Question 01

> Assume that we roll two fair six-sided dice. Let E be the event that the two dice's outcomes sum to 8. What is the probability of E?

The sample space of the experiment is the set of all possible outcomes of rolling two dice. Since each die has 6 possible outcomes, there are 6 outcomes for the first die and 6 outcomes for the second die. Therefore, the total number of possible outcomes is:

$$|\Omega| = 6 \times 6 = 36$$

The event (E) is the set of outcomes where the sum of the two dice is 8. The following outcomes satisfy this condition:

$$E = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$

Thus, there are 5 favorable outcomes in (E). Therefore, the probability of (E) is:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{5}{36}$$

Question 02

Continuing with question 1: Initialize the random seed to 2024 using `numpy.random.seed`. Using `numpy.random.randint`, simulate $1,000$ throws of two fair six-sided dice. Paste your code here.

From these simulations, what is the empirical frequency of E (i.e., the percentage of times this event occurred in simulation)?

```
import numpy as np

np.random.seed(2024)
n = 1000
dice = np.random.randint(1, 7, (n, 2))
E = (dice.sum(axis=1) == 8).sum()
P_E = E / n
P_E
# >> 0.144
```

From these simulations, the empirical frequency of E is 0.144.

Question 03

> Continuing with question 2: Reset the random seed to 2024 and repeat the above simulation a total of 10 times and report the empirical frequency of E for each of the 10 runs. Paste your code here.
>
> The empirical frequency of E from each simulation will differ. Why do these numbers differ?
>
> Yet, the probability of E is fixed and was calculated in part (a) above. Why does the probability disagree with the empirical frequencies?

```python
import numpy as np
np.random.seed(2024)

def simulate(n):
    dice = np.random.randint(1, 7, (n, 2))
    E = (dice.sum(axis=1) == 8).sum()
    return E / n

frequencies = [simulate(1000) for _ in range(10)]
frequencies
# >> [0.144, 0.147, 0.127, 0.144, 0.123, 0.133, 0.15, 0.13, 0.127, 0.153]
```

The empirical frequencies of E from each simulation differ because the simulation is based on random outcomes. The probability of E is fixed and was calculated in part (a) above. The probability disagrees with the empirical frequencies because the empirical frequencies are based on a finite number of simulations. As the number of simulations increases, the empirical frequencies will converge to the true probability of E.

Question 04

Recall that A, B are independent if $P(A \cap B) = P(A)P(B)$.

Consider tossing a fair die. Let $A = \{2, 4, 6\}$, $B = \{1, 2, 3, 4\}$. Simulate draws from the sample space and verify whether the frequencies verify independence of A, B or not. Paste your code.

Also, verify whether the events are independent or not theoretically.

Empirical Verification:

```
import numpy as np

np.random.seed(2024)
n = 1000000
dice = np.random.randint(1, 7, n)

A = np.isin(dice, [2, 4, 6])
B = np.isin(dice, [1, 2, 3, 4])

P_A = A.sum() / n
P_B = B.sum() / n
P_A_and_B = (A & B).sum() / n

P_A, P_B, P_A_and_B, P_A * P_B
# >> (0.499416, 0.666206, 0.332734, 0.332713935696)
```

The simulated frequencies of A, B, and $A \cap B$ are approximately 0.499, 0.666, 0.333, and 0.333, respectively. This suggests that the events A and B are independent.

Theoretical Verification:

For a sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$, the events $A = \{2, 4, 6\}$ and $B = \{1, 2, 3, 4\}$ are independent if $P(A \cap B) = P(A)P(B)$.

- For $A = \{2, 4, 6\}$, $P(A) = \frac{3}{6} = \frac{1}{2}$
- For $B = \{1, 2, 3, 4\}$, $P(B) = \frac{4}{6} = \frac{2}{3}$
- For $A \cap B = \{2, 4\}$, $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$

Using these probabilities, we can check whether the events A and B are independent or not by checking whether $P(A \cap B) = P(A)P(B)$:

$$P(A \cap B) = \frac{1}{3}$$

$$P(A)P(B) = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$$

Since $P(A \cap B) = P(A)P(B)$, the events A and B are independent.

