1. Let X be a discrete random variable with the following PMF

$$P_X(x) = \begin{cases} \frac{1}{2} & \text{for } x = 0 \\ \frac{1}{3} & \text{for } x = 1 \\ \frac{1}{6} & \text{for } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find R_X, the range of the random variable X.

(b) Find $P(X \geq 1.5)$.

(c) Find $P(0 < X < 2)$.

(d) Find $P(X = 0|X < 2)$

Solution:

(a) The range of X can be found from the PMF. The range of X consists of possible values for X. Here we have

$$R_X = \{0, 1, 2\}.$$

(b) The event $X \geq 1.5$ can happen only if X is 2. Thus,

$$\begin{aligned} P(X \geq 1.5) &= P(X = 2) \\ &= P_X(2) = \frac{1}{6} \end{aligned} \tag{1}$$

(c) Similarly, we have

$$\begin{aligned} P(0 < X < 2) &= P(X = 1) \\ &= P_X(1) = \frac{1}{3} \end{aligned} \tag{2}$$

(d) This is a conditional probability problem, so we can use our famous formula

$P(A|B) = \frac{P(A \cap B)}{P(B)}$. We have

$$
\begin{aligned}
P(X=0|X<2) &= \frac{P\big(X=0, X<2\big)}{P(X<2)} \\
&= \frac{P(X=0)}{P(X<2)} \\
&= \frac{P_X(0)}{P_X(0)+P_X(1)} \\
&= \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}} = \frac{3}{5}
\end{aligned}
$$

2. Let X be the number of the cars being repaired at a repair shop. We have the following information:

   - At any time, there are at most 3 cars being repaired.
   - The probability of having 2 cars at the shop is the same as the probability of having one car.
   - The probability of having no car at the shop is the same as the probability of having 3 cars.
   - The probability of having 1 or 2 cars is half of the probability of having 0 or 3 cars.

   Find the PMF of X.

   Solution:

   The range of X is possible values for the number of cars being repaired. Based on the information:

$$R_X = \{0, 1, 2, 3\}.$$

$$
\begin{aligned}
P_X(1) &= P_X(2) \\
P_X(0) &= P_X(3) \\
P_X(2) &= \frac{1}{2}P_X(3) \\
P_X(1)+P_X(2) &= \frac{1}{2}(P_X(0)+P_X(3)) \\
&= \frac{1}{2}(2 \times P_X(0)) = \frac{1}{2}(2 \times P_X(3)) \qquad (3) \\
&= P_X(0) \qquad (4)
\end{aligned}
$$

Let $P_X(1) = \alpha$.

Then:

$$\begin{cases} P_X(1) & = P_X(2) = \alpha \\ P_X(0) & = P_X(3) = 2\alpha \end{cases}$$

Then, we have the following equation:

$$\begin{aligned} \sum_{k=0}^{3} P_X(k) = 1 \quad & \longrightarrow \quad 2\alpha + \alpha + \alpha + 2\alpha = 1 \\ & \longrightarrow \quad \alpha = \frac{1}{6} \end{aligned} \tag{5}$$

$$\begin{cases} P_X(0) & = P_X(3) = \frac{1}{3} \\ P_X(1) & = P_X(2) = \frac{1}{6} \end{cases}$$

3. I roll two dice and observe two numbers X and Y. If $Z = X - Y$, find the range and PMF of Z.

Solution:

Note $R_X = R_Y = \{1, 2, 3, 4, 5, 6\}$ and $P_X(k) = P_Y(k) = \begin{cases} \frac{1}{6} & \text{for } k = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$

Since $Z = X - Y$, we conclude:

$R_Z = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$

$$\begin{aligned} P_Z(-5) &= P(X = 1, Y = 6) \\ &= P(X = 1) \cdot P(Y = 6) \quad \text{(Since } X \text{ and } Y \text{ are independent)} && (6) \\ &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} && (7) \end{aligned}$$

$$
\begin{aligned}
P_Z(-4) &= P(X=1,Y=5)+P(X=2,Y=6) \\
&= P(X=1)\cdot P(Y=5)+P(X=2)\cdot P(Y=6)(\text{Since } X \text{ and } Y \text{ are independent}) \quad (8) \\
&= \frac{1}{6}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{18} \quad (9)
\end{aligned}
$$

Similarly:

$$
\begin{aligned}
P_Z(-3) &= P(X=1,Y=4)+P(X=2,Y=5)+P(X=3,Y=6) \\
&= P(X=1)\cdot P(Y=4)+P(X=2)\cdot P(Y=5)+P(X=3)\cdot P(Y=6) \quad (10) \\
&= 3.\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{12} \quad (11)
\end{aligned}
$$

$$
\begin{aligned}
P_Z(-2) &= P(X=1,Y=3)+P(X=2,Y=4)+P(X=3,Y=5)+P(X=4,Y=6) \\
&= P(X=1)\cdot P(Y=3)+P(X=2)\cdot P(Y=4) \quad (12) \\
&+ P(X=3)\cdot P(Y=5)+P(X=4)\cdot P(Y=6) \quad (13) \\
&= 4.\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{9} \quad (14)
\end{aligned}
$$

$$
\begin{aligned}
P_Z(-1) &= P(X=1,Y=2)+P(X=2,Y=3)+P(X=3,Y=4) \\
&+ P(X=4,Y=5)+P(X=5,Y=6) \quad (15) \\
&= P(X=1)\cdot P(Y=2)+P(X=2)\cdot P(Y=3)+P(X=3)\cdot P(Y=4) \quad (16) \\
&+ P(X=4)\cdot P(Y=5)+P(X=5)\cdot P(Y=6) \quad (17) \\
&= 5.\frac{1}{6}\cdot\frac{1}{6}=\frac{5}{36} \quad (18)
\end{aligned}
$$

$$
\begin{aligned}
P_Z(0) &= P(X=1,Y=1)+P(X=2,Y=2)+P(X=3,Y=3) \\
&+ P(X=4,Y=4)+P(X=5,Y=5)+P(X=6,Y=6) \quad (19) \\
&= P(X=1)\cdot P(Y=1)+P(X=2)\cdot P(Y=2)+P(X=3)\cdot P(Y=3) \quad (20) \\
&+ P(X=4)\cdot P(Y=4)+P(X=5)\cdot P(Y=5)+P(X=6)\cdot P(Y=6) \quad (21) \\
&= 6.\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{6} \quad (22)
\end{aligned}
$$

Note by symmetry, we have :

$P_Z(k) = P_Z(-k)$

So,

$$\begin{cases} P_Z(0) = \frac{1}{6} \\ P_Z(1) = P_Z(-1) = \frac{5}{36} \\ P_Z(2) = P_Z(-2) = \frac{1}{9} \\ P_Z(3) = P_Z(-3) = \frac{1}{12} \\ P_Z(4) = P_Z(-4) = \frac{1}{18} \\ P_Z(5) = P_Z(-5) = \frac{1}{36} \end{cases}$$

4. Let X and Y be two independent discrete random variables with the following PMFs:

$$P_X(k) = \begin{cases} \frac{1}{4} & \text{for } k = 1 \\ \frac{1}{8} & \text{for } k = 2 \\ \frac{1}{8} & \text{for } k = 3 \\ \frac{1}{2} & \text{for } k = 4 \\ 0 & \text{otherwise} \end{cases}$$

and

$$P_Y(k) = \begin{cases} \frac{1}{6} & \text{for } k = 1 \\ \frac{1}{6} & \text{for } k = 2 \\ \frac{1}{3} & \text{for } k = 3 \\ \frac{1}{3} & \text{for } k = 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X \leq 2 \text{ and } Y \leq 2)$.

(b) Find $P(X > 2 \text{ or } Y > 2)$.

(c) Find $P(X > 2 | Y > 2)$.

(d) Find $P(X < Y)$.

