Probability Exercises

Gianmaria Romano

ACSAI A.Y. 2024/2025 (First Semester)

Disclaimer

This document contains the solutions to the exercise sheets provided by Professor Silvestri during the ACSAI Probability course that was given jointly with Professor Bertini during the Academic Year 2024/2025.
Keep in mind, however, that some exercises, marked by $(*)$ have **not** been discussed during the exercise session, and therefore the solution provided in the document may be wrong.
This document is free to use/share but please remember to credit me as the author and therefore do not hide/remove this page.

Contents

Chapter 1

Exercise Sheet 1: Set theory and probability spaces

1.1 Exercise 1

Prove the following identities between sets:

1. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $(*)$
   The statement is proved by double inclusion, which consists in proving that $(A \cup B) \cap C \subseteq (A \cap C) \cup (B \cap C)$ and $(A \cap C) \cup (B \cap C) \subseteq (A \cup B) \cap C$. Start by proving that $(A \cup B) \cap C \subseteq (A \cap C) \cup (B \cap C)$: pick a point $\omega \in (A \cup B) \cap C$ and consider the following equivalence:

   $$\omega \in (A \cup B) \cap C \Leftrightarrow \omega \in (A \text{ or } B) \text{ and } C \Leftrightarrow \omega \in (A \text{ and } C) \text{ or } (B \text{ and } C) \Leftrightarrow \omega \in (A \cap C) \cup (B \cap C)$$

   It is instead possible to prove that $(A \cap C) \cup (B \cap C) \subseteq (A \cup B) \cap C$ by reading the first equation backwards

2. $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ $(*)$
   The statement is proved by double inclusion, which consists in proving that $(A \cap B) \cup C \subseteq (A \cup C) \cap (B \cup C)$ and $(A \cup C) \cap (B \cup C) \subseteq (A \cap B) \cup C$. Start by proving that $(A \cap B) \cup C \subseteq (A \cup C) \cap (B \cup C)$: pick a point $\omega \in (A \cap B) \cup C$ and consider the following equivalence:

   $$\omega \in (A \cap B) \cup C \Leftrightarrow \omega \in (A \text{ and } B) \text{ or } C \Leftrightarrow \omega \in (A \text{ or } C) \text{ and } (B \text{ or } C) \Leftrightarrow \omega \in (A \cup C) \cap (B \cup C)$$

   It is instead possible to prove that $(A \cup C) \cap (B \cup C) \subseteq (A \cap B) \cup C$ by reading the first equation backwards.

3. $(A \cap B)^C = A^C \cup B^C$
   The statement is proved by double inclusion, which consists in proving

that $(A \cap B)^C \subseteq A^C \cup B^C$, but, at the same time, $A^C \cup B^C \subseteq (A \cap B)^C$. Start by proving that $(A \cap B)^C \subseteq A^C \cup B^C$: pick a point $\omega \notin (A \cap B)$ and consider the following equivalence:

$$\omega \in (A \cap B)^C \Leftrightarrow \omega \notin A \text{ or } \omega \notin B \Leftrightarrow \omega \in A^C \text{ or } \omega \in B^C \Leftrightarrow \omega \in (A^C \cup B^C)$$

On the other hand, it is possible to prove that $A^C \cup B^C \subseteq (A \cap B)^C$ by reading the first equation backwards.

1.2 Exercise 2 $(*)$

Let A, B be two sets. Prove that the following statements are equivalent:

1. $A \subset B$.
2. $A \cap B = A$.
3. $A \cup B = B$.

It is possible to show that the statements are equivalent by proving that they cyclically imply each other:

- $A \subset B \Rightarrow A \cap B = A$
  Pick a point $\omega \in A$ and observe that, if $A \subset B$, then $\omega \in B$ as well, meaning that, since $\omega \in A$ and B, it must hold that $\omega \in A \cap B$, implying that $A \subset A \cap B$ as this is true $\forall\ \omega \in A$.
  However, it should noted that, trivially $A \cap B \subset A$ as well, meaning that, by double inclusion, it is possible to conclude that, if $A \subset B$, then $A \cap B = A$.

- $A \cap B = A \Rightarrow A \cup B = B$
  If $A \cap B = A$, then let $A \cup B = (A \cap B) \cup B$, which, by applying distributivity, can be rewritten as $A \cup B = (A \cup B) \cap (B \cup B) = (A \cup B) \cap B$.
  However, since it trivially holds that $B \subset A \cup B$, it is possible to exploit the first implication in order to conclude that $A \cup B = (A \cup B) \cap B = B$.

- $A \cup B = B \Rightarrow A \subset B$
  Generally speaking, given two non-empty sets A and B, if $\omega \in A$, then it must hold that $\omega \in A \cup B$ as well.
  However, since it trivially holds that $A \subset A \cup B$, it is possible to exploit the assumption that $A \cup B = B$ to conclude that, actually, $A \subset B$.

Therefore, it is possible to conclude that, indeed, the three statements are equivalent

1.3 Exercise 3

Let A, B, C be sets, not necessarily disjoint. Write $|A \cup B \cup C|$ in terms of the cardinalities of A, B, C and of their intersections.

- Logical approach
  Suppose $|A \cup B \cup C| = |A| + |B| + |C|$: this computation is not always correct because, a priori, it is not know whether the sets are disjoint or not, so eventual elements in $|A \cap B|$, $|B \cap C|$ and $|A \cap C|$ are counted twice, whereas elements in $|A \cap B \cap C|$ would be counted three times.
  For this reason, start to adjust the value by letting:

  $$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C|$$

  While this computation allows to get rid of repetitions in sets $A \cap B$, $B \cap C$ and $A \cap C$, it is still incorrect because elements in $A \cap B \cap C$, which were initially counted three times, have now been subtracted three times, meaning that the actual counting will be:

  $$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|$$

  Later on, it will be possible to generalize this process for n sets via inclusion/exclusion principle.

- Formal approach
  By associativity, rewrite $A \cup B \cup C$ as a union of two sets $(A \cup B)$ and C, resulting in:

  $$A \cup B \cup C = (A \cup B) \cup C \Rightarrow |(A \cup B) \cup C| = |A \cup B| + |C| - |(A \cup B) \cap C|$$

  Most particularly, remember that $|A \cup B| = |A| + |B| - |A \cap B|$, and, by distributivity $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, resulting in:

  $$|(A \cup B) \cup C| = |A| + |B| - |A \cap B| - (|A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|)$$

  However, since $(A \cap C) \cap (B \cap C) = A \cap B \cap C$, the formula simplifies to:

  $|(A \cup B) \cup C| = |A| + |B| - |A \cap B| - (|(A \cap C) \cup (B \cap C)|)$, resulting in:

  $$|A \cup B \cup C| = |A| + |B| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

1.4 Exercise 4

After interviewing 50 students, the following data are collected: 25 students studied French, 20 studied German and 5 studied both languages. Compute:

1. The number of students who only studied French.
   Let $A = \{\text{students who only studied French}\} = F \setminus (F \cap G)$, meaning that: $|A| = |F| - |F \cap G| = 25 - 5 = 20$ students.

2. The number of students who only studied German.
   Let $B = \{\text{students who only studied French}\} = G \setminus (G \cap F)$, meaning that:
   $|B| = |G| - |G \cap F| = 20 - 5 = 15$ students (recall that $G \cap F = F \cap G$).

3. The number of students who did not study any of the two languages.
   Let $C = \{\text{students who studied neither language}\} = (F \cup G)^C$, meaning that:
   $|C| = |\Omega| - |(F \cup G)| = 50 - (25 + 20 - 5) = 10$ students.

1.5 Exercise 5 $(*)$

After interviewing 60 people, the following data are collected: 25 people read Mickey Mouse, 26 people read Tex, 23 people read Diabolik. Moreover, 9 people read both Mickey Mouse and Tex, 11 people read both Mickey Mouse and Diabolik, 8 people read both Tex and Diabolik. Finally, 3 people read all the three journals. Compute:

1. How many people read only Mickey Mouse.
   Let $A = \{\text{Mickey Mouse only}\} = M \cap (D \cup T)^C$, meaning that:
   $|A| = |M| - (|M \cap D| + |M \cap T| - |M \cap D \cap T|) = 25 - (11 + 9 - 3) = 8$ people.

2. How many people read only Tex.
   Let $B = \{\text{Tex only}\} = T \cap (M \cup D)^C$, meaning that:
   $|B| = |T| - (|T \cap M| + |T \cap D| - |T \cap M \cap D|) = 26 - (9 + 8 - 3) = 12$ people.

3. How many people read only Diabolik.
   Let $C = \{\text{Diabolik only}\} = D \cap (M \cup T)^C$, meaning that:
   $|C| = |D| - (|D \cap M| + |D \cap T| - |D \cap M \cap T|) = 23 - (11 + 8 - 3) = 7$ people.

4. How many people read at least one journal.
   Let $E = \{\text{at least one journal}\} = M \cup D \cup T$, meaning that, by inclusion/exclusion formula, it is possible to state that:
   $|E| = |M| + |D| + |T| - |M \cap D| - |M \cap T| - |D \cap T| + |M \cap D \cap T| = 49$ people.

5. How many people read exactly one journal.
   Let $F = \{\text{exactly one journal}\} = A \cup B \cup C$ and notice that, since the sets are all disjoint, the inclusion/exclusion formula simplifies to the additive case, meaning that:
   $|F| = |A| + |B| + |C| = 8 + 12 + 7 = 27$ people.

6. How many people read none of the journals.
   Let $G = \{\text{none of the journals}\} = \Omega \setminus E$, meaning that:
   $|G| = |\Omega| - |E| = 60 - 49 = 11$ people.

1.6 Exercise 6

Let $(\Omega, \mathbb{P})$ be a probability space, and let A, B, C be events. It is known that $A \cap B \cap C = \emptyset$, $\mathbb{P}(A \cap C) = \frac{1}{5}$ and $\mathbb{P}(B \cap C) = \frac{2}{5}$.

1. Compute $\mathbb{P}((A \cup B) \cap C)$.
   By distributivity, let $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, meaning that:

   $$\mathbb{P}((A \cup B) \cap C) = \mathbb{P}((A \cap C) \cup (B \cap C)) = \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}((A \cap C) \cap (B \cap C))$$

   However, since $(A \cap C) \cap (B \cap C) = A \cap B \cap C = \emptyset$, the equation simplifies to:

   $$\mathbb{P}((A \cup B) \cap C) = \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$$

2. What are the possible values of $\mathbb{P}(A \cap B)$? (For example, can it be that $\mathbb{P}(A \cap B) = 1$?)
   In this case, it is impossible to have $\mathbb{P}(A \cap B) = 1$ because it would violate $A \cap B \cap C = \emptyset$.
   For this reason, observe that $(A \cap B) \cap ((A \cup B) \cap C) = \emptyset$, meaning that it is possible to apply the definition of probability measure to state that:

   $$\mathbb{P}(A \cap B) + \mathbb{P}((A \cup B) \cap C) \leq 1 \Rightarrow \mathbb{P}(A \cap B) \leq \frac{2}{5}$$

   In addition, since the definition of probability measure requires that $\mathbb{P}(A \cap B) \geq 0$ (with $\mathbb{P}(A \cap B) = 0 \Leftrightarrow A \cap B = \emptyset$), it is possible to provide a lower bound for this value as well, meaning that:

   $$0 \leq \mathbb{P}(A \cap B) \leq \frac{2}{5}$$

1.7 Exercise 7

1. If $\mathbb{P}(A) = \frac{1}{3}$ and $\mathbb{P}(B^C) = \frac{1}{4}$, can A and B be disjoint events?
   Assume that $A \cap B = \emptyset$, meaning that:

   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{3} + (1 - \frac{1}{4}) = \frac{13}{12} > 1$$

   However, this would violate the condition that $0 \leq \mathbb{P}(A \cup B) \leq 1$, meaning that the two events cannot be disjoint.

2. If $\mathbb{P}(A) = \frac{1}{4}$ and $\mathbb{P}(A \cup B) = \frac{3}{4}$, what is the value of $\mathbb{P}(B)$ when A and B are disjoint?
   If $A \cap B = \emptyset$, then:

   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \Rightarrow \mathbb{P}(B) = \mathbb{P}(A \cup B) - \mathbb{P}(A) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

3. If $\mathbb{P}(A) = \mathbb{P}(B) = \frac{3}{8}$, can it be the case that $\mathbb{P}(A \cup B) = \frac{1}{4}$? How about $\mathbb{P}(A \cup B) = \frac{7}{8}$?
   Assume $\mathbb{P}(A \cup B) = \frac{1}{4}$: since $A \subseteq (A \cup B)$ and $B \subseteq (A \cup B)$, it must hold by monotonicity that:

   $$\mathbb{P}(A) \leq \mathbb{P}(A \cup B) \wedge \mathbb{P}(B) \leq \mathbb{P}(A \cup B) \Rightarrow \mathbb{P}(A \cup B) \geq \max\{\mathbb{P}(A), \mathbb{P}(B)\} \Rightarrow \mathbb{P}(A \cup B) \geq \frac{3}{8}$$

   For this reason, it is not possible to have $\mathbb{P}(A \cup B) = \frac{1}{4}$ because it would violate the monotonicity property.
   Now, assume instead that $\mathbb{P}(A \cup B) = \frac{7}{8}$: this means that:

   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = \frac{3}{8} + \frac{3}{8} - \frac{7}{8} = -\frac{1}{8}$$

   However, having $\mathbb{P}(A \cap B) < 0$ would violate the definition of probability measure, meaning that it is not possible to have $\mathbb{P}(A \cup B) = \frac{7}{8}$.

4. Let $\mathbb{P}(A) = \frac{3}{4}$ and $\mathbb{P}(B) = \frac{3}{8}$. Check that $\frac{1}{8} \leq \mathbb{P}(A \cap B) \leq \frac{3}{8}$.
   Notice that, by monotonicity, $A \cap B \subseteq A$ and $A \cap B \subseteq B$, meaning that:

   $$\mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\} \Rightarrow \mathbb{P}(A \cap B) \leq \frac{3}{8}$$

   In addition, remember that, by definition, it must hold that:

   $$\mathbb{P}(A \cap B) \leq 1 \Rightarrow \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq 1 \Rightarrow \mathbb{P}(A \cap B) \geq \frac{1}{8}$$

   Therefore, it is possible to conclude that:

   $$\frac{1}{8} \leq \mathbb{P}(A \cap B) \leq \frac{3}{8}$$

5. Prove the following inequality: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$.
   Since the probability measure definition requires $0 \leq \mathbb{P}(X) \leq 1 \ \forall \ X \in F_\Omega$, it is possible to rearrange the inequality in the following way:

   $$\mathbb{P}(A \cup B) \leq 1 \Rightarrow \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq 1 \Rightarrow \mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$$

1.8 Exercise 8

Two fair dice, one red and one blue, are tossed.

1. Describe the sample space Ω.

   $$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{(r, b) \mid r, b \in \{1, 2, 3, 4, 5, 6\}\}$$

   By product rule, this set has cardinality $|\Omega| = 6 \cdot 6 = 36$.

2. Describe the following events as subsets of Ω:

(a) $A = \{\text{The red die gives } 5\}$.
$A = \{(5, b) | b \in \{1, 2, 3, 4, 5, 6\}\}$.

(b) $B = \{\text{One of the dice gives } 5\}$.
$B = \{(5, b) \mid b \in \{1, 2, 3, 4, 5, 6\}\} \cup \{(r, 5) \mid r \in \{1, 2, 3, 4, 5, 6\}\}$.

(c) $C = \{\text{Both dice give } 5\}$.
$C = \{(5, 5)\}$.

(d) $D = \{\text{No die gives } 5\}$.
$D = \{(r, b) \mid r, b \in \{1, 2, 3, 4, 6\}\}$.

(e) $E = \{\text{The sum of the two dice equals } 5\}$.
$E = \{(r, b) \mid r, b \in \{1, 2, 3, 4, 5, 6\} \wedge r + b = 5\}$.

3. Compute the probability of the above events.

(a) $A = \{\text{The red die gives } 5\}$.
Let $|A| = 6$, meaning that:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

(b) $B = \{\text{One of the dice gives } 5\}$.
Observe that:

$$\{(r, 5) \mid r \in \{1, 2, 3, 4, 5, 6\}\} \cap \{(5, b) \mid b \in \{1, 2, 3, 4, 5, 6\}\} = \{(5, 5)\}$$

This means that $|B| = 6 + 6 - 1 = 11$, which results in:

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{11}{36}$$

(c) $C = \{\text{Both dice give } 5\}$.
Observe that C is an elementary event, meaning that $|C| = 1$, which results in:

$$\mathbb{P}(C) = \frac{|C|}{|\Omega|} = \frac{1}{36}$$

(d) $D = \{\text{No die gives } 5\}$.
Observe that, actually, $D = B^C$, meaning that:

$$\mathbb{P}(D) = 1 - \mathbb{P}(B) = 1 - \frac{11}{36} = \frac{25}{36}$$

Alternatively, compute $|D| = 25$ and obtain:

$$\mathbb{P}(D) = \frac{|D|}{|\Omega|} = \frac{25}{36}$$

(e) $E = \{\text{The sum of the two dice equals } 5\}$.
Let $|E| = 4$, meaning that:

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{4}{36} = \frac{1}{9}$$

Chapter 2

Exercise Sheet 2: Combinatorics

2.1 Exercise 1

A company consists of 25 people. Among these people, a president and a secretary have to be selected.

