UNIVERSITY OF TEHRAN

Statistical Inference Report

Hadiseh Mesbah
Student ID: 810102253

November 7, 2023

Introduction

This document presents detailed mathematical and statistical solutions to the assigned problems, focusing on probability theory and statistical inference.

Problem 1: Alarm System Effectiveness

Given an alarm system with the following probabilities:

- Probability of alarm when there is a dangerous condition $(A|D)$: 0.95.
- Probability of alarm when conditions are normal $(A|N)$: 0.005.
- Probability of normal conditions (N): 0.995.
- Probability of dangerous conditions (D): 0.005.

Task 1: Probability of a False Alarm

Find the probability of a false alarm $(P(N|A))$, i.e., the alarm sounds, but conditions are normal.

Solution: Using Bayes' Theorem:

$$P(N|A) = \frac{P(A|N)P(N)}{P(A|N)P(N) + P(A|D)P(D)}$$

Inserting the given probabilities, calculate $P(N|A)$.

$$P(N|A) = \frac{0.005 \times 0.995}{0.005 \times 0.995 + 0.95 \times 0.005} \tag{1}$$

And,

$$P(N|A) \approx 0.511. \tag{2}$$

Task 2: Probability of an Unidentified Critical Condition

Find the probability of failing to identify a dangerous condition $(P(D|\overline{A}))$.

Solution:

$$P(D|\overline{A}) = \frac{P(\overline{A}|D)P(D)}{P(\overline{A}|D)P(D) + P(\overline{A}|N)P(N)}$$

where,

$$P(\overline{A}|D) = 1 - P(A|D) and P(\overline{A}|N) = 1 - P(A|N). \tag{3}$$

$$P(\overline{A}) = 0.990025 \tag{4}$$

And,

$$P(D|\overline{A}) = \frac{0.05 \times 0.005}{0.990025} \approx 0.0002525188757. \tag{5}$$

Task 3: Probability of false alarms and unidentified critical conditions expected to occur during a 10-year period

Find the probability of Expected False Alarms ($P(D|\overline{A})$).

Solution:

To find the probability in the new way that conditions are dangerous when the alarm does not trigger, $P(D|\overline{A})$, we use Bayes' theorem:

$$P(D|\overline{A}) = \frac{P(\overline{A}|D)P(D)}{P(\overline{A})}. \tag{6}$$

Expectations Over 10 Years

The expected number of false alarms and unidentified critical conditions over ten years can be calculated by multiplying the daily probabilities by the number of days in 10 years (3650 days).

$$\text{Expected False Alarms} = \frac{0.005 \times 0.995}{0.009725} \times 3650 \approx 1867 \tag{7}$$

$$\text{Expected Unidentified Critical Conditions} = \frac{0.005 \times 0.005}{0.990275} \times 3650 \approx 0.92 \tag{8}$$

the Effectiveness of the Alarming System

1867 expected false alarms over 10 years is a significant number. False alarms can lead to desensitization and a lack of trust in the system, potentially causing unnecessary disruption and inconvenience. In this case, a probability of 0.92 for the system to detect dangerous conditions over 10 years implies a relatively high level of effectiveness in identifying and alerting to critical situations. In summary, the alarming system appears to be effective in terms of detecting real dangers (low missed alarms), but the high rate of false alarms is a significant drawback. The effectiveness of the system could be improved by reducing false alarms through more accurate sensors, improved algorithms, or a refined set of triggering conditions to enhance the system's reliability and user trust. Additionally, it's crucial to validate the specific safety requirements and adjust the alarming system's sensitivity accordingly to ensure that it aligns with the desired level of safety and risk tolerance.

Problem 2: Random Variables and Their Properties

Consider two independent and identically distributed random variables X_1 and X_2, each with mean m and variance σ^2.

Task 1: Mean and Variance of $Y_1 = X_1 + X_2$

Solution:

$$E[Y_1] = E[X_1 + X_2] = E[X_1] + E[X_2] = m + m = 2m$$

$$\text{Var}(Y_1) = \text{Var}(X_1) + \text{Var}(X_2) = \sigma^2 + \sigma^2 = 2\sigma^2$$

Task 2: Mean and Variance of $Y_2 = 2X_1$

Solution:

$$E[Y_2] = E[2X_1] = 2E[X_1] = 2m$$
$$\text{Var}(Y_2) = \text{Var}(2X_1) = 4\text{Var}(X_1) = 4\sigma^2$$

Task 3: Comparison of Variances

Solution: The variances of Y_1 and Y_2 are $2\sigma^2$ and $4\sigma^2$ respectively, thus they are different.

