BM20A9301 Statistics – Exercise set 2

To be done by 15.–19.1.2024

Text in blue or red is not part of the problem or its solution. It's there as extra information to help you learn.

To get points, you need to participate to exercise sessions, and have something to show for each exercise you have marked. If you cheat (mark a problem done but don't show anything when called), you don't get points for the session and get -5 points on your score.

Exercise 1 (Counting elements)**.** Denote by $|A|$ the number of elements of the set A.

(a) If $|A| = 20$, $|B| = 40$ and $|A \cap B| = 5$, what is $|A \cup B|$?

(b) If $|A| = 200$, $|B| = 23$ and $|A \cup B| = 203$, what is $|A \cap B|$?

(c) If $|A| = 34$, $|A \cap B| = 10$ and $|A \cup B| = 37$, what is $|B|$?

(d) If $|A| = 10$, $|A \cap B| = 4$, what is $|A - B|$?

Solution. Looking at Venn diagrams for inspiration, and remembering the inclusion exclusion principle $|X \cup Y| = |X| + |Y| - |X \cap Y|$, we get:

(a) We have $|A \cup B| = |A| + |B| - |A \cap B| = 20 + 40 - 5 = 55$.

(b) Now $203 = |A \cup B| = |A| + |B| - |A \cap B| = 200 + 23 - |A \cap B|$ so $|A \cap B| = 223 - 203 = 20$.

(c) This time $37 = |A \cup B| = |A| + |B| - |A \cap B| = 34 + |B| - 10$ so $|B| = 37 - (34 - 10) = 13$.

(d) For this last one recall that $A - B = A - (A \cap B)$ and $A \cap B \subseteq A$. So $|A - B| = |A| - |A \cap B| = 10 - 4 = 6$.

Exercise 2 (Calculating probabilities)**.** Let A and B be events such that $A \subseteq B$. We know that $P(A) = 0.3$ and $P(B) = 0.5$. Calculate the following probabilities.

(a) $P(A \cup B)$

(b) $P(A \cap B)$

(c) $P(B - A)$

Solution.

(a) Since $A \subseteq B$ we have $A \cup B = B$, so $P(A \cup B) = P(B) = 0.5$.

(b) Similarly $A \cap B = A$ when $A \subseteq B$, so $P(A \cap B) = P(A) = 0.3$.

(c) Recall that in general $P(B-A) = P(B) - P(A \cap B)$ so by (b) we have $P(B-A) = 0.5 - 0.3 = 0.2$.

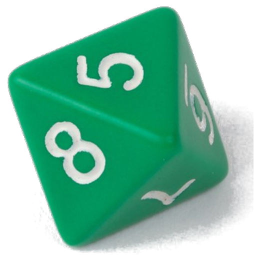
Exercise 3 (Probability modelling)**.** Two fair 8-sided dice are rolled. Model this random experiment mathematically and answer the following questions.

(a) Describe the sample space Ω of your model.

(b) What are the probabilities of each $s \in \Omega$?

(c) What is the probability that the sum of dices is greater than 13?

(d) What is the probability the sum is an even number?



Solution.

(a) If the dice are named A and B then we could set

$$\Omega = \{(a, b) \mid a, b \in \{1, 2, 3, 4, 5, 6, 7, 8\}\}$$

meaning that A rolled a and B rolled b.

(b) There is no reason to suspect that any number is more likely than any other on a single dice. Also there is no reason to suspect that the dice affect each other's probabilities. Hence all the outcomes should be equally likely. We have $|\Omega| = 8 \times 8 = 64$ so $P(s) = 1/64$ for each $s \in \Omega$.

(c) The event $S =$"sum of dices is greater than 13" is satisfied by these occurences:

$$\begin{array}{l}(8,6),(8,7),(8,8)\\(7,7),(7,8)\\(6,8)\end{array}$$

of which there are 6. Since every occurence is equally likely we have $P(S) = |S|/|\Omega| = 6/64 = 3/32 \approx 0.09375$.

(d) Let's draw Ω in a table and color the cells where the sum is even:

(8,1)	(8,2)	(8,3)	(8,4)	(8,5)	(8,6)	(8,7)	(8,8)
(7,1)	(7,2)	(7,3)	(7,4)	(7,5)	(7,6)	(7,7)	(7,8)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)	(6,7)	(6,8)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)	(5,7)	(5,8)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)	(4,7)	(4,8)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)	(3,7)	(3,8)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)	(1,8)

We see that there are just as many coloured and non-coloured occurences. Hence the probability is 1/2.

Exercise 4 (Combinatorial probability)**.** A fair coin is flipped **10 times** and it lands on heads or its opposite side tails.

(a) Describe the sample space.

(b) What is the probability to get heads every flip?

(c) What is the probability to get exactly one tails?

(d) What is the probability to get exactly two tails?

Solution.

(a) The sample space Ω consists of all strings of 10 characters, each of which if a 'H' or a 'T', for "heads" and "tails". The left-most character corresponds to the first throw etc.

(b) There is only one occurence that satisfies this event: $HHHHHHHHHH \in \Omega$. In total Ω has $2^{10} = 1024$ elements, because it's a choice of 2 items (H or T) repeated 10 times, with order mattering and repetition of a selection allowed. Each outcome is equally likely so the probability is 1/1024.