1. In how many different ways can the president and secretary be selected? Assuming the election is carried out as an ordered selection without repetitions, there will be 25 possible choices for president, followed by 24 remaining options for secretary, meaning that there will be:

   $$25 \cdot 24 = 600 \text{ possible ways to perform the selection}$$

   Alternatively, consider that it is possible to view the selection as an unordered selection of 2 people among 25 candidates, which can be done in:

   $$\binom{25}{2} = \frac{25!}{(25-2)! \cdot 2!} = 300 \text{ possible ways}$$

   Said positions can actually be ordered in $2! = 2$ ways, resulting in a total of:

   $$2 \cdot \binom{25}{2} = 2 \cdot 300 = 600 \text{ possible selections}$$

2. If the president and secretary are chosen uniformly at random, what is the probability that a given person is selected for either of the positions? Assume a fixed person is chosen as president: it is possible to choose the secretary among the remaining people, which can be done in $\binom{24}{1}$ ways, resulting in:

$$\mathbb{P}(\{\text{the person is chosen for either position}\}) = \frac{\binom{24}{1}}{\binom{25}{2}} = \frac{\frac{24!}{(24-1)!\cdot 1!}}{\frac{25!}{(25-2)!\cdot 2!}} = \frac{2}{25}$$

Alternatively, observe that it is possible to rewrite the event as:

$$\{\text{the person is chosen for either position}\} = \{\text{chosen as president}\} \cup \{\text{chosen as secretary}\}$$

Since the two events are disjoint, it is possible to recover that that:

$$\mathbb{P}(\{\text{the person is chosen for either position}\}) = \frac{1}{25} + (1 - \frac{1}{25})\frac{1}{24} = \frac{2}{25}$$

2.2 Exercise 2

Count the number of anagrams (even with no meaning) of the words: RICE, PASTA, POTATOES.

- RICE
  This is a simple permutation of 4 letters, meaning that there are $4! = 24$ anagrams.

- PASTA
  Observe that considering a simple permutation of 5! anagrams would be incorrect because the letter A gets repeated twice, meaning that you need to partition the permutation with respect to this repetition, resulting in:

$$\binom{5}{2} = \frac{5!}{2!} = 60 \text{ anagrams}$$

- POTATOES
  Observe that considering a simple permutation of 8! anagrams would be incorrect because the letters O and T get repeated twice each, meaning that you need to partition the permutation with respect to these repetitions, resulting in:

$$\binom{8}{2\ 2} = \frac{8!}{2! \cdot 2!} = 10080 \text{ anagrams}$$

2.3 Exercise 3

Let S be a set of cardinality n. Count the number of subsets of S of cardinality k (with $k = 0, \dots, n$).
Observe that, since sets are unordered collections, creating a subset of k elements can be compared to picking k items from the n elements without ordering and repetitions, meaning that there are:

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \text{ subsets of cardinality } k$$

N.B.: It is possible to use the Newton binomial to show that a set S of cardinality n has a total of 2^n subsets because, assuming $1 = 1^k 1^{n-k}$, it must always hold that:

$$|P(S)| = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \frac{n!}{(n-k)! \cdot k!} = 2^n$$

2.4 Exercise 4 (∗)

To pass an exam, students have to provide answers to (exactly) 10 out of the 13 proposed questions.

1. In how many ways can the 10 questions be selected?
   By definition, the selection will be carried out without replacement and by discarding ordering, meaning that the questions can be chosen in:
   $$\binom{13}{10} = \frac{13!}{(13-10)! \cdot 10!} \text{ possible ways.}$$
2. Assuming that the first two questions are compulsory, in how many ways can the 10 questions be selected?
   If the first two questions are compulsory, then it is possible to fix them and choose the remaining questions without replacement and ordering, meaning that, by product rule, the questions can now be chosen in:
   $$\binom{2}{2}\binom{13-2}{10-2} = \binom{11}{8} = \frac{11}{(11-8)! \cdot 8!} \text{ possible ways.}$$
3. Assuming that students are required to answer either the first question or the second one, but not both questions, in how many ways can the 10 questions be selected?
   In this case, students must choose one question among the first two and choose the others among the remaining questions without replacement and ordering, meaning that, by product rule, the questions can now be chosen in:
   $$\binom{2}{1}\binom{13-2}{10-1} = \binom{2}{1}\binom{11}{9} = 2\frac{11!}{(11-9)! \cdot 9!} \text{ possible ways.}$$

2.5 Exercise 5

Bob and Alice go out with 5 friends. They start the evening at a bar. In front of the bar counter there are 7 empty stools in a row, and each person chooses one stool at random.

1. What is the probability that Bob and Alice sit next to each other?
   Start by considering the sample space, which is basically a bijection from the friends to the seats, which can be done in:

   $$|\Omega| = 7! \text{ possible ways.}$$

   Assume now that Alice and Bob sit at the first two seats: this implies that Alice and Bob can organize themselves in 2 possible ways, while their friends can form a permutation, meaning that:

   $$\mathbb{P}(\{\text{Alice and Bob sit in the first two chairs}\}) = \frac{2 \cdot 5!}{7!} = \frac{1}{21}$$

   However, Alice and Bob can actually sit next to each other in six possible ways, meaning that:

   $$\mathbb{P}(\{\text{Alice and Bob sit next to each other}\}) = \frac{6}{21} = \frac{2}{7}$$

2. After the bar, they head to a restaurant, where they are given a round table with 7 chairs. Each person chooses one chair uniformly at random. What is the probability that Bob and Alice sit next to each other?
   Start by considering the sample space, which is basically a bijection from the friends to the seats, which can be done in:

   $$|\Omega| = 7! \text{ possible ways.}$$

   Assume now that Alice and Bob sit at the first two seats: this implies that Alice and Bob can organize themselves in 2 possible ways, while their friends can form a permutation, meaning that:

   $$\mathbb{P}(\{\text{Alice and Bob sit in the first two chairs}\}) = \frac{2 \cdot 5!}{7!} = \frac{1}{21}$$

   However, Alice and Bob can actually sit next to each other in seven possible ways, meaning that:

   $$\mathbb{P}(\{\text{Alice and Bob sit next to each other}\}) = \frac{7}{21} = \frac{1}{3}$$

2.6 Exercise 6

5 cards are randomly picked from a deck of 52 cards. Compute the probability of getting:

1. Four of a kind.
   Start by considering that the cards can be distributed in:

   $$|\Omega| = \binom{52}{5} \text{ possible ways}$$

A player is said to have a "four of a kind" if they have four cards of the same ranking and a fifth card of another ranking.
Suppose you want to fix the four cards to get the four of a kind: there are 13 possible rankings and, for each of these, the four cards can be chosen in $\binom{4}{4} = 1$ way, meaning that these cards can be given in a total of $13 \cdot \binom{4}{4} = 13$ ways.
On the other hand, the fifth card can be chosen without any constraint, which can be done in $\binom{52-4}{1} = \binom{48}{1} = 48$ ways, meaning that:

$$\mathbb{P}(\{\text{four of a kind}\}) = \frac{13 \cdot 48}{\binom{52}{5}}$$

2. Flush.
A player is said to have a "flush" if they have five cards of the same suit.
For each of the four suits, it is possible to choose the cards in $\binom{13}{5}$ possible ways, for a total of $4 \cdot \binom{13}{5}$ possible flushes, meaning that:

$$\mathbb{P}(\{\text{flush}\}) = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}}$$

3. Full house.
A player is said to have a "full house" if they have three cards of a ranking and two cards of another ranking.
Most particularly, the first three cards can be chosen in $13 \cdot \binom{4}{3}$ possible ways, whereas the other two cards can be chosen in $(13-1) \cdot \binom{4}{2} = 12 \cdot \binom{4}{2}$ possible ways (remember that the first ranking can be chosen without constraints, whereas the second ranking should be a different one), meaning that:

$$\mathbb{P}(\{\text{full house}\}) = \frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}}{\binom{52}{5}}$$

4. Two pairs (but not full house).
A player is said to have a "two pairs" if they have two cards of a given ranking and two cards of a second ranking, but, since you want to avoid having a full house as well, the last card should be chosen among the cards with different rankings.
Most particularly, the first two cards can be chosen in $13 \cdot \binom{4}{2}$ possible ways, the other two cards can be chosen in $(13-1) \cdot \binom{4}{2} = 12 \cdot \binom{4}{2}$, followed by $\binom{52-(4+4)}{1} = \binom{44}{1} = 44$ ways to choose the last card, meaning that:

$$\mathbb{P}(\{\text{two pairs but not a full house}\}) = \frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} \cdot 44}{\binom{52}{5}}$$

5. Three of a kind (but not four of a kind nor full house).
A player is said to have a "three of a kind" if they have at least three

cards of the same ranking, but, since you want to avoid having a four of a kind or a full house as well, it should be observed that exactly three cards must have that ranking, followed by two other cards of different rankings. For this reason, the first three cards can be chosen in $13 \cdot \binom{4}{3}$ ways, followed by $12 \cdot \binom{4}{1}$ and $11 \cdot \binom{4}{1}$ ways, respectively, for the remaining cards: by noticing that the other two cards can be grouped in $2! = 2$ ways, you must discard ordering, meaning that:

$$\mathbb{P}(\{\text{three of a kind but not a four of a kind or full house}\}) = \frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} \cdot 11 \cdot \binom{4}{1}}{2! \cdot \binom{52}{2}}$$

2.7 Exercise 7 $(*)$

In a Computer Science department, Rooms I and II can host 50 students each, while Room III can host 100 students. In order to attend the Probability course in those rooms, 200 students are divided into three groups (of 50, 50 and 100 students).

1. In how many ways can these groups be formed, and in how many ways can they be assigned to the different rooms?
   Start by stating that, if selection is done without replacement and without considering student order within a group, then it is possible to assign 50 out of 200 students in Room I, followed by assigning 50 out of 150 students in Room II and, lastly, the remaining 100 students are placed in Room III.
   Therefore, by applying the product rule, the groups can be formed in:
   $$\binom{200}{50}\binom{150}{50}\binom{100}{100} = \binom{200}{50}\binom{150}{50} \text{ possible ways.}$$
   Keep in mind, however, that, when assigning the groups to a room, room order does matter, meaning that, since Rooms I and II can both host 50 students, the smaller groups can be placed in $2!$ ways, whereas, since Room III is the only one that can host 100 students, the larger group can be assigned to only one room, meaning that the number of ways in which the groups can be assigned to the rooms is:
   $$|\Omega| = 2 \cdot \binom{200}{50}\binom{150}{50}$$

2. Alice and Bob would like to attend the course together, but they would like to avoid being in the same class as Will. Compute the probability that their wish comes true.
   Observe that Alice and Bob's wish can come true under one of the following scenarios:

- In the first case, Alice and Bob are together in one of the smaller rooms, which requires to Will to be either in the small room or in the large room.
  For this reason, if Alice and Bob are already fixed in one of the small rooms, it is possible to choose the remaining 48 students in that room among 197 people (remember that Will must be taken out in this selection), whereas the selection for the remaining rooms can be done normally, meaning that, by applying the product rule and by taking into account the fact that there are two small rooms, this scenario can happen in:

$$2\binom{197}{48}\binom{150}{50}\binom{100}{100} = 2\binom{197}{48}\binom{150}{50} \text{ possible ways.}$$

- In the second case, Alice and Bob are in the large room, which requires Will to be in one of the smaller rooms.
  For this reason, if Alice and Bob are already fixed in the large room, it is possible to choose the remaining 98 students in the large room among 197 people (remember that Will must be taken out in this selection), whereas the selection for the smaller room can be done normally, meaning that, by product rule, this scenario can happen in:

$$\binom{197}{98}\binom{100}{50}\binom{50}{50} = \binom{197}{98}\binom{100}{50} \text{ possible ways.}$$

Most particularly, since the aforementioned cases are mutually exclusive, it is possible to apply additivity in order to conclude that:

$$\mathbb{P}(\{\text{Alice and Bob's wish comes true}\}) = \frac{2\binom{197}{48}\binom{150}{50} + \binom{197}{98}\binom{100}{50}}{2\binom{200}{50}\binom{150}{50}}$$

2.8 Exercise 8

Let Ω be a finite, non-empty set and let $H : \Omega \to \mathbb{R}$ denote a given function. For each $\beta \geq 0$, define the probability $\mathbb{P}_\beta$ on Ω by setting, for every $\omega \in \Omega$,

$$\mathbb{P}_\beta(\{\omega\}) = \frac{e^{-\beta H(\omega)}}{Z_\beta}$$

where Z_β is a positive real number (recall that on finite sample spaces the probability measure is uniquely determined by its values on the single outcomes $\omega \in \Omega$). Define further:

$$m = \min_{\omega \in \Omega} H(\omega)$$

$$E_m = \{\omega \in \Omega : H(\omega) = m\} = H^-1(\{m\})$$

1. Write Z_β in terms of β (and H).
   Knowing that $\mathbb{P}(\Omega) = 1$, it is possible to state that:

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \frac{e^{-\beta H(\omega)}}{Z_\beta} = 1 \Rightarrow \frac{1}{Z_\beta} \cdot \sum_{\omega \in \Omega} e^{-\beta H(\omega)} = 1 \Rightarrow Z_\beta = \sum_{\omega \in \Omega} e^{-\beta H(\omega)}$$

2. Check that if $\beta = 0$ then $\mathbb{P}_\beta$ is the uniform probability on Ω.
   For $\beta = 0$, the equation will simplify to $\mathbb{P}_0(\omega) = \frac{1}{Z_0}$. (recall that $e^0 = 1$)

3. Check that

$$\lim_{\beta \to \infty} \mathbb{P}_\beta(E_m) = 1$$

   (check that $\mathbb{P}_\beta$ concentrates on the minima of H as $\beta \to \infty$).
   Assuming $\omega \in E_m$, consider the following limit:

$$\lim_{\beta \to \infty} \sum_{\omega \in E_m} \frac{e^{-\beta H(\omega)}}{Z_\beta} = \lim_{\beta \to \infty} \sum_{\omega \in E_m} \frac{e^{-\beta m}}{Z_\beta} = \lim_{\beta \to \infty} \frac{e^{-\beta m}|E_m|}{Z_\beta}$$

   Most particularly, notice that:

$$Z_\infty = \lim_{\beta \to \infty} \sum_{\omega \in \Omega} e^{-\beta(H(\omega)+m-m)} = \begin{cases} 0 & \text{if H}(\omega) > \text{m} \\ 1 & \text{if H}(\omega) = \text{m} \end{cases}$$

   Therefore, it is possible to let $Z_\beta \approx e^{-\beta m}|E_m|$ by asymptotic comparison, allowing to conclude that:

$$\lim_{\beta \to \infty} \frac{e^{-\beta m}|E_m|}{Z_\beta} = 1$$

Chapter 3

Exercise Sheet 3: Combinatorics and properties of probability measures

3.1 Exercise 1

A lecture course is attended by 9 students. The lecturer writes 3 exam papers, and each paper is assigned to 3 students.

1. In how many ways can the 9 students be paired with the 3 exam papers? Given exam papers A, B and C, the first group will get paper A, the second group will get paper B and the third group will get paper C: in this case, since assignment order matters (a group getting paper A will be different from that group getting paper B) but group ordering does not (the group (A, B, C) will be equivalent to (B, C, A)), it is possible to consider the task as a partition of the students with respect to the groups, resulting in:

   $$\binom{9}{3\ 3\ 3} = \frac{9!}{3! \cdot 3! \cdot 3!} \text{ possible anagrams.}$$

   Alternatively, by considering the aforementioned assumptions about ordering, it is possible to apply the product rule to obtain:

   $$\binom{9}{3} \cdot \binom{9-3}{3} \cdot \binom{9-(3+3)}{3} = \binom{9}{3} \cdot \binom{6}{3} \cdot \binom{3}{3} \text{ possible assignments.}$$

2. If instead the lecturer prepares 9 different exam papers, in how many ways can the 9 students be paired with the 9 exam papers?
   Similarly to the previous point, consider a partition with respect to the

groups resulting in:

$$\binom{9}{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1} = 9! \text{ possible assignments.}$$

Alternatively, observe that, in this case, the ordering will simply be a permutation, which can be carried out in 9! possible ways.

3. Consider the same class of 9 students: In how many ways can the students be partitioned into 3 groups, each made of 3 students?
   Consider an ordered group partition, which could be done in $\binom{9}{3\ 3\ 3}$ ways and then discard its ordering: in fact, since the three groups of three students can be ordered in 3! possible ways, it is possible to partition the students in:

$$\frac{1}{3!} \cdot \binom{9}{3\ 3\ 3} = \frac{1}{3!} \cdot \frac{9!}{3! \cdot 3! \cdot 3!} = \frac{9!}{3! \cdot 3! \cdot 3! \cdot 3!} \text{ possible ways.}$$

4. In how many ways can the students be partitioned into 3 groups, one made of 5 students and 2 made of 2 students?
   Similarly to the previous point, consider an ordered group partition, which could be done in $\binom{9}{5\ 2\ 2}$ ways and then discard its ordering: in fact, since the two groups of two students can be ordered in 2! possible ways, it is possible to partition the students in:

$$\frac{1}{2!} \cdot \binom{9}{5\ 2\ 2} = \frac{1}{2!} \cdot \frac{9!}{5! \cdot 2! \cdot 2! \cdot 2!} = \frac{9!}{2! \cdot 5! \cdot 2! \cdot 2!} \text{ possible ways.}$$

5. In how many ways can the students be partitioned into 9 groups, each made of exactly one student?
   Observe that, in this case, the partition is actually a simple permutation, which can be done in 9! possible ways, but, again, ordering must be discarded, meaning that, since the nine groups can be ordered in 9! possible ways, it is possible to partition the students in:

$$\frac{1}{9!} \cdot 9! = \frac{9!}{9!} = 1 \text{ way.}$$

3.2 Exercise 2

An Italian deck of cards is made of 40 cards of 4 different suits, numbered from 1 (ace) to 10.
In a card game called "tresette" there are 4 players. Each player is given 10 cards. A player is said to get a "napoletana" if they get an ace, a 2 and a 3 of the same suit.
You are sitting at the table, and are given your 10 cards.

