# Probability and Statistics: MA6.101

# Homework 3

Topics Covered: Discrete Random Variables

Q1: Let $X$ be a discrete random variable that can take the values $\{-2, -1, 1, 2\}$. The probability mass function (PMF) of $X$ is given by $p_X(X = x) = c(x + 4)$ for some constant $c$.

(a) Find the value of the constant $c$.

(b) Find the expected value of $X$, $E[X]$.

(c) Find the variance of $X$, $Var(X)$.

**Solution:**

(a) For any PMF, the sum of probabilities over all possible values of the random variable must equal 1. Therefore, we must have

$$\sum_x p_X(X = x) = 1$$

. By substituting the given PMF, we get:

$$c(-2 + 4) + c(-1 + 4) + c(1 + 4) + c(2 + 4) = 1$$
$$c(2) + c(3) + c(5) + c(6) = 1$$
$$16c = 1 \implies c = \tfrac{1}{16}$$

Thus, the PMF is

$$p_X(X = x) = \frac{x + 4}{16}$$

.

(b) The expected value $E[X]$ is calculated as the sum of each value multiplied by its probability,

$$E[X] = \sum_x x \cdot p_X(X = x)$$

.

$$\begin{aligned} E[X] &= (-2)\left(\frac{2}{16}\right) + (-1)\left(\frac{3}{16}\right) + (1)\left(\frac{5}{16}\right) + (2)\left(\frac{6}{16}\right) \\ &= \frac{-4 - 3 + 5 + 12}{16} = \frac{10}{16} = \frac{5}{8} \end{aligned}$$

(c) The variance $Var(X)$ is given by the formula $Var(X) = E[X^2] - (E[X])^2$. We first need to calculate $E[X^2]$.

$$E[X^2] = \sum_x x^2 \cdot p_X(X = x)$$

$$E[X^2] = (-2)^2\left(\frac{2}{16}\right) + (-1)^2\left(\frac{3}{16}\right) + (1)^2\left(\frac{5}{16}\right) + (2)^2\left(\frac{6}{16}\right)$$
$$= \frac{4(2)+1(3)+1(5)+4(6)}{16} = \frac{8+3+5+24}{16} = \frac{40}{16} = \frac{5}{2}$$

Now, we can compute the variance:

$$Var(X) = E[X^2] - (E[X])^2 = \frac{5}{2} - \left(\frac{5}{8}\right)^2 = \frac{5}{2} - \frac{25}{64} = \frac{160-25}{64} = \frac{135}{64}$$

Q2: A tech company manufactures microchips in batches of 20. In each batch, 5 chips come from a premium production line and are guaranteed to be flawless. The other 15 chips come from a standard production line, where each chip has a 10% chance of being defective, independent of the others. Let $Z$ be the total number of flawless chips in a randomly selected batch.

(a) Find the PMF of $Z$.

(b) A batch is considered "high-quality" if it contains at least 18 flawless chips. What is the probability of this?

**Solution:**

(a) The total number of flawless chips, $Z$, is the sum of the 5 guaranteed flawless chips and the number of flawless chips from the standard line. Let $W$ be the number of flawless chips from the 15 on the standard line. Then, $Z = 5 + W$.

For the standard line, the probability of a chip being defective is 0.1, so the probability of it being flawless is $p = 1 - 0.1 = 0.9$. Since each of the 15 chips is an independent trial, $W$ follows a binomial distribution, $W \sim \text{Binomial}(n = 15, p = 0.9)$. The PMF of $W$ is:

$$p_W(W = w) = \binom{15}{w}(0.9)^w(0.1)^{15-w} \quad \text{for } w = 0, 1, \ldots, 15$$

The possible values for the total flawless chips $Z$ are from 5 (if $W = 0$) to 20 (if $W = 15$). The PMF of $Z$ for any value $k$ in this range $\{5, \ldots, 20\}$ is:

$$p_Z(Z = k) = \mathbb{P}(Z = k) = \mathbb{P}(5 + W = k) = \mathbb{P}(W = k - 5)$$

Substituting $w = k - 5$ into the PMF of $W$, we get the PMF of $Z$:

$$p_Z(Z = k) = \binom{15}{k-5}(0.9)^{k-5}(0.1)^{15-(k-5)} = \binom{15}{k-5}(0.9)^{k-5}(0.1)^{20-k}$$

This formula is valid for $k \in \{5, 6, \ldots, 20\}$.

(b) A batch is "high-quality" if $Z \geq 18$. We express this in terms of $W$:

$$\mathbb{P}(Z \geq 18) = \mathbb{P}(5 + W \geq 18) = \mathbb{P}(W \geq 13)$$

This means we need to sum the probabilities for $W = 13, 14, 15$.

$$\mathbb{P}(W \geq 13) = \mathbb{P}(W = 13) + \mathbb{P}(W = 14) + \mathbb{P}(W = 15)$$

Calculating the terms:

$$\begin{aligned}
\mathbb{P}(W=13) &= \binom{15}{13}(0.9)^{13}(0.1)^2 = 105 \cdot (0.25418) \cdot (0.01) \approx 0.2669 \\
\mathbb{P}(W=14) &= \binom{15}{14}(0.9)^{14}(0.1)^1 = 15 \cdot (0.22877) \cdot (0.1) \approx 0.3432 \\
\mathbb{P}(W=15) &= \binom{15}{15}(0.9)^{15}(0.1)^0 = 1 \cdot (0.20589) \cdot (1) \approx 0.2059
\end{aligned}$$

Summing these gives the final probability:

$$\mathbb{P}(Z \geq 18) \approx 0.2669 + 0.3432 + 0.2059 = 0.816$$

So there is an 81.6% chance that a batch is high-quality.

Q3: You have come to attend the Probability and Statistics Lecture on Saturday. It is raining outside so you have kept your umbrellas outside H105. Suppose that the strength of the class is $N$. After the tutorial has ended, you have to find your umbrella, but you are running late for Kadamba Biryani, so you pick one umbrella randomly from the pile. Similarly everyone else picks an umbrella randomly independent from each other.

(a) What is the probability that at least one of you receives his/her own umbrella?

(b) Let $X_N$ be the number of people who receive their own umbrella. What is

$$lim_{n\to\infty}\mathbb{P}(X_N \geq 1)$$

(c) Find the PMF of $X_N$.