Solution:

(a) X and Y are two independent random variables. So:

$$\begin{aligned} P(X \leq 2 \text{ and } Y \leq 2) &= P(X \leq 2) \cdot P(Y \leq 2) \\ &= \big(P_X(1) + P_X(2)\big) \cdot \big(P_Y(1) + P_Y(2)\big) \quad (23) \\ &= (\frac{1}{4} + \frac{1}{8})(\frac{1}{6} + \frac{1}{6}) \quad (24) \\ &= \frac{3}{8} \cdot \frac{1}{3} = \frac{1}{8} \quad (25) \end{aligned}$$

(b) Using $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and the fact that X and Y are two independent random variables:

$$\begin{aligned} P(X > 2 \text{ or } Y > 2) &= P(X > 2) + P(Y > 2) - P(X > 2 \text{ and } Y > 2) \quad (26) \\ &= (\frac{1}{8} + \frac{1}{2}) + (\frac{1}{3} + \frac{1}{3}) - (\frac{1}{8} + \frac{1}{2}) \cdot (\frac{1}{3} + \frac{1}{3}) \quad (27) \\ &= \frac{5}{8} + \frac{2}{3} - \frac{5}{8} \cdot \frac{2}{3} = \frac{21}{24} = \frac{7}{8} \quad (28) \end{aligned}$$

(c) Since X and Y are two independent random variables:

$$\begin{aligned} P(X > 2 | Y > 2) &= P(X > 2) \\ &= \frac{1}{8} + \frac{1}{2} = \frac{5}{8} \end{aligned} \quad (29)$$

(d) Using the law of total probability:

$$\begin{aligned} P(X < Y) &= \sum_{k=1}^{4} P(X < Y | Y = k) \cdot P(Y = k) \\ &= P(X < 1 | Y = 1) \cdot P(Y = 1) + P(X < 2 | Y = 2) \cdot P(Y = 2) \\ &+ P(X < 3 | Y = 3) \cdot P(Y = 3) + P(X < 4 | Y = 4) \cdot P(Y = 4) \\ &= 0 \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + (\frac{1}{4} + \frac{1}{8}) \cdot \frac{1}{3} + (\frac{1}{4} + \frac{1}{8} + \frac{1}{8}) \cdot \frac{1}{3} \\ &= \frac{1}{24} + \frac{1}{8} + \frac{1}{6} = \frac{8}{24} = \frac{1}{3} \quad (30) \end{aligned}$$

5. 50 students live in a dormitory. The parking lot has the capacity for 30 cars. If each student has a car with probability $\frac{1}{2}$ (independently from other students), what is the probability that there won't be enough parking spaces for all the cars.

Solution:

If X is the number of the cars owned by 50 students in the dormitory, then:

$X \sim Binomial(50, \frac{1}{2})$

Thus:

$$\begin{aligned}
P(X > 30) &= \sum_{k=31}^{50} \binom{50}{k} (\frac{1}{2})^k (\frac{1}{2})^{50-k} \\
&= \sum_{k=31}^{50} \binom{50}{k} (\frac{1}{2})^{50} \qquad (31) \\
&= (\frac{1}{2})^{50} \sum_{k=31}^{50} \binom{50}{k} \qquad (32)
\end{aligned}$$

6. (The matching problem) N guests arrive at a party. Each person is wearing a hat. We collect all the hats and then randomly redistribute the hats, giving each person one of the N hats randomly. Let X_N be the number of people who receive their own hats. Find PMF of X_N.

   Hint: We previously found that

   $P(X_N = 0) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots (-1)^N \frac{1}{N!}.$ for $N = 1, 2, \cdots.$

   Using this, find $P(X_N = k)$ for all $k \in \{0, 1, \cdots N\}$

   Solution:

   Let $a_N = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + (-1)^N \frac{1}{N!} = P(X_N = 0)$.

   What is $P(X_N = 1)$?

   This is the probability that exactly one person receives his/her hat. We can fix this person. Therefore, the rest of the people ($N - 1$ people) won't receive their own hats.

   So, there are N ways to choose the person who gets his/her own hat.

   The probability that the chosen person gets his/her hat is equal to $\frac{1}{N}$.

   The probability that none of the other $N - 1$ people receive their own hats is $P(X_{N-1} = 0)$.

   So, we have the following equation:

$$\begin{aligned}
P(X_N = 1) &= N \cdot \frac{1}{N} \cdot P(X_{N-1} = 0) \\
&= P(X_{N-1} = 0) = a_{N-1} \qquad (33)
\end{aligned}$$

Similarly for calculating $P(X_N = 2)$, there are $\binom{N}{2}$ ways to choose two people who get their own hats.

$\frac{1}{N} \cdot \frac{1}{N-1}$ is the probability that the two chosen people receive their own hats.

And $P(X_{N-2} = 0)$ is the probability that none of the other $N - 2$ people receive their own hats.

$$\begin{aligned} P(X_N = 2) &= \binom{N}{2} \cdot \frac{1}{N} \cdot \frac{1}{N-1} \cdot P(X_{N-2} = 0) \\ &= \frac{1}{2} P(X_{N-2} = 0) = \frac{1}{2} a_{N-2} \end{aligned} \tag{34}$$

And in general:

$$\begin{aligned} P(X_N = k) &= \binom{N}{k} \cdot \frac{1}{N} \cdot \frac{1}{N-1} \cdots \frac{1}{N-k+1} \cdot P(X_{N-k} = 0) \\ &= \frac{1}{k!} P(X_{N-k} = 0) = \frac{1}{k!} a_{N-k} \quad \text{for} \quad k = 0, 1, 2, \cdots, N \end{aligned} \tag{35}$$

7. For each of the following random variables, find $P(X > 5)$, $P(2 < X \leq 6)$ and $P(X > 5|X < 8)$. You do not need to provide the numerical values for your answers. In other words, you can leave your answers in the form of sums.

   (a) $X \sim Geometric(\frac{1}{5})$
   (b) $X \sim Binomial(10, \frac{1}{3})$
   (c) $X \sim Pascal(3, \frac{1}{2})$
   (d) $X \sim Hypergeometric(10, 10, 12)$
   (e) $X \sim Poisson(5)$

Solution:

First note that if $R_X \subset \{0, 1, 2, \cdots\}$, then

- $P(X > 5) = \sum_{k=6}^{\infty} P_X(k) = 1 - \sum_{k=0}^{5} P_X(k)$.
- $P(2 < X \leq 6) = P_X(3) + P_X(4) + P_X(5) + P_X(6)$.
- $P(X > 5|X < 8) = \frac{P(5<X<8)}{P(X<8)} = \frac{P_X(6)+P_X(7)}{\sum_{k=0}^{7} P_X(k)}$.

So,

(a) $X \sim Geometric(\frac{1}{5}) \longrightarrow \quad P_X(k) = (\frac{4}{5})^{k-1}(\frac{1}{5}) \quad$ for $k = 1, 2, 3, \cdots$
Therefore,

$$\begin{aligned} P(X > 5) &= 1 - \sum_{k=1}^{5}(\frac{4}{5})^{k-1}(\frac{1}{5}) \\ &= 1 - (\frac{1}{5}) \cdot \big(1 + (\frac{4}{5}) + (\frac{4}{5})^2 + (\frac{4}{5})^3 + (\frac{4}{5})^4\big) && (36) \\ &= 1 - (\frac{1}{5}) \cdot \frac{1 - (\frac{4}{5})^5}{1 - (\frac{4}{5})} = (\frac{4}{5})^5. && (37) \end{aligned}$$

