COMP 2804 — Solutions Assignment 4

Question 1: Write your name and student number.

Solution:

- Name: Saki Kumagai
- Student number: 5

Question 2: Alexa[1], Tri[2], and Zoltan[3] play the OddPlayer game: In one round, each player flips a fair coin.

1. Assume that not all flips are equal. Then the coin flips of exactly two players are equal. The player whose coin flip is different is called the *odd player*. In this case, the odd player wins the game. For example, if Alexa flips tails, Tri flips heads, and Zoltan flips tails, then Tri is the odd player and wins the game.
2. If all three coin flips are equal, then the game is repeated.

Below, this game is presented in pseudocode:

Algorithm OddPlayer**:**

```
    // all coin flips made are mutually independent
    each player flips a fair coin;
    if not all coin flips are equal
    then the game terminates and the odd player wins
    else OddPlayer
    endif
```

- What is the sample space?
- Define the event

$$A = \text{“Alexa wins the game”}.$$

  Express this event as a subset of the sample space.
- Use your expression from the previous part to determine $\Pr(A)$.
- Use symmetry to determine $\Pr(A)$. Explain your answer in plain English and a few sentences.

  Hint: What is the probability that Tri wins the game? What is the probability that Zoltan wins the game?

[1] your friendly TA
[2] another friendly TA
[3] yet another friendly TA

Solution: We will use the sets

$$R = \{HHH, TTT\}$$

and

$$D = \{HHT, HTH, THH, HTT, THT, TTH\},$$

where R stands for "repeat" and D stands for "done".

In one round, each player flips a fair coin. This is equivalent to choosing a uniformly random element in the set $R \cup D$. For example, if the chosen element is HTT, then Alexa flipped heads, Tri flipped tails, and Zoltan flipped tails.

If the chosen element belongs to D, then there is an odd player, and this player wins the game. Otherwise, the chosen element belongs to R, in which case the game is repeated. Based on this this, the sample space is the set

$$S = \{F_1 F_2 \cdots F_n F_{n+1} : n \geq 0, F_1, \ldots, F_n \in R, F_{n+1} \in D\}.$$

We know the winner of the game as soon as an element of D is chosen. Alexa wins if and only if this element is one of THH and HTT. Based on this, the event A that Alexa wins the game is equal to

$$A = \{F_1 F_2 \cdots F_n F_{n+1} : n \geq 0, F_1, \ldots, F_n \in R, F_{n+1} \in \{THH, HTT\}\}.$$

In one round, the probability of choosing an element of R is equal $2/8 = 1/4$ and the probability of choosing an element of $\{THH, HTT\}$ is also equal to $2/8 = 1/4$. Since the rounds are independent, we have, for any $n \geq 0$, $F_1, \ldots, F_n \in R$, $F_{n+1} \in \{THH, HTT\}\}$,

$$\Pr(F_1 F_2 \cdots F_n F_{n+1}) = (1/4)^{n+1}.$$

Remember from class that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

for any real number x with $0 < x < 1$. This gives,

$$\begin{aligned}
\Pr(A) &= \sum_{n=0}^{\infty} \Pr(F_1 F_2 \cdots F_n F_{n+1}) \\
&= \sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^{n+1} \\
&= \frac{1}{4} \sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^{n} \\
&= \frac{1}{4} \cdot \frac{1}{1 - 1/4} \\
&= \frac{1}{3}.
\end{aligned}$$

For the last part of this question, here is an easier way to show that $\Pr(A) = 1/3$. We define the two additional events

$$T = \text{“Tri wins the game”}$$

and

$$Z = \text{“Zoltan wins the game”}.$$

By symmetry,

$$\Pr(A) = \Pr(T) = \Pr(Z).$$

Since there always is exactly one winner, we have

$$\Pr(A) + \Pr(T) + \Pr(Z) = 1.$$

This implies that $\Pr(A) = 1/3$.

Question 3: Consider the set $S = \{2, 3, 5, 30\}$. We choose a uniformly random element x from this set. Define the random variables

$$\begin{aligned} X &= \begin{cases} 1 & \text{if } x \text{ is divisible by } 2, \\ 0 & \text{otherwise,} \end{cases} \\ Y &= \begin{cases} 1 & \text{if } x \text{ is divisible by } 3, \\ 0 & \text{otherwise,} \end{cases} \\ Z &= \begin{cases} 1 & \text{if } x \text{ is divisible by } 5, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

- Is the sequence X, Y, Z of random variables pairwise independent? As always, justify your answer.
- Is the sequence X, Y, Z of random variables mutually independent? As always, justify your answer.