**A:**

(a) Let $A_i$ be the event that $i$-th person receives his/her own hat. Then we are interested in finding $\mathbb{P}(E_N)$, where $E_i = A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_i$. To find $\mathbb{P}(E_N)$, we use the inclusion-exclusion principle. We have

$$\begin{aligned}
\mathbb{P}(E_N) &= \mathbb{P}\left(\bigcup_{i=1}^{N} A_i\right) \\
&= \sum_{i=1}^{N} \mathbb{P}(A_i) - \sum_{i,j:i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i,j,k:i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \\
&\quad + (-1)^{N-1}\mathbb{P}\left(\bigcap_{i=1}^{N} A_i\right).
\end{aligned}$$

Note that there is complete symmetry here, that is, we can write

$$\begin{aligned}
\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \ldots = \mathbb{P}(A_N); \\
\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \ldots = \mathbb{P}(A_2 \cap A_4) = \ldots; \\
\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_2 \cap A_4) = \ldots = \mathbb{P}(A_2 \cap A_4 \cap A_5) = \ldots; \\
\vdots
\end{aligned}$$

Thus, we have

$$\sum_{i=1}^{N} \mathbb{P}(A_i) = N\mathbb{P}(A_1);$$
$$\sum_{i,j:i<j} \mathbb{P}(A_i \cap A_j) = \binom{N}{2}\mathbb{P}(A_1 \cap A_2);$$
$$\sum_{i,j,k:i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) = \binom{N}{3}\mathbb{P}(A_1 \cap A_2 \cap A_3);$$
$$\vdots$$

Therefore, we have

$$\mathbb{P}(E_N) = N\mathbb{P}(A_1) - \binom{N}{2}\mathbb{P}(A_1 \cap A_2) + \binom{N}{3}\mathbb{P}(A_1 \cap A_2 \cap A_3) - \ldots + (-1)^{N-1}\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_N)$$

Now, we only need to find $\mathbb{P}(A_1)$, $\mathbb{P}(A_1 \cap A_2)$, $\mathbb{P}(A_1 \cap A_2 \cap A_3)$, etc. to finish solving the problem. To find $\mathbb{P}(A_1)$, we have

$$\mathbb{P}(A_1) = \frac{|A_1|}{|S|}.$$

Here, the sample space $S$ consists of all possible permutations of $N$ objects (umbrellas). Thus, we have

$$|S| = N!$$

On the other hand, $A_1$ consists of all possible permutations of $N-1$ objects (because the first object is fixed). Thus

$$|A_1| = (N-1)!$$

Therefore, we have

$$\mathbb{P}(A_1) = \frac{|A_1|}{|S|} = \frac{(N-1)!}{N!} = \frac{1}{N}$$

Similarly, we have

$$|A_1 \cap A_2| = (N-2)!$$

Thus,

$$\mathbb{P}(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|S|} = \frac{(N-2)!}{N!} = \frac{1}{N(N-1)}.$$

Similarly,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{|A_1 \cap A_2 \cap A_3|}{|S|} = \frac{(N-3)!}{N!} = \frac{1}{N(N-1)(N-2)};$$
$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{|A_1 \cap A_2 \cap A_3 \cap A_4|}{|S|} = \frac{(N-4)!}{N!} = \frac{1}{N(N-1)(N-2)(N-3)};$$
$$\vdots$$

Thus, we have

$$\mathbb{P}(E_N) = N \cdot \frac{1}{N} - \binom{N}{2} \cdot \frac{1}{N(N-1)} + \binom{N}{3} \cdot \frac{1}{N(N-1)(N-2)} - \ldots + (-1)^{N-1}\frac{1}{N!} \tag{2.6}$$

By simplifying a little bit, we obtain

$$\mathbb{P}(E_N) = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{N-1}\frac{1}{N!}.$$

(b) It is interesting to note what happens when $N$ becomes large. To see that, we should remember the Taylor series expansion of $e^x$. In particular,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Letting $x = -1$, we have

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots$$

Thus, we conclude that as $N$ becomes large, $\mathbb{P}(X_N \geq 1)$ approaches $1 - \frac{1}{e}$.

(c) Let $X_N$ be the number of people who receive their own umbrellas when $N$ umbrellas are randomly distributed to $N$ people. We want to find the probability mass function (PMF) of $X_N$, which is $p_{X_N}(k)$ for $k \in \{0, 1, \ldots, N\}$.
The total number of ways to distribute the umbrellas is $N!$, as each person can receive one of the $N$ umbrellas in a unique permutation. To find the number of ways that exactly $k$ people receive their own umbrellas, we can use the principle of inclusion-exclusion. We can break down the problem into two parts:

i. **Choose the $k$ people** who receive their own umbrellas. This is a combination problem, and there are $\binom{N}{k}$ ways to choose these people.

ii. **Arrange the remaining $N-k$ umbrellas** for the remaining $N-k$ people such that none of them receive their own hat. This is a derangement problem. The number of derangements of $m$ items, denoted by $D_m$, is given by the formula:

$$D_m = m! \sum_{j=0}^{m} \frac{(-1)^j}{j!}$$

For our case, $m = N - k$.

Therefore, the number of ways to have exactly $k$ matches is $\binom{N}{k} \cdot D_{N-k}$.

The PMF of $X_N$ is the ratio of this number to the total number of permutations:

$$\begin{aligned} p_{X_N}(k) &= \frac{\binom{N}{k} \cdot D_{N-k}}{N!} \\ &= \frac{\frac{N!}{k!(N-k)!} \cdot D_{N-k}}{N!} \\ &= \frac{1}{k!(N-k)!} \cdot D_{N-k} \\ &= \frac{1}{k!} \cdot \frac{D_{N-k}}{(N-k)!} \end{aligned}$$

Using the formula for derangements, the term $\frac{D_{N-k}}{(N-k)!}$ can be expressed as:

$$\frac{D_{N-k}}{(N-k)!} = \sum_{j=0}^{N-k} \frac{(-1)^j}{j!}$$

Substituting this back into the expression for $p_{X_N}(k)$, we get the final form of the PMF:

$$p_{X_N}(k) = \frac{1}{k!} \sum_{j=0}^{N-k} \frac{(-1)^j}{j!}$$

This formula holds for $k = 0, 1, \ldots, N$. The probability is zero for any other value of $k$.