Note that we can obtain this result directly from the random experiment behind the geometric random variable:
$P(X < 5) = P(\text{No heads in 5 coin tosses}) = (\frac{4}{5})^5$

$$\begin{aligned} P(2 < X \leq 6) &= P_X(3) + P_X(4) + P_X(5) + P_X(6) \\ &= (\frac{1}{5})(\frac{4}{5})^2 + (\frac{1}{5})(\frac{4}{5})^3 + (\frac{1}{5})(\frac{4}{5})^4 + (\frac{1}{5})(\frac{4}{5})^5 && (38) \\ &= (\frac{1}{5})(\frac{4}{5})^2 \cdot \big(1 + \frac{4}{5} + (\frac{4}{5})^2 + (\frac{4}{5})^3\big) && (39) \\ &= (\frac{4}{5})^2\big(1 - (\frac{4}{5})^4\big). && (40) \end{aligned}$$

$$\begin{aligned} P(X > 5 | X < 8) &= \frac{P(5 < X < 8)}{P(X < 8)} = \frac{P_X(6) + P_X(7)}{\sum_{k=1}^{7} P_X(k)} \\ &= \frac{(\frac{1}{5})\big((\frac{4}{5})^5 + (\frac{4}{5})^6\big)}{(\frac{1}{5})\sum_{k=1}^{7}(\frac{4}{5})^{k-1}} && (41) \\ &= \frac{(\frac{4}{5})^5 + (\frac{4}{5})^6}{1 + (\frac{4}{5}) + \cdots (\frac{4}{5})^6} && (42) \end{aligned}$$

(b) $X \sim Binomial(10, \frac{1}{3}) \longrightarrow \quad P_X(k) = \binom{10}{k}(\frac{1}{3})^k(\frac{2}{3})^{10-k} \quad$ for $k = 0, 1, 2, \cdots, 10$
So,

$$\begin{aligned}
P(X>5) &= 1-\sum_{k=0}^{5}\binom{10}{k}(\frac{1}{3})^k(\frac{2}{3})^{10-k} \\
&= 1-\big[\binom{10}{0}(\frac{1}{3})^0(\frac{2}{3})^{10}+\binom{10}{1}(\frac{1}{3})^1(\frac{2}{3})^9+\binom{10}{2}(\frac{1}{3})^2(\frac{2}{3})^8 && (43)\\
&+\binom{10}{3}(\frac{1}{3})^3(\frac{2}{3})^7+\binom{10}{4}(\frac{1}{3})^4(\frac{2}{3})^6+\binom{10}{5}(\frac{1}{3})^5(\frac{2}{3})^5\big] && (44)\\
&. && (45)
\end{aligned}$$

We can also solve it in a direct way:

$$\begin{aligned}
P(X>5) &= \sum_{k=6}^{10}\binom{10}{k}(\frac{1}{3})^k(\frac{2}{3})^{10-k} \\
&= \binom{10}{6}(\frac{1}{3})^6(\frac{2}{3})^4+\binom{10}{7}(\frac{1}{3})^7(\frac{2}{3})^3+\binom{10}{8}(\frac{1}{3})^8(\frac{2}{3})^2 && (46)\\
&+\binom{10}{9}(\frac{1}{3})^9(\frac{2}{3})^1+\binom{10}{10}(\frac{1}{3})^{10}(\frac{2}{3})^0 && (47)\\
&= (\frac{1}{3})^{10}\cdot\Big(\binom{10}{6}2^4+\binom{10}{7}2^3+\binom{10}{8}2^2+\binom{10}{9}2+\binom{10}{10}\Big) && (48)\\
&= (\frac{1}{3})^{10}\cdot\Big(\binom{10}{6}2^4+\binom{10}{7}2^3+\binom{10}{8}2^2+21\Big). && (49)
\end{aligned}$$

$$\begin{aligned}
P(2<X\le 6) &= P_X(3)+P_X(4)+P_X(5)+P_X(6) \\
&= \binom{10}{3}(\frac{1}{3})^3(\frac{2}{3})^7+\binom{10}{4}(\frac{1}{3})^4(\frac{2}{3})^6 && (50)\\
&+\binom{10}{5}(\frac{1}{3})^5(\frac{2}{3})^5+\binom{10}{6}(\frac{1}{3})^6(\frac{2}{3})^4 && (51)\\
&= (\frac{1}{3})^{10}[\binom{10}{3}2^7+\binom{10}{4}2^6+\binom{10}{5}2^5+\binom{10}{6}2^4] && (52)\\
&= 2^4(\frac{1}{3})^{10}[\binom{10}{3}2^3+\binom{10}{4}2^2+\binom{10}{5}2+\binom{10}{6}] && (53)\\
&. && (54)
\end{aligned}$$

$$P(X>5|X<8)=\frac{P(5<X<8)}{P(X<8)}=\frac{P_X(6)+P_X(7)}{\sum_{k=0}^{7}P_X(k)}$$

$$=\frac{P_X(6)+P_X(7)}{1-P_X(8)-P_X(9)-P_X(10)} \tag{55}$$

$$=\frac{\binom{10}{6}(\frac{1}{3})^6(\frac{2}{3})^4+\binom{10}{7}(\frac{1}{3})^7(\frac{2}{3})^3}{1-((\binom{10}{8}(\frac{1}{3})^8(\frac{2}{3})^2+\binom{10}{9}(\frac{1}{3})^9(\frac{2}{3})^1+\binom{10}{10}(\frac{1}{3})^{10}(\frac{2}{3})^0)} \tag{56}$$

$$=\frac{(\frac{1}{3})^{10}(2^4\binom{10}{6}+2^3\binom{10}{7})}{1-\big((\frac{1}{3})^{10}(2^2\binom{10}{8}+2\binom{10}{9}+\binom{10}{10})\big)} \tag{57}$$

$$=\frac{(\frac{1}{3})^{10}(2^4\binom{10}{6}+2^3\binom{10}{7})}{1-\big((\frac{1}{3})^{10}(2^2\times 45+2\times 10+1)\big)} \tag{58}$$

$$=\frac{(\frac{1}{3})^{10}\times 2^3(2\binom{10}{6}+\binom{10}{7})}{1-\big((\frac{1}{3})^{10}\times 201\big)} \tag{59}$$

$$=\frac{2^3(2\binom{10}{6}+\binom{10}{7})}{3^{10}-201} \tag{60}$$

(c) $X\sim Pascal(3,\frac{1}{2})\longrightarrow \quad P_X(k)=\binom{k-1}{2}(\frac{1}{2})^k \quad$ for $k=3,4,5,\cdots$
So,

$$P(X>5)=1-\sum_{k=3}^{5}\binom{k-1}{2}(\frac{1}{2})^k$$

$$=1-(\binom{2}{2}(\frac{1}{2})^3+\binom{3}{2}(\frac{1}{2})^4+\binom{4}{2}(\frac{1}{2})^5\big) \tag{61}$$

$$=1-\big((\frac{1}{2})^3+3(\frac{1}{2})^4+6(\frac{1}{2})^5\big) \tag{62}$$

$$=1-(\frac{1}{2})^5\big(4+6+6\big) \tag{63}$$

$$=1-((\frac{1}{2})^5\times 2^4)=\frac{1}{2}. \tag{64}$$

$$P(2<X\le 6)=P_X(3)+P_X(4)+P_X(5)+P_X(6)$$

$$=\binom{2}{2}(\frac{1}{2})^3+\binom{3}{2}(\frac{1}{2})^4+\binom{4}{2}(\frac{1}{2})^5+\binom{5}{2}(\frac{1}{2})^6 \tag{65}$$

$$=(\frac{1}{2})^3+3(\frac{1}{2})^4+6(\frac{1}{2})^5+10(\frac{1}{2})^6 \tag{66}$$

$$=(\frac{1}{2})^6(8+3\times 4+6\times 2+10)=42\times(\frac{1}{2})^6=\frac{21}{32}. \tag{67}$$

$$P(X>5|X<8) = \frac{P(5<X<8)}{P(X<8)} = \frac{P_X(6)+P_X(7)}{\sum_{k=3}^{7} P_X(k)}$$