Solution: We first note that

$$X = 1 \text{ if and only if } x \in \{2, 30\},$$

$$Y = 1 \text{ if and only if } x \in \{3, 30\},$$

$$Z = 1 \text{ if and only if } x \in \{5, 30\}.$$

Therefore,

$$\Pr(X = 1) = \Pr(X = 0) = 1/2,$$

$$\Pr(Y = 1) = \Pr(Y = 0) = 1/2,$$

$$\Pr(Z = 1) = \Pr(Z = 0) = 1/2.$$

We are going to show that the sequence X, Y, Z is pairwise independent. For this, we have to verify lots of equations, but they are all more or less the same:

$$\Pr(X = 0 \wedge Y = 0) = \Pr(x = 5) = 1/4 = \Pr(X = 0) \cdot \Pr(Y = 0),$$

$$\Pr(X = 0 \wedge Y = 1) = \Pr(x = 3) = 1/4 = \Pr(X = 0) \cdot \Pr(Y = 1),$$

$$\Pr(X = 1 \wedge Y = 0) = \Pr(x = 2) = 1/4 = \Pr(X = 1) \cdot \Pr(Y = 0),$$

$$\Pr(X = 1 \wedge Y = 1) = \Pr(x = 30) = 1/4 = \Pr(X = 1) \cdot \Pr(Y = 1).$$

The four equations above imply that X and Y are independent.

$$\Pr(X = 0 \wedge Z = 0) = \Pr(x = 3) = 1/4 = \Pr(X = 0) \cdot \Pr(Z = 0),$$

$$\Pr(X = 0 \wedge Z = 1) = \Pr(x = 5) = 1/4 = \Pr(X = 0) \cdot \Pr(Z = 1),$$

$$\Pr(X = 1 \wedge Z = 0) = \Pr(x = 2) = 1/4 = \Pr(X = 1) \cdot \Pr(Z = 0),$$

$$\Pr(X = 1 \wedge Z = 1) = \Pr(x = 30) = 1/4 = \Pr(X = 1) \cdot \Pr(Z = 1).$$

The four equations above imply that X and Z are independent.

$$\Pr(Y = 0 \wedge Z = 0) = \Pr(x = 2) = 1/4 = \Pr(Y = 0) \cdot \Pr(Z = 0),$$

$$\Pr(Y = 0 \wedge Z = 1) = \Pr(x = 5) = 1/4 = \Pr(Y = 0) \cdot \Pr(Z = 1),$$

$$\Pr(Y = 1 \wedge Z = 0) = \Pr(x = 3) = 1/4 = \Pr(Y = 1) \cdot \Pr(Z = 0),$$

$$\Pr(Y = 1 \wedge Z = 1) = \Pr(x = 30) = 1/4 = \Pr(Y = 1) \cdot \Pr(Z = 1).$$

The four equations above imply that Y and Z are independent.

Is the sequence X, Y, Z mutually independent? We observe the following: If $X = 1$ and $Y = 1$, then $x = 30$, which implies that $Z = 1$. This suggests that they are not mutually independent. This is correct, because

$$\Pr(X = 1 \wedge Y = 1 \wedge Z = 1) = \Pr(x = 30) = 1/4,$$

whereas

$$\Pr(X = 1) \cdot \Pr(Y = 1) \cdot \Pr(Z = 1) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8.$$

Question 4: Let a and b be real numbers. You flip a fair and independent coin three times. For $i = 1, 2, 3$, let

$$f_i = \begin{cases} a & \text{if the } i\text{-th coin flip results in heads,} \\ b & \text{if the } i\text{-th coin flip results in tails.} \end{cases}$$

Define the random variables

$$\begin{aligned} X &= f_1 \cdot f_2, \\ Y &= f_2 \cdot f_3. \end{aligned}$$

For each of the following questions, justify your answer.

- Assume that $a = b$. Are the random variables X and Y independent?
- Assume that $a = 0$ and $b \neq a$. Are the random variables X and Y independent?
- Assume that $a \neq 0$ and $b = -a$. Are the random variables X and Y independent?
- Assume that $a \neq 0$, $b \neq 0$, $a \neq b$, and $b \neq -a$. Are the random variables X and Y independent?

Solution:

- Assume that $a = b$.

  The following table gives the different possibilities:

f_1	f_2	X	f_2	f_3	Y
a	a	a^2	a	a	a^2
a	a	a^2	a	a	a^2
a	a	a^2	a	a	a^2
a	a	a^2	a	a	a^2

  We see that both X and Y are always equal to a^2. Because of this, in order to determine if X and Y are independent, we have to verify only one equation:

$$\Pr(X = a^2 \wedge Y = a^2) = 1 = 1 \cdot 1 = \Pr(X = a^2) \cdot \Pr(Y = a^2).$$

  Thus, X and Y are independent.