Task 4: Covariance between Y_1 and Y_2

Solution:

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(X_1 + X_2, 2X_1) = 2\text{Cov}(X_1, X_1) + \text{Cov}(X_2, X_1)$$

Since X_1 and X_2 are independent, $\text{Cov}(X_2, X_1) = 0$, and hence:

$$\text{Cov}(Y_1, Y_2) = 2\sigma^2$$

Problem 3: Birthday Problem

We are analyzing the probability of shared birthdays among n individuals, assuming a uniform distribution.

Task 1: probability function P for Ω

Solution: To define the probability function P for the sample space Ω, which represents all possible sequences of n birthdays (one for each person), we can consider that each birthday can be assigned a number from 1 to 365, excluding leap days.

$$P(\text{specific sequence}) = \frac{1}{365} \times \frac{1}{365} \times \frac{1}{365} \times \ldots \times \frac{1}{365} = \left(\frac{1}{365}\right)^n$$

Task 2: Probability Function P for n Birthdays

Solution: The probability of at least one person sharing your birthday (Event A) is:

$$P(A) = 1 - \left(\frac{364}{365}\right)^n$$

For at least two people sharing a birthday (Event B) and at least three sharing a birthday (Event C), probabilities can be computed similarly but are more complex and may require combinatorial calculations or simulations.

Task 3: Minimum n for $P(A) > 0.5$

Solution:
Solve $1 - \left(\frac{364}{365}\right)^n > 0.5$ for n to find the minimum number of people required.

$$\begin{aligned}
1 - \left(\frac{364}{365}\right)^n &> 0.5 \\
\left(\frac{364}{365}\right)^n &< 0.5 \\
\ln\left(\left(\frac{364}{365}\right)^n\right) &< \log(0.5) \\
n \cdot \ln\left(\frac{364}{365}\right) &< \log(0.5) \\
n &> \frac{\ln(0.5)}{\ln\left(\frac{364}{365}\right)} \\
n &\approx 252.6519 \\
n &= 253
\end{aligned}$$

Task 4: Rationale for $n > \frac{365}{2}$

Explanation: The birthday paradox explains why n does not need to be as large as $\frac{365}{2}$ to have a significant probability of shared birthdays. It's a counterintuitive result arising from the principles of probability.

Task 6: probability for $P(B)$

Solution:
To calculate the probability P(B) that at least two people in a group share a birthday, one can use the complementary probability approach. This involves first finding the probability that no two people share a birthday, and then subtracting that from 1 to get the probability that at least two people share a birthday. So, $P(\neg B)$ is:

$$P(\neg B) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365 - n + 1}{365} = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365 - (n-1)}{365}$$

then,

$$P(B) = 1 - P(\neg B)$$

Thus, $P(B)$ is:

$$P(B) = 1 - \frac{365!}{365^n \cdot (365 - n)!}$$

Task 5 and 7: python code

Solution:
The code for each problem of the course is contained in a zip file with the format $[courseName - problemNumber]$.

Task 8: the frequency of birthdays varies

Explanation:
When estimating the minimum number of people who share the same birthday, we assume that each day in a year has the same probability. However, in reality, this assumption may not be accurate as certain months, such as October, may have a higher chance than others while some months, like April, may have a lower chance. It is possible that certain events or situations, such as weather or social circumstances, could contribute to these outcomes. In any case, in real-world scenarios, we often only need a small number of individuals to find two people who share the same birthday.

Problem 4: Classic Probability Questions

Task 1: Probability, Both Children, are Girls

Given that the first child is a girl, find the probability that both children are girls.

Solution: Assuming equal probability for boy or girl birth: can represent this as:

$$P(\text{Both girls}|\text{First child is a girl})$$

then,

$$P(\text{Both girls}|\text{First child is a girl}) = \frac{P(\text{Both girls} \cap \text{First child is a girl})}{P(\text{First child is a girl})}$$

Probability of both children being girls (Both girls): $P(\text{Both girls}) = \frac{1}{4}$ since there are 4 equally likely possibilities (GG, GB, BG, BB), and one of them is "Both girls."
Probability of the First child being a girl (First child is a girl): $P(\text{First child is a girl}) = \frac{2}{4}$ because there are two equally likely possibilities where the older child is a girl (GG, GB).

$$P(\text{Both girls}|\text{First child is a girl}) = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

Task 2: Probability, Both Children, are Boys

Given that at least one of the children is a boy find the probability that both are boys.

