)^2 - 2 * \mu * X]$$

$$Var(X) = E[X^2] - 2E[X * \mu] + E[\mu^2]$$

$$Var(X) = E[X^2] - 2\mu * E[X] + \mu^2$$

$$Var(x) = E[X^2] - \mu^2$$

$$Var(aX + b) = E[(aX + b)^2] - (E[aX + b])^2$$

$$Var(aX + b) = E[a^2 X^2 + b^2 + 2aXb] - (aE[X] + b)(aE[X] + b)$$

$$Var(aX + b) = a^2 E[X^2] + b^2 + 2abE[X] - a^2 E[X]^2 - b^2 - 2abE[X]$$

$$Var(aX + b) = a^2 E[X^2] - a^2 E[X]^2$$

$$Var(aX + b) = a^2 (E[X^2] - E[X]^2)$$

$$Var(aX + b) = a^2 (Var(X))$$

2 10 Marks

1. Let random variable \mathcal{X} denote the outcome of a dice. Plot the cumulative distribution function (CDF) of \mathcal{X} . Also find the mean and variance of \mathcal{X} . Additionally prove that (prove! don not numerically verify. Start with either RHS or LHS and prove the other side.)

$$\sum_{x \in \{1,2,\dots,6\}} xp_X(x) = 1 + \sum_{x \in \{1,2,\dots,6\}} (1 - F_X(x))$$

(Hint: Write the CDF on the rhs in terms of PMF). The RHS is an alternative formula to get the expectation of non-negative random variables using the CDF.)

10 marks

Q1. Let random variable X denote the outcome of a dice...
... prove that:

$$\sum_{x \in \{1, 2, \dots, 6\}} x P_X(x) = 1 + \sum_{x \in \{1, 2, \dots, 6\}} (1 - F_X(x))$$

Ans. Assuming we have a fair dice, the PMF of X is given by:

I
$$P_X(x=k) = \frac{1}{6}, \quad k \in \{1, 2, 3, \dots, 6\}$$

→ The CDF of X can be computed by sequentially summing up these probabilities:

$$F_X(x) = P_X(X \leq x)$$

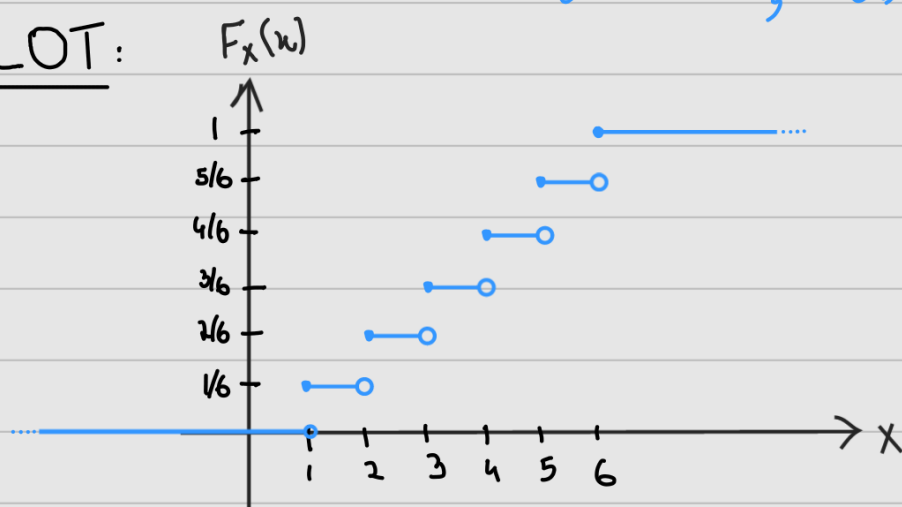
Since X is a discrete random variable;

$$F_X(x) = \sum_{k=1}^n P_X(X = x_k \mid x_k \leq x)$$

→ So, we have:

$$F_X(x) = \begin{cases} 0 & ; x < 1 \\ z/6 & ; z \leq x < z+1, z \in \{1, 2, 3, \dots, 5\} \\ 1 & ; x \geq 6 \end{cases}$$

PLOT:



II
$$\text{mean}(\mu) = E[X] = \sum_{x \in \{1, 2, \dots, 6\}} x P_X(x)$$

$$\rightarrow \mu = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6}$$

$$\rightarrow \boxed{\mu = 3.5}$$

$$\text{Var}(X) = E[X^2] - \mu^2 = \sum_{x \in \{1, 2, \dots, 6\}} x^2 P_X(x) - \mu^2$$

$$\rightarrow \text{Var}(X) = \frac{1}{6} [1^2 + 2^2 + \dots + 6^2] - 3.5^2$$

$$\rightarrow \boxed{\text{Var}(X) = \frac{35}{12} = 2.92}$$

III TP: $\sum_{x \in \{1, 2, \dots, 6\}} x P_X(x) = 1 + \sum_{x \in \{1, 2, \dots, 6\}} (1 - F_X(x)) \quad - \textcircled{1}$

Proof: Method-1

$$\text{Let } S = \sum_{x \in \{1, 2, \dots, 6\}} x P_X(x)$$

In LHS of $\textcircled{1}$, we have:

$$\rightarrow S = \begin{bmatrix} 1 P_X(1) \\ + 2 P_X(2) \\ + 3 P_X(3) \\ + 4 P_X(4) \\ + 5 P_X(5) \\ + 6 P_X(6) \end{bmatrix} = \begin{bmatrix} P_X(1) \\ + P_X(2) + P_X(2) \\ + P_X(3) + P_X(3) + P_X(3) \\ + P_X(4) + P_X(4) + P_X(4) + P_X(4) \\ + P_X(5) + P_X(5) + P_X(5) + P_X(5) + P_X(5) \\ + P_X(6) + P_X(6) + P_X(6) + P_X(6) + P_X(6) + P_X(6) \end{bmatrix}$$

$\underbrace{\hspace{1cm}}_1 \quad \underbrace{\hspace{1cm}}_{P_X(X>1)} \quad \underbrace{\hspace{1cm}}_{P_X(X>2)} \quad \underbrace{\hspace{1cm}}_{P_X(X>3)} \quad \underbrace{\hspace{1cm}}_{P_X(X>4)} \quad \underbrace{\hspace{1cm}}_{P_X(X>5)}$

$$\rightarrow S = 1 + P_X(X>1) + P(X>2) + \dots + P(X>6) \quad [P(X>6)=0]$$

$$\rightarrow S = 1 + (1 - P_X(X \leq 1)) + (1 - P_X(X \leq 2)) + \dots + (1 - P_X(X \leq 6))$$

$$\rightarrow S = 1 + (1 - F_X(1)) + (1 - F_X(2)) + \dots + (1 - F_X(6))$$

$$\rightarrow S = 1 + \sum_{x \in \{1, 2, \dots, 6\}} (1 - F_X(x))$$

$$\Rightarrow \underline{S = \text{RHS}}$$

$\Rightarrow \text{HP}$

Method-2

→ In RHS,

$$S = 1 + \sum_{x \in \{1, 2, \dots, 6\}} (1 - F_x(x))$$

$$\rightarrow S = \sum_{x \in \{1, 2, \dots, 6\}} (P_x(x)) + \sum_{x \in \{1, 2, \dots, 6\}} (1 - F_x(x))$$

$$\rightarrow S = \sum_{x \in \{1, 2, \dots, 6\}} (P_x(x)) + \sum_{x \in \{1, 2, \dots, 6\}} P_x(X > x)$$

$$\rightarrow S = \sum_{x \in \{1, 2, \dots, 6\}} (P_x(x)) + \sum_{x \in \{1, 2, \dots, 6\}} \sum_{k \in \{1, 2, \dots, 6\}} P_x(X = x_k | x_k > x)$$

→ Now, in the term $\sum_{x \in \{1, 2, \dots, 6\}} \sum_{k \in \{1, 2, \dots, 6\}} P_x(X = x_k | x_k > x)$;

for $x=1$, we get terms $P_x(2), P_x(3), \dots, P_x(6)$
 $x=2$, we get terms $P_x(3), \dots, P_x(6)$
 \vdots
 $x=6$, we get no term

→ So in this term, the weight of:

$$P_x(1) : 0$$

$$P_x(2) : 1$$

\vdots

$$P_x(6) : 5$$

$$\Rightarrow P_x(x) : x - 1$$

$$\Rightarrow S = \sum_{x \in \{1, 2, \dots, 6\}} (P_x(x)) + \sum_{x \in \{1, 2, \dots, 6\}} (x-1)(P_x(x))$$

$$\Rightarrow S = \sum_{x \in \{1, 2, \dots, 6\}} x P_x(x)$$

$$\Rightarrow \underline{S = \text{LHS}}$$

$$\Rightarrow \underline{\underline{\text{HP}}}$$