(c) There are 10 different strings of characters in Ω satisfying this event, namely $THHHHHHHHH$, $HTHHHHHHHH$, ..., $HHHHHHHHHT$. Hence the probability is 10/1024.

(d) To create the occurences of this event start from a string of 10 H's, and select two characters which will becomes a T. For this we have to select 2 spots starting from 10 options, no repetition allowed (there must be two T's, not one), and order not mattering (only the positions occupiend by T's matter). There are $\binom{10}{2} = 10 \times 9/2 = 45$ such strings. Hence the probability is 45/1024.

Exercise 5 (Independence and conditional probability)**.** Someone throws two fair dice, a red and a blue one. Consider the events B_6 = "the blue dice threw a 6", R_3 = "the red dice threw a 3" and $S_{\leq 9}$ = "the sum of the dice is at most 9".

(a) Are B_6 and R_3 idenpendent events?

(b) Draw the sample space for this problem and colour the occurences in which $S_{\leq 9}$ happens.

(c) Calculate $P(B_6 \mid S_{\leq 9})$ and $P(R_3 \mid S_{\leq 9})$.

(d) Are B_6 and R_3 still independent if you know for sure that $S_{\leq 9}$ happens? In other words are they conditionally independent conditioned on $S_{\leq 9}$, meaning $P(B_6 \cap R_3 \mid S_{\leq 9}) = P(B_6 \mid S_{\leq 9}) \cdot P(R_3 \mid S_{\leq 9})$?

Solution.

(a) Yes they are independent: the dice do not affect each other's probabilities.

(b) The row each have a fixed value of the blue dice and the columns have a fixed value of the red dice:

(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)

(c) Method 1: We have $P(S_{\leq 9}) = 30/36$ because there are 30 green cells and each has probability $1/36$. The event B_6 corresponds to the top row and R_3 to the column number 3. Hence $P(B_6 \cap S_{\leq 9}) = 3/36$ and $P(R_3 \cap S_{\leq 9}) = 6/36$. Now

$$P(B_6 \mid S_{\leq 9}) = \frac{P(B_6 \cap S_{\leq 9})}{P(S_{\leq 9})} = \frac{3/36}{30/36} = \frac{1}{10},$$
$$P(R_3 \mid S_{\leq 9}) = \frac{P(R_3 \cap S_{\leq 9})}{P(S_{\leq 9})} = \frac{6/36}{30/36} = \frac{1}{5}.$$

Method 2: There are 30 occurences in $S_{\leq 9}$ and 3 in $B_6 \cap S_{\leq 9}$, namely half of the top row which corresponds to the blue dice rolling a 6. Hence $P(B_6 \mid S_{\leq 9}) = 3/30 = 1/10$. Similarly, in the green region there are 6 occurences with the red dice having thrown a 3, so $P(R_3 \mid S_{\leq 9}) = 6/30 = 1/5$.

(d) We have $(B_6 \cap R_3) \cap S_{\leq 9} = \{(6,3)\}$ which is just one occurence, and $|S_{\leq 9}| = 30$, so $P(B_6 \cap R_3 \mid S_{\leq 9}) = 1/30$. From (c) we see that

$$P(B_6 \mid S_{\leq 9}) \cdot P(R_3 \mid S_{\leq 9}) = \frac{1}{10} \cdot \frac{1}{5} = \frac{1}{50} \neq \frac{1}{30} = P(B_6 \cap R_3 \mid S_{\leq 9}).$$

Hence they are conditionally dependent of $S_{\leq 9}$. This means that if you know that $S_{\leq 9}$ happens, then you can deduce something about the chance of B_6 happening once you observe R_3 happening. Imagine you know that the red dice threw a 4. Then there is no way for the blue dice to give a 6 because $S_{\leq 9}$ has to happen. So knowing that $S_{\leq 9}$ happens for sure "removes" the independence of the two dice.

Exercise 6 (Law of total probability)**.** The numbers in this exercise are completely fictional. Do not try to deduce how well people actually did in the Mathematics A exam from this exercise!

Students take two maths classes: Mathematics A and Statistics. A student gets a high score in Mathematics A with probability 40%. Those students that did well in that course have an 80% change of getting a high score in Statistics. Those that did not get a high score in Mathematics A have only a 30% chance to get a high score in Statistics. What is the probability that a student will get a high score in Statistics?

Solution. Write the following events:

- $H_A =$"high score in Mathematics A"
- $H_S =$"high score in Statistics"

Then

- $H_A^C =$"not a high score in Mathematics A"
- $H_S^C =$"not a high score in Statistics"

The problem statement gives

$$P(H_A) = 0.4, \qquad P(H_S \mid H_A) = 0.8, \qquad P(H_S \mid H_A^C) = 0.3$$

from which we also conclude that $P(H_A^C) = 1 - P(H_A) = 0.6$. The law of total probability gives

$$P(H_S) = P(H_S \mid H_A)P(H_A) + P(H_S \mid H_A^C)P(H_A^C) = 0.8 \cdot 0.4 + 0.3 \cdot 0.6 = 0.5.$$

Surprising, isn't it!? Even though high-scorers only have an 80% chance of getting a high score, overall it's easier getting a high schore in the second course than in the first one.