- Assume that $a = 0$ and $b \neq a$.

  The following table gives the different possibilities:

f_1	f_2	X	f_2	f_3	Y
0	0	0	0	0	0
0	b	0	0	b	0
b	0	0	b	0	0
b	b	b^2	b	b	b^2

  In this case, X is equal to b^2 with probability $1/4$, and equal to 0 with probability $3/4$. The same is true for Y. We have

$$\begin{aligned}
\Pr(X = b^2 \wedge Y = b^2) &= \Pr(f_1 = b \wedge f_2 = b \wedge f_3 = b) \\
&= \Pr(f_1 = b) \cdot \Pr(f_2 = b) \cdot \Pr(f_3 = b) \\
&= 1/2 \cdot 1/2 \cdot 1/2 \\
&= 1/8,
\end{aligned}$$

whereas

$$\Pr(X = b^2) \cdot \Pr(Y = b^2) = 1/4 \cdot 1/4 = 1/16.$$

Therefore, X and Y are not independent.

- Assume that $a \neq 0$ and $b = -a$.

  The following table gives the different possibilities:

f_1	f_2	X	f_2	f_3	Y
a	a	a^2	a	a	a^2
a	$-a$	$-a^2$	a	$-a$	$-a^2$
$-a$	a	$-a^2$	$-a$	a	$-a^2$
$-a$	$-a$	a^2	$-a$	$-a$	a^2

  We have

$$\begin{aligned} & \Pr(X = a^2 \wedge Y = a^2) \\ = \; & \Pr((f_1, f_2, f_3) = (a, a, a) \vee (f_1, f_2, f_3) = (-a, -a, -a)) \\ = \; & 2/8 \\ = \; & 1/4, \end{aligned}$$

  whereas

$$\Pr(X = a^2) \cdot \Pr(Y = a^2) = 1/2 \cdot 1/2 = 1/4.$$

  Thus,

$$\Pr(X = a^2 \wedge Y = a^2) = \Pr(X = a^2) \cdot \Pr(Y = a^2).$$

  The exact same reasoning shows that

$$\Pr(X = a^2 \wedge Y = -a^2) = \Pr(X = a^2) \cdot \Pr(Y = -a^2),$$

$$\Pr(X = -a^2 \wedge Y = a^2) = \Pr(X = -a^2) \cdot \Pr(Y = a^2),$$

  and

$$\Pr(X = -a^2 \wedge Y = -a^2) = \Pr(X = -a^2) \cdot \Pr(Y = -a^2).$$

  Therefore, X and Y are independent.

- Assume that $a \neq 0$, $b \neq 0$, $a \neq b$, and $b \neq -a$.

  The following table gives the different possibilities:

f_1	f_2	X	f_2	f_3	Y
a	a	a^2	a	a	a^2
a	b	ab	a	b	ab
b	a	ab	b	a	ab
b	b	b^2	b	b	b^2

Note that a^2, ab, and b^2 are three different numbers.

We have
$$\Pr(X = a^2 \wedge Y = b^2) = 0,$$
whereas
$$\Pr(X = a^2) \cdot \Pr(Y = b^2) = 1/4 \cdot 1/4 \neq 0.$$
Therefore, X and Y are not independent.

Question 5: You are given three fair dice. One die is red, one die is blue, and one die is green. You roll each die once, independently of the other two dice. Define the random variables
$$\begin{aligned} X_r &= \text{the value of the red die,} \\ X_b &= \text{the value of the blue die,} \\ X_g &= \text{the value of the green die,} \\ Y &= \max(X_r, X_b, X_g), \end{aligned}$$
and, for each integer k with $1 \leq k \leq 6$, the events
$$\begin{aligned} A_k &= \text{“}Y = k\text{”}, \\ B_k &= \text{“}Y \leq k\text{”}. \end{aligned}$$

- Determine the expected values $\mathbb{E}(X_r)$, $\mathbb{E}(X_b)$, and $\mathbb{E}(X_g)$.

  Solution: We have seen in class that
  $$\mathbb{E}(X_r) = \mathbb{E}(X_b) = \mathbb{E}(X_g) = 7/2.$$

- Let k be an integer with $1 \leq k \leq 6$. Determine $\Pr(B_k)$.

  Solution: We observe that
  $$B_k \text{ if and only if } X_r \leq k \text{ and } X_b \leq k \text{ and } X_g \leq k.$$
  Since the rolls of the dice are independent, we get
  $$\begin{aligned} \Pr(B_k) &= \Pr(X_r \leq k \wedge X_b \leq k \wedge X_g \leq k) \\ &= \Pr(X_r \leq k) \cdot \Pr(X_b \leq k) \cdot \Pr(X_g \leq k) \\ &= (k/6) \cdot (k/6) \cdot (k/6) \\ &= (k/6)^3. \end{aligned}$$

- Determine $\Pr(A_1)$.