Solution:

$$P(\text{Both boys}|\text{At least one is a boy})$$

Probability of both children being boys (Both boys): $P(\text{Both boys}) = \frac{1}{4}$ (There are 4 equally likely possibilities, and only one of them is "Both boys").
Probability of at least one child being a boy (At least one is a boy): $P(\text{At least one is a boy}) = 1 - P(\text{Both girls}) = 1 - \frac{1}{4} = \frac{3}{4}$.
Then,

$$P(\text{Both boys}|\text{At least one is a boy}) = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Problem 5: Taxi Identification Problem

Scenario

A taxi was involved in a hit-and-run accident at night. Two taxi companies operate in the city: the Green Taxis (99% of all taxis) and the Blue Taxis (1%). A witness identified the taxi involved in the accident as a Blue Taxi. The court tested the reliability of the witness under the same conditions that prevailed on the night of the accident and concluded that the witness correctly identified each one of the two colors 99% of the time and failed 20% of the time.

Task:

deliver a brief speech to the the jury, aiming to provide them with sufficient doubt regarding your client's guilt

Defendant's Case for Reasonable Doubt

Good day, ladies and gentlemen of the jury. I am here today to clarify the uncertainties and doubts surrounding this case. Though we must recognize the importance of eyewitness testimony, it is essential to understand the limitations and potential errors that come with such testimony.

As instructed by the courts, when assessing eyewitness identification testimony, we must consider several factors. These include the eyewitness's ability and chance to observe the suspect, the conditions during the observation, and the possibility of subsequent influence or suggestiveness.

The witness saw a blue taxi leaving the accident scene. In our city, only 1 percent of taxis are blue. The witness was tested and correctly identified blue taxis 99 percent of the time and misidentified green taxis as blue 2percent of the time.

Let's represent this information in a table:

Witness says	Taxi is Blue	Taxi is Green
Blue	99%	1%
Green	2%	98%

From this table, we can see that the witness can mistake green taxis for blue ones. Now, imagine the witness observing 100 taxis passing by, one being blue and 99 being green. This is what we could expect:

Witness says	Taxi is Blue	Taxi is Green
Blue	1 (99% of 1)	0 (1% of 1)
Green	$\sim$2 (2% of 99)	$\sim$97 (98% of 99)

In the case where the witness says "Blue," we find that the taxi could be one of the $\sim$2 green taxis mistaken as blue or the one blue taxi. Therefore, the probability that the taxi was green, despite the witness's claim, is $\sim 2/(1+\sim 2)$, about 67 percent. Based on the test, Given that the witness identified the taxi as blue, the probability that it is blue is only 33 percent, which may not provide enough certainty to make a confident decision.

Although we respect the witness's testimony, it is equally important to consider the scientific evidence that suggests a high probability of misidentification. Our responsibility is to examine all evidence with scrutiny as a person's freedom is at stake. It is vital to remember that evidence that seems irrefutable at first glance may still be subject to scrutiny.

In conclusion, the data obtained from the witness perception test raises reasonable doubt regarding my client's guilt. This case highlights the significance of interpreting eyewitness testimony cautiously and recognizing its limitations in light of scientific evidence.

Thank you for your attention.

Mathematical and Statistical Analysis

1. **Definitions and Notations**: Let B be the event that the taxi is Blue, and G be the event that the taxi is Green. Let W_B be the event that the witness identifies the taxi as Blue and W_G be the event that the witness identifies the taxi as Green.

2. **Given Probabilities**: $P(B) = 0.1$ and $P(G) = 0.99$ (probabilities of actual taxi colors). $P(W_B|B) = 0.99$ and $P(W_G|G) = 0.98$ (probabilities that the witness correctly identifies the taxi color). $P(W_B|G) = 0.2$ and $P(W_G|B) = 0.1$ (probabilities that the witness incorrectly identifies the taxi color).