$$= \frac{\binom{5}{2}(\frac{1}{2})^6 + \binom{6}{2}(\frac{1}{2})^7}{\binom{2}{2}(\frac{1}{2})^3 + \binom{3}{2}(\frac{1}{2})^4 + \binom{4}{2}(\frac{1}{2})^5 + \binom{5}{2}(\frac{1}{2})^6 + \binom{6}{2}(\frac{1}{2})^7} \tag{68}$$

$$= \frac{10(\frac{1}{2})^6 + 15(\frac{1}{2})^7}{(\frac{1}{2})^3 + 3(\frac{1}{2})^4 + 6(\frac{1}{2})^5 + 10(\frac{1}{2})^6 + 15(\frac{1}{2})^7} \tag{69}$$

$$= \frac{20+15}{16+24+24+20+15} = \frac{35}{99} \tag{70}$$

(d) $X \sim Hypergeometric(10, 10, 12)$ $b = r = 10, k = 12$
$R_X = \{max(0, k-r), \cdots, min(k, b)\} = \{2, 3, 4, \cdots, 10\}$
So:
$P_X(k) = \frac{\binom{10}{k}\binom{10}{12-k}}{\binom{20}{12}}$ for $k = 2, 3, \cdots, 10$

$$P(X>5) = 1 - \sum_{k=2}^{5} \frac{\binom{10}{k}\binom{10}{12-k}}{\binom{20}{12}}$$

$$= 1 - \left(\frac{\binom{10}{2}\binom{10}{10}}{\binom{20}{12}} + \frac{\binom{10}{3}\binom{10}{9}}{\binom{20}{12}} + \frac{\binom{10}{4}\binom{10}{8}}{\binom{20}{12}} + \frac{\binom{10}{5}\binom{10}{7}}{\binom{20}{12}}\right) \tag{71}$$

$$= 1 - \frac{1}{\binom{20}{12}}\left(\binom{10}{2}\binom{10}{10} + \binom{10}{3}\binom{10}{9} + \binom{10}{4}\binom{10}{8} + \binom{10}{5}\binom{10}{7}\right) \tag{72}$$

$$= 1 - \frac{1}{\binom{20}{12}}\left(\binom{10}{2} + 10 \cdot \binom{10}{3} + \binom{10}{4}\binom{10}{8} + \binom{10}{5}\binom{10}{7}\right). \tag{73}$$

$$P(2<X\leq 6) = P_X(3)+P_X(4)+P_X(5)+P_X(6)$$

$$= \frac{\binom{10}{3}\binom{10}{9}}{\binom{20}{12}} + \frac{\binom{10}{4}\binom{10}{8}}{\binom{20}{12}} + \frac{\binom{10}{5}\binom{10}{7}}{\binom{20}{12}} + \frac{\binom{10}{6}\binom{10}{6}}{\binom{20}{12}} \tag{74}$$

$$= \frac{1}{\binom{20}{12}}\left(\binom{10}{3}\binom{10}{9} + \binom{10}{4}\binom{10}{8} + \binom{10}{5}\binom{10}{7} + \binom{10}{6}\binom{10}{6}\right) \tag{75}$$

$$= \frac{1}{\binom{20}{12}}\left(10 \times \binom{10}{3} + \binom{10}{4}\binom{10}{8} + \binom{10}{5}\binom{10}{7} + \binom{10}{6}\binom{10}{6}\right) \tag{76}$$

$$. \tag{77}$$

$$
\begin{aligned}
P(X>5|X<8) = \frac{P(5<X<8)}{P(X<8)} &= \frac{P_X(6)+P_X(7)}{\sum_{k=2}^{7} P_X(k)} \\
&= \frac{\frac{\binom{10}{6}\binom{10}{6}}{\binom{20}{12}} + \frac{\binom{10}{7}\binom{10}{5}}{\binom{20}{12}}}{\frac{\binom{10}{2}\binom{10}{10}}{\binom{20}{12}} + \frac{\binom{10}{3}\binom{10}{9}}{\binom{20}{12}} + \frac{\binom{10}{4}\binom{10}{8}}{\binom{20}{12}} + \frac{\binom{10}{5}\binom{10}{7}}{\binom{20}{12}} + \frac{\binom{10}{6}\binom{10}{6}}{\binom{20}{12}} + \frac{\binom{10}{7}\binom{10}{5}}{\binom{20}{12}}} \quad (78) \\
&= \frac{\binom{10}{6}\binom{10}{6} + \binom{10}{7}\binom{10}{5}}{\binom{10}{2}\binom{10}{10} + \binom{10}{3}\binom{10}{9} + \binom{10}{4}\binom{10}{8} + \binom{10}{5}\binom{10}{7} + \binom{10}{6}\binom{10}{6} + \binom{10}{7}\binom{10}{5}} \quad (79) \\
& \quad (80)
\end{aligned}
$$

(e) $X \sim Poisson(5)$

$P_X(k) = \frac{e^{-5}5^k}{k!} \qquad$ for $k = 0, 1, 2, \cdots$

$$
\begin{aligned}
P(X>5) &= 1 - \sum_{k=0}^{5} \frac{e^{-5}5^k}{k!} \\
&= 1 - \left(\frac{5^0e^{-5}}{0!} + \frac{5^1e^{-5}}{1!} + \frac{5^2e^{-5}}{2!} + \frac{5^3e^{-5}}{3!} + \frac{5^4e^{-5}}{4!} + \frac{5^5e^{-5}}{5!}\right) \quad (81) \\
&= 1 - \left(e^{-5} + 5e^{-5} + \frac{25e^{-5}}{2} + \frac{5^3e^{-5}}{3!} + \frac{5^4e^{-5}}{4!} + \frac{5^5e^{-5}}{5!}\right) \quad (82) \\
&= 1 - e^{-5}\left(6 + \frac{25}{2} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!}\right). \quad (83)
\end{aligned}
$$

$$
\begin{aligned}
P(2<X\leq 6) &= P_X(3) + P_X(4) + P_X(5) + P_X(6) \\
&= \frac{e^{-5}5^3}{3!} + \frac{e^{-5}5^4}{4!} + \frac{e^{-5}5^5}{5!} + \frac{e^{-5}5^6}{6!} \quad (84) \\
&= e^{-5}\left(\frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!}\right) \quad (85) \\
&. \quad (86)
\end{aligned}
$$

$$P(X > 5|X < 8) = \frac{P(5 < X < 8)}{P(X < 8)} = \frac{P_X(6) + P_X(7)}{\sum_{k=0}^{7} P_X(k)}$$

$$= \frac{e^{-5}(\frac{5^6}{6!} + \frac{5^7}{7!})}{e^{-5}(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!} + \frac{5^7}{7!})} \tag{87}$$

$$= \frac{\frac{5^6}{6!} + \frac{5^7}{7!}}{\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!} + \frac{5^7}{7!}} \tag{88}$$

$$= \frac{\frac{5^6}{6!} + \frac{5^7}{7!}}{6 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!} + \frac{5^7}{7!}} \tag{89}$$

8. Suppose you take a pass-fail test repeatedly. Let S_k be the event that you are successful in your k^{th} try, and F_k be the event that you fail the test in your k^{th} try. On your first try, you have 50 percent chance of passing the test:

$$P(S_1) = 1 - P(F_1) = \frac{1}{2}.$$

Assume that as you take the test more often, your chance of failing the test goes down. In particular,

$$P(F_k) = \frac{1}{2} \cdot P(F_{k-1}), \quad \text{for } k = 2, 3, 4, \cdots$$

However, the result of different exams are independent. Suppose you take the test repeatedly until you pass the test for the first time. Let X be the total number of tests you take, so $Range(X) = \{1, 2, 3, \cdots\}$.

(a) Find $P(X = 1), P(X = 2)$, and $P(X = 3)$.

(b) Find a general formula for $P(X = k)$ for $k = 1, 2, \cdots$.

(c) Find the probability that you take the test more than 2 times.

(d) Given that you take the test more than once, find the probability that you take the test exactly twice.