  Solution: Since $A_1 = B_1$, we have
  $$\Pr(A_1) = (1/6)^3.$$

- Let k be an integer with $2 \leq k \leq 6$. Express the event B_k in terms of the events A_k and B_{k-1}.

  Solution: Since $Y \leq k$ if and only if $Y = k$ or $Y \leq k-1$, we have

  $$B_k \text{ if and only if } A_k \vee B_{k-1}.$$

- Let k be an integer with $2 \leq k \leq 6$. Determine $\Pr(A_k)$.

  Solution: From the previous part, we get

  $$\Pr(B_k) = \Pr(A_k \vee B_{k-1}).$$

  Since A_k and B_{k-1} are disjoint, we get

  $$\Pr(B_k) = \Pr(A_k) + \Pr(B_{k-1}).$$

  Therefore,

  $$\begin{aligned} \Pr(A_k) &= \Pr(B_k) - \Pr(B_{k-1}) \\ &= \left(\frac{k}{6}\right)^3 - \left(\frac{k-1}{6}\right)^3. \end{aligned}$$

  Note that this is also true if $k = 1$.

- Determine the expected value $\mathbb{E}(Y)$ of the random variable Y.

  Solution: The random variable Y can take any of the values $1, 2, 3, 4, 5, 6$. Therefore,

  $$\begin{aligned} \mathbb{E}(Y) &= \sum_{k=1}^{6} k \cdot \Pr(Y = k) \\ &= \sum_{k=1}^{6} k \cdot \Pr(A_k) \\ &= \sum_{k=1}^{6} k \cdot \left(\left(\frac{k}{6}\right)^3 - \left(\frac{k-1}{6}\right)^3 \right). \end{aligned}$$

  Using Wolfram Alpha, we get

  $$\mathbb{E}(Y) = \frac{119}{24} = 4.958333333\cdots$$

- Is the following true or false?

  $$\mathbb{E}(\max(X_r, X_b, X_g)) = \max(\mathbb{E}(X_r), \mathbb{E}(X_b), \mathbb{E}(X_g)).$$

Solution: The left-hand side is

$$\mathbb{E}\left(\max\left(X_r, X_b, X_g\right)\right) = \mathbb{E}(Y) = 119/24,$$

whereas the right-hand side is

$$\max\left(\mathbb{E}\left(X_r\right), \mathbb{E}\left(X_b\right), \mathbb{E}\left(X_g\right)\right) = \max(7/2, 7/2, 7/2) = 7/2.$$

Thus, the left-hand side is not equal to the right-hand side.

Question 6: Let $n \geq 1$ be an integer and let $A[1 \ldots n]$ be an array that stores a permutation of the set $\{1, 2, \ldots, n\}$. If the array A is sorted, then $A[k] = k$ for $k = 1, 2, \ldots, n$ and, thus,

$$\sum_{k=1}^{n} |A[k] - k| = 0. \tag{1}$$

If the array A is not sorted and $A[k] = i$, where $i \neq k$, then $|A[k]-k|$ is equal to the "distance" that the value i must move in order to make the array sorted. Thus, the summation in (1) is a measure for the "sortedness" of the array A: If the summation is small, then A is "close" to being sorted. On the other hand, if the summation is large, then A is "far away" from being sorted. In this exercise, you will determine the expected value of the summation in (1).

Assume that the array stores a uniformly random permutation of the set $\{1, 2, \ldots, n\}$. For each $k = 1, 2, \ldots, n$, define the random variable

$$X_k = |A[k] - k|,$$

and let

$$X = \sum_{k=1}^{n} X_k.$$

- Assume that $n = 1$. Determine the expected value $\mathbb{E}(X)$.
- Assume that $n \geq 2$. Is the sequence $X_1, X_2, \ldots, X_n$ of random variables pairwise independent?
- Assume that $n \geq 1$. Let k be an integer with $1 \leq k \leq n$. Prove that

$$\mathbb{E}\left(X_k\right) = \frac{n+1}{2} + \frac{k^2 - k - kn}{n}.$$

Hint: Assume $A[k] = i$. If $1 \leq i \leq k$, then $|A[k] - k| = k - i$. If $k + 1 \leq i \leq n$, then $|A[k] - k| = i - k$. For any integer $m \geq 1$,

$$1 + 2 + 3 + \cdots + m = \frac{m(m+1)}{2}.$$

- Assume that $n \geq 1$. Prove that

$$\mathbb{E}(X) = \frac{n^2 - 1}{3}.$$

Hint:

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Solution: Before we start, consider two integers i and k with $1 \leq i \leq n$ and $1 \leq k \leq n$. What is $\Pr(A[k] = i)$, i.e., the probability that in a uniformly random permutation of $\{1, 2, \ldots, n\}$, the value i is at position k? Since the permutation is random, the value i has the same probability to be in any of the positions $1, 2, \ldots, n$. Therefore,

$$\Pr(A[k] = i) = 1/n.$$

A more formal proof goes as follows: There are $n!$ many possible permutations. In exactly $(n-1)!$ of these, the value i is at position k. Therefore,

$$\Pr(A[k] = i) = (n-1)!/n! = 1/n.$$

If $n = 1$, then the array A stores only the number 1 and it is at position 1. In this case, we always have

$$X = X_1 = |A[1] - 1| = |1 - 1| = 0.$$

Therefore,

$$\mathbb{E}(X) = \mathbb{E}(0) = 0.$$

Assume that $n \geq 2$. To determine if the sequence $X_1, X_2, \ldots, X_n$ is pairwise independent, we observe the following: If $X_1 = 0$, then $A[1] = 1$, which implies that $A[n] \geq 2$, which implies that $X_n = |A[n] - n| \leq n - 2$. This suggests that $X_1, X_2, \ldots, X_n$ is not pairwise independent. Indeed,

$$\Pr(X_1 = 0 \wedge X_n = n - 1) = 0,$$

whereas

$$\Pr(X_1 = 0) \cdot \Pr(X_n = n - 1) = \Pr(A[1] = 1) \cdot \Pr(A[n] = 1) = 1/n \cdot 1/n \neq 0.$$

Next we assume that $n \geq 1$ and choose an integer k with $1 \leq k \leq n$. We are going to determine

$$\mathbb{E}(X_k) = \mathbb{E}(|A[k] - k|).$$

We know that $A[k] \in \{1, 2, \ldots, n\}$. Therefore,

$$\begin{aligned}
\mathbb{E}(X_k) &= \sum_{i=1}^{n} |i - k| \cdot \Pr(A[k] = i) \\
&= \sum_{i=1}^{n} |i - k| \cdot 1/n \\
&= \frac{1}{n} \sum_{i=1}^{n} |i - k|.
\end{aligned}$$

We use the hint to determine the summation:

$$\begin{aligned}
\sum_{i=1}^{n} |i-k| &= \sum_{i=1}^{k}(k-i) + \sum_{i=k+1}^{n}(i-k) \\
&= (k-1)+(k-2)+\cdots+2+1+0 \\
&\quad +1+2+3+\cdots+(n-k) \\
&= \frac{(k-1)k}{2} + \frac{(n-k)(n-k+1)}{2},
\end{aligned}$$

which, after some algebra, is equal to

$$\frac{n^2+n+2k^2-2k-2kn}{2}.$$

We conclude that

$$\begin{aligned}
\mathbb{E}\left(X_k\right) &= \frac{1}{n} \cdot \frac{n^2+n+2k^2-2k-2kn}{2} \\
&= \frac{n+1}{2} + \frac{k^2-k-kn}{n}.
\end{aligned}$$

Next, we are going to determine $\mathbb{E}(X)$. For this, we use the linearity of expectation and the hint:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{k=1}^{n} X_k\right) \\
&= \sum_{k=1}^{n} \mathbb{E}\left(X_k\right) \\
&= \sum_{k=1}^{n} \left(\frac{n+1}{2} + \frac{k^2-k-kn}{n}\right) \\
&= \frac{n(n+1)}{2} + \frac{1}{n}\sum_{k=1}^{n} k^2 - \frac{1}{n}\sum_{k=1}^{n} k - \sum_{k=1}^{n} k \\
&= \frac{n(n+1)}{2} + \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \frac{1}{n} \cdot \frac{n(n+1)}{2} - \frac{n(n+1)}{2}.
\end{aligned}$$

After some algebra, this simplifies to

$$\mathbb{E}(X) = \frac{n^2-1}{3}.$$

Remark: This was painful! If you know of an easier/shorter way to determine $\mathbb{E}(X)$, please let me know.

Question 7: Let $b \geq 1$, $c \geq 1$, and $w \geq 1$ be integers, and let $n = b + c + w$. You are given b beer bottles $B_1, B_2, \ldots, B_b$, c cider bottles $C_1, C_2, \ldots, C_c$, and w wine bottles $W_1, W_2, \ldots, W_w$. Let $m \geq 1$ be an integer with $m \leq b$ and $m \leq n - b$.