3. **Required Probability**: We need to find $P(B|W_B)$, the probability that the taxi is Blue, given that the witness identified it as Blue.

4. **Application of Bayes' Theorem**:

$$P(B|W_B) = \frac{P(W_B|B) \times P(B)}{P(W_B)}$$

   Where $P(W_B)$ is the total probability of the witness identifying the taxi as Blue, given by:

$$P(W_B) = P(W_B|B) \times P(B) + P(W_B|G) \times P(G)$$

   Thus:

$$P(B|W_B) = \frac{0.99 \times 0.1}{0.99 \times 0.1 + 0.2 \times 0.99} = 0.333$$

5. **Conclusion**: The calculation updates the probability of the taxi being Blue, considering the witness's testimony and prior possibilities. This result highlights the impact of evidence reliability and prior probabilities on posterior beliefs, demonstrating Bayesian inference in practical scenarios.

Problem 6: Dice Rolling

Part 1: Probability Mass Function (PMF) for S

Calculate the probability that the sum of the dice is seven.

Solution: The probability mass function (PMF) is a function that calculates the likelihood of each possible outcome of a discrete random variable. In this case, the random variable S represents the number of sides on the chosen die, which can only take on the values of 4, 6, or 8. The total number of dice is 4. The tetrahedron (4 sides) and cube (6 sides) each have a 1/4 chance of being chosen, while the octahedra (8 sides) have a 2/4 or 1/2 chance (since there are two of them). So, the PMF of S is as follows:

$$P(S=4)=\frac{1}{4}$$
$$P(S=6)=\frac{1}{4}$$
$$P(S=8)=\frac{1}{2}$$

This text satisfies the conditions for a probability mass function (PMF): the probability of each outcome is between 0 and 1 (inclusive), and the sum of all possibilities is equal to 1.

Task 2:Bayes' Rule and PMF of S given R=3

Solution: probability of S being 4, 6, or 8 given R = 3: We want to find the pmf of S given R = 3. We'll use Bayes' rule for this.
Bayes' rule states:

$$P(S|R)=\frac{P(R|S)\cdot P(S)}{P(R)}$$

The probability of rolling a 3 depends on the die:

$$P(R=3|S=4)=\frac{1}{4},\ \text{because a 4-sided die has one 3 out of four possible outcomes.}$$
$$P(R=3|S=6)=\frac{1}{6},\ \text{because a 6-sided die has one 3 out of six possible outcomes.}$$
$$P(R=3|S=8)=\frac{1}{8},\ \text{because an 8-sided die has one 3 out of eight possible outcomes.}$$

Now, apply Bayes' rule for each value of S:

$$P(S=4|R=3)=\frac{P(R=3|S=4)\cdot P(S=4)}{P(R=3)}=\frac{1/4\cdot 1/4}{P(R=3)}=\frac{1/16}{P(R=3)}$$
$$P(S=6|R=3)=\frac{P(R=3|S=6)\cdot P(S=6)}{P(R=3)}=\frac{1/6\cdot 1/4}{P(R=3)}=\frac{1/24}{P(R=3)}$$
$$P(S=8|R=3)=\frac{P(R=3|S=8)\cdot P(S=8)}{P(R=3)}=\frac{1/8\cdot 1/2}{P(R=3)}=\frac{1/16}{P(R=3)}$$

We don't know P(R=3) directly, but we can calculate it using the law of total probability:

$$\begin{aligned}P(R=3)&=P(R=3|S=4)P(S=4)+P(R=3|S=6)P(S=6)+P(R=3|S=8)P(S=8)\\&=\frac{1}{4}\cdot\frac{1}{4}+\frac{1}{6}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{2}=\frac{1}{24}+\frac{2}{16}=\frac{4}{24}\end{aligned}$$

Thus,

$$
\begin{aligned}
P(S=4|R=3) &= \frac{1/16}{4/24} = \frac{6}{16} \approx 0.37 \\
P(S=6|R=3) &= \frac{1/24}{4/24} = \frac{1}{16} \approx 0.25 \\
P(S=8|R=3) &= \frac{1/16}{4/24} = \frac{6}{16} \approx 0.37
\end{aligned}
$$

The die with eight sides (S = 8) and the die with four sides (S = 4) are the most probable choices when R = 3 due to their high conditional probability.