Solution:

(a) First note that:

$P(F_1) = \frac{1}{2} \quad P(F_2) = \frac{1}{2} \cdot \frac{1}{2} = (\frac{1}{2})^2 \quad P(F_3) = (\frac{1}{2})^3$

$P(F_k) = (\frac{1}{2})^k \quad$ for $k = 1, 2, \cdots$

$P(S_k) = 1 - P(F_k) = 1 - (\frac{1}{2})^k \quad$ for $k = 1, 2, \cdots$

$$\begin{aligned} P(F_k) &= (\frac{1}{2})^k && \text{for } k = 1, 2, \cdots \\ P(S_k) &= 1 - P(F_k) = 1 - (\frac{1}{2})^k && \text{for } k = 1, 2, \cdots \end{aligned} \tag{90}$$

Thus:

$$\begin{aligned} P(X = 1) &= P(S_1) = \frac{1}{2} \\ P(X = 2) &= P(F_1)P(S_2) \\ &= \frac{1}{2} \cdot \big(1 - (\frac{1}{2})^2\big) = \frac{3}{8} && (91) \\ P(X = 3) &= P(F_1)P(F_2)P(S_3) \\ &= \frac{1}{2} \cdot \frac{1}{4} \cdot \big(1 - (\frac{1}{2})^3\big) = \frac{1}{8} \cdot \frac{7}{8} = \frac{7}{64} && (92) \\ & && (93) \end{aligned}$$

(b)

$$\begin{aligned} P(X = k) &= P(F_1)P(F_2) \cdots P(S_k) \\ &= \frac{1}{2}(\frac{1}{2})^2 \cdots (\frac{1}{2})^{k-1}\big(1 - (\frac{1}{2})^k\big) && (94) \\ &= (\frac{1}{2})^{\frac{k(k-1)}{2}}\big(1 - (\frac{1}{2})^k\big) && (95) \end{aligned}$$

Therefore:

$$\begin{aligned} P(X = k) &= P(F_1)P(F_2) \cdots P(S_k) \\ &= (\frac{1}{2})^{\frac{k(k-1)}{2}}\big(1 - (\frac{1}{2})^k\big) \text{ for } k = 1, 2, \cdots \end{aligned} \tag{96}$$

(c)

$$\begin{aligned} P(X > 2) &= 1 - P(X = 1) - P(X = 2) \\ &= 1 - \frac{1}{2} - \frac{3}{8} = \frac{1}{8} \end{aligned} \tag{97}$$

(d)

$$
\begin{aligned}
P(X=2|X>1) &= \frac{P(X=2 \text{ and } X>1)}{P(X>1)} \\
&= \frac{P(X=2)}{P(X>1)} = \frac{\frac{3}{8}}{1-\frac{1}{2}} && (98) \\
&= \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{4} && (99)
\end{aligned}
$$

9. In this problem, we would like to show that the geometric random variable is **memoryless**. Let $X \sim \textit{Geometric}(p)$. Show that

$$
P(X > m + l | X > m) = P(X > l), \qquad \text{for } m, l \in \{1, 2, 3, \cdots\}
$$

We can interpret this in the following way: remember that a geometric random variable can be obtained by tossing a coin repeatedly until observing the first heads. If we toss the coin several times and do not observe a heads, from now on it is like we start all over again. In other words, the failed coin tosses do not impact the distribution of waiting time from now on. The reason for this is that the coin tosses are independent.

Solution:

Since $X \sim \textit{Geometric}(p)$, we have:

$P_X(k) = (1-p)^{k-1}p$ for $k = 1, 2, ...$

Thus:

$$
\begin{aligned}
P(X > m) &= \sum_{k=m+1}^{\infty} (1-p)^{k-1} p \\
&= (1-p)^m p \sum_{k=0}^{\infty} (1-p)^k \\
&= p(1-p)^m \frac{1}{1-(1-p)} && (100) \\
&= (1-p)^m && (101)
\end{aligned}
$$

Similarly

$$
P(X > m + l) = (1-p)^{m+l} \qquad (102)
$$

Therefore:

$$
\begin{aligned}
P(X > m + l | X > m) &= \frac{P(X > m + l \text{ and } P(X > m))}{P(X > m)} \\
&= \frac{P(X > m + l)}{P(X > m)} \qquad (103) \\
&= \frac{(1-p)^{m+l}}{(1-p)^m} \\
&= (1-p)^l \qquad (104) \\
&= P(X > l) \qquad (105)
\end{aligned}
$$

10. (20%) An urn consists of 20 red balls and 30 green balls. We choose 10 balls at random from the urn. The sampling is done **without** replacement (repetition not allowed).

    (a) (10%) What is the probability that there will be exactly 4 red balls among the chosen balls?

    (b) (10%) Given that there are at least 3 red balls among the chosen balls, what is the probability that there are exactly 4 red balls?

Solution:

a)

$$P(A) = \frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10}}$$

b)

$$
\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(A)}{P(B)} \qquad (106)
\end{aligned}
$$

$$P(B) = \sum_{k=3}^{10} \frac{\binom{20}{k}\binom{30}{10-k}}{\binom{50}{10}} \qquad (107)$$

Therefore:

$$P(B|A) = \frac{\binom{20}{4}\binom{30}{6}}{\sum_{k=3}^{10}\binom{20}{k}\binom{30}{10-k}}$$

11. The number of emails that I get in a weekday (Monday through Friday) can be modeled by a Poisson distribution with an average of $\frac{1}{6}$ emails per minute. The number of emails that I receive on weekends (Saturday and Sunday) can be modeled by a Poisson distribution with an average of $\frac{1}{30}$ emails per minute.

    (a) What is the probability that I get no emails in on an interval of length 4 hours on a Sunday?

    (b) A random day is chosen (all days of the week are equally likely to be selected), and a random interval of length one hour is selected in the chosen day. It is observed that I did not receive any emails in that interval. What is the probability that the chosen day is a weekday?

Solution:

a)

$$T = 4 \times 60 = 240 \text{ min}$$
$$\lambda = 240 \times \frac{1}{30} = 8$$

Thus $X \sim Poisson(\lambda = 8)$

$$P(X = 0) = e^{-\lambda} = e^{-8} \tag{108}$$

b)

Let D be the event that a weekday is chosen and let E be the event that a Saturday or Sunday is chosen. Then:

$$\begin{aligned} P(D) &= \frac{5}{7} \\ P(E) &= \frac{2}{7} \end{aligned} \tag{109}$$

Let A be the event that I receive no emails during the chosen interval then:

$$P(A|D) = e^{-\lambda_1} = e^{-\frac{1}{6}\cdot 60} = e^{-10}$$
$$P(A|E) = e^{-\lambda_2} = e^{-\frac{1}{30}\cdot 60} = e^{-2}$$

Therefore:

$$\begin{aligned}
P(D|A) = \frac{P(A|D).P(D)}{P(A)} &= \frac{e^{-10}\frac{5}{7}}{P(A|D)P(D) + P(A|E)P(E)} \\
&= \frac{e^{-10}\frac{5}{7}}{e^{-10}\frac{5}{7} + e^{-2}\frac{2}{7}} && (110) \\
&= \frac{5}{5 + 2e^{8}} \approx 8.4 \times 10^{-4} && (111)
\end{aligned}$$

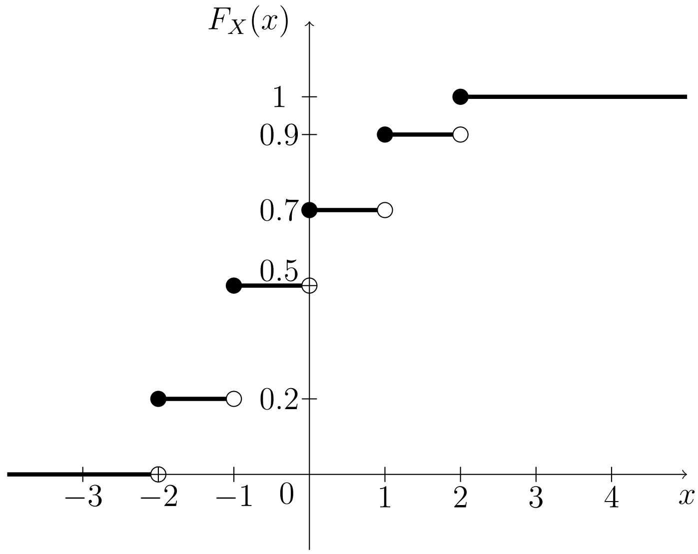
12. Let X be a discrete random variable with the following PMF

$$P_X(x) = \begin{cases} 0.2 & \text{for } x = -2 \\ 0.3 & \text{for } x = -1 \\ 0.2 & \text{for } x = 0 \\ 0.2 & \text{for } x = 1 \\ 0.1 & \text{for } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Find and Plot the CDF of X.