All n bottles are in a box. From this box, you choose a uniformly random subset consisting of m bottles. Define the random variables

$$\begin{aligned} X &= \text{the number of beer bottles in the chosen subset,} \\ Y &= \text{the number of cider bottles in the chosen subset,} \\ Z &= \text{the number of wine bottles in the chosen subset.} \end{aligned}$$

- Determine the expected value $\mathbb{E}(X + Y + Z)$.
- Let k be an integer with $0 \leq k \leq m$. Prove that
$$\Pr(X = k) = \frac{\binom{b}{k}\binom{n-b}{m-k}}{\binom{n}{m}}.$$
- For each $i = 1, 2, \ldots, b$ and $j = 1, 2, \ldots, c$, define the indicator random variables
$$X_i = \begin{cases} 1 & \text{if } B_i \text{ is in the chosen subset,} \\ 0 & \text{otherwise.} \end{cases}$$
and
$$Y_j = \begin{cases} 1 & \text{if } C_j \text{ is in the chosen subset,} \\ 0 & \text{otherwise.} \end{cases}$$
Prove that
$$\mathbb{E}(X_i) = \mathbb{E}(Y_j) = \frac{m}{n}.$$
- Prove that
$$\sum_{k=0}^{m} \frac{k\binom{b}{k}\binom{n-b}{m-k}}{\binom{n}{m}} = \frac{bm}{n}.$$
- Let i and j be integers with $1 \leq i \leq b$ and $1 \leq j \leq c$. Are the random variables X_i and Y_j independent?
- Let i and j be integers with $1 \leq i \leq b$ and $1 \leq j \leq c$. Determine $\mathbb{E}(X_i \cdot Y_j)$.
- Let i and j be integers with $1 \leq i \leq b$ and $1 \leq j \leq c$. Is the following true or false?
$$\mathbb{E}(X_i \cdot Y_j) = \mathbb{E}(X_i) \cdot \mathbb{E}(Y_j).$$

Solution: Since $X + Y + Z$ is equal to the total number of bottles that we choose, this sum is always equal to m. Therefore,

$$\mathbb{E}(X + Y + Z) = \mathbb{E}(m) = m.$$

Let k be an integer with $0 \leq k \leq m$. We are going to determine $\Pr(X = k)$, i.e., the probability that we choose exactly k beer bottles, when choosing m bottles:

1. There are $\binom{n}{m}$ ways to choose m bottles out of n bottles.

2. If we choose k beer bottles, then we also choose $m-k$ non-beer bottles. By the Product Rule, the number of ways to do this is equal to

$$\binom{b}{k}\binom{n-b}{m-k}.$$

3. It follows that

$$\Pr(X=k) = \frac{\binom{b}{k}\binom{n-b}{m-k}}{\binom{n}{m}}.$$

4. Later, we will need the expected value of X; we determine it here. Observe that X can take any value in $\{0, 1, 2, \ldots, m\}$. Therefore,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{m} k \cdot \Pr(X=k) \\ &= \sum_{k=0}^{m} \frac{k\binom{b}{k}\binom{n-b}{m-k}}{\binom{n}{m}}.\end{aligned}$$

Below, we are going to use indicator random variables to determine $\mathbb{E}(X)$ in an easier way. Since X_i is an indicator random variable (i.e., its value is 0 or 1), we have

$$\mathbb{E}(X_i) = \Pr(X_i = 1) = \Pr(B_i \text{ is among the } m \text{ chosen bottles}).$$

1. There are $\binom{n}{m}$ ways to choose m bottles out of n bottles.

2. The number of ways to choose m bottles out of n, one of which is equal to B_i, is equal to $\binom{n-1}{m-1}$.

3. It follows that

$$\begin{aligned}\mathbb{E}(X_i) &= \frac{\binom{n-1}{m-1}}{\binom{n}{m}} \\ &= \frac{(n-1)!}{(m-1)!(n-m)!} \cdot \frac{m!(n-m)!}{n!} \\ &= m/n.\end{aligned}$$

By exactly the same reasoning, we have

$$\mathbb{E}(Y_j) = m/n.$$

Since $X = \sum_{i=1}^{b} X_i$, we get

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{b} X_i\right) \\
&= \sum_{i=1}^{b} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{b} m/n \\
&= bm/n.
\end{aligned}$$

Thus, we have two expressions for $\mathbb{E}(X)$. These expressions must be equal, implying that

$$\sum_{k=0}^{m} \frac{k\binom{b}{k}\binom{n-b}{m-k}}{\binom{n}{m}} = \frac{bm}{n}.$$