Part 3: Most likely die when R=6

Solution: probability of S being 4, 6, or 8 given R = 6: We want to find the pmf of S given R = 3. We'll use Bayes' rule for this.
Bayes' rule states:

$$
P(S|R) = \frac{P(R|S) \cdot P(S)}{P(R)}
$$

The probability of rolling a 3 depends on the die:

$$
\begin{aligned}
&P(R=6|S=4) = 0, \text{ because a 4-sided die does not have a 6 side} \\
&P(R=6|S=6) = \frac{1}{6}, \text{ because a 6-sided die has one 3 out of six possible outcomes.} \\
&P(R=6|S=8) = \frac{1}{8}, \text{ because an 8-sided die has one 3 out of eight possible outcomes.}
\end{aligned}
$$

Now, apply Bayes' rule for each value of S:

$$
\begin{aligned}
P(S=4|R=6) &= \frac{P(R=6|S=4) \cdot P(S=4)}{P(R=6)} = \frac{0 \cdot 1/4}{P(R=6)} = 0 \\
P(S=6|R=6) &= \frac{P(R=6|S=6) \cdot P(S=6)}{P(R=6)} = \frac{1/6 \cdot 1/4}{P(R=6)} = \frac{1/24}{P(R=6)} \\
P(S=8|R=6) &= \frac{P(R=6|S=8) \cdot P(S=8)}{P(R=6)} = \frac{1/8 \cdot 1/2}{P(R=6)} = \frac{1/16}{P(R=6)}
\end{aligned}
$$

We don't know P(R=6) directly, but we can calculate it using the law of total probability:

$$
\begin{aligned}
P(R=6) &= P(R=6|S=4)P(S=4) + P(R=6|S=6)P(S=6) + P(R=6|S=8)P(S=8) \\
&= \frac{1}{4} \cdot 0 + \frac{1}{6} \cdot \frac{1}{4} + \frac{1}{8} \cdot \frac{1}{2} = \frac{1}{24} + \frac{1}{16} = \frac{5}{48}
\end{aligned}
$$

Thus,

$$P(S = 4|R = 6) = 0$$
$$P(S = 6|R = 6) = \frac{1/24}{5/48} = \frac{1}{16} \approx 0.2$$
$$P(S = 8|R = 6) = \frac{1/16}{5/48} = \frac{6}{16} \approx 0.3$$

The die with eight sides (S = 8) is the most probable choice when R = 6 due to its high conditional probability.

Part 4: Most likely die when R=7

If we want to roll a 7, only the 8-sided die can produce this result. In this case, the 8-sided die is the most likely choice.

Problem 7: Dices Rolling

Consider rolling a 4-side and 6-side dice.

Task 1: Standard Deviation for X, Y, and Z

Calculate the calculate the standard deviation for X, Y, and Z.

Solution:
For X (4-sided die): The possible outcomes of X are 1, 2, 3, and 4, each with a probability of 1/4. The mean (μ) of X is:

$$\mu_X = \frac{1}{4} \sum_{i=1}^{4} i = \frac{1}{4} \cdot 10 = 2.5$$

The variance (σ^2) of X is:

$$\sigma_X^2 = \frac{1}{4} \sum_{i=1}^{4} (i - \mu_X)^2 = \frac{1}{4} \cdot \big((1 - 2.5)^2 + (2 - 2.5) + (3 - 2.5)^2 + (4 - 2.5)\big) = \frac{5}{4}$$

Thus, calculate the standard deviation for X (σ_X) as follows:

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\frac{5}{4}} = 1.118033989$$

For Y (6-sided die): The possible outcomes of Y are 1, 2, 3, 4, 5, and 6, each with a probability of 1/6. The mean (μ) of Y is:

$$\mu_Y = \frac{1}{6} \sum_{i=1}^{6} i = \frac{1}{6} \cdot 21 = 3.5$$

The variance (σ^2) of Y is:

$$\sigma_Y^2 = \frac{1}{6}\sum_{i=1}^{6}(i-\mu_Y)^2 =$$

$$\frac{1}{6}\cdot\left((1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2\right) = \frac{35}{12}$$

So, the standard deviation (σ_Y) of Y is:

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\frac{35}{12}} = 1.707825128$$