Solution:

$$F_X(x) = \begin{cases} 0 & \text{for } x < -2 \\ 0.2 & \text{for } -2 \leq x < -1 \\ 0.5 & \text{for } -1 \leq x < 0 \\ 0.7 & \text{for } 0 \leq x < 1 \\ 0.9 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$



13. Let X be a discrete random variable with the following CDF:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{6} & \text{for } 0 \le x < 1 \\ \frac{1}{2} & \text{for } 1 \le x < 2 \\ \frac{3}{4} & \text{for } 2 \le x < 3 \\ 1 & \text{for } x \ge 3 \end{cases}$$

Find the range and PMF of X.

Solution:

$$R_X = \{0, 1, 2, 3\}.$$

$$P_X(x) = F_X(x) - F_X(x - \epsilon)$$

$$
\begin{aligned}
P_X(0) &= F_X(0) - F_X(0-\epsilon) = \frac{1}{6} - 0 = \frac{1}{6} \\
P_X(1) &= F_X(1) - F_X(1-\epsilon) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3} \\
P_X(2) &= F_X(2) - F_X(2-\epsilon) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4} \\
P_X(3) &= F_X(3) - F_X(3-\epsilon) = 1 - \frac{3}{4} = \frac{1}{4}
\end{aligned}
\tag{112}
$$

$$
P_X(x) = \begin{cases} \frac{1}{6} & \text{for } x = 0 \\ \\ \frac{1}{3} & \text{for } x = 1 \\ \\ \frac{1}{4} & \text{for } x = 2 \\ \\ \frac{1}{4} & \text{for } x = 3 \\ \\ 0 & \text{otherwise} \end{cases}
$$

14. Let X be a discrete random variable with the following PMF

$$
P_X(k) = \begin{cases} 0.5 & \text{for } k = 1 \\ 0.3 & \text{for } k = 2 \\ 0.2 & \text{for } k = 3 \\ 0 & \text{otherwise} \end{cases}
$$

(a) Find EX.

(b) Find $\mathrm{Var}(X)$, and $SD(X)$.

(c) If $Y = \frac{2}{X}$, find EY.

Solution:

(a)

$$
\begin{aligned}
EX &= \sum_k x_k P_x(x_k) \\
&= 1 \times 0.5 + 2 \times 0.3 + 3 \times 0.2 = 0.5 + 0.6 + 0.6 = 1.7
\end{aligned}
\tag{113}
$$

(b)

$$\begin{aligned} EX^2 &= \sum_k x_k^2 P_x(x_k) \text{ LOTUS} \\ &= (1)^2 \times 0.5 + (2)^2 \times 0.3 + (3)^2 \times 0.2 = 0.5 + 1.2 + 1.8 = 3.5 \end{aligned} \tag{114}$$

Thus,

$$var(X) = EX^2 - (EX)^2 = 3.5 - (1.7)^2 = 0.61$$

$$SD(X) = \sqrt{var(X)} = \sqrt{0.61} = 0.781$$

(c)

$$\begin{aligned} E[\frac{2}{X}] &= \sum_k (\frac{2}{x_k}) P_x(x_k) \quad \text{by LOTUS} \\ &= (\frac{2}{1}) \times 0.5 + (\frac{2}{2}) \times 0.3 + (\frac{2}{3}) \times 0.2 && (115) \\ &= 1 + 0.3 + \frac{0.4}{3} = 1 + \frac{3}{10} + \frac{2}{15} = \frac{43}{30} && (116) \end{aligned}$$

15. Let $X \sim Geometric(\frac{1}{3})$ and let $Y = |X - 5|$. Find the range and PMF of Y.

Solution:

$$R_X = \{1, 2, 3, ...\} \tag{117}$$

$$P_X(k) = \frac{1}{3}\left(\frac{2}{3}\right)^{k-1}, \qquad \text{for } k = 1, 2, 3, ... \tag{118}$$

Thus:

$$R_Y = \{|X-5| \big| X \in R_X\} = 0, 1, 2, ... \tag{119}$$

Thus,

$$\begin{aligned} P_Y(0) &= P(Y=0) = P(|X-5|=0) = P(X=5) \\ &= (\frac{2}{3})^4(\frac{1}{3}) \end{aligned} \tag{120}$$

For $k = 1, 2, 3, 4$

$$\begin{aligned} P_Y(k) &= P(Y=k) = P(|X-5|=k) = P(X=5+k \quad \text{or} \quad X=5-k) \\ &= P_X(5+k) + P_X(5-k) = [(\frac{2}{3})^{4+k} + (\frac{2}{3})^{4-k}](\frac{1}{3}) \end{aligned} \tag{121}$$

For $k \geq 5$,

$$\begin{aligned} P_Y(k) &= P(Y=k) = P(|X-5|=k) = P(X=5+k) \\ &= P_X(5+k) = (\frac{2}{3})^{4+k}(\frac{1}{3}) \end{aligned} \tag{122}$$

So, in summary:

$$P_Y(k) = \begin{cases} (\frac{2}{3})^{k+4}(\frac{1}{3}) & \text{for } k = 0, 5, 6, 7, 8, ... \\ ((\frac{2}{3})^{k+4} + (\frac{2}{3})^{4-k})(\frac{1}{3}) & \text{for } k = 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

16. Let X be a discrete random variable with the following PMF

$$P_X(k) = \begin{cases} \frac{1}{21} & \text{for } k \in \{-10, -9, \cdots, -1, 0, 1, \cdots, 9, 10\} \\ 0 & \text{otherwise} \end{cases}$$

The random variable $Y = g(X)$ is defined as

$$Y = g(X) = \begin{cases} 0 & \text{if } X \leq 0 \\ X & \text{if } 0 < X \leq 5 \\ 5 & \text{otherwise} \end{cases}$$

Find the PMF of Y.

Solution:

$$R_X = \{-10, -9, ..., 9, 10\} \tag{123}$$

$Y = g(X)$ Thus:

$$R_Y = \{0, 1, 2, 3, 4, 5\} \tag{124}$$

Thus,

$$P_Y(0) = P(X \leq 0) = \sum_{k=-10}^{0} P_X(k) = \frac{11}{21}$$

For $k = 1, 2, 3, 4$

$$\begin{aligned} P_Y(1) &= P(Y = 1) = P(X = 1) = \frac{1}{21} \\ P_Y(2) &= P_Y(3) = P_Y(4) = \frac{1}{21} \\ P_Y(5) &= P(X \geq 5) \\ &= P_X(5) + P_X(6) + P_X(7) + P_X(8) + P_X(9) + P_X(10) && (125) \\ &= \frac{6}{21} && (126) \\ & && (127) \end{aligned}$$

Thus:

$$P_Y(k) = \begin{cases} \frac{11}{21} & \text{for } k = 0 \\ \frac{1}{21} & \text{for } k = 1, 2, 3, 4 \\ \frac{6}{21} & \text{for } k = 5 \\ 0 & \text{otherwise} \end{cases}$$

17. Let $X \sim Geometric(p)$. Find $\mathrm{Var}(X)$.

Solution:

First note:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad \text{for } |x| < 1 \tag{128}$$

Taking the derivative:

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2} \quad \text{for } |x| < 1 \tag{129}$$

Taking another derivative:

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3} \quad \text{for } |x| < 1 \tag{130}$$

Now if $X \sim Geometric(p)$, then

$$P_X(k) = p(1-p)^{k-1} = pq^{k-1} \text{ for } k = 1, 2, ...$$

Thus:

$$\begin{aligned} EX &= p\sum_{k=1}^{\infty} kq^{k-1} \\ &= p\frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned} \tag{131}$$

$$\begin{aligned} E[X(X-1)] &= p\sum_{k=1}^{\infty} k(k-1)q^{k-1} \quad \text{by LOTUS} \\ &= pq\sum_{k=2}^{\infty} k(k-1)q^{k-2} = pq\frac{2}{(1-q)^3} \qquad (132) \\ &= \frac{2pq}{p^3} = \frac{2q}{p^2} \qquad (133) \end{aligned}$$

Thus:

$$\begin{aligned} EX^2 - EX &= \frac{2q}{p^2} \\ EX^2 &= \frac{2q}{p^2} + \frac{1}{p} \end{aligned} \tag{134}$$

Therefore:

$$\begin{aligned} \text{Var}(X) = EX^2 - (EX)^2 &= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2(1-p)+p-1}{p^2} = \frac{1-p}{p^2} \end{aligned} \tag{135}$$

18. Let $X \sim Pascal(m, p)$. Find $\text{Var}(X)$.

Solution:

Note that a $Pascal(m, p)$ random variable can be written as a sum of m independent $Geometric(p)$ random variables. That is, if $X_i \sim Geometric(p)$ for $i = 1, 2, ..., m$, then the random variable X, defined as $X = X_1 + X_2 + ... + X_m$ has $Pascal(m, p)$ distribution.

Thus:

$$var(X) = var(X_1) + var(X_2) + ... + var(X_m) \text{ since } x_i\text{'s are independent} \tag{136}$$

But

$$var(X_i) = \frac{(1-p)}{p^2} \tag{137}$$

Thus:

$$var(X) = \frac{m(1-p)}{p^2} \tag{138}$$

19. Suppose that $Y = -2X + 3$. If we know $EY = 1$ and $EY^2 = 9$, find EX and $\text{Var}(X)$.

Solution:

$$Y = -2X + 3 \tag{139}$$

$$EY = -2EX + 3 \quad \text{linearity of expectation} \tag{140}$$

$$1 = -2EX + 3 \quad \rightarrow \qquad EX = 1 \tag{141}$$

$$\text{Var}(Y) = 4 \times \text{Var}(X) = EY^2 - (EY)^2 = 9 - 1 = 8 \tag{142}$$

$$\rightarrow \quad \text{Var}(X) = 2 \tag{143}$$

20. There are 1000 households in a town. Specifically, there are are 100 households with one member, 200 households with 2 members, 300 households with 3 members, 200 households with 4 members, 100 households with 5 members, and 100 households with 6 members. Thus, the total number of people living in the town is

$$N = 100 \cdot 1 + 200 \cdot 2 + 300 \cdot 3 + 200 \cdot 4 + 100 \cdot 5 + 100 \cdot 6 = 3300 \tag{144}$$

   (a) We pick a household at random, and define the random variable X as the number of people in the chosen household. Find PMF and the expected value of X.

   (b) We pick a person in the town at random, and define the random variable Y as the number of people in the household where the chosen person lives. Find the PMF and the expected value of Y.

Solution:

(a)

$$\begin{aligned}
R_X &= \{1,2,3,4,5,6\} && (145)\\
P_X(1) &= \frac{100}{1000} = \frac{1}{10} \qquad P_X(2) = \frac{2}{10} \qquad P_X(3) = \frac{3}{10} && (146)\\
P_X(4) &= \frac{2}{10} \qquad P_X(5) = P_X(6) = \frac{1}{10} && (147)\\
& && (148)
\end{aligned}$$

$$\begin{aligned}
EX &= 1\times\frac{1}{10} + 2\times\frac{2}{10} + 3\times\frac{3}{10} + 4\times\frac{2}{10} + 5\times\frac{1}{10} + 6\times\frac{1}{10} && (149)\\
&= 3.3 && (150)
\end{aligned}$$

Another way to find EX is to note that

$$EX = \frac{\text{number of people}}{\text{number of households}} = \frac{3300}{1000} = 3.3 \tag{151}$$

(b)

$$\begin{aligned}
R_Y &= \{1,2,3,4,5,6\} && (152)\\
P_Y(1) &= \frac{\text{number of people in households of size 1}}{\text{number of people}} = \frac{100}{3300} = \frac{1}{33} && (153)\\
P_Y(2) &= \frac{400}{3300} = \frac{4}{33} \qquad P_Y(3) = \frac{9}{33} && (154)\\
P_Y(4) &= \frac{8}{33} \qquad P_Y(5) = \frac{5}{33} && (155)\\
P_Y(6) &= \frac{6}{33} && (156)\\
& && (157)
\end{aligned}$$

$$\begin{aligned}
EY &= 1\times\frac{1}{33} + 2\times\frac{4}{33} + 3\times\frac{9}{33} + 4\times\frac{8}{33} + 5\times\frac{5}{33} + 6\times\frac{6}{33} && (158)\\
&\approx 3.9 && (159)
\end{aligned}$$

21. (Coupon collector's problem [1]) Suppose that there are N different types of coupons. Each time you get a coupon, it is equally likely to be any of the N possible types. Let X be the number of coupons you will need to get before having observed each coupon at least once.

(a) Show that you can write $X = X_0+X_1+\cdots+X_{N-1}$, where $X_i \sim Geometric(\frac{N-i}{N})$.

(b) Find EX.

Solution:

(a) After you have already collected i distinct coupons, define X_i to be the number of additional coupons you need to collect in order to get the $i+1$'th distinct coupon. Then, we have $X_0 = 1$, since the first coupon you collect is always a new one. Then, X_1 will be a geometric random variable with success probability of $p_2 = \frac{N-1}{N}$. More generally, $X_i \sim Geometric(\frac{N-i}{N})$, for $i = 0, 1, ..., N-1$. But note that by definition write $X = X_0 + X_2+\cdots+X_{N-1}$.

(b) By linearity of expectation, we have

$$\begin{aligned} EX &= EX_0 + EX_1 + \cdots + EX_{N-1} \\ &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1} \\ &= N\left[1 + \frac{1}{2} + \cdots + \frac{1}{N-1} + \frac{1}{N}\right] \end{aligned}$$

22. (St. Petersburg paradox) Here is a famous problem called the St. Petersburg paradox. Wikipedia states the problem as follows [2]: "A casino offers a game of chance for a single player in which a fair coin is tossed at each stage. The pot starts at 1 dollar and is doubled every time a head appears. The first time a tail appears, the game ends and the player wins whatever is in the pot. Thus the player wins 1 dollar if a tail appears on the first toss, 2 dollars if a head appears on the first toss and a tail on the second, 4 dollars if a head appears on the first two tosses and a tail on the third, 8 dollars if a head appears on the first three tosses and a tail on the fourth, and so on. In short, the player wins 2^{k-1} dollars if the coin is tossed k times until the first tail appears. What would be a fair price to pay the casino for entering the game?".

(a) Let X be the amount of money (in dollars) that the player wins. Find EX.

(b) What is the probability that the player wins more than 65 dollars?

(c) Now suppose that the casino only has a finite amount of money. Specifically, suppose that the maximum amount of the money that the casino will pay you is 2^{30} dollars (around 1.07 billion dollars). That is if you win more than 2^{30} dollars, the casino is going to pay you only 2^{30} dollars. Let Y be the money that the player wins in this case. Find EY.