A nicer way to write this is

$$\sum_{k=0}^{m} k\binom{b}{k}\binom{n-b}{m-k} = \frac{bm}{n}\binom{n}{m}.$$

Are the random variables X_i and Y_j independent? Using the results obtained above, we have

$$\Pr(X_i = 1) \cdot \Pr(Y_j = 1) = m/n \cdot m/n = (m/n)^2.$$

On the other hand,

$$\begin{aligned}
\Pr(X_i = 1 \wedge Y_j = 1) &= \Pr(\text{both } B_i \text{ and } C_j \text{ are chosen}) \\
&= \frac{\binom{n-2}{m-2}}{\binom{n}{m}} \\
&= \frac{(n-2)!}{(m-2)!(n-m)!} \cdot \frac{m!(n-m)!}{n!} \\
&= \frac{m(m-1)}{n(n-1)}.
\end{aligned}$$

It follows that

$$\Pr(X_i = 1 \wedge Y_j = 1) \neq \Pr(X_i = 1) \cdot \Pr(Y_j = 1).$$

Therefore, X_i and Y_j are not independent.

Next we determine $\mathbb{E}(X_i \cdot Y_j)$. Since the value of the random variable $X_i \cdot Y_j$ is 0 or 1, we have

$$\begin{aligned}
\mathbb{E}(X_i \cdot Y_j) &= \Pr(X_i \cdot Y_j = 1) \\
&= \Pr(X_i = 1 \wedge Y_j = 1) \\
&= \frac{m(m-1)}{n(n-1)}.
\end{aligned}$$

For the last part, we have

$$\mathbb{E}\left(X_i \cdot Y_j\right) = \frac{m(m-1)}{n(n-1)},$$

whereas

$$\mathbb{E}\left(X_i\right) \cdot \mathbb{E}\left(Y_j\right) = m/n \cdot m/n.$$

These two expressions are not equal.

Question 8: As you all know, Elisa Kazan is the President of the Carleton Computer Science Society. Elisa's neighborhood pub serves three types of drinks: cider, wine, and beer. Elisa likes cider[4] and wine[5], but does not like beer[6].

After a week of hard work, Elisa goes to this pub and repeatedly orders a random drink (the results of the orders are mutually independent). If she gets a glass of cider or a glass of wine, then she drinks it and places another order. As soon as she gets a pint of beer, she drinks it and takes a taxi home.

When Elisa orders one drink, she gets a glass of cider with probability 2/5, a glass of wine with probability 2/5, and a pint of beer with probability 1/5.

Define the random variables

$$\begin{aligned} X &= \text{the number of drinks that Elisa orders,} \\ Y &= \text{the number of different types that Elisa drinks.} \end{aligned}$$

If we denote cider by C, wine by W, and beer by B, then a possible sequence of drinks is $CCWCB$; for this case $X = 5$ and $Y = 3$. For the sequence $WWWB$, we have $X = 4$ and $Y = 2$.

- Determine the expected value $\mathbb{E}(X)$.

  Solution: If we say that one order is a *success* if Elisa gets a beer, and a *failure* otherwise, then Elisa orders until the first success. The success probability is $p = 1/5$. We have seen in class that

  $$\mathbb{E}(X) = 1/p = 5.$$

- Describe the sample space in terms of strings consisting of characters C, W, and B.

  Solution: We have seen a very similar thing in class:

  $$S = \{D_1 D_2 \cdots D_n B : n \geq 0, D_1, \ldots, D_n \in \{C, W\}\}.$$

- Describe the event "$Y = 1$" in terms of a subset of the sample space.

  Solution: The event $Y = 1$ happens if and only if the first drink is a beer. Therefore, this event corresponds to the subset $\{B\}$ of the sample space S.

[4] I know this is true

[5] I am not sure if this is true

[6] I know this is true

- Use the result of the previous part to determine $\Pr(Y = 1)$.

  Solution: We have
$$\Pr(Y = 1) = \Pr(B) = 1/5.$$

- Describe the event "$Y = 2$" in terms of a subset of the sample space.

  Solution: The event $Y = 2$ happens if and only if Elisa gets one or more ciders, followed by one beer, or one or more wines, followed by one beer. As a subset of the sample space S, the event $Y = 2$ is equal to
$$\{C^n B : n \geq 1\} \cup \{W^n B : n \geq 1\}.$$

- Use the result of the previous part to determine $\Pr(Y = 2)$.