1. Compute the probability that you get a "napoletana" of a given suit.
   Start by considering that the player's hand can be formed in:

$$|\Omega| = \binom{40}{10} \text{ possible ways.}$$

   Supposing that the given suit is fixed, it is possible to get the corresponding napoletana by fixing the needed cards and choosing the others at random, meaning that there will be $\binom{1}{1} = 1$ way to get the ace, $\binom{1}{1} = 1$ way to get the two and $\binom{1}{1} = 1$ way to get the three, followed by $\binom{37}{7}$ possible ways to get the other cards, resulting in:

$$\mathbb{P}(\{\text{napoletana of a given suit}\}) = \frac{1 \cdot 1 \cdot 1 \cdot \binom{37}{7}}{\binom{40}{10}} = \frac{\binom{37}{7}}{\binom{40}{10}}$$

2. Compute the probability that you get a "napoletana" of two different suits.
   Similarly to the previous point, assume that, if the two suits are fixed, it is possible to get the respective napoletanas by fixing the six required cards (for each of them, the choice can be done in one way), and then take the other cards at random, which can be done in $\binom{34}{4}$ possible ways, resulting in:

$$\mathbb{P}(\{\text{napoletana of two different suits}\}) = \frac{1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot \binom{34}{4}}{\binom{40}{10}} = \frac{\binom{34}{4}}{\binom{40}{10}}$$

3. Compute the probability that you get at least one "napoletana".
   Reformulate the event as:

$$A = \{\text{get at least one napoletana}\} = \bigcup_{i=1}^{4} A_i = \{\text{get exactly } i \text{ napoletanas}\}$$

   However, since these sub-events are not disjoint, the probability of getting at least one napoletana must be calculated by using the inclusion/exclusion formula, resulting in:

$$\mathbb{P}(A) = \sum_{i=1}^{4} \mathbb{P}(A_i) - \sum_{\substack{i=1 \\ i<j}}^{4} \sum_{j=1}^{4} \mathbb{P}(A_i \cap A_j) + \sum_{\substack{i=1 \\ }}^{4} \sum_{\substack{j=1 \\ i<j<k}}^{4} \sum_{k=1}^{4} \mathbb{P}(A_i \cap A_j \cap A_j) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$$

   Most particularly, notice that:

   - For the event $A_1 = \{\text{get exactly 1 napoletana}\}$, there are $\binom{4}{1}$ ways of getting exactly one napoletana due to suit choice, meaning that:

$$\mathbb{P}(\{\text{get exactly 1 napoletana}\}) = \binom{4}{1} \cdot \frac{\binom{37}{7}}{\binom{40}{10}}$$

- For the event $A_2 = \{\text{get exactly 2 napoletanas}\}$, there are $\binom{4}{2}$ ways of getting exactly two napoletanas due to suit choice, meaning that:

$$\mathbb{P}(\{\text{get exactly 2 napoletanas}\}) = \binom{4}{2} \cdot \frac{\binom{34}{4}}{\binom{40}{10}}$$

- For the event $A_3 = \{\text{get exactly 3 napoletanas}\}$, there are $\binom{4}{3}$ ways of getting exactly three napoletanas due to suit choice, meaning that:

$$\mathbb{P}(\{\text{get exactly 3 napoletanas}\}) = \binom{4}{3} \cdot \frac{\binom{31}{1}}{\binom{40}{10}}$$

- The event $A_4 = \{\text{get exactly 4 napoletanas}\}$ is impossible because it would require 12 cards, so $\mathbb{P}(\{\text{get exactly 4 napoletanas}\}) = 0$.

Therefore, it is possible to conclude that:

$$\mathbb{P}(\{\text{get at least one napoletana}\}) = (\binom{4}{1} \cdot \frac{\binom{37}{7}}{\binom{40}{10}}) - (\binom{4}{2} \cdot \frac{\binom{34}{4}}{\binom{40}{10}}) + (\binom{4}{3} \cdot \frac{\binom{31}{1}}{\binom{40}{10}})$$

3.3 Exercise 3

Bob must take the Maths exam. The pool of exercises consists of 50 differential equations exercises, 30 geometry exercises and 10 statistic exercises.
Bob has no knowledge of these subjects, so he decides to memorise 20 differential equations exercises, 10 geometry exercises and 5 statistic exercises. During the exam, Bob only solves the exercises that he has memorised.

1. If the Maths teacher prepares the exam by randomly choosing, in the pool of exercises, 4 geometry exercises, what is the probability that Bob solves all 4 exercises?
   While the professor can choose the geometry questions in $\binom{30}{4}$ possible ways, Bob can answer all the geometry questions only if he memorised them, which happens in $\binom{10}{4}$ cases, therefore resulting in:

$$\mathbb{P}(\{\text{Bob solves all geometry questions}\}) = \frac{\binom{10}{4}}{\binom{30}{4}}$$

2. Assume instead that the Maths teacher prepares the exam by randomly choosing, in the pool of exercises, 5 differential equations exercises, 4 geometry exercises and 1 statistics exercises: How many different exam papers can the teacher prepare? (Exam papers containing the same exercises in different order are considered identical)
   Since ordering does not matter, the professor can choose the differential equation exercises in $\binom{50}{5}$ ways, the geometry exercises in $\binom{30}{4}$ ways and the statistics exercise in $\binom{10}{1}$ ways, meaning that, by applying the product

rule, he can prepare:

$$|\Omega| = \binom{50}{5} \cdot \binom{30}{4} \cdot \binom{10}{1} \text{ possible exam papers.}$$

3. What is the probability that Bob solves all 10 exercises?
   The event will happen only if all the questions were chosen among the ones Bob memorised, which can happen in $\binom{20}{5}$ cases for differential equation exercises, in $\binom{10}{4}$ cases for geometry exercises and in $\binom{5}{1}$ cases for statistics exercises, meaning that:

$$\mathbb{P}(\{\text{Bob solves all exercises}\}) = \frac{\binom{20}{5} \cdot \binom{10}{4} \cdot \binom{5}{1}}{\binom{50}{5} \cdot \binom{30}{4} \cdot \binom{10}{1}}$$

4. What is the probability that Bob solves 3 differential equations exercises, 2 geometry exercises and 1 statistics exercise?
   This event can happen if some questions are chosen among the ones Bob memorised whereas the remaining exercises are chosen among the ones Bob did not memorise.
   This choice can be done in $\binom{20}{3} \cdot \binom{30}{2}$ ways for differential equation exercises, in $\binom{10}{2} \cdot \binom{20}{4}$ ways for geometry exercises and in $\binom{5}{1} \cdot \binom{5}{0}$ ways for statistics exercises, meaning, that:

$$\mathbb{P}(\{\text{Bob solves exactly 3 DE, 2 G, 1 S}\}) = \frac{(\binom{20}{3} \cdot \binom{30}{2}) \cdot (\binom{10}{2} \cdot \binom{20}{4}) \cdot (\binom{5}{1} \cdot \binom{5}{0})}{\binom{50}{5}\binom{30}{4}\binom{10}{1}}$$

3.4 Exercise 4

Prove the inclusion/exclusion principle: if $A_1, \ldots, A_n$ are arbitrary events, then

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})$$

The statement is proved by induction: start by considering the base cases where $n \leq 3$:

1. Let $n = 1$: $\mathbb{P}(A_1) = (-1)^0 \cdot \mathbb{P}(A_1) = \mathbb{P}(A_1)$, so the statement is true.
2. Let $n = 2$.

   $$\mathbb{P}(A_1 \cup A_2) = ((-1)^0 \cdot \mathbb{P}(A_1)) + ((-1)^0 \cdot \mathbb{P}(A_2)) + ((-1)^1 \cdot \mathbb{P}(A_1 \cap A_2)) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$$

   Again, the statement is true.
3. By following the same logic, now let $n = 3$.

   $$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$$

   Again, the statement is true.

Now, assuming that the statement is true for some n, it must hold true for $n+1$ as well.

$$\mathbb{P}(\bigcup_{i=1}^{n+1} A_i) = \mathbb{P}((\bigcup_{i=1}^{n} A_i) \cup A_{n+1}) = \mathbb{P}(\bigcup_{i=1}^{n} A_i) + \mathbb{P}(A_{n+1}) - \mathbb{P}((\bigcup_{i=1}^{n}) \cap A_{n+1})$$

By applying distributivity, it is possible to re-write the equation above as follows:

$$\mathbb{P}(\bigcup_{i=1}^{n+1} A_i) = \mathbb{P}((\bigcup_{i=1}^{n} A_i) \cup A_{n+1}) = \mathbb{P}(\bigcup_{i=1}^{n} A_i) + \mathbb{P}(A_{n+1}) - \mathbb{P}(\bigcup_{i=1}^{n}(A_i \cap A_{n+1}))$$

Since the statement is true for n, it follows that:

$$\mathbb{P}((\bigcup_{i=1}^{n} A_i) \cup A_{n+1}) = (\sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbb{P}(\bigcap_{j=1}^{k} A_{i_j})) + \mathbb{P}(A_{n+1}) + (\sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbb{P}((\bigcap_{j=1}^{k} A_{i_j}) \cap A$$

Since the statement is proved to be true for $n+1$ as well, the inclusion/exclusion formula is therefore correct.

3.5 Exercise 5

Let S, S' be finite sets, with $|S| = n$ and $|S'| = k$. Answer the following questions for any $n, k \in \mathbb{N}$.

1. How many functions from S to S' are there?
   Generally speaking, a function $f : S \to S'$ associates to every input $s \in S$ an image $s' \in S'$: most particularly, for each of these n inputs there are k possible outputs, meaning that it is possible to build:

   $$k^n \text{ possible functions from } S \text{ to } S'.$$

2. How many strictly increasing functions from S to S' are there?
   Start by noticing that it is necessary to have $k \geq n$ because, otherwise, it would not be possible to create a strictly increasing function as, eventually, it will be possible to find some values $s_1, s_2 \in S$ such that $s_1 < s_2 \not\Rightarrow f(s_1) < f(s_2)$.
   However, just using combinatorics is not trivial, meaning that, given the set $A = \{f : S \to S' \mid f \text{ is strictly increasing}\}$, the best strategy is to find some set B such that, by creating a bijection between sets A and B, it is possible to recover $|A| = |B|$.
   Observe that it is possible to choose the set $B = \{\text{images of } n\text{-tuples}\}$ and that, since creating a n-tuple from the set S' of images is equivalent to choosing n values from a set of k elements, it is possible to state that $|B| = \binom{k}{n} = |A|$, meaning that it is possible to conclude that there are:

   $$\binom{k}{n} \text{ strictly increasing functions from } S \text{ to } S'.$$

3. How many injective functions from S to S' are there?
Generally speaking, a function $f : S \to S'$ is said to be injective if it holds that $f(s_1) = f(s_2) \Leftrightarrow s_1 = s_2$, which requires setting $k \geq n$ as, otherwise, injection is violated as it would be possible to find some $s_1 \neq s_2 \in S$ such that $f(s_1) = f(s_2)$.
For this reason, it is possible to state that, when creating an injective function, there are k possible output choices for the first input, $k-1$ output choices for the second input and so on and so forth until $k-(n-1)$ output choices for the n^{th} input, meaning that, by applying the product rule, it is possible to build:
$$k \cdot (k-1) \cdot \dots \cdot (k-n+1) = \frac{k!}{(k-n)!} \text{ injective functions from } S \text{ to } S'.$$

4. How many bijective functions from S to S' are there?
Generally speaking, a function $f : S \to S'$ is said to be bijective if it is both injective and surjective: since injection requires $k \geq n$ and surjection requires $k \leq n$, it should be noticed that a bijective function can be built if and only if $k = n$.
Most particularly, observe that, by assigning inputs and outputs like in the previous point, it is possible to obtain that the first input has k output choices, the second input has $k-1$ output choices and so on and so forth until $k-(n-1) = 1$ output choice for the n^{th} input (recall $k = n$).
For this reason, it is possible to conclude that a bijective function is basically built via a simple permutation, meaning that there are:
$$n! \text{ bijective functions from } S \text{ to } S'.$$

5. How many non-decreasing functions from S to S' are there?
Again, just using combinatorics would be too hard, meaning that, given the set $C = \{f : S \to S' | f \text{ is non-decreasing}\}$, the goal is to find some set D such that, by creating a bijection between C and D, it is possible to recover $|C| = |D|$.
Consider now the set of functions $D = \{g : S \to S''\}$, where g is obtained by maintaining $g(1) = f(1)$ and then, $\forall\ 1 < s \leq n$, $g(s) = f(s)+1$, meaning that the function g will have $k+(n-1)$ possible images, resulting in $|S''| = k+(n-1)$.
Most particularly, observe that, in this case, it is useful to define another set $D = \{g : S \to S'' \mid g \text{ is strictly increasing}\}$, meaning that, as seen in the second point, it is possible to recover that $|D| = \binom{k+(n-1)}{n}$ (stick-and-stars method) and therefore, since it is possible to create a bijection between sets C and D, it is possible to state that, actually, $|D| = \binom{k+(n-1)}{n} = |C|$, allowing to conclude that there are:
$$\binom{k+(n-1)}{n} \text{ non-decreasing functions from } S \text{ to } S'.$$

6. How many surjective functions from S to S' are there?
Generally speaking, a function $f : S \to S'$ is said to be surjective if it holds that, $\forall\ s' \in S'$, $\exists\ s \in S$ (not necessarily unique) such that $f(s) = s'$, which requires $k \leq n$ because otherwise surjection would be violated as some values would not be taken on by the function.
Most particularly, given the set $\Omega = \{f : S \to S'\}$, it is possible to view it as a union of the disjoint sets $F_S = \{f : S \to S' \mid f \text{ is surjective}\}$ and $F_N = \{f : S \to S' \mid f \text{ is not surjective}\}$, allowing to state that $|\Omega| = |F_S| + |F_N|$.
For this reason, let $F_{N_i} = \{f \in F_N \mid f \text{ does not take value } i\}$ and rewrite the set F_N as $F_N = \bigcup_{i=1}^{k} F_{N_i}$ and find its cardinality by using the inclusion/exclusion formula:

$$|F_N| = \sum_{i=1}^{k} (-1)^{i-1} \sum_{1 \leq i_1 < \cdots < i_j \leq k} |F_{N_{i_1}} \cap \cdots \cap F_{N_{i_j}}|$$

Therefore, it is possible to state that there are $|F_S| = |\Omega| - |F_N|$ surjective functions from S to S', with this quantity also being expressed as:

$$k^n - \sum_{i=1}^{k} (-1)^{i-1} \sum_{1 \leq i_1 < \cdots < i_j \leq k} |F_{N_{i_1}} \cap \cdots \cap F_{N_{i_j}}|$$

3.6 Exercise 6

Alice (A), Bob (B) and Caleb (C) play a tournament with the following rules. In the first round, A and B play against each other. The winner then plays against C, and if they win the game they win the tournament. If, instead C wins the game, then C plays against the loser of the first round, and so on and so forth. The first player who wins two consecutive rounds wins the tournament.
It is known that A, B and C have the same skills, so each game is equally likely to be won by any of the two players.

1. Is any of the players favoured by the tournament rules?
Yes: Alice and Bob are slightly more favoured to win because they have an extra round to play in compared to Caleb.

2. Compute the probability that the game after n rounds, $n \geq 2$.
Assume Alice wins the first round: if she also wins the second round, then she wins the tournament, whereas if she loses the second round, then the tournament will continue with a new leader (in this case Caleb) and so on and so forth.
For this reason, for the tournament to end after n rounds, it must hold that the first round can have any outcome, then, from round 2 to round $n-2$, the current leader must always lose and, in the end, the leader must

win both rounds $n-1$ and n.
Knowing that each player is equally likely to win the first round, it is possible to let:

$$\mathbb{P}(\{\text{end after } n \text{ rounds}\}) = \mathbb{P}(\{\text{leader loses 2 to } n-2 \text{ and wins } n-1 \text{ and } n\})$$

Since each player is equally likely to win each round, this results in:

$$\mathbb{P}(\{\text{the game ends after } n \text{ rounds}\}) = (\frac{1}{2})^{n-2} \cdot (\frac{1}{2}) = \frac{1}{2^{n-1}}$$

3. Compute the winning probability for A, B and C.
   Let $p_A, p_B, p_C \in (0,1)$ be the winning probabilities for player A, B, C respectively: by the symmetry recovered in the first point, it is possible to state that $p_A = p_B = p$ but $p_C \neq p$ as it was shown that Alice and Bob are slightly more favoured than Caleb.
   So, knowing that the first round is always played between Alice and Bob, assume without loss of generality, that Alice won the first round: Caleb will win the tournament only if he beats Alice in the second round and then beats Bob as well in the third round, but, if Caleb loses against Bob, then this pattern will repeat itself.
   Therefore, since each player is equally likely to win a single round, it is possible to express Caleb's winning probability as:

$$p_C = (\frac{1}{2} \cdot \frac{1}{2}) + (\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot p_C) \Rightarrow p_C = \frac{1}{4} + \frac{1}{8} \cdot p_C \Rightarrow p_C = \frac{2}{7}$$

   Most particularly, knowing that p_A, p_B, p_C are mutually exclusive events, it is possible to state that:

$$p_A + p_B + p_C = 1 \Rightarrow p + p + \frac{2}{7} = 1 \Rightarrow 2p = \frac{5}{7} \Rightarrow p_A = p_B = p = \frac{5}{14}$$

   Therefore, it is possible to recover that Alice and Bob are indeed slightly more likely to win the tournament compared to Caleb.