Solution:

(a)

$$R_X = \{1, 2, 4, 8, \cdots\} \tag{160}$$

$$P_X(1) = \frac{1}{2} \qquad\qquad P_X(2) = \frac{1}{4} \qquad\qquad \cdots \tag{161}$$

$$\tag{162}$$

In general, $P_X(2^{k-1}) = \frac{1}{2^k}$ for $k = 1, 2, \cdots$

$$EX = 1\frac{1}{2} + 2\frac{1}{4} + 4\frac{1}{8} + \cdots \tag{163}$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots = \infty \tag{164}$$

(b)

$$P(X > 65) = P_X(128) + P_X(256) + \cdots \tag{165}$$

$$= \frac{1}{2^8} + \frac{1}{2^9} + \frac{1}{2^{10}} + \cdots \tag{166}$$

$$= \frac{1}{2^8}(1 + \frac{1}{2} + \cdots) = \frac{1}{2^8} \cdot 2 = \frac{1}{2^7} = \frac{1}{128} \tag{167}$$

c)

$$P_Y(2^{k-1}) = \begin{cases} \frac{1}{2^k} & \text{for } k = 1, 2, \cdots, 30 \\ \frac{1}{2^{31}} + \frac{1}{2^{32}} + \cdots = \frac{1}{2^{30}} & \text{for } k = 31 \\ 0 & \text{otherwise} \end{cases}$$

$$EY = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + \cdots + 2^{29} \times \frac{1}{2^{30}} + 2^{30} \times \frac{1}{2^{30}} = 16 \tag{168}$$

23. Let X be a random variable with mean $EX = \mu$. Define the function $f(\alpha)$ as

$$f(\alpha) = E[(X - \alpha)^2].$$

Find the value of α that minimizes f.

Solution:

$$
\begin{aligned}
f(\alpha) &= E(X^2 - 2\alpha X + \alpha^2) && (169)\\
&= EX^2 - 2\alpha EX + \alpha^2 && (170)
\end{aligned}
$$

Thus:

$$
\begin{aligned}
f(\alpha) &= \alpha^2 - 2(EX)\alpha + EX^2 && (171)\\
& && (172)
\end{aligned}
$$

$f(\alpha)$ is a polynomial of degree 2 with positive coefficient for α^2

$$
\begin{aligned}
\frac{\partial f(\alpha)}{\partial \alpha} = 0 \quad &\rightarrow \quad 2\alpha - 2EX = 0 && (173)\\
&\rightarrow \quad \alpha = EX && (174)
\end{aligned}
$$

24. You are offered to play the following game. You roll a fair die once and observe the result which is shown by the random variable X. At this point, you can stop the game and win X dollars. You can also choose to roll the die for the second time to observe the value Y. In this case, you will win Y dollars. Let W be the value that you win in this game. What strategy do you use to maximize EW? What is the maximum EW you can achieve using your strategy?

    Solution:

    If you roll the die for the second time, then your expected winning is:

    $$EY = 3.5 \qquad (175)$$

    Thus, here is the strategy:

    If $X > 3.5$ stop and take X dollars.
    If $X < 3.5$ roll the die for the second time and take Y dollars.

    $$
    \begin{aligned}
    P_W(1) &= P(X < 3.5 \text{ and } Y = 1) && (176)\\
    &= P(X < 3.5)P(Y = 1) \quad \text{since X and Y are independent} && (177)\\
    &= \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} && (178)
    \end{aligned}
    $$

Similarly, $P_W(2) = P_W(3) = \frac{1}{12}$
$W = 4$ if either ($X = 4$) or ($X < 3.5$ and $Y = 4$), thus:

$$P_W(4) = P_X(4) + P(X < 3.5)P(Y = 4) \tag{179}$$
$$= \frac{1}{6} + \frac{1}{2}.\frac{1}{6} = \frac{1}{6} + \frac{1}{12} = \frac{3}{12} \tag{180}$$

Similarly, $P_W(5) = P_W(6) = \frac{3}{12}$

$$EW = 1 \times \frac{1}{12} + 2 \times \frac{1}{12} + 3 \times \frac{1}{12} + 4 \times \frac{3}{12} + 5 \times \frac{3}{12} + 6 \times \frac{3}{12} \tag{181}$$
$$= \frac{17}{4} \tag{182}$$

25. The **median** of a random variable X is defined as any number m that satisfies both of the following conditions:

$$P(X \geq m) \geq \frac{1}{2} \qquad \text{and} \qquad P(X \leq m) \geq \frac{1}{2}.$$

Note that the median of X is not necessarily unique. Find the median of X if

(a) The PMF of X is given by

$$P_X(k) = \begin{cases} 0.4 & \text{for } k = 1 \\ 0.3 & \text{for } k = 2 \\ 0.3 & \text{for } k = 3 \\ 0 & \text{otherwise} \end{cases}$$

(b) X is the result of a rolling of a fair die.

(c) $X \sim \mathit{Geometric}(p)$, where $0 < p < 1$.

Solution: (a) $m = 2$, since

$$P(X \geq 2) = 0.6 \text{ and } P(X \leq 2) = 0.7 \tag{183}$$

(b)

$$P_X(k) = \frac{1}{6} \text{ for } k = 1,2,3,4,5,6 \tag{184}$$
$$\rightarrow 3 \leq m \leq 4 \tag{185}$$

Thus, we conclude $3 \leq m \leq 4$. Any value $\in [3,4]$ is a median for X.
(c)

$$P_X(k) = (1-p)^{k-1}p = q^{k-1}p \quad \text{ where } q = 1-p \tag{186}$$
$$P(X \leq m) = \sum_{k=1}^{\lfloor m \rfloor} q^{k-1}p = p(1+q+\cdots q^{m-1}) \tag{187}$$
$$= p\frac{1-q^{\lfloor m \rfloor}}{1-q} = 1 - q^{\lfloor m \rfloor} \tag{188}$$

We need $1 - q^{\lfloor m \rfloor} \geq \frac{1}{2}$.
Therefore:

$$q^{\lfloor m \rfloor} \leq \frac{1}{2} \qquad \rightarrow \lfloor m \rfloor log_2(q) \leq -1 \qquad \rightarrow \lfloor m \rfloor log_2 \frac{1}{q} \geq 1 \tag{189}$$
$$\rightarrow \lfloor m \rfloor \geq \frac{1}{log_2 \frac{1}{q}} \tag{190}$$

Also

$$P(X \geq m) = \sum_{k=\lceil m \rceil}^{\infty} q^{k-1}p = pq^{\lceil m \rceil - 1}(1+q+\cdots) \tag{191}$$
$$= p\frac{q^{\lceil m \rceil} - 1}{1-q} = q^{\lceil m \rceil - 1} \tag{192}$$

Thus:

$$q^{\lceil m \rceil - 1} \geq \frac{1}{2} \quad \rightarrow (\lceil m \rceil - 1)log_2 q \geq -1 \tag{193}$$
$$\rightarrow (\lceil m \rceil - 1)log_2(\frac{1}{q}) \leq 1 \quad \rightarrow \lceil m \rceil - 1 \leq \frac{1}{log_2(\frac{1}{q})} \tag{194}$$
$$\rightarrow \lceil m \rceil \leq \frac{1}{log_2(\frac{1}{q})} + 1 \tag{195}$$
$$\tag{196}$$

Thus any m satisfying:

$$\lfloor m \rfloor \geq \frac{1}{log_2 \frac{1}{q}} \text{ and } \lceil m \rceil \leq \frac{1}{log_2(\frac{1}{q})} + 1 \tag{197}$$

$$\tag{198}$$

is a median for X. For example if $p = \frac{1}{5}$ then $\lfloor m \rfloor \geq 3.1$ and $\lceil m \rceil \leq 4.1$. So $m = 4$.

References

[1] `http://en.wikipedia.org/wiki/Coupon_collector's_problem`

[2] `http://en.wikipedia.org/wiki/St._Petersburg_paradox`