Question 05

Repeat question 4 for $A = \{2, 3, 4, 6\}$, $B = \{1, 2, 3, 4\}$

Empirical Verification:

```
import numpy as np

np.random.seed(2024)
n = 1000000
dice = np.random.randint(1, 7, n)

A = np.isin(dice, [2, 3, 4, 6])
B = np.isin(dice, [1, 2, 3, 4])

P_A = A.sum() / n
P_B = B.sum() / n
P_A_and_B = (A & B).sum() / n

P_A, P_B, P_A_and_B, P_A * P_B
# >> (0.66637, 0.666206, 0.499688, 0.44393969222)
```

The simulated frequencies of A, B, and $A \cap B$ are approximately 0.666, 0.666, 0.500, and 0.444, respectively. This suggests that the events A and B are not independent.

Theoretical Verification:

For a sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$, the events $A = \{2, 3, 4, 6\}$ and $B = \{1, 2, 3, 4\}$ are independent if $P(A \cap B) = P(A)P(B)$.

- For $A = \{2, 3, 4, 6\}$, $P(A) = \frac{4}{6} = \frac{2}{3}$
- For $B = \{1, 2, 3, 4\}$, $P(B) = \frac{4}{6} = \frac{2}{3}$
- For $A \cap B = \{2, 3, 4\}$, $P(A \cap B) = \frac{3}{6} = \frac{1}{2}$

Using these probabilities, we can check whether the events A and B are independent or not by checking whether $P(A \cap B) = P(A)P(B)$:

$$P(A \cap B) = \frac{1}{2}$$

$$P(A)P(B) = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$$

Since $P(A \cap B) \neq P(A)P(B)$, the events A and B are not independent.

Question 06

> Use Bayes' Theorem to solve the following problem.
>
> Suppose that 80 percent of all computer scientists are shy, whereas only 15 percent of all data scientists are shy. Suppose also that 80 percent of the people at a large gathering are computer scientists and the other 20 percent are data scientists. If you meet a shy person at random at the gathering, what is the probability that the person is a computer scientist?

We are tasked with finding the probability of $A =$ Computer Scientist given that $B =$ Shy. We can use Bayes' Theorem to solve this problem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The probability of a person being shy given that they are a computer scientist is $P(B|A) = 0.8$. The probability of a person being a computer scientist is $P(A) = 0.8$.

By the Law of Total Probability, we can find the probability of a person being shy:

$$P(B) = P(B|A)P(A)+P(B|A^c)P(A^c) = 0.8\times0.8+0.15\times0.2 = 0.67$$

Therefore, the probability of a person being a computer scientist given that they are shy is:

$$P(A|B) = \frac{0.8 \times 0.8}{.8 \times 0.8 + 0.15 \times 0.2} = \frac{0.64}{0.67} \approx 0.9552$$

Question 07

> Suppose that two players Alice and Bob take turns rolling a pair of balanced dice and that the winner is the first player who obtains the sum of 6 on a given roll of the two dice. If Alice rolls first, what is the probability that Bob will win?

Let A be the event that Alice wins and B be the event that Bob wins. We are interested in finding $P(B)$, the probability that Bob wins.

Step 1: Analyze the Dice Rolls

For any given roll, the sample space Ω consists of all possible outcomes of rolling two dice. Since each die has 6 possible outcomes, the total number of outcomes is:

$$|\Omega| = 6 \times 6 = 36$$

The set of outcomes where the sum of the two dice is 6 is:

$$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

There are 5 favorable outcomes. Therefore, the probability that a player rolls a sum of 6 on any given roll is:

$$P(\text{rolling a 6}) = \frac{5}{36}$$

The probability of not rolling a 6 is:

$$P(\text{not rolling a 6}) = 1 - \frac{5}{36} = \frac{31}{36}$$

Step 2: First Round

Alice rolls first. The probability that Alice wins on her first roll is:

$$P(A) = \frac{5}{36}$$

If Alice does not win, Bob rolls next. The probability that Bob wins on his first roll, given that Alice did not win, is:

$$P(\text{Bob wins on first roll}) = \frac{31}{36} \times \frac{5}{36}$$

Step 3: Recursive Process

If neither Alice nor Bob wins in the first round, the process resets, and the game proceeds with Alice rolling again. The probability that neither Alice nor Bob wins in the first round is:

$$P(\text{Neither wins on first roll}) = \frac{31}{36} \times \frac{31}{36}$$

After this, the game repeats with the same structure as before, but now both players are one round further into the game. Therefore, we can define the probability that Bob wins recursively.