4. Can the tournament last forever?
   Yes, as long as no one manages to win two consecutive rounds.

Chapter 4

Exercise Sheet 4: Event independence and probability distributions

4.1 Exercise 1

Let A and B be events with $\mathbb{P}(A) = 0.3$, $\mathbb{P}(A \cup B) = 0.5$ and $\mathbb{P}(B) = p$.
Find the value of p in the following cases:

1. A and B are disjoint.
   Since $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, resulting in:

$$0.5 = 0.3 + p \Leftrightarrow p = 0.2$$

2. A and B are independent.
   Let $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$: by independence, it must hold that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, meaning that:

$$0.5 = 0.3 + p - 0.3p \Leftrightarrow 0.7p = 0.2 \Leftrightarrow p = \frac{2}{7}$$

3. A is a subset of B.
   If $A \subseteq B$, then $A \cup B = B$, meaning that $\mathbb{P}(A \cup B) = \mathbb{P}(B) \Rightarrow p = 0.5$.
   Alternatively, by following the same logic, $A \subseteq B \Rightarrow A \cap B = A$, resulting in:

$$0.5 = 0.3 + p - 0.3 \Leftrightarrow p = 0.5$$

4.2 Exercise 2

Let A, B and C be three independent events: Prove that the following events are independent:

1. A^C, B, C
   Start by showing that A^C and B are independent events: start by writing B as the union of disjoint sets so that, by additivity, it must hold that:

   $$B = (A \cap B) \cup (A^C \cap B) \Rightarrow \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^C \cap B)$$

   However, since A and B are independent events by assumption, it is possible to rewrite $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, allowing to rearrange the previous equation as:

   $$\mathbb{P}(A^C \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(A^C)\mathbb{P}(B)$$

   Therefore, the events A^C and B are independent, and, without loss of generality, this statement can be proved for any couple of events.
   At this point, it is possible to apply the same reasoning to show that the events A^C and $B \cap C$ are independent: start by writing $B \cap C$ as the union of disjoint sets, so that, by additivity, it must hold that:

   $$B \cap C = ((B \cap C) \cap A) \cup ((B \cap C) \cap A^C) \Rightarrow \mathbb{P}(B \cap C) = \mathbb{P}((B \cap C) \cap A) + \mathbb{P}((B \cap C) \cap A^C)$$

   However, since A, B and C are independent events by assumption, it is possible to recover that they are pairwise independent as well (so $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$), resulting in:

   $$\mathbb{P}(B)\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C)\mathbb{P}(A) + \mathbb{P}((B \cap C) \cap A^C), \text{ resulting in:}$$
   $$\mathbb{P}(A^C \cap (B \cap C)) = \mathbb{P}(B)\mathbb{P}(C)(1 - \mathbb{P}(A)) = \mathbb{P}(A^C)\mathbb{P}(B)\mathbb{P}(C)$$

2. A^C, B^C, C $(*)$
   Let $A^C \cap B^C \cap C = C \setminus (A \cup B)$, meaning that, by applying distributivity, it is possible to state that:

   $$\mathbb{P}(A^C \cap B^C \cap C) = \mathbb{P}(C \setminus (A \cup B)) = \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) = \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C))$$

   At this point, it is possible to apply the inclusion/exclusion formula and, by the assumption that A, B, C are independent events, it is possible to conclude that:

   $$\mathbb{P}(A^C \cap B^C \cap C) = \mathbb{P}(C) - (\mathbb{P}(A)\mathbb{P}(C) + \mathbb{P}(B)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)) = \mathbb{P}(A^C)\mathbb{P}(B^C)\mathbb{P}(C)$$

3. A^C, B^C, C^C $(*)$
   Start by showing that A^C and B^C are pairwise independent: by applying De Morgan's laws, rewrite $A^C \cap B^C = (A \cup B)^C$ and, by applying the inclusion/exclusion formula and by exploiting the assumption that A, B, C are

independent events, observe that the probability of this event factorises:

$$\mathbb{P}(A^C \cap B^C) = \mathbb{P}((A \cup B)^C) = 1 - \mathbb{P}(A \cup B)$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(A^C \cap B^C) = 1-(\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A)\mathbb{P}(B)) = 1-\mathbb{P}(A)-\mathbb{P}(B)(1-\mathbb{P}(A)) = \mathbb{P}(A^C)\mathbb{P}(B^C)$$

Similarly, it is possible to show this statement for three events: by applying again De Morgan's laws, let $A^C \cap B^C \cap C^C = (A \cup B \cup C)^C$ and apply inclusion/exclusion formula:

$$\mathbb{P}(A^C \cap B^C \cap C^C) = \mathbb{P}((A \cup B \cup C)^C) = 1 - \mathbb{P}(A \cup B \cup C)$$

Ultimately, the computations will factorise like in the case with two sets, allowing to conclude that:

$$\mathbb{P}(A^C \cap B^C \cap C^C) = (1-\mathbb{P}(A))(1-\mathbb{P}(B))(1-\mathbb{P}(C)) = \mathbb{P}(A^C)\mathbb{P}(B^C)\mathbb{P}(C^C)$$

4.3 Exercise 3

Two standard dice are rolled.

1. Show that the event "the sum of the dice is 9" is not independent of the outcome of the first die.
   Given the event $A = \{\text{the sum of the dice is } 9\}$, consider a counterexample with the event $B = \{\text{the first die gives } 1\}$.
   Notice that, in this case, $A \cap B = \emptyset$, meaning that $\mathbb{P}(A \cap B) = 0$, which, however, contradicts the condition of independence because:

   $$\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{9} \cdot \frac{1}{6} = \frac{1}{54} \neq 0 = \mathbb{P}(A \cap B)$$

   N.B.: The same reasoning can be if $B = \{\text{the first die gives } 2\}$.

2. Show that the event "the sum of the dice is 7" is independent of the outcome of the first die.
   Given the event $A = \{\text{the sum of the dice is } 7\}$, the goal is to show that, given the event $B = \{\text{the first die gives } k\}$, it must hold that:

   $$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \;\forall\; k \in \{1, 2, 3, 4, 5, 6\}$$

   Most particularly, observe that, for all values of k, $\mathbb{P}(B) = \frac{1}{6}$, and that it is possible to apply additivity by rewriting the event A as the union of the following mutually exclusive events:

   $$\mathbb{P}(A) = \mathbb{P}(\{1, 6\} \cup \{2, 5\} \cup \{3, 4\} \cup \{4, 3\} \cup \{5, 2\} \cup \{6, 1\}) = \frac{1}{6}$$

   Furthermore, observe that $A \cap B = \{(k, 7-k)\} \;\forall\; k \in \{1, 2, 3, 4, 5, 6\}$, meaning that $\mathbb{P}(A \cap B) = \frac{1}{36}$, allowing to conclude, for all the chosen k, that:

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}(A \cap B)$$

By showing that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, it is therefore possible to conclude that the probability of the two dice summing up to 7 is independent of the result of the first die.

3. Give an intuitive explanation of why the two above cases are different. Just look at the combinations that allow the events to happen.

4.4 Exercise 4

When 3 horses compete in a race, their winning probabilities are 0.3, 0.5 and 0.2 respectively. They compete in 3 consecutive races. Compute the probability of the following events:

1. The same horse wins all races
   Write the event $A = \{\text{the same horse wins all three races}\}$ as the union of disjoint events in order to apply additivity, meaning that:

   $$A = \bigcup_{i=1}^{3} A_i = \{\text{horse } i \text{ wins all three races}\} \Rightarrow \mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$$

   Most particularly, observe that each of the three events can be modelled as a Bernoulli distribution of three successes, meaning that:

   $$\mathbb{P}(A_i) = p_i^3, \text{ therefore resulting in:}$$

   $$\mathbb{P}(\{\text{the same horse wins all three races}\}) = (0.3)^3 + (0.5)^3 + (0.2)^3$$

2. Each horse wins exactly one race.
   Assume that horse A wins the first race, horse B wins the second race and horse C wins the third race: since these events are independent, it is possible to state that:

   $$\mathbb{P}(\{1\text{A}, 2\text{B}, 3\text{C}\}) = 0.3 \cdot 0.5 \cdot 0.2$$

   However, the horses can actually win a competition each in any order: this ordering can be done as a simple permutation of three elements which can be done in $3! = 6$ possible ways, resulting in:

   $$\mathbb{P}(\{\text{each horse wins exactly one race}\}) = 6 \cdot 0.3 \cdot 0.5 \cdot 0.2$$

4.5 Exercise 5

A rocket hits its target with probability $\frac{1}{3}$.

1. If 3 rockets are fired, what is the probability that at least one of them hits the target?
   Write the event in terms of its complement, meaning that:

   $$\mathbb{P}(\{\text{at least one rocket hits the target}\}) = 1 - \mathbb{P}(\{\text{all rockets miss the target}\})$$

   Most particularly, the event that all rockets miss the target can be modelled as a Bernoulli distribution of three failures, meaning that:

   $$\mathbb{P}(\{\text{all rockets miss the target}\}) = (1 - \frac{1}{3})^3 = (\frac{2}{3})^3 = \frac{8}{27}, \text{ resulting in:}$$
   $$\mathbb{P}(\{\text{at least one rocket hits the target}\}) = 1 - \frac{8}{27} = \frac{19}{27}$$

2. Find the minimum number of rockets which need to be fired in order to guarantee that the probability that at least one of them hits the target is at least $\frac{9}{10}$.
   Consider the event in terms of its complement:

   $$\mathbb{P}(\{\text{at least one rocket hits the target}\}) \geq \frac{9}{10} \Leftrightarrow \mathbb{P}(\{\text{all rockets miss the target}\}) \leq \frac{1}{10}$$

   Most particularly, assuming n denotes the number of rockets to fire such that the probability that at least one of them will hit the target is at least $\frac{9}{10}$, it is possible to model the event that all of them miss the target as a Bernoulli distribution of n failures, meaning that:

   $$(\frac{2}{3})^n \leq \frac{1}{10} \Rightarrow n \geq \frac{\ln 10}{\ln \frac{3}{2}}$$

   Therefore, at least $\lceil \frac{\ln 10}{\ln \frac{3}{2}} \rceil$ rockets must be fired to satisfy the condition.

4.6 Exercise 6

Albert plays 10 rounds at roulette betting on red 1 euro each round. According to the casino rules, the probability of winning each round is $\frac{18}{37}$.

1. Compute the probability that Albert wins for the first time at the fifth round.
   Since this event can be modelled as a geometric distribution in which the first success is recorded at the fifth attempt (or, similarly, as a Bernoulli distribution of four failures followed by a success), it is possible to state that:

   $$\mathbb{P}(\{\text{Albert first wins at the fifth round}\}) = (1 - \frac{18}{37})^4 \cdot \frac{18}{37} = (\frac{19}{37})^4 \cdot \frac{18}{37}$$

2. Compute the probability that Albert wins at least two rounds.
   In this case, it is easier to reason in terms of the complement, meaning that:

   $$\mathbb{P}(\{\text{Albert wins at least two rounds}\}) = 1-\mathbb{P}(\{\text{Albert wins less than two rounds}\})$$

   Most particularly, if Albert wins less than two rounds, he either wins one round only or no rounds at all, meaning that, since the two cases are mutually exclusive, it is possible to compute:

   $\mathbb{P}(\{\text{less than two rounds}\}) = \mathbb{P}(\{\text{no rounds}\})+\mathbb{P}(\{\text{exactly one round}\})$, so:

   $\mathbb{P}(\{\text{Albert wins less than two rounds}\}) = (\frac{19}{37})^{10}+10\cdot\frac{18}{37}\cdot(\frac{19}{37})^{9}$, resulting in:

   $$\mathbb{P}(\{\text{Albert wins at least two rounds}\}) = 1 - [(\frac{19}{37})^{10} + 10 \cdot \frac{18}{37} \cdot (\frac{19}{37})^{9}]$$

   Alternatively, it is possible to represent the event that Albert wins at least two rounds as the following union of disjoint events:

   $$\mathbb{P}(\{\text{Albert wins at least two rounds}\}) = \bigcup_{k=2}^{10} A_k = \{\text{Albert wins exactly } k \text{ rounds}\}$$

   Most particularly, each of these sub-events can be modelled as a binomial distribution, meaning that:

   $\mathbb{P}(\{\text{Albert wins exactly } k \text{ rounds}\}) = \binom{10}{k}\cdot(\frac{18}{37})^{k}\cdot(\frac{19}{37})^{10-k}$, resulting in:

   $$\mathbb{P}(\{\text{Albert wins at least two rounds}\}) = \sum_{k=2}^{10}\binom{10}{k} \cdot (\frac{18}{37})^{k} \cdot (\frac{19}{37})^{10-k}$$

3. Compute the probability that after the 10 rounds the net gain by Albert is 2 euros.
   Assuming Albert gains one euro by winning a round and loses one euro by losing a round, if w denotes the rounds Albert won and l denotes the rounds Albert lost, it must hold that:

   $$\begin{cases} w - l = 2 \\ w + l = 10 \end{cases} \Leftrightarrow \begin{cases} w = 6 \\ l = 4 \end{cases}$$

   Therefore, it is possible to rewrite the event as:

   $\mathbb{P}(\{\text{Albert's net gain is 2 euros}\}) = \mathbb{P}(\{\text{Albert wins exactly six rounds}\})$, so:

   $$\mathbb{P}(\{\text{Albert's net gain is 2 euros}\}) = \binom{10}{6} \cdot (\frac{18}{37})^{6} \cdot (\frac{19}{37})^{4}$$

4.7 Exercise 7

For $n \in \mathbb{N}$ and $p \in (0, 1)$ consider the binomial distribution (number of heads in n biased coin tosses):

$$\mathbb{P}(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, \dots, n.$$

Prove that $\mathbb{P}(k)$ is increasing for $k \leq \overline{k}$, with $\overline{k} = \overline{k}(n, p)$ to be determined, and decreasing for $k > \overline{k}$.
By definition, $\mathbb{P}(k)$ is non-decreasing if $\mathbb{P}(k) \leq \mathbb{P}(k+1)$, so consider the following inequality:

$$\binom{n}{k} p^k (1-p)^{n-k} \leq \binom{n}{k+1} p^{k+1} (1-p)^{n-(k+1)} \Rightarrow \frac{1-p}{n-k} \leq \frac{p}{k+1}, \text{ from which:}$$

$$(k+1)(1-p) \leq p(n-k) \Rightarrow k(1-p+p) \leq np - (1-p) \Rightarrow k \leq np - (1-p)$$

Therefore, by setting the mode $\overline{k} = np - (1-p)$ to be an integer, it is possible to recover that, indeed, $\mathbb{P}(k)$ is non-decreasing for $k \leq \overline{k}$.

4.8 Exercise 8

1. Let B, N, $n \in \mathbb{N}$. Prove, via a probabilistic interpretation, the identity:

$$\sum_{k=0}^{n} \binom{B}{k} \binom{N}{n-k} = \binom{N+B}{n}$$

   The left side of the equation represents all the possible ways of picking $0 \leq k \leq n$ elements from B items, followed by picking $n-k$ elements from N items, with both choices being made without ordering nor repetitions. On the other hand, the right side of the equation represents the way of choosing n elements from a set of $N + B$ items, but, supposing that k of these elements are chosen from the B items whereas the remaining $n - k$ elements are chosen from the N elements, it is possible to state that:

$$\binom{B+N}{k+(n-k)} = \binom{N+B}{n} = \sum_{k=0}^{n} \binom{B}{k} \binom{N}{n-k}$$

2. Alice and Bob toss a fair coin n times each. Compute the probability that they obtain the same number of heads.
   Rewrite the event as a union of mutually exclusive events:

$$\{\text{they see the same number of heads}\} = \bigcup_{k=0}^{n} A_k = \{\text{they both see } k \text{ heads}\}$$

By additivity, it holds that:

$$\mathbb{P}(\text{they see the same number of heads}) = \sum_{k=0}^{n} \mathbb{P}(A_k)$$

Most particularly, Alice's coin tosses are independent from Bob's coin tosses, meaning that it is possible to state that:

$$\mathbb{P}(\{\text{they both see } k \text{ heads}\}) = \mathbb{P}(\{\text{Alice sees } k \text{ heads}\})\mathbb{P}(\{\text{Bob sees } k \text{ heads}\})$$

These sub-events can be modelled by means of a binomial distribution:

$$\mathbb{P}(\{\text{Alice sees } k \text{ heads}\}) = \mathbb{P}(\{\text{Bob sees } k \text{ heads}\}) = \binom{n}{k} \cdot (\frac{1}{2})^k (\frac{1}{2})^{n-k}$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(\{\text{they see the same number of heads}\}) = \sum_{k=0}^{n} [\binom{n}{k} (\frac{1}{2})^k \cdot (\frac{1}{2})^{n-k}]^2, \text{ so:}$$

$$\mathbb{P}(\{\text{they see the same number of heads}\}) = (\frac{1}{2})^{2n} \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot \binom{n}{n-k}$$

N.B.: By using the identity that was proved in the first point, it is possible to further simplify the solution as:

$$\mathbb{P}(\{\text{they see the same number of heads}\}) = (\frac{1}{2})^{2n} \cdot \binom{2n}{n}$$

Chapter 5

Exercise Sheet 5: Conditional probability

5.1 Exercise 1

A binary signal is transmitted via a channel. Due to the background noise, it may happen that when 0 is transmitted 1 is received, and similarly it may happen that when 1 is transmitted 0 is received. Assume that:

- The probability that 0 is received correctly is 0.94.
- The probability that 1 is received correctly is 0.91.