Q4: A bag contains 8 red balls and 7 blue balls (all identical within color). You draw 4 balls without replacement. Define the random variable

$$X = \begin{cases} \text{number of ways to arrange the drawn balls in a row such that} \\ \quad \text{no two balls of the same color are adjacent.} \end{cases}$$

(a) Find the probability distribution of $X$, and $E[X]$

(b) For general $n$ red balls and $m$ blue balls, where you draw $k$ balls without replacement, provide a general formula for $E[X]$

[Mrudani]

**A:**

(a) Observe that for any combination of red and blue balls (such that the total number of balls drawn is 4), two balls of the same color will always be adjacent in at least one position for any case, except when you draw 2 red balls and 2 blue balls and arrange them in a certain way, that being:

$$RBRB \quad \text{or} \quad BRBR$$

All other arrangements are invalid Let $R$ be the number of red balls drawn from the bag. The random variable $X$ then takes the values

$$X = \begin{cases} 2 & R = 2, \\ 0 & \text{otherwise} \end{cases}$$

Also:

$$P(R = r) = \frac{\binom{8}{r} \cdot \binom{7}{4-r}}{\binom{15}{r}} \quad \text{for } r = 1, 2 \ldots 4$$

$$P(X = 2) = P(R = 2) = \frac{\binom{8}{2} \cdot \binom{7}{2}}{\binom{15}{4}} = \frac{28 \cdot 21}{1365} = \frac{28}{65}$$

$$P(X = 0) = 1 - P(X = 2) = \frac{37}{65}$$

Hence,

$$E[X] = 2 \cdot P(X = 2) + 0 \cdot P(X = 0) = 2 \cdot \frac{28}{65} = \frac{56}{65}$$

(b) For $n$ red balls, $m$ blue balls and a draw of $k$ balls, we get the following cases: Let the number of drawn red balls be $r$ and blue balls be $b$ (such that $r + b = k$)

i. if r = b then there are two possible arrangements. This case is possible only if $k$ is even.

ii. if $|r - b| = 1$ then an arrangement is possible when the color with more balls is at the end (eg. RBRBR). Note that this was not possible in the case where $k = 4$ since $k$ is even. Hence this case in general will only be possible if k is odd.

iii. Otherwise there are 0 possible arrangements.

**Case 1: $k$ is even**

$$P\left(R = \frac{k}{2}\right) = \frac{\binom{n}{\frac{k}{2}} \cdot \binom{m}{\frac{k}{2}}}{\binom{m+n}{k}}$$

Then:

$$E[X] = 2 \cdot \frac{\binom{n}{\frac{k}{2}} \cdot \binom{m}{\frac{k}{2}}}{\binom{m+n}{k}}$$

**Case 2: $k$ is odd**

$$X = \begin{cases} 1 & R = \frac{k+1}{2}, \\ 1 & B = \frac{k+1}{2} \quad \text{(or equivalently } R = \frac{k-1}{2}\text{)}, \\ 0 & \text{otherwise} \end{cases}$$

$$P\left(R = \frac{k+1}{2}\right) = \frac{\binom{n}{\frac{k+1}{2}} \cdot \binom{m}{\frac{k-1}{2}}}{\binom{m+n}{k}}$$

$$P\left(R = \frac{k-1}{2}\right) = \frac{\binom{n}{\frac{k-1}{2}} \cdot \binom{m}{\frac{k+1}{2}}}{\binom{m+n}{k}}$$

$$\begin{aligned} E[X] &= 1 \cdot P\left(R = \frac{k+1}{2}\right) + 1 \cdot P\left(R = \frac{k-1}{2}\right) \\ &= \frac{\binom{n}{\frac{k+1}{2}} \cdot \binom{m}{\frac{k-1}{2}} + \binom{n}{\frac{k-1}{2}} \cdot \binom{m}{\frac{k+1}{2}}}{\binom{m+n}{k}} \end{aligned}$$

Q5: Let $X \sim$ Geometric($\frac{1}{4}$), and let $Y = |X - 4|$. Find the range and PMF of $Y$. [Prabhas]

**A:**
We are given a random variable $X$ that follows a Geometric distribution with a success probability of $p = \frac{1}{4}$. range of X is $\{1, 2, 3, \dots\}$. The Probability Mass Function (PMF) of $X$ is:

$$P(X = k) = (1-p)^{k-1}p = \left(\frac{3}{4}\right)^{k-1}\left(\frac{1}{4}\right) \quad \text{for } k = 1, 2, 3, \dots$$

A new random variable $Y$ is defined as $Y = |X - 4|$.
We need to find $p_Y(y) = P(Y = y)$ for each value $y$ in the range of $Y$. We analyze this in cases.
**Case 1:** $y = 0$
The event $Y = 0$ occurs if and only if $|X - 4| = 0$, which means $X = 4$.

$$p_Y(0) = P(Y = 0) = P(X = 4) = \left(\frac{3}{4}\right)^{4-1}\left(\frac{1}{4}\right) = \left(\frac{3}{4}\right)^{3}\left(\frac{1}{4}\right) = \frac{27}{64}\cdot\frac{1}{4} = \frac{27}{256}$$

**Case 2:** $y > 0$
For any integer $y > 0$, the event $Y = y$ occurs if $|X - 4| = y$. This implies two possibilities:

$$X - 4 = y \quad \text{or} \quad X - 4 = -y$$
$$X = 4 + y \quad \text{or} \quad X = 4 - y$$

Since these two outcomes for $X$ are mutually exclusive, we can add their probabilities:

$$p_Y(y) = P(X = 4 + y) + P(X = 4 - y)$$

However, we must ensure that these values of $X$ are in its range $\{1, 2, 3, \dots\}$.

- The value $4 + y$ is always greater than or equal to 5 (since $y \geq 1$), so it is always in the range of $X$.
- The value $4 - y$ is in the range of $X$ only if $4 - y \geq 1$, which means $y \leq 3$.