For Z (average of X and Y): The mean (μ) of Z is the average of the means of X and Y:

$$\mu_Z = \frac{1}{2}\cdot(\mu_X+\mu_Y) = \frac{1}{2}\cdot(2.5+3.5) = 3.0$$

The variance (σ^2) of Z is the sum of the conflicts of X and Y divided by 4 (since X and Y are independent):

$$\sigma_Z^2 = \frac{\sigma_X^2+\sigma_Y^2}{4} = \frac{\frac{5}{4}+\frac{35}{12}}{4} = \frac{25}{24}$$

So, the standard deviation (σ_Z) of Z is:

$$\sigma_Z = \sqrt{\sigma_Z^2} = \sqrt{\frac{25}{24}} = 1.020620726$$

Task 2: Comprehensive probability mass function pmf and cdf for Z

Calculate the probability mass function (pmf) and cumulative distribution function (cdf)

Solution:

To derive the Probability Mass Function (PMF) for Z, need to calculate the probability of each possible outcome for Z. These outcomes are the averages of results from X and Y. For every value of Z, add the possibilities of all the combinations of outcomes from X and Y that lead to the given value of Z.

$$\begin{cases}
P(X=1,Y=1) = \frac{1}{24} & \text{if } z=1 \\
P(X=1,Y=2)+P(X=2,Y=1) = \frac{2}{24} & \text{if } z=1.5 \\
P(X=1,Y=3)+P(X=2,Y=2)+P(X=3,Y=1) = \frac{3}{24} & \text{if } z=2 \\
P(X=1,Y=4)+P(X=2,Y=3)+P(X=3,Y=2)+P(X=4,Y=1) = \frac{4}{24} & \text{if } z=2.5 \\
P(X=1,Y=5)+P(X=2,Y=4)+P(X=3,Y=3)+P(X=4,Y=2) = \frac{4}{24} & \text{if } z=3 \\
P(X=1,Y=6)+P(X=2,Y=5)+P(X=3,Y=4)+P(X=4,Y=3) = \frac{4}{24} & \text{if } z=3.5 \\
P(X=2,Y=6)+P(X=3,Y=5)+P(X=4,Y=4) = \frac{3}{24} & \text{if } z=4 \\
P(X=3,Y=6)+P(X=4,Y=5) = \frac{2}{24} & \text{if } z=4.5 \\
P(X=4,Y=6) = \frac{1}{24} & \text{if } z=5
\end{cases}$$

To obtain the Cumulative Distribution Function (CDF) for Z, calculate the cumulative probabilities for each possible outcome of Z. For a particular value of z, the cumulative probability is the sum of probabilities of all effects that are less than or equal to z.

$$F_Z(z) = P(Z \leq z) = \sum_x \sum_y P(X = x, Y = y) \text{ for } x \in \{1, 2, 3, 4\}, y \in \{1, 2, 3, 4, 5, 6\}$$

$$F_Z(z) = P(Z \leq z) = \begin{cases} \frac{1}{24} & \text{if } z = 1 \\ \frac{3}{24} & \text{if } z = 1.5 \\ \frac{6}{24} & \text{if } z = 2 \\ \frac{10}{24} & \text{if } z = 2.5 \\ \frac{14}{24} & \text{if } z = 3 \\ \frac{18}{24} & \text{if } z = 3.5 \\ \frac{21}{24} & \text{if } z = 4 \\ \frac{23}{24} & \text{if } z = 4.5 \\ \frac{24}{24} & \text{if } z = 5 \end{cases}$$

Task 3: Calculate the expected overall gain or loss after 60 rounds of the game

In each round of the game, calculate the expected gain (or loss) based on the probabilities of X being greater than Y and X being less than or equal to Y, and then sum these expected values over all 60 rounds.

Solution:

First, find $P(X > Y)$ and $P(X \leq Y$ based on the probability mass function (pmf) for Z:

$$P(X > Y) = \sum_{z=1}^{24} P(Z = z)$$

- X can be 1, 2, 3, or 4.
- Y can be 1, 2, 3, 4, 5, or 6.
- X greater than Y can occur in specific cases, such as $(X = 2, Y = 1)$, $(X = 3, Y = 1)$, $(X = 3, Y = 2)$, and $(X = 4, Y = 1)$, $(X = 4, Y = 2)$, $(X = 4, Y = 3)$.
- There are 6 cases out of a total of (4×6) possible outcomes, so the probability is $\frac{6}{24}$.
- There are cases where X is equal to Y (e.g., $X = 1, Y = 1$) and cases where Y is greater than X (e.g., $X = 1, Y = 2$, $X = 2, Y = 3$, etc.).
- The probability is $\frac{18}{24}$.
- When $X > Y$, you win $2X$ dollars with a probability of $\frac{6}{24}$.