A single bit is transmitted, which is 0 with probability 0.45 and 1 with probability 0.55. Compute:

1. The probability of receiving 1.
   A "one" can be received either if a "one" was transmitted correctly or a "zero" was flipped during transmission, so, by applying the law of total probability, it is possible to state that:

   $\mathbb{P}(R = 1) = \mathbb{P}(R = 1|T = 1)\mathbb{P}(T = 1) + \mathbb{P}(R = 1|T = 0)\mathbb{P}(T = 0)$, resulting in:

   $\mathbb{P}(R = 1) = (0.91 \cdot 0.55) + ((1 - 0.94) \cdot 0.45) = (0.91 \cdot 0.55) + (0.06 \cdot 0.45)$

2. The probability of receiving 0.
   Similarly to the previous point, a "zero" can be received either if a "zero" was transmitted correctly or a "one" was flipped during transmission, so, by applying the law of total probability, it is possible to state that:

   $\mathbb{P}(R = 0) = \mathbb{P}(R = 0|T = 0)\mathbb{P}(T = 0) + \mathbb{P}(R = 0|T = 1)\mathbb{P}(T = 1)$, resulting in:

   $\mathbb{P}(R = 0) = (0.94 \cdot 0.45) + ((1 - 0.91) \cdot 0.55) = (0.94 \cdot 0.45) + (0.09 \cdot 0.55)$

3. The probability that 1 was transmitted, given that 1 was received.
   Apply Bayes' theorem and reason in terms of the "reverse condition":

$$\mathbb{P}(T=1|R=1) = \frac{\mathbb{P}(R=1|T=1)\mathbb{P}(T=1)}{\mathbb{P}(R=1)} = \frac{0.91 \cdot 0.55}{(0.91 \cdot 0.55) + (0.06 \cdot 0.45)}$$

4. The probability that 0 was transmitted, given that 0 was received.
   Similarly to the previous point, apply again Bayes' theorem and consider the "reverse condition":

$$\mathbb{P}(T=0|R=0) = \frac{\mathbb{P}(R=0|T=0)\mathbb{P}(T=0)}{\mathbb{P}(R=0)} = \frac{0.94 \cdot 0.45}{(0.94 \cdot 0.45) + (0.09 \cdot 0.55)}$$

5. The probability that the signal is wrongly received.
   By using the law of total probability, write the event as a union of disjoint events, meaning that:

   $\mathbb{P}(R \neq T) = \mathbb{P}(R=1|T=0)\mathbb{P}(T=0)+\mathbb{P}(R=0|T=1)\mathbb{P}(T=1)$, resulting in:

$$\mathbb{P}(R \neq T) = (0.06 \cdot 0.45) + (0.09 \cdot 0.55)$$

5.2 Exercise 2

Alice and Bob each toss a biased coin: Alice's coin gives head with probability $\frac{1}{3}$, while Bob's coin gives head with probability $\frac{1}{4}$.

1. Compute the probability that both Alice and Bob get head.
   Since the two tosses are independent, it is possible to factorise the event as:

$$\mathbb{P}(\{\text{both see H}\}) = \mathbb{P}(\{\text{Alice sees H}\}) \cdot \mathbb{P}(\{\text{Bob sees H}\}) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$$

2. Compute the probability that they get exactly one head.
   Write the event as a union of mutually exclusive cases and apply additivity:

$$\mathbb{P}(\{\text{exactly one head}\}) = \mathbb{P}(\{\text{Alice H}\}\cap\{\text{Bob T}\})+\mathbb{P}(\{\text{Alice T}\}\cap\{\text{Bob H}\})$$

   However, since Alice and Bob's coin tosses are independent from each other, it is possible to factorise the two events, resulting in:

$$\mathbb{P}(\{\text{exactly one head}\}) = \frac{1}{3} \cdot (1-\frac{1}{4}) + (1-\frac{1}{3}) \cdot \frac{1}{4} = \frac{5}{12}$$

3. Knowing that the coin toss resulted in one head and one tail, compute the probability that Alice got head.
   Apply the definition of conditional probability:

$$\mathbb{P}(\{\text{Alice H}|\text{exactly one head}\}) = \frac{\mathbb{P}(\{\text{Alice H}\} \cap \{\text{Bob T}\})}{\mathbb{P}(\{\text{exactly one head}\})} = \frac{\frac{1}{3} \cdot (1-\frac{1}{4})}{\frac{5}{12}} = \frac{3}{5}$$

5.3 Exercise 3

It is known that twins can be identical, in which case they are necessarily of the same sex, or non-identical, in which case they are of the same sex in 50% of the cases. Let p denote the probability that the twins are identical.

1. Compute, as a function of p, the probability that two twins are identical, knowing that they are of the same sex.
   Apply Bayes' theorem and consider the reverse condition:

   $$\mathbb{P}(\{\text{identical}|\text{same sex}\}) = \frac{\mathbb{P}(\{\text{same sex}|\text{identical}\})\mathbb{P}(\{\text{identical}\})}{\mathbb{P}(\{\text{same sex}\})}$$

   Most particularly, let $\mathbb{P}(\{\text{same sex}|\text{identical}\}) = 1$ since it represents a necessary condition, while, by applying the law of total probability, it is possible to obtain that:

   $$\mathbb{P}(\{\text{same sex}\}) = \mathbb{P}(\{\text{same sex and identical}\}) + \mathbb{P}(\{\text{same sex and not identical}\})$$

   By applying the definition of conditional probability, it is possible to state that:

   $$\mathbb{P}(\{\text{same sex and identical}\}) = \mathbb{P}(\{\text{same sex}|\text{identical}\})\mathbb{P}(\{\text{identical}\}) = p$$

   $$\mathbb{P}(\{\text{same sex and not identical}\}) = \mathbb{P}(\{\text{same sex}|\text{not identical}\})\mathbb{P}(\{\text{not identical}\}) = \frac{1-p}{2}$$

   It is therefore possible to conclude that:

   $$\mathbb{P}(\text{identical}|\text{same sex}) = \frac{p}{p + \frac{1-p}{2}} = \frac{2p}{p+1}$$

2. Compute, as a function of p, the probability that two twins are not of the same sex.
   By applying the law of total probability, write the event as the union of two mutually exclusive cases:

   $$\mathbb{P}(\{\text{opposite sex}\}) = \mathbb{P}(\{\text{opposite sex and identical}\}) + \mathbb{P}(\{\text{opposite sex and not identical}\})$$

   However, since $\mathbb{P}(\{\text{opposite sex and identical}\}) = 0$ because it is an impossible event, the previous equation simplifies to:

   $$\mathbb{P}(\{\text{opposite sex}\}) = \mathbb{P}(\{\text{opposite sex}|\text{not identical}\})\mathbb{P}(\{\text{not identical}\}) = \frac{1-p}{2}$$

5.4 Exercise 4

In a factory three machines A, B, and C make respectively 40%, 10% and 50% of the produced items. The respective percentages of faulty items are 2%, 3% and 4%. Pick an item at random.

1. Compute the probability that the item is faulty.
   By applying the law of total probability, it holds that:

$$\mathbb{P}(F) = \mathbb{P}(F|A)\mathbb{P}(A) + \mathbb{P}(F|B)\mathbb{P}(B) + \mathbb{P}(F|C)\mathbb{P}(C) = \frac{31}{1000}$$

2. Knowing that the item is faulty, compute the probability that it was produced by machine A, B or C.
   Apply Bayes' theorem to consider the reverse condition for all three cases:

   - Case A

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(F|A)\mathbb{P}(A)}{\mathbb{P}(F)} = \frac{\frac{2}{100} \cdot \frac{40}{100}}{\frac{31}{1000}} = \frac{8}{31}$$

   - Case B

$$\mathbb{P}(B|F) = \frac{\mathbb{P}(F|B)\mathbb{P}(B)}{\mathbb{P}(F)} = \frac{\frac{3}{100} \cdot \frac{10}{100}}{\frac{31}{1000}} = \frac{3}{31}$$

   - Case C

$$\mathbb{P}(C|F) = \frac{\mathbb{P}(F|C)\mathbb{P}(C)}{\mathbb{P}(F)} = \frac{\frac{4}{100} \cdot \frac{50}{100}}{\frac{31}{1000}} = \frac{20}{31}$$

5.5 Exercise 5

An urn contains three coins: the first coin is fair, and has head (H) one one side and tail (T) on the other, the second coin has H on both sides and the third coin has T on both sides. A coin is chosen at random from the urn, and it is tossed without looking at which one it is.

1. Compute the probability that the coin toss gives H.
   Assume HT denotes the fair coin, HH denotes the coin with two heads and TT denotes the coin with two tails: by applying the law of total probability, it is possible to rewrite the event as the union of mutually exclusive events:

$$\mathbb{P}(H) = \mathbb{P}(H|HT)\mathbb{P}(HT) + \mathbb{P}(H|HH)\mathbb{P}(HH) + \mathbb{P}(H|TT)\mathbb{P}(TT)$$

   Most particularly, observe that $\mathbb{P}(H|HH) = 1$ because it is a certain event, whereas $\mathbb{P}(H|TT) = 0$ because it is an impossible event.
   In addition. let $\mathbb{P}(HT) = \mathbb{P}(HH) = \mathbb{P}(TT) = \frac{1}{3}$ by symmetry (remember that the coin is chosen at random), meaning that it is possible to conclude that:

$$\mathbb{P}(H) = (\frac{1}{2} \cdot \frac{1}{3}) + (1 \cdot \frac{1}{3}) + (0 \cdot \frac{1}{3}) = \frac{1}{2}$$

2. Given that the coin toss gave H, compute the probability that on the other side of the coin there is T.
   Apply Bayes' theorem and reason in terms of the reverse condition:

$$\mathbb{P}(HT|H) = \frac{\mathbb{P}(H|HT)\mathbb{P}(HT)}{\mathbb{P}(H)} = \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

3. Knowing that the coin toss gave H, it is picked up and, without looking at the other side, tossed again. Compute the probability that the coin gives H again.
   Apply the definition of conditional probability:

$$\mathbb{P}(\{\text{two } H|\text{first } H\}) = \frac{\mathbb{P}(\text{two } H \text{ and first } H)}{\mathbb{P}(\text{first } H)}$$

   Most particularly, by applying the law of total probability, it is possible to express the event of getting head followed by another head as the union of two mutually exclusive events:

$$\mathbb{P}(\{\text{two } H \text{ and first } H\}) = \mathbb{P}(\{\text{two } H|HT\})\mathbb{P}(HT)+\mathbb{P}(\{\text{two } H|HH\})\mathbb{P}(HH), \text{ so:}$$

$$\mathbb{P}(\{\text{two } H \text{ and first } H\}) = (\frac{1}{2}\cdot\frac{1}{2})\cdot\frac{1}{3} + (1\cdot 1)\cdot\frac{1}{3} = \frac{5}{12}$$

   Therefore, it is possible to conclude that:

$$\mathbb{P}(\{\text{two } H \text{ and first } H\}) = \frac{\frac{5}{12}}{\frac{1}{2}} = \frac{5}{6}$$

5.6 Exercise 6

Let S be a set of cardinality n. Pick two subsets of S at random.
Compute the probability that the first set is a subset of the second set.
Start by noticing that the two subsets of S can be chosen at random in $\binom{2^n}{2}$ possible ways.
Now, assuming S_1 and S_2 are the chosen subsets, it is possible to notice that the event "$S_1 \subseteq S_2$" can be written as the following union of disjoint events:

$$\mathbb{P}(\bigcup_{k=0}^{n}\{S_1 \subseteq S_2 \wedge |S_2| = k\}) = \sum_{k=0}^{n}\mathbb{P}(\{S_1 \subseteq S_2\} \cap \{|S_2| = k\})$$

By applying the law of total probability, it is also possible to write the aforementioned event as:

$$\mathbb{P}(\bigcup_{k=0}^{n}\{S_1 \subseteq S_2 \wedge |S_2| = k\}) = \sum_{k=0}^{n}\mathbb{P}(S_1 \subseteq S_2||S_2| = k)\cdot\mathbb{P}(|S_2| = k)$$

Mots particularly, knowing that a subset of S containing k elements can be built in $\binom{n}{k}$ possible ways, it is possible to state that the probability of choosing a subset S_2 of cardinality k can be computed in the following way:

$$\mathbb{P}(|S_2| = k) = \frac{\binom{n}{k}}{2^n}$$

Furthermore, it is possible to state that $S_1 \subseteq S_2$ if and only if S_1 is chosen among the subsets of S_2: knowing that $|S_2| = k \Rightarrow |P(S_2)| = 2^k$, it is possible to recover that:

$$\mathbb{P}(S_1 \subseteq S_2 || S_2| = k) = \frac{2^k}{2^n}$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(S_1 \subseteq S_2) = \sum_{k=0}^{n} \frac{2^k}{2^n} \cdot \frac{\binom{n}{k}}{2^n} = \frac{1}{2^{2n}} \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot 2^k$$

N.B.: By writing $2^k = 2^k \cdot 1^{n-k}$, the professor exploited the Newton binomial to further simplify the result as the following:

$$\mathbb{P}(S_1 \subseteq S_2) = \frac{(2+1)^n}{2^{2n}} = (\frac{3}{4})^n$$

5.7 Exercise 7

Three roads connect the houses A, B and C in such a way that from each house one can get to any other house with a direct path. Due to bad weather, the roads may be (temporarily) closed.
Let $p_{AB} \in (0,1)$ (respectively p_{BC}, p_{AC}) denote the probability that the road linking A and B (respectively, B and C, A and C) is open. You can assume that each road is open or closed independently of the state of the other roads.
You are at house A.

1. Compute the probability that you can get to house C.
   There are two ways to reach house C: either road $\vec{AC}$ is open and it is possible to directly reach house C, or road $\vec{AC}$ is closed so you need roads $\vec{AB}$ and $\vec{BC}$ to be both open to reach house C from house B, resulting in, by applying conditional probability and event independence, in the following result:

   $$\mathbb{P}(A \to C) = p_{AC} + p_{AB}p_{BC}(1 - p_{AC})$$

2. Someone told you that it is not possible to get to house C due to bad weather. Compute the probability that you can get to house B.
   In this case, it must hold that both roads $\vec{BC}$ and $\vec{AC}$ are closed, whereas road $\vec{AB}$ must be open.
   Therefore, by applying conditional probability, it is possible to obtain:

   $$\mathbb{P}(A \to B | A \nrightarrow C) = \frac{\mathbb{P}(A \to B \cap A \nrightarrow C)}{\mathbb{P}(A \nrightarrow C)}$$

Most particularly, observe that, by independence, it is possible to recover that:

$$\mathbb{P}(A \to B | A \nrightarrow C) = p_{AB}(1-p_{BC})(1-p_{AC}) \text{ and } \mathbb{P}(A \nrightarrow C) = 1-\mathbb{P}(A \to C)$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(A \to B | A \nrightarrow C) = \frac{p_{AB}(1-p_{BC})(1-p_{AC})}{1-(p_{AC}+p_{AB}p_{BC}(1-p_{AC}))}$$

3. Now suppose that between A and B there are three direct paths, each one open with probability q independently from the others: compute again the above probabilities, without redoing the computations from scratch.
   In this case, the value p_{AB} will be different, and it is possible to reason in terms of the event's complement, resulting in:

   $$p_{AB} = 1 - \mathbb{P}(\{\text{all roads } \vec{AB} \text{ are closed}\}) = 1-(1-q)^3$$

   At this point, it will be possible to find the above probabilities by plugging in the new value of p_{AB} in the previous formulae.

Chapter 6

Exercise Sheet 6: Conditional probability

6.1 Exercise 1

Your goal is to collect n coupons for your album. What is the probability that you will do so by buying k coupons ($k \geq n$)? You may use the uniform probability measure on the k-ple of coupons you buy.
Your goal is to collect n coupons for your album. What is the probability that you will do so by buying k coupons ($k \geq n$)? (You may use the uniform probability measure on the k-ple of coupons you buy.)
Start by noticing that each purchased coupon allows for n possible choices, meaning that there will be n^k possible combinations of coupons that you can buy.
At this point, consider the event $A = \{\text{all coupons}\}$ in terms of the complement:

$$A^C = \bigcup_{j=1}^{n} A_j^C = \{\text{coupon } j \text{ is missing}\}$$

However, since these sub-events are not mutually exclusive (it is possible to miss both coupon 1 and 3, for instance), the probability of A^C has to be calculated using the inclusion/exclusion formula, meaning that:

$$\mathbb{P}(A^C) = \sum_{j=1}^{n} (-1)^{j-1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_j})$$

Most particularly, observe that, for a generic $A_{i_1} \cap \cdots \cap A_{i_j}$, it holds that:

$$|A_{i_1} \cap \cdots \cap A_{i_j}| = (n-j)^k, \text{ resulting in:}$$

$$\mathbb{P}(A^C) = \sum_{j=1}^{n} (-1)^{j-1} \sum_{i \leq i_1 < \cdots < i_j \leq n} \frac{(n-j)^k}{n^k} = \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} \cdot (\frac{n-j}{n})^k$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^C) = 1 - [\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j} \cdot (\frac{n-j}{n})^k]$$

6.2 Exercise 2 ($*$)

Two fair dice are rolled.