This requires us to split Case 2 into two sub-cases.
**Sub-case 2a:** $y \in \{1, 2, 3\}$
In this case, both $4 + y$ and $4 - y$ are valid outcomes for $X$.

$$\begin{aligned} p_Y(y) &= P(X = 4 + y) + P(X = 4 - y) \\ &= \left(\frac{3}{4}\right)^{(4+y)-1}\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right)^{(4-y)-1}\left(\frac{1}{4}\right) \\ &= \frac{1}{4}\left[\left(\frac{3}{4}\right)^{y+3} + \left(\frac{3}{4}\right)^{3-y}\right] \end{aligned}$$

**Sub-case 2b:** $y > 3$
In this case, $4-y$ is less than 1, so it is not a valid outcome for $X$. Thus, $P(X = 4 - y) = 0$.

$$\begin{aligned} p_Y(y) &= P(X = 4+y) + P(X = 4-y) \\ &= \left(\frac{3}{4}\right)^{(4+y)-1}\left(\frac{1}{4}\right) + 0 \\ &= \frac{1}{4}\left(\frac{3}{4}\right)^{y+3} \end{aligned}$$

Combining all cases, the PMF of $Y$ is:

$$p_Y(y) = \begin{cases} \frac{27}{256} & \text{for } y = 0, \\ \frac{1}{4}\left[\left(\frac{3}{4}\right)^{y+3} + \left(\frac{3}{4}\right)^{3-y}\right] & \text{for } y \in \{1, 2, 3\}, \\ \frac{1}{4}\left(\frac{3}{4}\right)^{y+3} & \text{for } y \in \{4, 5, 6, \dots\}, \\ 0 & \text{otherwise.} \end{cases}$$

Q6: Refer to the median of a random varibale definition in tutorial. 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) $X \sim \text{Geometric}(p)$, where $0 < p < 1$.

**A:**

$$\mathbb{P}(X \geq m) \geq \frac{1}{2} \quad \mathbb{P}(X \leq m) \geq \frac{1}{2}$$

$$\mathbb{P}(X > m) + p_X(X = m) \geq \frac{1}{2} \quad F_X(m) \geq \frac{1}{2}$$

$$1 - F_X(m) + p_X(X = m) \geq \frac{1}{2}$$

$$\boxed{\frac{1}{2} \leq F_X(m) \leq \frac{1}{2} + p_X(X = m)}$$

(a)

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 0.4 & 1 \leq x < 2 \\ 0.7 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

**Case 1:** $m$ is not a discrete number:

$$p_X(X = m) = 0 \quad \text{(since } X \text{ is a discrete random variable)}$$

$$F_X(m) = \frac{1}{2} \quad \text{for} \quad m \notin \mathbb{Z}^+$$

**Case 2:** $m$ is discrete:

$$p_X(X = m) \neq 0$$

$$\frac{1}{2} \leq F_X(m) \leq \frac{1}{2} + p_X(X = m)$$

For $m \geq 2$:

$$m \geq 2 \quad \text{will always satisfy condition (2), but}$$

$$m > 3 \quad \text{will not satisfy it}$$

$$\implies m \text{ might lie in } [2, 3]$$

For non-integer values of $m$ in $[2, 3]$:

$$F_X(m) = \frac{1}{2}, \quad \text{which is not true.}$$

$\therefore$ We only consider discrete points where PMF is non-zero and satisfies (1) and (2).

$$\boxed{m = 2}$$

(b) Let $X \sim \text{Geometric}(p)$.

$$p_X(X = k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots$$

$$F_X(x = k) = 1 - (1-p)^k$$

For $F_X(m) \geq \frac{1}{2}$:

$$1 - (1-p)^m \geq \frac{1}{2}$$

$$(1-p)^m \leq \frac{1}{2}$$

$$m \ln(1-p) \leq \ln \frac{1}{2}$$

$$m \geq \frac{-\ln 2}{\ln(1-p)}$$

For $F_X(m) \leq \frac{1}{2} + p_X(m)$:

$$1 - (1-p)^m \leq \frac{1}{2} + p(1-p)^{m-1}$$

$$\frac{1}{2} \leq p + (1-p)^{m-1}$$

$$(1-p)^{m-1} \geq \frac{1}{2}$$

$$(m-1)\ln(1-p) \leq \ln\frac{1}{2}$$

$$m \leq 1 + \frac{-\ln 2}{\ln(1-p)}$$

For $p = \frac{1}{2}$, there will be two discrete points that satisfy this:

$$m = \{1, 2\}$$

For the interval $[1, 2)$:

$$F_X(m) = p_X(1) = (1-p)\frac{1}{2} = \frac{1}{2}$$

$$\implies m \in [1, 2) \cup \{2\}$$

$$\boxed{m \in [1, 2]}$$

For $p \neq \frac{1}{2}$, there will be only one discrete point that satisfies this condition.
The answer might not entirely be the range, but only discrete points

Q7: Let $s > 0$ be a real parameter and let $X$ be a non-negative integer valued random variable whose cumulative distribution function is given by

$$F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < 0, \\ 1 - (\lfloor x \rfloor + 2)^{-s}, & x \geq 0. \end{cases}$$

Equivalently, for integer $n \geq 0$,

$$F(n) = 1 - (n+2)^{-s}.$$

(a) Show that the PMF of $X$ is

$$\mathbb{P}(X = n) = (n+1)^{-s} - (n+2)^{-s}, \qquad n = 0, 1, 2, \ldots$$

and verify that the probabilities sum to 1.

(b) (i) Show that the tail probabilities satisfy

$$\mathbb{P}(X \geq k) = (k+1)^{-s}, \qquad k = 0, 1, 2, \ldots$$

(ii) Deduce the related strict-tail formula

$$\mathbb{P}(X > k) = (k+2)^{-s}, \qquad k = 0, 1, 2, \ldots$$

(c) Prove that for any nonnegative integer-valued random variable $X$,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k),$$

i.e. the expectation equals the sum of its tail probabilities.

[Solution]

**(a)** For integer $n \geq 0$,

$\mathbb{P}(X = n) = F(n) - F(n^-) = \big(1-(n+2)^{-s}\big) - \big(1-(n+1)^{-s}\big) = (n+1)^{-s} - (n+2)^{-s}.$

Thus the PMF is as stated. Summing the probabilities:

$$\sum_{n=0}^{\infty} \big[(n+1)^{-s} - (n+2)^{-s}\big] = 1^{-s} - \lim_{N\to\infty} (N+2)^{-s} = 1,$$

so the distribution is normalized.