Let P_B be the probability that Bob eventually wins. We can express it as:

$$P_B = P(\text{Bob wins on first roll}) + P(\text{Neither wins in first round}) \times P_B$$

Substitute the values:

$$P_B = \frac{31}{36} \times \frac{5}{36} + \frac{31}{36} \times \frac{31}{36} \times P_B$$

Step 4: Solve the Equation

To solve for P_B, we rearrange the equation:

$$P_B = \frac{31}{36} \times \frac{5}{36} + \frac{31}{36} \times \frac{31}{36} \times P_B$$

Rearranging:

$$P_B - \frac{31}{36} \times \frac{31}{36} \times P_B = \frac{31}{36} \times \frac{5}{36}$$

Factor out P_B:

$$P_B \left(1 - \frac{31}{36} \times \frac{31}{36}\right) = \frac{31}{36} \times \frac{5}{36}$$

Simplify the left-hand side:

$$P_B \times \frac{215}{1296} = \frac{31}{36} \times \frac{5}{36}$$

Solve for P_B:

$$P_B = \frac{\frac{31}{36} \times \frac{5}{36}}{\frac{215}{1296}} = \frac{31 \times 5 \times 1296}{36 \times 36 \times 215} = \frac{31 \times 5}{36 \times 215} = \frac{155}{7740} = \frac{31}{155}$$

Final Answer The probability that Bob wins, given that Alice rolls first, is:

$$P(B) = \frac{31}{155} \approx 0.2$$

Question 08

Suppose that a box contains 8 red balls and 2 blue balls. If five balls are selected at random, without replacement, determine the probability mass function of the number of red balls that will be obtained.

Step 1: Define the Random Variable

Let the random variable X represent the number of red balls obtained in the sample of 5 balls. Since the sample is drawn without replacement, X follows a **hypergeometric distribution** because: - There is a finite population of 10 balls (8 red and 2 blue). - The sample size is 5. - The draws are without replacement, meaning that each selected ball affects the subsequent probabilities.

The number of red balls in the sample can range from 3 to 5, because we must draw at least 3 red balls (since there are only 2 blue balls) and can draw at most 5 red balls (the entire sample).

Step 2: Formula for Hypergeometric Distribution

The probability mass function (PMF) for a hypergeometric distribution is given by the formula:

$$P(X = k) = \frac{\binom{8}{k}\binom{2}{5-k}}{\binom{10}{5}}$$

Where: - $\binom{8}{k}$ represents the number of ways to choose k red balls from the 8 available red balls. - $\binom{2}{5-k}$ represents the number of ways to choose $5-k$ blue balls from the 2 available blue balls. - $\binom{10}{5}$ represents the total number of ways to choose 5 balls from the 10 total balls.

Step 3: Calculate the Denominator

First, calculate the total number of ways to choose 5 balls from 10:

$$\binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} = 252$$

Step 4: Calculate the Numerator for Each Value of k

Now, calculate the probability for each possible value of k, where k represents the number of red balls in the sample. The possible values of k are 3, 4, and 5.

Case 1: $X = 3$ (3 red balls) For $k = 3$:

$$P(X = 3) = \frac{\binom{8}{3}\binom{2}{2}}{\binom{10}{5}}$$

Where: - $\binom{8}{3} = \frac{8\times7\times6}{3\times2\times1} = 56$ - $\binom{2}{2} = 1$

Thus:

$$P(X = 3) = \frac{56 \times 1}{252} = \frac{56}{252} = \frac{2}{9}$$

Case 2: $X = 4$ (4 red balls) For $k = 4$:

$$P(X = 4) = \frac{\binom{8}{4}\binom{2}{1}}{\binom{10}{5}}$$

Where: - $\binom{8}{4} = \frac{8\times7\times6\times5}{4\times3\times2\times1} = 70$ - $\binom{2}{1} = 2$

Thus:

$$P(X = 4) = \frac{70 \times 2}{252} = \frac{140}{252} = \frac{5}{9}$$

Case 3: $X = 5$ (5 red balls) For $k = 5$:

$$P(X = 5) = \frac{\binom{8}{5}\binom{2}{0}}{\binom{10}{5}}$$

Where: - $\binom{8}{5} = \frac{8\times7\times6}{3\times2\times1} = 56$ - $\binom{2}{0} = 1$

Thus:

$$P(X = 5) = \frac{56 \times 1}{252} = \frac{56}{252} = \frac{2}{9}$$

Step 5: Final Probability Mass Function (PMF)

The probability mass function for X, the number of red balls in the sample, is:

$$P(X = k) = \begin{cases} \frac{2}{9} & \text{if } k = 3 \\ \frac{5}{9} & \text{if } k = 4 \\ \frac{2}{9} & \text{if } k = 5 \\ 0 & \text{otherwise} \end{cases}$$