1. What is the (conditional) probability that one shows 2, given that they show different numbers?
   Start by applying the definition of conditional probability:

   $$\mathbb{P}(\{\text{one shows 2}|\text{different}\}) = \frac{\mathbb{P}(\{\text{one shows 2}\} \cap \{\text{different}\})}{\mathbb{P}(\{\text{different}\})}$$

   Most particularly, observe that $\mathbb{P}(\{\text{one shows 2}\} \cap \{\text{different}\}) = \frac{10}{36}$ as there are 10 combinations satisfying the condition, whereas, by reasoning in terms of the complement, $\mathbb{P}(\{\text{different}\}) = 1 - \mathbb{P}(\{\text{same}\}) = 1 - \frac{6}{36} = \frac{30}{36}$, allowing to conclude that:

   $$\mathbb{P}(\{\text{one shows 2}|\text{different}\}) = \frac{\frac{10}{36}}{\frac{30}{36}} = \frac{10}{30} = \frac{1}{3}$$

2. What is the (conditional) probability that the first shows 6, given that the sum is k?
   Start by applying the definition of conditional probability:

   $$\mathbb{P}(\{\text{first is 6}|\text{sum is } k\}) = \frac{\mathbb{P}(\{\text{first is 6}\} \cap \{\text{sum is } k\})}{\mathbb{P}(\{\text{sum is } k)\}} = \begin{cases} 0 & \text{if } 2 \leq k \leq 6 \\ ??? & \text{if } 7 \leq k \leq 12 \end{cases}$$

   Most particularly, observe that, if $7 \leq k \leq 12$, it is possible to generalize the probability that the sum of the dice is k in the following way:

   $$\mathbb{P}(\{\text{sum is } k\} \cap \{7 \leq k \leq 12\}) = \frac{13-k}{36}$$

   Observe that since, $\forall\ 7 \leq k \leq 12$, there is only one combination summing up to k whose first value is 6 (with said combination being $(6, k-6)$), it is possible to conclude that:

   $$\mathbb{P}(\{\text{first is 6}|\text{sum is } k\}) = \begin{cases} 0 & \text{if } 2 \leq k \leq 6 \\ \frac{1}{13-k} & \text{if } 7 \leq k \leq 12 \end{cases}$$

3. What is the (conditional) probability that at least one shows 6, given that the sum is k?

Start by applying the definition of conditional probability:

$$\mathbb{P}(\{\text{at least one is 6}|\text{sum is } k\}) = \frac{\mathbb{P}(\{\text{at least one is 6}\} \cap \{\text{sum is } k\})}{\mathbb{P}(\{\text{sum is } k)\}} = \begin{cases} 0 & \text{if } 2 \leq k \leq 6 \\ ??? & \text{if } 7 \leq k \leq 12 \end{cases}$$

Unlike the previous case, however, ordering does not matter, meaning that, by applying the stick-and-starts argument, there are only 21 possible combinations and therefore, if $7 \leq k \leq 12$, it is possible to obtain:

$$\mathbb{P}(\{\text{sum is } k\} \cap \{7 \leq k \leq 12\}) = \begin{cases} \frac{3}{21} & \text{if } 7 \leq k \leq 8 \\ \frac{2}{21} & \text{if } 9 \leq k \leq 10 \\ \frac{1}{21} & \text{if } 11 \leq k \leq 12 \end{cases}$$

Most particularly, since, $\forall\ 7 \leq k \leq 12$, there is exactly one combination summing up to k and with at least one of the values being 6 (with that combination being $\{6, k-6\}$), it is possible to conclude that:

$$\mathbb{P}(\{\text{at least one is 6}|\text{sum is } k\}) = \begin{cases} 0 & \text{if } 2 \leq k \leq 6 \\ \frac{1}{3} & \text{if } 7 \leq k \leq 8 \\ \frac{1}{2} & \text{if } 9 \leq k \leq 10 \\ 1 & \text{if } 11 \leq k \leq 12 \end{cases}$$

6.3 Exercise 3

An urn contains a red ball and a green ball. One ball is picked at random from the urn, its colour is observed, and the ball is placed back in the urn together with a new ball of the same colour. Let R_i, for $i = 1, 2, 3$ denote the event "the i^{th} picked ball is red".

1. Compute $\mathbb{P}(R_1|R_2)$.
   Start by considering the following probability tree:

First box	First draw	Second box	Second draw	Third box	Third draw
RG	Red ($\frac{1}{2}$)	RRG	Red ($\frac{2}{3}$)	RRRG	Red ($\frac{3}{4}$)
RG	Red ($\frac{1}{2}$)	RRG	Red ($\frac{2}{3}$)	RRRG	Green ($\frac{1}{4}$)
RG	Red ($\frac{1}{2}$)	RRG	Green ($\frac{1}{3}$)	RRGG	Red ($\frac{1}{2}$)
RG	Red ($\frac{1}{2}$)	RRG	Green ($\frac{1}{3}$)	RRGG	Green ($\frac{1}{2}$)
RG	Green ($\frac{1}{2}$)	RGG	Red ($\frac{1}{3}$)	RGRG	Red ($\frac{1}{2}$)
RG	Green ($\frac{1}{2}$)	RGG	Red ($\frac{1}{3}$)	RGRG	Green ($\frac{1}{2}$)
RG	Green ($\frac{1}{2}$)	RGG	Green ($\frac{2}{3}$)	RGGG	Red ($\frac{1}{4}$)
RG	Green ($\frac{1}{2}$)	RGG	Green ($\frac{2}{3}$)	RGGG	Green ($\frac{3}{4}$)

   Starting from the definition of conditional probability, recover the reverse condition with Bayes' theorem and then apply the law of total probability:

$$\mathbb{P}(R_1|R_2) = \frac{\mathbb{P}(R_1 \cap R_2)}{\mathbb{P}(R_2)} = \frac{\mathbb{P}(R_1)\mathbb{P}(R_2|R_1)}{\mathbb{P}(R_2|R_1)\mathbb{P}(R_1) + \mathbb{P}(R_2|G_1)\mathbb{P}(G_1)}$$

$$\mathbb{P}(R_1|R_2) = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

2. Compute $\mathbb{P}(R_3|R_2)$.
   Again, start from the definition of conditional probability and apply the law of total probability:

$$P(R_3|R_2) = \frac{\mathbb{P}(R_3 \cap R_2)}{\mathbb{P}(R2)} = \frac{\mathbb{P}(R_1 \cap R_2 \cap R_3) + \mathbb{P}(G_1 \cap R_2 \cap R_3)}{\mathbb{P}(R_2|R_1)\mathbb{P}(R_1) + \mathbb{P}(R_2|G_1)\mathbb{P}(G_1)}$$

   Most particularly, by applying composite probability, it is possible to recover that:

$$P(R_2 \cap R_3) = P(R_1)P(R_2|R_1)P(R_3|R_1 \cap R_2) + P(G_1)P(R_2|G_1)P(R_3|G_1 \cap R_2) = \frac{1}{3}$$

   Therefore, the probability of this event will be given by:

$$\mathbb{P}(R_3|R_2) = \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4}}{\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

3. Compute $\mathbb{P}(R_1|R_3)$.
   Again, starting from the definition conditional probability, apply the law of total probability:

$$\mathbb{P}(R_1|R_3) = \frac{\mathbb{P}(R_1 \cap R_3)}{\mathbb{P}(R_3)} = \frac{\mathbb{P}(R1 \cap R2 \cap R3) + P(R1 \cap G2 \cap R3)}{\mathbb{P}(R_3)}$$

   Most particularly, notice that, by symmetry of the experiment, it is possible to retrieve that $\mathbb{P}(R_3) = \frac{1}{2}$, resulting in:

$$\mathbb{P}(R_1|R_3) = \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4}}{\frac{1}{2}} = \frac{2}{3}$$

6.4 Exercise 4

A biased coin is given, with bias p unknown and to be determined. The "Maximum Likelihood Estimator" $\hat{p}$ for p is defined by requiring that $\hat{p}$ maximises the probability of the observed event. Compute the "Maximum Likelihood Estimator" $\hat{p}$ for the following two experiments:

1. Toss the coin 200 times and obtain 67 heads. $(*)$
   Notice that this event can be represented using a binomial distribution:

$$f(p) = \mathbb{P}(67 \text{ H in } 200 \text{ tosses}) = \binom{200}{67} p^{67}(1-p)^{133}$$

At this point, it is possible to find $\hat{p}$ by studying the sign of the derivative of the previous value:

$$\frac{d(f(p))}{dp} = 67p^{66}(1-p)^{133} - 133p^{67}(1-p)^{132} = p^{66}(1-p)^{132}(67(1-p)-133p) > 0 \Rightarrow p < \frac{67}{200}$$

Therefore, $\hat{p} = \frac{67}{200}$ will be this experiment's maximum likelihood estimator.

2. The first head is observed on the fifth coin toss. $(*)$
   Notice that this event can be represented using a geometric distribution:

$$f(p) = \mathbb{P}(\text{first H at the fifth toss}) = (1-p)^4 p$$

At this point, it is possible to find $\hat{p}$ by studying the sign of the derivative of the previous value:

$$\frac{d(f(p))}{dp} = -4(1-p)^3 p + (1-p)^4 = (1-p)^3(-4p + (1-p)) > 0 \Rightarrow p < \frac{1}{5}$$

Therefore, $\hat{p} = \frac{1}{5}$ will be this experiment's maximum likelihood estimator.

3. Toss the coin n times and obtain k heads.
   Model the event as a generic binomial distribution:

$$f(p) = \mathbb{P}(k \text{ H in } n \text{ tosses}) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

By monotonicity between a function f and its logarithm $\ln f$, consider this approach:

$$\ln f(p) = \ln[\binom{n}{k} p^k (1-p)^{n-k}] = \ln(\binom{n}{k}) + k\ln(p) + (n-k)\ln(1-p)$$

At this point, it is possible to find $\hat{p}$ by studying the sign of the derivative of $\ln f(p)$:

$$\frac{d(\ln f(p))}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = \frac{k(1-p)-(n-k)p}{p(1-p)} = k - np > 0 \Rightarrow p < \frac{k}{n}$$

Therefore, the maximum likelihood estimator for this experiment will be:

$$\hat{p} = \frac{k}{n}$$

Alternatively. it is possible to directly study the derivative of the distribution function.

4. The first head is observed on the h^{th} coin toss.
   Model the event as a generic geometric distribution up to and including the k^{th} toss:

$$f(p) = \mathbb{P}(\text{first H at the } h^{\text{th}} \text{ toss}) = (1-p)^{h-1} p$$

By monotonicity between a function f and its logarithm $\ln f$, consider:

$$\ln f(p) = \ln[(1-p)^{h-1}p] = -(h-1)\ln(1-p) + \ln(p)$$

At this point, it is possible to find $\hat{p}$ by studying the sign of the derivative of $\ln f(p)$:

$$\frac{d(\ln f(p))}{dp} = \frac{h-1}{1-p} + \frac{1}{p} = \frac{-p(h-1)+1-p}{p(1-p)} = -p(h-1)+1-p > 0 \Rightarrow p < \frac{1}{h}$$

Therefore, the maximum likelihood estimator for this experiment will be:

$$\hat{p} = \frac{1}{h}$$

Alternatively, it is possible to directly study the derivative of the distribution function.

6.5 Exercise 5

A biased coin, with probability of head $p \in (0,1)$, is tossed repeatedly.
Given $a, b \geq 1$, compute the probability that the coin gives a heads before b tails.
Solve the problem recursively, and define:

$$f(a,b) = \mathbb{P}(a \text{ H before } b \text{ T}) = \mathbb{P}(\{\text{first H}\} \cap \{a \text{ H before } b \text{ T}\}) + \mathbb{P}(\{\text{first T}\} \cap \{a \text{ H before } b \text{ T}\})$$

By applying conditional probability, rewrite this function as:

$$f(a,b) = \mathbb{P}(a \text{ H before } b \text{ T}|\text{first H})\mathbb{P}(\text{first H}) + \mathbb{P}(a \text{ H before } b \text{ T}|\text{first T})\mathbb{P}(\text{first T})$$

This means that, according to the outcome of the first coin toss, it is possible to rewrite the conditional probabilities as:

- $\mathbb{P}(a \text{ H before } b \text{ T}|\text{first H}) = \mathbb{P}(a-1 \text{ H before } b \text{ T})$.
- $\mathbb{P}(a \text{ H before } b \text{ T}|\text{first T}) = \mathbb{P}(a \text{ H before } b-1 \text{ T})$.

This allows to rewrite the recursive function as:

$$f(a,b) = p \cdot f(a-1,b) + (1-p) \cdot f(a,b-1)$$

However, since this recursion is non-trivial, it is better to solve it by using an alternative approach: observe that it is enough to check at most $a+b-1$ coin tosses because this implies that either at least a heads were observed (meaning that at most $b-1$ tails were observed) or at least b tails were observed (meaning that at most $a-1$ heads were observed), and, since these cases are mutually exclusive, it is possible to define:

$$\mathbb{P}(a \text{ H before } b \text{ T}) = \mathbb{P}(\text{at least } a \text{ H in the first } a+b-1 \text{ tosses})$$

Most particularly, observe that:

$$\mathbb{P}(\text{at least } a \text{ H in the first } a+b-1 \text{ tosses}) = \sum_{k=a}^{a+b-1} \mathbb{P}(\text{exactly } k \text{ H in the first } a+b-1 \text{ tosses})$$

Observe that each of the events "exactly k H on the first $a+b-1$ tosses" can be modelled as a binomial distribution, resulting in:

$$\mathbb{P}(\text{at least } a \text{ H in the first } a+b-1 \text{ tosses}) = \sum_{k=a}^{a+b-1} \binom{a+b-1}{k} \cdot p^k (1-p)^{a+b-1-k}$$

6.6 Exercise 6

A referendum is called in a population of n individuals, all of them having the right to vote. Each individual will vote with probability $\frac{1}{2}$, independently of the others. Moreover, if the individual votes, they vote "YES" with probability $\frac{1}{2}$ independently of the others.

1. Compute the probability that a given individual goes to vote and votes "YES".
   Consider the following decision tree:

Vote?	Choice	Total probability
Yes	Yes	$\frac{1}{4}$
Yes	No	$\frac{1}{4}$
No	-	$\frac{1}{2}$

   Apply composite probability:

   $$\mathbb{P}(\text{vote YES}) = \mathbb{P}(\text{YES}|\text{vote})\mathbb{P}(\text{vote}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

2. Compute the probability that the number of "YES" votes is k, for $0 \leq k \leq n$.
   Model the event as a binomial distribution on $p = \frac{1}{4}$ and $1 - p = \frac{3}{4}$ (remember that the "failure" event of not voting YES accounts for both voting NO and not voting at all):

   $$\mathbb{P}(\text{exactly } k \text{ YES}) = \binom{n}{k} \cdot (\frac{1}{4})^k (\frac{3}{4})^{n-k}$$

3. Knowing that there have been exactly k "YES" votes, compute the probability that exactly m individuals voted, for $k \leq m \leq n$.
   Apply Bayes' theorem and revert the condition, resulting in:

   $$\mathbb{P}(m \text{ votes}|k \text{ YES}) = \frac{\mathbb{P}(k \text{ YES}|m \text{ votes})\mathbb{P}(m \text{ votes})}{\mathbb{P}(k \text{ YES})} = \frac{[\binom{m}{k} \cdot (\frac{1}{2})^k (\frac{1}{2})^{m-k}] \cdot [\binom{n}{m} \cdot (\frac{1}{2})^m (\frac{1}{2})^{n-m}]}{\binom{n}{k} \cdot (\frac{1}{4})^k (\frac{3}{4})^{n-k}}$$

The professor further simplified the result to:

$$\mathbb{P}(m \text{ votes}|k \text{ YES}) = \frac{(n-k)!}{(n-m)! \cdot (m-k)!} \cdot (\frac{1}{3})^{n-k} = \binom{n-k}{n-m} \cdot (\frac{1}{3})^{n-k}$$

6.7 Exercise 7

In a TV show, a guest is asked to choose among three doors. Behind one door there is a car while behind the other two there are two goats, denoted goat A and goat B. The guest wins what there is behind the chosen door. After the guest has made his initial choice, that door is not opened.
Instead, one of two unchosen doors which reveals a goat is opened. The guest is then offered the possibility of exchanging the door he initially chose with the one still closed. Let p be the (conditional) probability that the car is behind the last offered door. Show that:

1. $p = \frac{2}{3}$ if the opened door is chosen to always reveal a goat without bias between the two goats.
   Assume, from now on, that Door 1 contains the car, Door 2 contains Goat A and Door 3 contains Goat B.
   Apply conditional probability and let:

   $$p = \mathbb{P}(\text{remaining is car}|\text{open goat}) = \frac{\mathbb{P}(\{\text{remaining is car}\} \cap \{\text{open goat}\})}{\mathbb{P}(\text{open goat})}$$

   However, since, in this case, the presenter always opens a door containing a goat, $\mathbb{P}(\text{open goat}) = 1$, thus simplifying the probability of the event to:

   $$p = \mathbb{P}(\text{remaining is car}) = \mathbb{P}(\text{pick Door 2}) + \mathbb{P}(\text{pick Door 3}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

   Alternatively, it is possible to solve the problem by setting the probability of the event as $p = \mathbb{P}(\text{switch the door}|\text{win the car})$.

2. $p = \frac{1}{2}$ if the opened door is chosen at random.
   Again, apply conditional probability:

   $$p = \mathbb{P}(\text{remaining is car}|\text{open goat}) = \frac{\mathbb{P}(\{\text{remaining is car}\} \cap \{\text{open goat}\})}{\mathbb{P}(\text{open goat})}$$

   However, since the presenter now chooses the door at random, the result will be slightly different because $\mathbb{P}(\text{open goat}) = \mathbb{P}(\text{open Door 2 or Door 3})$ will now depend on which door is chosen first, resulting in:

   $$\mathbb{P}(\text{goat}) = \mathbb{P}(\text{goat}|\text{Door 1})\mathbb{P}(\text{Door 1}) + \mathbb{P}(\text{goat}|\text{Door 2})\mathbb{P}(\text{Door 2}) + \mathbb{P}(\text{goat}|\text{Door 3})\mathbb{P}(\text{Door 3})$$

   Therefore, the probability $p = \mathbb{P}(\text{remaining is car}|\text{open goat})$ will be given by:

   $$p = \frac{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

3. $p = \frac{1}{1+a}$ if the goat seen is goat A and the opened door is selected so that a goat is always chosen and, in the case there are two goats, the door revealing goat A is opened with probability a.
Again, apply conditional probability:

$$p = \mathbb{P}(\text{remaining is car}|\text{open goat A}) = \frac{\mathbb{P}(\{\text{remaining is car}\} \cap \{\text{open goat A}\})}{\mathbb{P}(\text{open goat A})}$$

However, the bias and the fact the presenter always reveals Goat A results in a slight change in a computation, meaning that:

$$p = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{a}{3}} = \frac{1}{1+a}$$

Chapter 7

Exercise Sheet 7: Random variables

7.1 Exercise 1

Write a random word made of 10 characters by choosing a character uniformly at random 10 times independently, from an alphabet of 26 characters.
Let X be the random variable that counts the number of A's in the resulting word.