**(b)(i)** For integer $k \geq 0$,

$$\mathbb{P}(X \geq k) = 1 - \mathbb{P}(X \leq k-1) = 1 - F(k-1) = (k+1)^{-s}.$$

(For $k = 0$ this gives $\mathbb{P}(X \geq 0) = 1$ as required.)

**(b)(ii)** Using the previous formula,

$$\mathbb{P}(X > k) = 1 - \mathbb{P}(X \leq k) = 1 - F(k) = (k+2)^{-s},$$

valid for $k \geq 0$.

**(c) Tail-sum identity — detailed indicator proof.**
*Step 1: Express $\mathbb{P}(X > k)$ in terms of $\mathbb{P}(X = n)$.* We begin by noting that the tail probability $\mathbb{P}(X > k)$ can be expressed as the sum of the probabilities that $X$ takes values greater than $k$. Specifically, for $k \geq 0$, we have:

$$\mathbb{P}(X > 0) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \dots$$

$$\mathbb{P}(X > 1) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \dots$$

and similarly for higher values of $k$:

$$\mathbb{P}(X > k) = \mathbb{P}(X = k+1) + \mathbb{P}(X = k+2) + \mathbb{P}(X = k+3) + \dots$$

*Step 2: Express the expectation as a sum.* The expectation $\mathbb{E}(X)$ of a non-negative integer-valued random variable $X$ is given by:

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X = n),$$

which is the standard definition of the expectation for a discrete random variable.
*Step 3: Relate the sum of tail probabilities to the expectation.* We now sum the tail probabilities $\mathbb{P}(X > k)$ over all $k \geq 0$:

$$\sum_{k=0}^{\infty} \mathbb{P}(X > k) = \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) + 3 \cdot \mathbb{P}(X = 3) + \dots$$

This expression is exactly the formula for the expectation of $X$, since:

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X = n).$$

Thus, we have shown that:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

*Conclusion.* This completes the proof that for any nonnegative integer-valued random variable $X$,

$$\boxed{\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)}.$$

Q8: Let $X$ be a random variable with mean $\mathrm{E}[X] = \mu$. Define the function $f(\alpha)$ as

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

Find the value of $\alpha$ that minimizes $f$.

**A:**

$$f(\alpha) = \mathrm{E}[(X - \alpha)^2] = \mathrm{E}[X^2 + \alpha^2 - 2X\alpha]$$
$$\implies f(\alpha) = \mathrm{E}[X^2] + \mathrm{E}[\alpha^2] + \mathrm{E}[-2X\alpha] \quad \text{(Linearity of Expectation)}$$
$$\implies f(\alpha) = \mathrm{E}[X^2] + \alpha^2 - 2\alpha\mathrm{E}[X]$$
$$\text{To minimize } f(\alpha), \quad f'(\alpha) = 0 \implies 2\alpha - 2\mathrm{E}[X] = 0$$
$$\implies \alpha = \mathrm{E}[X] = \mu$$

Q9: You roll m independent fair n-sided dice.

(a) What is the expected value of the minimum of the m rolls?

(b) What will be the approximated value for large n for the expectation calculated in the question of tutorial?

(c) What is the value for $m = 3$ and $n = 100$.

**A:**

(a) Let $X_1, X_2, \ldots, X_m$ be the outcomes of $m$ independent fair $n$-sided dice rolls. The possible outcomes for each die are the integers from 1 to $n$, and each outcome has a probability of $\frac{1}{n}$.
Let $Y$ be the random variable representing the minimum of the $m$ rolls, i.e.,

$$Y = \min(X_1, X_2, \ldots, X_m)$$

The expected value of a discrete random variable can be calculated using the formula:

$$\mathbb{E}[Y] = \sum_k k \cdot p_Y(Y = k)$$

However, a more elegant approach for non-negative integer-valued random variables is using the tail sum formula:

$$\mathbb{E}[Y] = \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k)$$

In our case, since the maximum roll is $n$, the summation goes up to $n$:

$$\mathbb{E}[Y] = \sum_{k=1}^{n} \mathbb{P}(Y \geq k)$$

The event $Y \geq k$ means that the minimum of the $m$ rolls is at least $k$. This is equivalent to all $m$ rolls being at least $k$:

$$\mathbb{P}(Y \geq k) = \mathbb{P}(\min(X_1, X_2, \ldots, X_m) \geq k)$$
$$\mathbb{P}(Y \geq k) = \mathbb{P}(X_1 \geq k \text{ and } X_2 \geq k \text{ and } \ldots \text{ and } X_m \geq k)$$

Since the dice rolls are independent, we can multiply the individual probabilities:

$$\mathbb{P}(Y \geq k) = \mathbb{P}(X_1 \geq k) \cdot \mathbb{P}(X_2 \geq k) \cdots \mathbb{P}(X_m \geq k)$$

For a single $n$-sided die, the outcomes greater than or equal to $k$ are $\{k, k+1, \ldots, n\}$. The number of such outcomes is $n - k + 1$. The total number of outcomes is $n$.

$$\mathbb{P}(X_i \geq k) = \frac{n-k+1}{n} = \frac{n+1-k}{n}$$

Now, substitute this back into the equation for $\mathbb{P}(Y \geq k)$:

$$\mathbb{P}(Y \geq k) = \left(\frac{n+1-k}{n}\right)^m$$

Now we can compute the expected value using the tail sum formula:

$$\mathbb{E}[Y] = \sum_{k=1}^{n} \left(\frac{n+1-k}{n}\right)^m$$

To simplify the summation, let's change the index of summation. Let $j = n + 1 - k$.
When $k = 1$, $j = n + 1 - 1 = n$. When $k = n$, $j = n + 1 - n = 1$.
The sum becomes:

$$\mathbb{E}[Y] = \sum_{j=1}^{n} \left(\frac{j}{n}\right)^m = \frac{1}{n^m} \sum_{j=1}^{n} j^m$$

The expected value of the minimum of $m$ independent $n$-sided dice rolls is:

$$\boxed{\mathbb{E}[Y] = \frac{1}{n^m} \sum_{j=1}^{n} j^m}$$

(b) For large $n$, the sum $\sum_{j=1}^{n} j^m$ can be approximated using the integral:

$$\sum_{j=1}^{n} j^m \approx \int_1^n x^m \, dx = \frac{n^{m+1} - 1}{m+1} \approx \frac{n^{m+1}}{m+1}$$

Therefore, for large $n$:

$$\mathbb{E}[Y] \approx \frac{1}{n^m} \cdot \frac{n^{m+1}}{m+1} = \frac{n}{m+1}$$

(c) With $m = 3$ rolls of $n = 100$-sided dice:

$$\mathbb{E}[Y] = \frac{1}{100^3} \sum_{j=1}^{100} j^3$$

Using the formula for the sum of cubes: $\sum_{j=1}^{n} j^3 = \left(\frac{n(n+1)}{2}\right)^2$

$$\sum_{j=1}^{100} j^3 = \left(\frac{100 \cdot 101}{2}\right)^2 = (50 \cdot 101)^2 = 5050^2 = 25,502,500$$

Therefore:

$$\mathbb{E}[Y] = \frac{25,502,500}{1,000,000} = 25.5025$$

The approximation formula gives: $\frac{100}{3+1} = 25$, which is close to the exact value.