$$\frac{1}{24} \cdot 4 + \frac{2}{24} \cdot 6 + \frac{3}{24} \cdot 8 = \frac{5}{3}$$

- When X is not greater than Y, you lose 1 dollar with a probability of $\frac{18}{24}$.

$$\frac{16}{24} \cdot (-1) = -\frac{18}{24}$$

- The expected value of a single round (EV(round)) can be calculated as:

$$EV(\text{round}) = \frac{2X}{24} - \frac{18}{24}$$

- Calculate the expected value of the entire game (60 rounds) by multiplying the EV of a single round by 60:

$$EV(\text{game}) = 60 \cdot EV(\text{round}) = 60 \cdot \left(\frac{5}{3} - \frac{18}{24}\right) = 55.00000002 \text{ dollars}$$

Problem 8: Raisin Box Probability

Task 1: Number of Raisins in the Cereal Box:

Solution: The number of raisins in the box can be determined by integrating the density function with respect to height. We integrate the density function from 0 to 30 cm:

$$\begin{aligned} N &= \int_0^{30} (40 - h)\, dh \\ &= \left[40h - \frac{1}{2}h^2\right]_0^{30} \\ &= \left[40(30) - \frac{1}{2}(30^2) - 0\right] \\ &= (1200 - 450) \\ &= 750 \text{ raisins} \end{aligned}$$

Task 2: Probability Density Function (pdf) for H:

Solution:
the density of raisins varies linearly from the bottom to the top of the box; normalize the density function to obtain the pdf g(h) for H.

$$g(h) = \frac{f(h)}{N}$$

Where N is the total number of raisins (750, as calculated above).
Therefore,

$$g(h) = \frac{40 - h}{750}$$

For illustration, see Figure 1

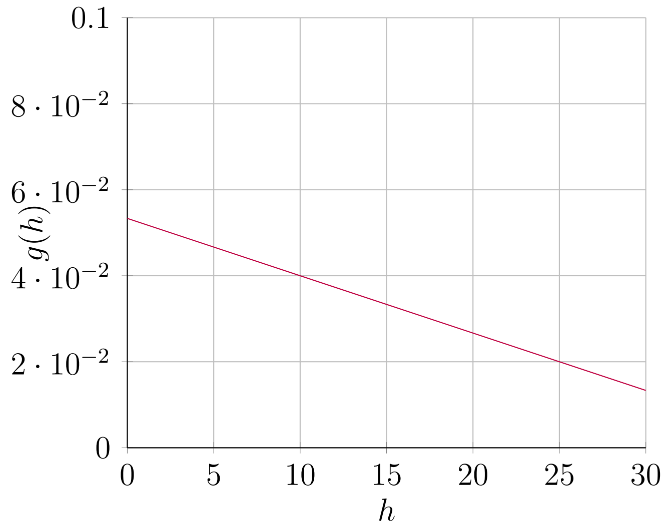

Figure 1: Probability Density Function (pdf) $g(h)$ for H

Task 3: Cumulative Distribution Function (CDF) for H:

Solution:
The cumulative distribution function is the integral of the pdf.

$$G(h) = \int_0^h g(x)\,dx$$

$$G(h) = \int_0^h \frac{40 - x}{750}\,dx$$

Thus,

$$G(h) = \frac{1}{750}\left[40x - \frac{1}{2}x^2\right]_0^h$$

$$G(h) = \frac{1}{750}\left(40h - \frac{1}{2}h^2 - 0\right)$$

Finally:

$$G(h) = \frac{40h - \frac{1}{2}h^2}{750}$$

For illustration, see Figure 2

Task 4: Probability of a Raisin in the Bottom Third of the Box:

Solution:
The bottom third of the box corresponds to the height range [0, 10] cm.