1. What is the distribution of X? $(*)$
   Observe that $X \backsim Binomial(10, \frac{1}{26})$, meaning that:
   $$\mathbb{P}(X = k) = \binom{10}{k} \cdot (\frac{1}{26})^k (1 - \frac{1}{26})^{10-k} \text{ for } k = 0, \ldots, 10.$$
2. Compute $\mathbb{E}(X)$, that is the average number of A's in a random word of length 10. $(*)$
   Since $X \backsim Binomial(10, \frac{1}{26})$, it is possible to directly compute the expectation of X as:
   $$\mathbb{E}(X) = 10 \cdot \frac{1}{26} = \frac{10}{26}$$
3. Repeat the experiment with a word of length N.
   Consider X to be a binomial random variable with $p = \frac{1}{26}$, meaning that:
   $$X \backsim Binomial(N, \frac{1}{26}) \Rightarrow \mathbb{P}(X = k) = \binom{N}{k} \cdot (\frac{1}{26})^k (1 - \frac{1}{26})^{N-k}$$
   Therefore, since $X \backsim Binomial(N, p)$, it is possible to directly compute $\mathbb{E}(X) = Np$, resulting in:
   $$\mathbb{E}(X) = N \cdot \frac{1}{26} = \frac{N}{26}$$

N.B.: An alternative way to find $\mathbb{E}(X)$ is by noticing that it is possible to express X in the following way:

$$X = \sum_{i=1}^{N} X_i \text{ , where:}$$

$$X_i \backsim Bernoulli(\frac{1}{26}) \Rightarrow X_i = \begin{cases} 1 & \text{with probability } p = \frac{1}{26} \\ 0 & \text{with probability } 1-p = \frac{25}{26} \end{cases}$$

Knowing that $X_i \backsim Bernoulli(p) \Rightarrow \mathbb{E}(X_i) = p$, it is possible to exploit the linearity of expectation and compute:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathbb{E}(X_i) = \frac{N}{26}$$

7.2 Exercise 2

A fair 6-faced die is tossed, and let X denote the observed value.

1. Compute the probability distribution of X. $(*)$
   Since all six outcomes are equally likely, the probability distribution of X will be given by:

   $$\mathbb{P}(X = k) = \frac{1}{6} \text{ for } k = 1, \ldots, 6.$$

2. Compute the expected value of X. $(*)$
   Apply the definition of expectation to obtain that:

   $$\mathbb{E}(X) = \sum_{k=1}^{6} k\mathbb{P}(X = k) = \sum_{k=1}^{6} \frac{k}{6} = \frac{21}{6}$$

3. Compute the variance of X. $(*)$
   Consider the definition of variance to state that:

   Most particularly, since $\mathbb{E}(X)$ was calculated in the previous point, focus on finding $\mathbb{E}(X^2)$:

   $$\mathbb{E}(X^2) = \sum_{k=1}^{6} k^2\mathbb{P}(X = k) = \sum_{k=1}^{6} \frac{k^2}{6} = \frac{91}{6}$$

   Therefore, it is possible to conclude that:

   $$\mathbb{V}(X) = \frac{91}{6} - (\frac{21}{6})^2 = \frac{35}{12}$$

4. Answer the above questions in the case of a n-faced die ($n \in \mathbb{N}$).
Since all outcomes are equally likely, the probability distribution of X will be given by:

$$\mathbb{P}(X = k) = \frac{1}{n}$$

Most particularly, by applying the definition of expectation, it is possible to recover:

$$\mathbb{E}(X) = \sum_{k=1}^{n} k \cdot \mathbb{P}(X = k) = \frac{1}{n} \cdot \sum_{k=1}^{n} k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Similarly, consider the definition of variance to state that:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X), \text{ where:}$$

$$\mathbb{E}(X^2) = \sum_{k=1}^{n} k^2 \cdot \mathbb{P}(X = k) = \frac{1}{n} \cdot \sum_{k=1}^{n} k^2 = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{2n^2 + 3n + 1}{6}$$

Therefore:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2n^2 + 3n + 1}{6} - \frac{n^2 + 2n + 1}{4} = \frac{n^2 - 1}{12}$$

7.3 Exercise 3

Toss two fair six-faced dice, and let X denote the minimum between the observed values.

1. Compute the probability distribution of X.
If X denotes the minimum of the observed results, it is possible to write its distribution as:

$$\mathbb{P}(X = k) = \mathbb{P}(X \geq k) - \mathbb{P}(X \geq k + 1) \text{ , where:}$$

$$\mathbb{P}(X = k) = \mathbb{P}(\{X_1 \geq k\} \cap \{X_2 \geq k\})$$

Most particularly, since the events are independent, it is possible to rewrite the distribution as:

$$\mathbb{P}(X = k) = \mathbb{P}(\{X_1 \geq k\})\mathbb{P}(\{X_2 \geq k\}) = \mathbb{P}(\{X_1 \geq k\})^2 = (\frac{6 - (k-1)}{6})^2$$

Therefore:

$$\mathbb{P}(X = k) = (\frac{6 - (k-1)}{6})^2 - (\frac{6 - k}{6})^2 = \frac{13 - 2k}{36}$$

2. Compute the expected value of X.

$$\mathbb{E}(X) = \sum_{k=1}^{6} k \cdot \mathbb{P}(X = k) = \frac{1 \cdot 11}{36} + \frac{2 \cdot 9}{36} + \frac{3 \cdot 7}{36} + \frac{4 \cdot 5}{36} + \frac{5 \cdot 3}{36} + \frac{6 \cdot 1}{36} = \frac{91}{36}$$

7.4 Exercise 4

A box contains 10 transistors, 3 of which are broken. You check one transistor at a time (without replacement) until you find a broken one. Compute the expected value of checked transistors.
Let X denote the number of checked transistors, meaning that, if $X = k$ ($k \leq 8$), then the first $k-1$ transistors were good while the k^{th} transistor must be broken.
Knowing that 7 transistors are good and 3 transistors are broken, it is possible to choose the functioning transistors in $\frac{7!}{(7-(k-1))!}$ ways as an ordered sampling without repetition, while the broken transistor can be chosen in $\binom{3}{1}$ ways, also implying that the remaining unchecked transistors can be determined in $(10-k)!$ ways: since there are 10! possible orderings of the transistors, it is possible to state that:

$$\mathbb{P}(X = k) = \frac{7!}{(7-(k-1))!} \cdot 3 \cdot (10-k)! = \frac{(10-k)(9-k)}{240}$$

Therefore, it is possible to conclude that:

$$\mathbb{E}(X) = \sum_{k=1}^{8} k \cdot \mathbb{P}(X = k) = \frac{11}{4}$$

7.5 Exercise 5

Consider a multiple choice exam with the following rules: there are 10 questions, and for each question there are 4 possible answers, among which exactly one is correct.
The evaluation algorithm is as follows: each correct answer gets +3 marks and each wrong answer gets -1 mark.
Alice did not study, so she answers all 10 questions at random.

1. Compute the probability that Alice passes the exam (she scores at least $\frac{18}{30}$).
   Assuming X denotes the score that Alice got in her exam, it is possible to state that Alice answers k out of 10 questions correctly with probability:

   $$\mathbb{P}(\{\text{Alice gets } k \text{ questions right}\}) = \binom{10}{k} \cdot (\frac{1}{4})^k (\frac{3}{4})^{10-k}$$

   However, Alice passes the exam if and only if $X \geq 18$, which, due to the evaluation algorithm, happens if and only if the following is true:

   $$3k - (10-k) \geq 18 \Leftrightarrow 3k - 10 + k \geq 18 \Leftrightarrow 4k \geq 28 \Leftrightarrow k \geq 7$$

   For this reason, Alice passes the exam if and only if she gets at least 7 questions right, meaning that:

$$\mathbb{P}(\{\text{Alice passes the exam}\}) = \mathbb{P}(X \geq 18) = \mathbb{P}(k \geq 7)$$

By applying additivity, it is possible to conclude that:

$$\mathbb{P}(\{\text{Alice passes the exam}\}) = \sum_{k=7}^{10} \binom{10}{k} \cdot (\frac{1}{4})^k (\frac{3}{4})^{10-k}$$

2. Compute Alice's expected final grade.
If X denotes Alice's final grade, consider that it is possible to express this random variable in the following way:

$$X = \sum_{i=1}^{10} X_i \text{ , where:}$$

$$X_i = \begin{cases} +3 & \text{with probability } \mathbb{P}(X_i = 3) = \frac{1}{4} \\ -1 & \text{with probability } \mathbb{P}(X_i = -1) = \frac{3}{4} \end{cases}$$

For this reason, it is possible to state that:

$$\mathbb{E}(X_i) = \sum_{x \in Im(X_i)} x \cdot \mathbb{P}(X_i = x) = 3 \cdot \frac{1}{4} + (-1) \cdot \frac{3}{4} = 0$$

Therefore, by exploiting the linearity of expectation, it is possible to conclude that:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{E}(X_i) = 0$$

Alternatively, observe that, regardless of how many questions Alice gets right, it is always possible to recover that $-10 \leq X \leq 30$, meaning that, by applying the definition of expectation, it is also possible to directly compute:

$$\mathbb{E}(X) = \sum_{x=-10}^{30} x \cdot \mathbb{P}(X = x)$$

3. Compute the variance of Alice's final grade.
Assuming that X_i denotes the score that Alice got for question i, consider the variance:

$$\mathbb{V}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}^2(X_i) \text{ , where:}$$

$$\mathbb{E}(X_i^2) = \sum_{x \in Im(X_i)} x^2 \cdot \mathbb{P}(X_i = x) = (3)^2 \cdot \frac{1}{4} + (-1)^2 \cdot \frac{1}{4} = 3$$

For this reason, it is possible to state that:

$$\mathbb{V}(X_i) = 3 - (0)^2 = 3$$

However, since Alice answers the questions at random independently of the other answers, it is possible to state that $X_1, \dots, X_{10}$ are independent random variables, meaning that it is possible to simplify the computations to:

$$\mathbb{V}(X) = \mathbb{V}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{V}(X_i) = 30$$

7.6 Exercise 6

Consider an urn with a white balls and b black balls: you pick k balls sequentially without replacement (so $k \leq a + b$)
Let X_i (for $i = 1, \dots, k$) be the random variable taking value 1 if the i^{th} ball is white and 0 if it is black, and, moreover, let X denote the total number of white balls picked.

1. Compute the distribution of X.
   Start by considering the sample space Ω: since the k balls are selected without replacement from an urn containing $a + b$ balls in total, it is possible to discard ordering and consider:

   $$|\Omega| = \binom{a+b}{k}$$

   At this point, if X denotes the total number of white balls that were picked in the experiment, consider the event:

   $$A = \{j \text{ white balls out of } k \text{ balls were picked}\} = \{X = j\}$$

   Most particularly, since this event requires choosing j balls from the a white balls and the remaining $k - j$ balls from the b black balls, it is possible to state, by product rule, that:

   $$|A| = \binom{a}{j}\binom{b}{k-j}$$

   For this reason, it is possible to conclude that:

   $$\mathbb{P}(X = j) = \frac{|A|}{|\Omega|} = \frac{\binom{a}{j}\binom{b}{k-j}}{\binom{a+b}{j}}$$

   Therefore, it is possible to notice that, actually, X represents a hypergeometric distribution.

2. Compute the expected value of X (*you should do both the direct calculation using the distribution of X, and the calculation using the expectations of the X_i's*).

By applying the definition of expectation, consider:

$$\mathbb{E}(X) = \sum_{j=0}^{\min\{a,k\}} j \cdot \mathbb{P}(X = j) = \sum_{j=0}^{\min\{a,k\}} j \cdot \frac{\binom{a}{j}\binom{b}{k-j}}{\binom{a+b}{k}}$$

$$\mathbb{E}(X) = \frac{1}{\binom{a+b}{k}} \cdot \sum_{j=0}^{\min\{a,k\}} j \cdot \frac{a!}{(a-j)! \cdot j!} \frac{b!}{(b-(k-j))! \cdot (k-j)!}$$

Since $\frac{j}{j!} = \frac{1}{(j-1)!}$, let $j = 1, \ldots, \min\{a, k\}$ and simplify the computations to:

$$\frac{1}{\binom{a+b}{k}} \sum_{j=1}^{\min\{a,k\}} \frac{a(a-1)!}{((a-1)-(j-1))!(j-1)!} \frac{b!}{((b-(k-1))-(j-1))!((k-1)-(j-1))!}$$

$$\text{Therefore, } \mathbb{E}(X) = \frac{a}{\binom{a+b}{k}} \cdot \sum_{j=0}^{\min\{a,k\}} \binom{a-1}{j-1} \cdot \binom{b}{k-j}$$

At this point, let $h = j - 1$, and rewrite the formula as:

$$\mathbb{E}(X) = \frac{a}{\binom{a+b}{k}} \cdot \sum_{h=0}^{\min\{a,k-1\}} \frac{\binom{a-1}{h}\binom{b}{k-(h+1)}}{\binom{a+b-1}{k-1}} \cdot \binom{a+b-1}{k-1}$$

However, by applying the properties of hypergeometric variables, it is possible to observe that:

$$\sum_{h=0}^{\min\{a,k-1\}} \frac{\binom{a-1}{h}\binom{b}{k-(h+1)}}{\binom{a+b-1}{k-1}} = 1$$

This allows to simplify the final result to:

$$\mathbb{E}(X) = \frac{a}{\binom{a+b}{k}} \cdot \binom{a+b-1}{k-1} = \frac{ka}{a+b}$$

Alternatively, remember that, by definition of X and X_i, it is possible to state that:

$$X = \sum_{i=1}^{k} X_i \Rightarrow \mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{k} X_i)$$

Most particularly, remember that:

$$X_i \backsim Bernoulli(p) \Rightarrow \mathbb{E}(X_i) = p = \mathbb{P}(X = 1)$$

By symmetry, it is possible to recover that:

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = 1) = \cdots = \mathbb{P}(X_i = 1) = \cdots = \mathbb{P}(X_k = 1) = \frac{a}{a+b}$$

For this reason, by exploiting the linearity of expectation, it is possible to conclude that:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} \mathbb{E}(X_i) = \frac{ka}{a+b}$$

3. Compute the covariance between X_i and X_j, for $i, j = 1, \ldots, k$.
   Assuming $i \neq j$, remember that covariance between two random variables X_i and X_j is defined as:

   $$Cov(X_i, X_j) = \mathbb{E}(X_i, X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)$$

   However, since $X_i, X_j \backsim Bernoulli(p)$, it is possible to state that:

   $$\mathbb{E}(X_i X_j) = \mathbb{P}(X_i = 1 \wedge X_j = 1) = \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1 | X_i = 1) = \frac{a}{a+b} \cdot \frac{a-1}{a-1+b}$$

   Therefore, it is possible to conclude that:

   $$Cov(X_i, X_j) = \frac{a}{a+b} \cdot \frac{a-1}{a-1+b} - (\frac{a}{a+b})^2$$

   Observe that, trivially, if $i = j$, then $X_i = X_j$ and $Cov(X_i, X - j) = \mathbb{V}(X_i)$.

4. Compute the variance of X (*you should do both the direct calculation using the distribution of X, and the calculation using $X = \sum_{i=1}^{k} X_i$ and the information from the previous question*).
   Apply the definition of variance and consider:

   $$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

   Most particularly, notice that, by letting $j^2 = j(j-1) + j$, it is possible to recover that:

   $$\mathbb{E}(X^2) = \sum_{j=1}^{\min\{a,k\}} j^2 \cdot \mathbb{P}(X = j) = \sum_{j=1}^{\min\{a,k\}} j(j-1)\mathbb{P}(X = j) + \sum_{j=1}^{\min\{a,k\}} j\mathbb{P}(X = j)$$

   However, since $\sum j\mathbb{P}(X = j) = \mathbb{E}(X)$, focus on the other sum and consider:

   $$\sum_{j=1}^{\min\{a,k\}} j(j-1)\mathbb{P}(X = j) = \sum_{j=1}^{\min\{a,k\}} j(j-1) \cdot \frac{\binom{a}{j}\binom{b}{k-j}}{\binom{a+b}{k}}$$

By applying the same reasoning for $\mathbb{E}(X)$, it will be possible to conclude that:

$$\mathbb{V}(X) = \frac{abk(a+b-k)}{(a+b)^2(a+b-1)}$$

Alternatively, consider a simpler approach and remember that it is possible to consider the variance of X as:

$$\mathbb{V}(X) = \mathbb{V}(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} \mathbb{V}(X_i) + 2\sum_{i=1}^{k}\sum_{j>i} Cov(X_i, X_j)$$

However, remember that:

$$\mathbb{V}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}^2(X_i) = \frac{a}{a+b} - (\frac{a}{a+b})^2 \text{ (recall that } \mathbb{V}(Bernoulli(p)) = p(1-p))$$

$$Cov(X_i, X_j) = \frac{a}{a+b}\frac{a-1}{a-1+b} - (\frac{a}{a+b})^2 \text{ (for } \frac{k(k-1)}{2} \text{ pairings } i = 1, \ldots, k \wedge j > i)$$

Therefore, it is possible to recover that:

$$\mathbb{V}(X) = k[\frac{a}{a+b} - (\frac{a}{a+b})^2] + 2 \cdot \frac{k(k-1)}{2}[\frac{a}{a+b}\frac{a-1}{a-1+b} - (\frac{a}{a+b})^2]$$

So, by simplifying the expression, it is possible to conclude that:

$$\mathbb{V}(X) = \frac{abk(a+b-k)}{(a+b)^2(a+b-1)}$$

7.7 Exercise 7

For $a, b, k \in \mathbb{N}$, consider the hypergeometric distribution:

$$\mathbb{P}_{a,b,k}(X = h) = \frac{\binom{a}{h}\binom{b}{k-h}}{\binom{a+b}{k}} \text{ , where } h = 0, \ldots, k$$

Compute the limit of $\mathbb{P}_{a,b,k}(X = h)$ as $a, b \to \infty$ with $\frac{a}{a+b} \to p \in (0,1)$ by providing a probabilistic interpretation of the result.
Observe that the hypergeometric distribution is used when picking items sequentially but without repetitions, but, if $N = a + b \to \infty$, it is possible to ignore the fact that the selection is done without repetitions and approximate the hypergeometric distribution to a binomial distribution of h elements out of k picked items, meaning that:

$$\lim_{(a,b)\to\infty} \mathbb{P}_{a,b,k}(X = h) = \lim_{(a,b)\to\infty} \frac{\binom{a}{h}\binom{b}{k-h}}{\binom{a+b}{k}} \approx \binom{k}{h} \cdot p^h (1-p)^{k-h} \text{ , where } p = \frac{a}{a+b}$$

Chapter 8

Exercise Sheet 8: Random variables

8.1 Exercise 1

Let X_1 and X_2 be independent random variables, uniformly distributed in $\{1, \ldots, n\}$.