Q10: A frog begins at position 100 on a number line. Every second, if the frog is at position $x > 1$, it jumps to a randomly chosen integer position between 1 and $x - 1$ (inclusive). What is the expected number of seconds to reach to position 1?

**A:** Let $\mathbb{E}(x)$ be the expected number of seconds for the frog to reach position 1, starting from position $x$. The frog is currently at position 100, so we want to find $\mathbb{E}(100)$.

We can set up a recursive relation for $\mathbb{E}(x)$.

If the frog is at position $x$, it can jump to any integer position from 1 to $x - 1$ with equal probability. There are $x - 1$ possible positions.

Let's analyze the transitions: From position $x$, the frog jumps to a position $j$, where $j \in \{1, 2, \ldots, x - 1\}$. The probability of jumping to any specific position $j$ is $\frac{1}{x-1}$.

The expected number of jumps from position $x$ is 1 (the next jump) plus the expected number of jumps from the new position. This leads to the recurrence relation:

$$\mathbb{E}(x) = 1 + \sum_{j=1}^{x-1} \frac{1}{x-1} \mathbb{E}(j)$$

We are looking for $\mathbb{E}(100)$.

Let's find the values of $\mathbb{E}(x)$ for small $x$:

- **Base case:** If the frog is at position 1, it has reached its destination. Thus, $\mathbb{E}(1) = 0$.
- For $x = 2$: The frog is at position 2. It can only jump to position 1.

$$\mathbb{E}(2) = 1 + \frac{1}{2-1}\mathbb{E}(1) = 1 + \frac{1}{1} \cdot 0 = 1.$$

  This makes sense: it takes one jump to get from 2 to 1.
- For $x = 3$: The frog is at position 3. It can jump to position 1 or 2 with probability 1/2.

$$\mathbb{E}(3) = 1 + \frac{1}{2}\mathbb{E}(1) + \frac{1}{2}\mathbb{E}(2) = 1 + \frac{1}{2}(0) + \frac{1}{2}(1) = 1.5.$$

- For $x = 4$: The frog is at position 4. It can jump to position 1, 2, or 3 with probability 1/3.

$$\begin{aligned}
\mathbb{E}(4) &= 1 + \frac{1}{3}\mathbb{E}(1) + \frac{1}{3}\mathbb{E}(2) + \frac{1}{3}\mathbb{E}(3) && (1)\\
&= 1 + \frac{1}{3}(0) + \frac{1}{3}(1) + \frac{1}{3}(1.5) && (2)\\
&= 1 + \frac{1}{3}(2.5) = 1 + \frac{5}{6} = \frac{11}{6} \approx 1.833. && (3)
\end{aligned}$$

Now, let's look for a general pattern. From the recursive formula, we can write:
For $\mathbb{E}(x)$:

$$(x-1)\mathbb{E}(x) = (x-1) + \sum_{j=1}^{x-1} \mathbb{E}(j)$$

For $\mathbb{E}(x-1)$:

$$\mathbb{E}(x-1) = 1 + \frac{1}{x-2}\sum_{j=1}^{x-2} \mathbb{E}(j)$$

$$(x-2)\mathbb{E}(x-1) = (x-2) + \sum_{j=1}^{x-2} \mathbb{E}(j)$$

By subtracting the two equations, we can find a simpler recurrence for $\mathbb{E}(x)$:

$$(x-1)\mathbb{E}(x) - (x-2)\mathbb{E}(x-1) = (x-1) - (x-2) + \mathbb{E}(x-1)$$

$$(x-1)\mathbb{E}(x) - (x-2)\mathbb{E}(x-1) = 1 + \mathbb{E}(x-1)$$

$$(x-1)\mathbb{E}(x) = (x-1)\mathbb{E}(x-1) + 1$$

$$\mathbb{E}(x) = \mathbb{E}(x-1) + \frac{1}{x-1}$$

This is a telescoping sum. We can use this to find a general formula for $\mathbb{E}(x)$:

$$\mathbb{E}(x) = \mathbb{E}(x-1) + \frac{1}{x-1}$$

$$\mathbb{E}(x-1) = \mathbb{E}(x-2) + \frac{1}{x-2}$$

$$\vdots$$

$$\mathbb{E}(2) = \mathbb{E}(1) + \frac{1}{1}$$

Summing these equations, we get:

$$\mathbb{E}(x) = \mathbb{E}(1) + \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{x-1}$$

Since $\mathbb{E}(1) = 0$, the formula simplifies to:

$$\mathbb{E}(x) = \sum_{j=1}^{x-1} \frac{1}{j}$$

This sum is the $(x-1)$-th harmonic number, $H_{x-1}$.
We want to find $\mathbb{E}(100)$, which corresponds to the harmonic number $H_{99}$.

$$\mathbb{E}(100) = \sum_{j=1}^{99} \frac{1}{j} = H_{99}$$

The value of $H_n$ can be approximated by $\ln(n) + \gamma$, where $\gamma$ is the Euler-Mascheroni constant ($\approx 0.57721$).

$$\mathbb{E}(100) = H_{99} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{99}$$

While the exact value is a sum, this is the final formula for the expected number of seconds. The exact numerical value can be calculated as approximately 5.1776.
The expected number of seconds to reach position 1 is the 99th harmonic number, $H_{99}$.