$$P(0 \leq H \leq 10) = G(10) - G(0)$$

Then,

$$G(0) = 0$$

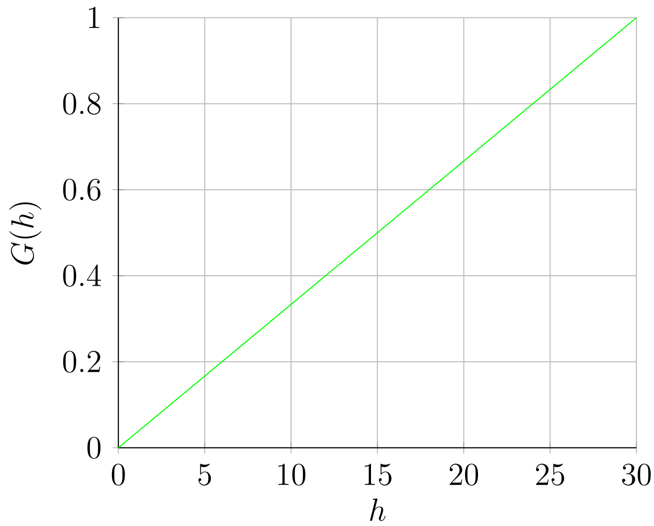

Figure 2: Cumulative Distribution Function (cdf) $G(h)$ for H

Thus,

$$P(0 \leq H \leq 10) = \frac{40(10) - \frac{1}{2}(10^2)}{750} \approx 0.466$$

Problem 9: distribution of random variables X and Y

Task 1: covariance (X, Y), and the correlation Cor(X, Y).

Solution: To calculate the joint distribution of random variables X and Y,use the probabilities provided for each value:

X and Y	$Y = 1$	$Y = -1$	$MarginalSum$
$X = 1$	c	$\frac{1}{2} - c$	$\frac{1}{2}$
$X = -1$	$\frac{1}{2} - c$	c	$\frac{1}{2}$
$MarginalSum$	$\frac{1}{2}$	$\frac{1}{2}$	1

Now, let's calculate the covariance (Cov(X, Y)):

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

First, calculate E[XY]:

$$E[XY] = 1 \cdot P(X = -1, Y = -1) + (-1) \cdot P(X = -1, Y = 1) + (-1) \cdot P(X = 1, Y = -1) + (-1) \cdot P(X = 1, Y = 1)$$

Calculate the expectations:

$$E[X] = 1 \cdot P(X = 1) + (-1) \cdot P(X = -1) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot -1 = 0$$

$$E[Y] = 1 \cdot P(Y = 1) + (-1) \cdot P(Y = -1) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot -1 = 0$$

Plug values into the covariance formula:

$$Cov(X,Y) = c + (\frac{1}{2} - c) \cdot -1 + (\frac{1}{2} - c) \cdot -1 + c$$
$$Cov(X,Y) = 4c - 1$$

calculate the correlation (Cor(X, Y)):

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

Since X and Y are discrete random variables, you can calculate their variances as follows:

$$Var(X) = E[X^2] - (E[X])^2$$
$$E[X^2] = \sigma(X^2)p(x) = -1^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1$$
$$Var(X) = 1 - 0 = 1$$

The variance of Y can also be calculated in the same manner.

$$Var(Y) = E[Y^2] - (E[Y])^2 = 1$$

$$Cor(X,Y) = \frac{Cov(X,Y)}{1 \cdot 1} = 4c - 1$$

Task 2: Determine the values of c

Solution: determine the values of c for which X and Y are independent and for which they are 100% correlated.
1. Independent (Uncorrelated) Variables: X and Y are independent then :

$$P(X \cup Y) = P(X) \cdot P(Y)$$
$$\text{for } (X = 1, Y = 1)$$
$$P(X = 1, Y = 1) = c$$
$$P(X = 1) \cdot P(Y = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$c = \frac{1}{4}$$

2. 100% Correlated Variables: When two variables, X and Y, have a correlation coefficient of 1, it means that they are perfectly correlated. To determine the values that will make X and Y 100% correlated, you can set Cor(X, Y) = 1. From our previous calculation :

$$Cor(X,Y) = 4c - 1$$
$$4c - 1 = 1 \text{ then, } C = \frac{1}{2}$$

Problem 11: Random variable

Task 1:Random variable

Solution:

- The sample means: This is a random variable. The value of $\bar{x}$ will change from sample to sample.
- The largest value in the sample: This is a random variable, and the value can change from one sample to the next.

Problem 10, 12 and 13: Python code

Solution: I have attached the .py code and a PDF report for the codes.