  Solution: We have
$$\Pr(Y = 2) = \sum_{n=1}^{\infty} \Pr\left(C^n B\right) + \sum_{n=1}^{\infty} \Pr\left(W^n B\right).$$

  Since both $\Pr(C)$ and $\Pr(W)$ are equal to $2/5$, both summations are equal, and we get
$$\begin{aligned}
\Pr(Y = 2) &= 2 \sum_{n=1}^{\infty} \left(\frac{2}{5}\right)^n \cdot \frac{1}{5} \\
&= \frac{4}{25} \sum_{m=0}^{\infty} \left(\frac{2}{5}\right)^m \\
&= \frac{4}{25} \cdot \frac{1}{1 - 2/5} \\
&= 4/15.
\end{aligned}$$

- Determine $\Pr(Y = 3)$.

  Solution: Since the value of Y is 1, 2, or 3, we have
$$\Pr(Y = 1) + \Pr(Y = 2) + \Pr(Y = 3) = 1,$$

  which gives
$$\Pr(Y = 3) = 1 - \Pr(Y = 1) - \Pr(Y = 2) = 1 - 1/5 - 4/15 = 8/15.$$

- Use the results of the previous five parts to determine the expected value $\mathbb{E}(Y)$.

  Solution: We have
$$\begin{aligned}
\mathbb{E}(Y) &= 1 \cdot \Pr(Y = 1) + 2 \cdot \Pr(Y = 2) + 3 \cdot \Pr(Y = 3) \\
&= 1 \cdot 1/5 + 2 \cdot 4/15 + 3 \cdot 8/15 \\
&= 7/3.
\end{aligned}$$

- Define the random variable

$$Y_c = \begin{cases} 1 & \text{if Elisa drinks at least one glass of cider,} \\ 0 & \text{otherwise.} \end{cases}$$

Determine the expected value $\mathbb{E}(Y_c)$.

Solution: We have $Y_c = 1$ if and only if Elisa gets zero or more wines, followed by one cider. As a subset of the sample space, this is the set

$$\{W^n C : n \geq 0\}.$$

Therefore,

$$\begin{aligned} \mathbb{E}(Y_c) &= \Pr(Y_c = 1) \\ &= \sum_{n=0}^{\infty} \Pr(W^n C) \\ &= \sum_{n=0}^{\infty} \left(\frac{2}{5}\right)^{n+1} \\ &= \frac{2}{5} \sum_{n=0}^{\infty} \left(\frac{2}{5}\right)^{n} \\ &= \frac{2}{5} \cdot \frac{1}{1 - 2/5} \\ &= 2/3. \end{aligned}$$

- Define the random variable

$$Y_w = \begin{cases} 1 & \text{if Elisa drinks at least one glass of wine,} \\ 0 & \text{otherwise.} \end{cases}$$

Determine the expected value $\mathbb{E}(Y_w)$.

Solution: By changing the roles of cider and wine, the expected value of Y_w is determined in exactly the same way as for Y_c:

$$\mathbb{E}(Y_w) = 2/3.$$

- Express Y in terms of Y_c and Y_w.

Solution: After a little bit of thought, you will see that

$$Y = 1 + Y_c + Y_w.$$

- Use the results of the previous three parts to determine the expected value $\mathbb{E}(Y)$.

  Solution: We have

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}\left(1+Y_c+Y_w\right) \\ &= 1+\mathbb{E}\left(Y_c\right)+\mathbb{E}\left(Y_w\right) \\ &= 1+2/3+2/3 \\ &= 7/3.\end{aligned}$$

  (Of course, this is the same answer as we got before!)

Question 9: Let $n \geq 2$ be an integer and consider n people $P_1, P_2, \ldots, P_n$. Each of these people has a uniformly random birthday, and all birthdays are mutually independent. (We ignore leap years.) Define the random variable

$$X = \text{ the number of indices } i \text{ such that } P_i \text{ and } P_{i+1} \text{ have the same birthday.}$$

Determine the expected value $\mathbb{E}(X)$.
Hint: Use indicator random variables.

Solution: We will use the indicator random variables $X_1, X_2, \ldots, X_{n-1}$, where

$$X_i = \begin{cases} 1 & \text{if } P_i \text{ and } P_{i+1} \text{ have the same birthday,} \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned}\mathbb{E}\left(X_i\right) &= \Pr\left(X_i=1\right) \\ &= \Pr\left(P_i \text{ and } P_{i+1} \text{ have the same birthday}\right).\end{aligned}$$

For two people, there are 365^2 possible ordered pairs describing their birthdays, and there are 365 ways to have the same birthday. Therefore,

$$\mathbb{E}\left(X_i\right) = \frac{365}{365^2} = \frac{1}{365}.$$

Since $X = \sum_{i=1}^{n-1} X_i$, we get

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n-1} X_i\right) \\ &= \sum_{i=1}^{n-1} \mathbb{E}\left(X_i\right) \\ &= \sum_{i=1}^{n-1} \frac{1}{365} \\ &= \frac{n-1}{365}.\end{aligned}$$