Q11: While roaming around in one of the gaming zones in your favorite mall, you come across an interesting game that might help you win some lunch money. It's a game of chance for a single player in which a fair coin is tossed at each stage. The pot starts at 1 rupee 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 rupee if a tail appears on the first toss, 2 rupees if a head appears on the first toss and a tail on the second, 4 rupees if a head appears on the first two tosses and a tail on the third, 8 rupees if a head appears on the first three tosses and a tail on the fourth, and so on. In short, the player wins $2^{k-1}$ rupees if the coin is tossed $k$ times until the first tail appears. What would be a fair price to pay the gaming zone for entering the game?

(a) Let $X$ be the amount of money (in rupees) that the player wins. Find $\mathbb{E}[X]$.

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

(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 $2^{30}$ rupees (around 1.07 billion rupees). That is, if you win more than $2^{30}$ rupees, the casino is going to pay you only $2^{30}$ rupees. Let $Y$ be the money that the player wins in this case. Find $\mathbb{E}[Y]$.

**A:**

(a)

$$P(X = 2^{k-1}) = P(\text{The coin was tossed } k \text{ times}$$
$$\text{and the first tail appeared on the } k^{\text{th}} \text{ toss})$$

$$\therefore P(X = 2^{k-1}) = \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2}$$

(The coin is fair, so the probability of heads or tails is $\frac{1}{2}$.)
The expected value of $X$ is given by:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} 2^{k-1} \cdot P(X = 2^{k-1}) = \sum_{k=1}^{\infty} 2^{k-1} \cdot \left(\frac{1}{2}\right)^{k} = \sum_{k=1}^{\infty} \frac{2^{k-1}}{2^k} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty$$

Thus, the expected value of $X$ is infinite, indicating that the player could potentially win an unbounded amount of money.

(b) We first find the number of tosses that are required for the player to win at least 65 rupees. The player wins more than 65 rupees if the first tail appears after at least 7 tosses (since $2^6 = 64$ and $2^7 = 128$). Thus, we need to find the probability that the first tail appears after at least 7 tosses: Therefore, the event "winning more than 65 rupees" is identical to the event "the first 7 tosses are all Heads" Since the tosses are independent, the probability of getting 7 Heads in a row is:

$$P(\text{first 7 tosses are Heads}) = \left(\frac{1}{2}\right)^{7} = \frac{1}{128}$$

Hence the probability that the player wins more than 65 rupees is $\frac{1}{128}$.

(c) Define $Y$ as the amount of money that the player wins, with the condition that if the player wins more than $2^{30}$ rupees, they only receive $2^{30}$ rupees.

$$Y = \begin{cases} 2^{k-1} & \text{for } k = 1, 2 \ldots, 30 \\ 2^{30} & \text{for } k = 31, 32 \ldots \end{cases}$$

Now,

$$p_Y(Y = 2^{k-1}) = \frac{1}{2^k}$$

Hence,

$$p_Y(Y = 2^{30}) = \sum_{k=31}^{\infty} \frac{1}{2^k} = \frac{\frac{1}{2^{30}}}{1 - \frac{1}{2}} = \frac{1}{2^{29}}$$

(Using the formula for the sum of an infinite geometric series)

$$\therefore E[Y] = \sum_{k=1}^{30} 2^{k-1} \cdot \frac{1}{2^k} + 2^{30} \cdot \frac{1}{2^{29}} = 30 \cdot \frac{1}{2} + = 15 + 2 = 17$$

Q12: Let $X$ be a random variable with mean $\mu$ and variance $\sigma^2$. For arbitrary constants $a$ and $b$, show that $Var(aX + b) = a^2 Var(X)$.

**Solution:**
We begin with the definition of variance, $Var(Y) = E[(Y - E[Y])^2]$. Let's define a new random variable $Y = aX + b$.

First, find the expected value of $Y$ by applying the linearity of expectation:

$$E[Y] = E[aX + b] = aE[X] + b = a\mu + b$$

Next, substitute $Y$ and $E[Y]$ back into the variance definition:

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

Simplifying the term inside the expectation gives:

$$(aX + b) - (a\mu + b) = aX - a\mu = a(X - \mu)$$

Substituting this back:

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

Since $a^2$ is a constant, it can be factored out of the expectation:

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

The term $E[(X - \mu)^2]$ is, by definition, the variance of $X$. Therefore, we have successfully shown that:

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

Q13: You are playing a game with an urn that contains 10 black balls and 1 white ball. The balls are drawn one by one from the urn without replacement. The game ends when the white ball is drawn.

(a) For every black ball that is drawn before the game ends, you win \$5. What are your expected winnings?

(b) Suppose the rules are changed. Now, if you draw $k$ black balls before the game ends, your total payout is $\$2^k$. (Note: if you draw the white ball first, $k = 0$ and your payout is $2^0 = \$1$). What is the new expected payout?

**Solution:**

(a) Let $X$ be the number of black balls drawn before the white ball. Your winnings are $W = 5X$. To find $E[W]$, we can use linearity of expectation: $E[W] = 5E[X]$. So, the problem reduces to finding the expected number of black balls drawn.

Instead of calculating the PMF of $X$ directly, we can use a symmetry argument. There are 11 balls in total. Consider a random permutation of all 11 balls. The position of the single white ball is equally likely to be any of the 11 positions.

$$\mathbb{P}(\text{White ball is in position } i) = \frac{1}{11} \quad \text{for } i = 1, 2, \ldots, 11$$

The number of black balls drawn, $X$, is simply the number of balls drawn before the white one. If the white ball is in position $i$, then $X = i - 1$. Therefore, the number of black balls drawn, $X$, is uniformly distributed on the set $\{0, 1, 2, \ldots, 10\}$.

The expected value of a discrete uniform random variable is the average of its values:

$$E[X] = \frac{0 + 1 + 2 + \cdots + 10}{11} = \frac{\frac{10(11)}{2}}{11} = 5$$

Now we can find the expected winnings:

$$E[W] = 5 \cdot E[X] = 5 \cdot 5 = 25$$

Your expected winnings are \$25.

(b) Let the new payout be $Y$. The payout is now a function of the number of black balls drawn, $Y = g(X) = 2^X$. We need to find $E[Y] = E[2^X]$.