Conclusion:

The probability mass function for the number of red balls selected when five balls are chosen at random from a box containing 8 red balls and 2 blue balls is given by:

$$P(X = 3) = \frac{2}{9}, \quad P(X = 4) = \frac{5}{9}, \quad P(X = 5) = \frac{2}{9}$$

Question 09

Given that the pdf of X for some constant c is:

$$f(x) = \begin{cases} cx^2 & \text{for } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find c. The total area under the probability density function (pdf) curve must be equal to 1. Therefore, we can find the constant c by integrating the pdf over the entire range of x and setting it equal to 1:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Since the pdf is zero outside the interval $[1, 2]$, we can integrate the pdf over this interval:

$$\int_{1}^{2} cx^2 dx = 1$$

Integrating cx^2 with respect to x:

$$c\int_{1}^{2} x^2 dx = 1$$

$$c\left[\frac{x^3}{3}\right]_1^2 = 1$$

$$c\left(\frac{2^3}{3} - \frac{1^3}{3}\right) = 1$$

$$c\left(\frac{8}{3} - \frac{1}{3}\right) = 1$$

$$c \times \frac{7}{3} = 1$$

$$c = \frac{3}{7}$$

(b) Find $P(X > 1.75)$. The probability that X is greater than 1.75 is the area under the pdf curve to the right of $x = 1.75$. This can be calculated by integrating the pdf from 1.75 to 2:

$$P(X > 1.75) = \int_{1.75}^{2} f(x)dx$$

Substitute the pdf $f(x) = \frac{3}{7}x^2$:

$$P(X > 1.75) = \int_{1.75}^{2} \frac{3}{7}x^2 dx$$

Integrating $\frac{3}{7}x^2$ with respect to x:

$$\frac{3}{7}\int_{1.75}^{2} x^2 dx$$

$$\frac{3}{7}\left[\frac{x^3}{3}\right]_{1.75}^{2}$$

$$\frac{3}{7}\left(\frac{2^3}{3} - \frac{1.75^3}{3}\right)$$

$$\frac{3}{7}\left(\frac{8}{3} - \frac{1.75^3}{3}\right)$$

$$\frac{3}{7}\left(\frac{2.9375}{3}\right)$$

$$\frac{8.8125}{21} \approx 0.4196$$

Question 10

Given that the pdf of X is:

$$f(x) = \begin{cases} \frac{1}{4} & \text{for } 0 < x < 1 \\ \frac{3}{8} & \text{for } 3 < x < 5 \\ 0 & \text{otherwise} \end{cases}$$

Find the cdf of X.

The cumulative distribution function (CDF) of a continuous random variable X is defined as:

$$F(x) = P(X \leq x)$$

To find the CDF of X, we need to integrate the probability density function (pdf) of X over the range $(-\infty, x]$ for each value of x.

For $x \leq 0$:

Since the pdf is zero for $x \leq 0$, the CDF is also zero for $x \leq 0$:

$$F(x) = 0 \text{ for } x \leq 0$$

For $0 < x \leq 1$:

For $0 < x \leq 1$, the pdf is $\frac{1}{4}$. The CDF for this range is the integral of the pdf from 0 to x:

$$F(x) = \int_0^x \frac{1}{4} dx = \frac{1}{4} x \text{ for } 0 < x \leq 1$$

For $1 < x \leq 3$:

Since the pdf is zero for $1 < x \leq 3$, the CDF remains constant at $\frac{1}{4}$ for $1 < x \leq 3$:

$$F(x) = \frac{1}{4} \text{ for } 1 < x \leq 3$$

For $3 < x \leq 5$:

For $3 < x \leq 5$, the pdf is $\frac{3}{8}$. The CDF for this range is the integral of the pdf from 0 to x:

$$F(x) = \int_3^x \frac{3}{8} dx = \frac{3}{8}x - \frac{3}{8} \times 3 = \frac{3}{8}x - \frac{9}{8} \text{ for } 3 < x \leq 5$$

For $x > 5$:

Since the pdf is zero for $x > 5$, the CDF remains constant at 1 for $x > 5$:

$$F(x) = 1 \text{ for } x > 5$$

Conclusion:

The cumulative distribution function (CDF) of the random variable X is:

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{1}{4}x & \text{for } 0 < x \leq 1 \\ \frac{1}{4} & \text{for } 1 < x \leq 3 \\ \frac{3}{8}x - \frac{9}{8} & \text{for } 3 < x \leq 5 \\ 1 & \text{for } x > 5 \end{cases}$$