We use the formula for the expectation of a function of a random variable:

$$E[g(X)] = \sum_k g(k) \cdot p_X(X = k)$$

From part (a), we know that $X$ is uniformly distributed on $\{0, 1, \ldots, 10\}$, so $P(X = k) = 1/11$ for each $k$ in this range.

$$\begin{aligned} E[Y] &= \sum_{k=0}^{10} 2^k \cdot p_X(X = k) \\ &= \sum_{k=0}^{10} 2^k \cdot \frac{1}{11} \\ &= \frac{1}{11} \sum_{k=0}^{10} 2^k \end{aligned}$$

The sum is a finite geometric series: $\sum_{i=0}^{n} r^i = \frac{r^{n+1}-1}{r-1}$.

$$\sum_{k=0}^{10} 2^k = \frac{2^{11} - 1}{2 - 1} = 2048 - 1 = 2047$$

Substituting this back, we find the expected payout:

$$E[Y] = \frac{2047}{11} \approx 186.09$$

The new expected payout is approximately \$186.09.

Q14: In the coupon collector's problem from the tutorial, there are $N$ different types of coupons. Each time you collect a coupon, it is equally likely to be any of the $N$ types. Let $X$ be the total number of coupons you need to collect to have at least one of each type. We can write $X = \sum_{i=0}^{N-1} X_i$, where $X_i$ is the number of additional coupons needed to get a new type, given that you already have $i$ distinct types. Each $X_i$ is an independent geometric random variable, $X_i \sim \text{Geometric}(p_i)$, with success probability $p_i = \frac{N-i}{N}$. Find the variance of $X$.

**Solution:**
The total number of coupons is $X = \sum_{i=0}^{N-1} X_i$. Since the stages of collecting each new coupon are independent, the variance of the sum is the sum of the variances:

$$Var(X) = Var\left(\sum_{i=0}^{N-1} X_i\right) = \sum_{i=0}^{N-1} Var(X_i)$$

The variance of a geometric random variable $Y \sim \text{Geometric}(p)$ is $Var(Y) = \frac{1-p}{p^2}$. For our variable $X_i$, the parameter is $p_i = \frac{N-i}{N}$. The variance of $X_i$ is therefore:

$$Var(X_i) = \frac{1-p_i}{p_i^2} = \frac{1-\frac{N-i}{N}}{\left(\frac{N-i}{N}\right)^2} = \frac{\frac{i}{N}}{\frac{(N-i)^2}{N^2}} = \frac{iN}{(N-i)^2}$$

To find the total variance, we sum this expression from $i = 0$ to $N - 1$. Note that for $i = 0$, the variance is 0.

$$Var(X) = \sum_{i=1}^{N-1} \frac{iN}{(N-i)^2}$$

To simplify this sum, we can perform a change of index. Let $j = N - i$. This implies $i = N - j$. As $i$ goes from 1 to $N - 1$, $j$ goes from $N - 1$ down to 1. The sum becomes:

$$\begin{aligned} Var(X) &= \sum_{j=1}^{N-1} \frac{(N-j)N}{j^2} = N \sum_{j=1}^{N-1} \frac{N-j}{j^2} \\ &= N\left(\sum_{j=1}^{N-1} \frac{N}{j^2} - \sum_{j=1}^{N-1} \frac{j}{j^2}\right) \\ &= N^2 \sum_{j=1}^{N-1} \frac{1}{j^2} - N \sum_{j=1}^{N-1} \frac{1}{j} \end{aligned}$$

The second sum is the $(N-1)$-th harmonic number, denoted $H_{N-1}$. Thus, the final expression is:

$$Var(X) = N^2\left(\sum_{k=1}^{N-1} \frac{1}{k^2}\right) - NH_{N-1}$$

Q15: Rule a surface with parallel lines a distance d apart. What is the probability that a randomly dropped needle of length $l \leq d$ crosses a line?

This is a classic problem in probability known as Buffon's Needle problem. The probability that a randomly dropped needle of length $l$ crosses a line when the lines are spaced $d$ apart can be derived using geometric probability. To start with, let $E(1)$ be the expected number of times a needle of length $l$ will intersect the lines. Let $X$ be the random variable denoting the number of intersections of the needle with the lines. Then

$$E[l] = \sum_{k=0}^{\infty} k \cdot P(X = k) = 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \ldots$$

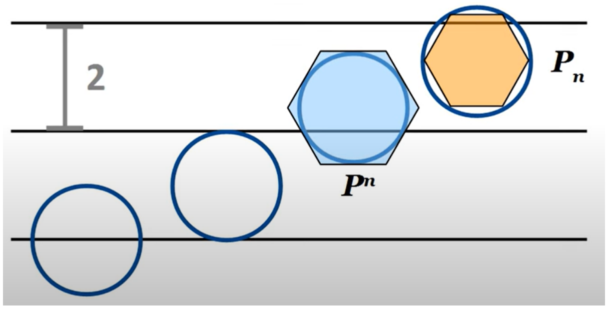
There are a few things to consider: In case of $l = 1$

(a) The needle will intersect mostly once with any of the lines.
(b) It will intersect twice in only one case (when the needle is perfect perpendicular to the lines)
(c) It will intersect infinite number of times only in one case, when the needle lies on one of the lines
(d) The last two cases happen with almost 0 probability (Think of all the other ways in which the needle can land, apart from those two cases)

Hence we can almost say that,

$$E[1] = P(X = 1)$$

By linearity of expectation,
$E[X + Y] = E[X] + E[Y]$
Therefore for a polygon with perimeter $x$, $E[x] = x \cdot E[1]$, that is the expected number of intersections is proportional to the polygon's perimeter
Now consider a circle with radius 1 instead. This circle will intersect exactly two times no matter the configuration.




With this circle, given a positive integer n, you can draw a polygon of n sides(each side of length 1), inscribing the circle, with perimeter $P_n$ and circumscribing it, with perimeter $P^n$.

$$P^n \leq Circumference(Circle) \leq P_n$$

$$\therefore E[P^n] \leq 2 \leq E[P_n]$$

As $n \to \infty$, The perimeters tend to the Circumference of the circle $2\pi$. Hence,

$$E[2\pi] \leq 2 \leq E[2\pi]$$

$$2\pi \cdot E[1] \leq 2 \leq 2\pi \cdot E[1]$$

$$\therefore 2\pi \cdot P(X = 1) = 2$$
$$P(X = 1) = \frac{1}{\pi}$$