1. Compute the probability distribution of $X_1 + X_2$.
   Start by defining the probability measure of X_1 and X_2:

   $$\mathbb{P}(X_1 = h) = \mathbb{P}(X_2 = h) = \frac{1}{n} \text{ for } h \in \{1, \ldots, n\}$$

   Most particularly, since X_1 and X_2 are assumed to be independent, it is possible to state that, if $X = X_1 + X_2$ and $X_1 = h$ is fixed, it is possible to compute:

   $$\mathbb{P}(X = k) = \mathbb{P}(X_1 = h)\mathbb{P}(X_2 = k - h) \text{ with } k \in \{2, \ldots, 2n\}$$

   At this point, assuming that $1 \leq h \leq n$ and $1 \leq k - h \leq n$, apply additivity and, in order to maintain positive probabilities, state that:

   $$\mathbb{P}(X = k) = \sum_{h=1}^{k-1} \mathbb{P}(X_1 = h, X_2 = k-h) = \sum_{h=\max\{1,k-n\}}^{\min\{k-1,n\}} \mathbb{P}(X_1 = h)\mathbb{P}(X_2 = k-h)$$

   Therefore, it is possible to conclude that:

   $$\mathbb{P}(X = k) = \frac{1}{n^2}(\min\{k - 1, n\} - \max\{1, k - n\} + 1)$$

   Alternatively, consider a case-by-case approach:

- Assume $k \leq n$:

$$\mathbb{P}(X = k) = \sum_{h=1}^{k-1} \mathbb{P}(X_1 = h)\mathbb{P}(X_2 = k - h) = \frac{k-1}{n^2}$$

- Assume $n < k \leq 2n$:

$$\mathbb{P}(X = k) = \sum_{h=1}^{n} \mathbb{P}(X_1 = k-h)\mathbb{P}(X_2 = h) = \sum_{h=k-n}^{n} \mathbb{P}(X_1 = k-h)\mathbb{P}(X_2 = h) = \frac{2n-k+1}{n^2}$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(X = k) = \begin{cases} \frac{k-1}{n^2} & \text{for } 2 \leq k \leq n+1 \\ \frac{2n-k+1}{n^2} & \text{for } n+1 < k \leq 2n \end{cases}$$

2. Compute the expected value of $X_1 + X_2$.
Exploit the linearity of expectation:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Most particularly, since both X_1 and X_2 are defined to be uniform random variables, it is possible to state that:

$$\mathbb{E}(X) = 2\mathbb{E}(X_1)$$

By definition of expectation, it is possible to recover that:

$$\mathbb{E}(X_1) = \sum_{k=1}^{n} k \cdot \mathbb{P}(X_1 = k) = \sum_{k=1}^{n} \frac{k}{n} = \frac{n(n-1)}{2n} = \frac{n+1}{2}$$

Therefore, it is possible to conclude that:

$$\mathbb{E}(X) = 2\mathbb{E}(X_1) = n + 1$$

3. Compute the variance of $X_1 + X_2$.
Knowing that $X = X_1 + X_2$ but X_1 and X_2 are independent random variables, it is possible to state that:

$$\mathbb{V}(X) = \mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + 2Cov(X_1, X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2)$$

However, since both X_1 and X_2 are uniform random variables, it is possible to state that:

$$\mathbb{V}(X_1) = \mathbb{V}(X_2) \Rightarrow \mathbb{V}(X) = 2\mathbb{V}(X_1)$$

By definition of variance, it is possible to state that:

$$\mathbb{V}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}^2(X_1), \text{ where:}$$

$$\mathbb{E}(X_1^2) = \sum_{k=1}^{n} k^2 \cdot \mathbb{P}(X_1 = k) = \sum_{k=1}^{n} \frac{k^2}{n} = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6}$$

Therefore, it is possible to discover that:

$$\mathbb{V}(X_1) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

It is thus possible to conclude that:

$$\mathbb{V}(X) = 2\mathbb{V}(X_1) = \frac{n^2-1}{6}$$

8.2 Exercise 2

A fair die is tossed repeatedly, until 5 or 6 is obtained. Let T denote the number of tosses and let X denote the number seen on the die in the last toss.

1. Compute $\mathbb{P}(T = 3, X = 5)$.
   This event means that the first two tosses resulted neither in 5 nor in 6, whereas the third and final toss gave 5: since each toss is independent, it is therefore possible to rewrite the probability of the event as:

   $$\mathbb{P}(T = 3, X = 5) = \mathbb{P}(\{1,2,3,4\}) \cdot \mathbb{P}(\{1,2,3,4\}) \cdot \mathbb{P}(\{5\}) = \frac{4}{6} \cdot \frac{4}{6} \cdot \frac{1}{6} = \frac{2}{27}$$

2. Compute the probability distribution of T.
   Since T denotes the number of tosses up to and including the firs 5/6, it is possible to model the variable as $T \backsim Geometric(p)$ with $p = \frac{2}{6}$, resulting in:

   $$\mathbb{P}(T = k) = (1 - \frac{2}{6})^{k-1} \cdot \frac{2}{6} = (\frac{4}{6})^{k-1} \cdot \frac{2}{6}$$

3. Compute the probability distribution of X.
   Since the experiment ends if and only if the last seen value is 5 or 6, it is possible to simply consider $X \in \{5, 6\}$ and state that:

   $$\mathbb{P}(X = 5) = \sum_{k=1}^{\infty} \mathbb{P}(X = 5, T = k) = \frac{1}{2} \sum_{k=1}^{\infty} (\frac{4}{6})^{k-1} \frac{2}{6} = \frac{1}{2}$$

   N.B.: This statement is symmetric and holds for $X = 6$ as well.

4. Are T and X independent random variables? Explain.
   Not always: the variables will be independent if and only if $X = 5$ or $X = 6$: assuming $X = 5$, it is indeed possible to recover that:

   $$\mathbb{P}(\{T = k\} \cap \{X = 5\}) = (\frac{4}{6})^{k-1} \cdot \frac{1}{6} = \mathbb{P}(T = k)\mathbb{P}(X = 5)$$

   N.B.: The statement is symmetric and holds for $X = 6$ as well.

8.3 Exercise 3

How many times should you roll, on average, a fair die in order to see all faces? If X denotes the number of tosses until all faces are seen, let X_k denote the number of tosses strictly after the $(k-1)^{\text{th}}$ new value is seen and until the k^{th} new value is seen, and define:

$$X = \sum_{k_1}^{6} X_k, \text{ with } X_k \backsim \mathit{Geometric}(p_k) \text{ for } p_k = \frac{7-k}{6}$$

Most particularly, since $X_k \backsim \mathit{Geometric}(p_k) \Rightarrow \mathbb{E}(X_k) = \frac{1}{p_k}$, it is possible to exploit the linearity of expectation to conclude that:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{k=1}^{6} X_k) = \sum_{k=1}^{6} \mathbb{E}(X_k) = \sum_{k=1}^{6} \frac{1}{p_k} = \frac{147}{10}$$

8.4 Exercise 4

The breaking time of component C_i is given by a random variable T_i (with $i = 1, \dots, k$). Assume that the random variables $T1, \dots, T_k$ are independent and that $T_i \backsim \mathit{Geometric}(p)$ with $p \in (0, 1)$.

1. Find the probability distribution of the breaking time of the circuit C_{ser} obtained by organising in series all the components $C_1, \dots, C_k$.
   If the circuit is organised in series, then it is sufficient for the circuit to break that just one component breaks, so, if T_S denotes the breaking time of the circuit, it must hold that:

   $$T_S = \min\{T_1, \dots, T_k\} \Rightarrow \mathbb{P}(T = n) = \mathbb{P}(T_1 \geq n) \cdot \ \dots \ \cdot \mathbb{P}(T_k \geq n)$$

   Most particularly, consider that:

   $$\mathbb{P}(T_S \geq n) = \mathbb{P}(\bigcap_{i=1}^{k} \{T_i \geq n\})$$

   However, since the random variables are all independent, this factorises to:

   $$\mathbb{P}(T_S \geq n) = \prod_{i=1}^{k} \mathbb{P}(T_i \geq n) = (\mathbb{P}(T_1 \geq n))^k = ((1-p)^{n-1})^k = (1-p)^{k(n-1)}$$

   At this point, observe that $T_S \backsim \mathit{Geometric}(1 - (1-p)^k)$ because:

   $$\mathbb{P}(T_S = n) = \mathbb{P}(T_S \geq n) - \mathbb{P}(T_S \geq n+1) = (1-p)^{k(n-1)}(1 - (1-p)^k)$$

   N.B.: If $q = 1 - (1-p)^k$, the result simplifies to $\mathbb{P}(T_S = n) = (1-q)^{n-1}$.

2. Find the probability distribution of the breaking time of the circuit C_{par} obtained by organising in parallel all the components $C_1, \dots, C_k$.
   If the circuit is organised in parallel, then the circuit breaks if and only if all components break, so, if T_P denotes the breaking time of the circuit, it must hold that:

$$T_P = \max\{T_1, \dots, T_k\} \Rightarrow \mathbb{P}(T = n) = \mathbb{P}(T_1 \leq n) \cdot \ \dots \ \cdot \mathbb{P}(T_k \leq n)$$

   Most particularly, consider that:

$$\mathbb{P}(T_P \leq n) = \mathbb{P}(\bigcap_{i=1}^{k} \{T_i \leq n\})$$

   However, since the random variables are all independent, this factorises to:

$$\mathbb{P}(T_P \leq n) = \prod_{i=1}^{k} \mathbb{P}(T_i \leq n) = (\mathbb{P}(T_1 \leq n))^k = (1 - \mathbb{P}(T_1 > n))^k = (1-(1-p)^n)^k$$

   At this point, observe that, while T_P cannot be reconducted to a geometric random variable, it is still possible to state that:

$$\mathbb{P}(T_p = n) = \mathbb{P}(T_P \leq n) - \mathbb{P}(T_P \leq n-1) = (1-(1-p)^n)^k - (1-(1-p)^{n-1})^k$$

8.5 Exercise 5

In a Bernoulli scheme with head probability $p \in (0, 1)$, let X denote the random variable counting the number outcomes identical to the first one: that is, $X = 1$ if the first coin toss gives head and the second gives tail or the first gives tail and the second gives head, $X = 2$ if two heads followed by a tail or two tails and a head, and so on.

1. Find the probability distribution of X.
   Consider the event as the union of two mutually exclusive cases:

$$\mathbb{P}(X = k) = \mathbb{P}(X = k, 1^{\text{st}} = H) + \mathbb{P}(X = k, 1^{\text{st}} = T)$$

   Both cases can be modelled using identical independent experiments, resulting in:

$$\mathbb{P}(X = k) = p^k(1-p) + (1-p)^k p \text{ (with } k \geq 1)$$

2. Compute the expected value of X.
   Consider the definition of expectation:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k[p^k(1-p) + (1-p)^k p]$$

Now, start to simplify this value by rewriting it as following:

$$\mathbb{E}(X) = p(1-p)\sum_{k=1}^{\infty} kp^{k-1} + p(1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Most particularly, observe that, by noticing differentiation, it is possible to exploit geometric series to simplify computations:

$$kp^{k-1} = \frac{d(p^k)}{dp} \Rightarrow \sum_{k=1}^{\infty} \frac{d(p^k)}{dp} = \frac{d(\frac{1}{1-p})}{dp} = \frac{1}{(1-p)^2}$$

$$k(1-p)^{k-1} = -\frac{d((1-p)^k)}{dp} \Rightarrow \sum_{k=1}^{\infty} -\frac{d((1-p)^k)}{dp} = -\frac{d(\frac{1}{1-(1-p)})}{dp} = \frac{1}{p^2}$$

Therefore, it is possible to conclude that:

$$\mathbb{E}(X) = \frac{p(p-1)}{(1-p)^2} + \frac{p(1-p)}{p^2} = \frac{p}{1-p} + \frac{1-p}{p}$$

N.B.: An alternative way to compute the expectation of X is by using indicator random variables: assuming N_H denotes the number of heads up to the first tail, whereas N_T denotes the number of tails up to the first head, it is possible to rewrite X as:

$$X = 1_{(\text{first H})}N_H + 1_{(\text{first T})}N_T$$

Therefore, by applying linearity of the expectation, it is possible to state that:

$$\mathbb{E}(X) = \mathbb{E}(1_{(\text{first H})}N_H + 1_{(\text{first T})}N_T) = \mathbb{E}(1_{(\text{first H})}N_H) + \mathbb{E}(1_{(\text{first T})}N_T)$$

Most particularly, observe that each random variable can be reconducted to known distributions:

- Since indicator random variables can be considered to be special instances of Bernoulli random variables, let $1_{(\text{first H})} \backsim \mathit{Bernoulli}(p)$ and $1_{(\text{first T})} \backsim \mathit{Bernoulli}(1-p)$.
- Notice that, by definition of the experiment, $N_H \backsim \mathit{Geometric}(1-p)$ and $N_T \backsim \mathit{Geometric}(p)$.

For this reason, it is possible to conclude that:

$$\mathbb{E}(X) = \mathbb{E}(1_{(\text{first H})}N_H + 1_{(\text{first T})}N_T) = \frac{p}{1-p} + \frac{1-p}{p}$$

3. Compute the variance of X.
Start by rewriting X using indicator random variables:

$$X = 1_{(\text{first H})}N_H + 1_{(\text{first T})}N_T$$

Since X is expressed as the sum of two random variables, apply the following definition of variance:

$$\mathbb{V}(X) = \mathbb{V}(1_{(\text{first H})}N_H + 1_{(\text{first T})}N_T) = \mathbb{V}(1_{(\text{first H})}N_H) + \mathbb{V}(1_{(\text{first T})}N_T) + 2Cov(1_{(\text{first H})}N_H, 1_{(\text{first T})}N_T)$$

At this point, start to compute the three values:

- $\mathbb{V}(1_{(\text{first H})}N_H)$
  Apply the definition of variance:

  $$\mathbb{V}(1_{(\text{first H})}N_H) = \mathbb{E}(1_{(\text{first H})}N_H^2) - \mathbb{E}^2(1_{(\text{first H})}N_H)$$

  By applying the definition of conditional probability, it is possible to rewrite the expectation of the squared variable in the following way:

  $$\mathbb{E}(1_{(\text{first H})}N_H^2) = p\mathbb{E}(N_H^2|\text{first is H})$$

  However, notice that, since $N_H \backsim Geometric(1-p)$, it is possible, in this case, to simplify computations by stating that:

  $$\mathbb{E}(N_H^2) = \mathbb{V}(N_H) + \mathbb{E}^2(N_H) = \frac{p}{(1-p)^2} + (\frac{1}{1-p})^2 = \frac{p+1}{(1-p)^2}$$

  Therefore, it is possible to conclude that:

  $$\mathbb{V}(1_{(\text{first H})}N_H) = p\frac{(p+1)}{(1-p)^2} - (\frac{p}{1-p})^2 = \frac{p}{(1-p)^2}$$

- $\mathbb{V}(1_{(\text{first T})}N_T)$
  By following the same logic as in the previous point, start by applying the definition of variance:

  $$\mathbb{V}(1_{(\text{first T})}N_T) = \mathbb{E}(1_{(\text{first T})}N_T^2) - \mathbb{E}^2(1_{(\text{first T})}N_T)$$

  By applying the definition of conditional probability, it is possible to rewrite the expectation of the squared variable in the following way:

  $$\mathbb{E}(1_{(\text{first T})}N_T^2) = (1-p)\mathbb{E}(N_T^2|\text{first is T})$$

  However, notice that, since $N_T \backsim Geometric(p)$, it is possible, in this case, to simplify computations by stating that:

  $$\mathbb{E}(N_T^2) = \mathbb{V}(N_T) + \mathbb{E}^2(N_T) = \frac{1-p}{p^2} + (\frac{1}{p})^2 = \frac{2-p}{p^2}$$

  Therefore, it is possible to conclude that:

  $$\mathbb{V}(1_{(\text{first T})}N_T) = (1-p)\frac{2-p}{p^2} - (\frac{1-p}{p})^2 = \frac{1-p}{p^2}$$

- $Cov(1_{(\text{first H})}N_H, 1_{(\text{first T})}N_T)$
  Apply the definition of covariance:

  $$Cov(1_{(\text{first H})}N_H, 1_{(\text{first T})}N_T) = \mathbb{E}(1_{(\text{first H})}N_H 1_{(\text{first T})}N_T) - \mathbb{E}(1_{(\text{first H})}N_H)\mathbb{E}(1_{(\text{first T})}N_T)$$

  However, since the events represented by the indicator random variables are mutually exclusive, it is possible to directly state that $\mathbb{E}(1_{(\text{first H})}N_H 1_{(\text{first T})}N_T) = 0$, allowing to conclude that:

  $$Cov(1_{(\text{first H})}N_H, 1_{(\text{first T})}N_T) = -(\frac{p}{1-p} \cdot \frac{1-p}{p}) = -1$$

Therefore, it is possible to conclude that:

$$\mathbb{V}(X) = \frac{p}{(1-p)^2} + \frac{1-p}{p^2} - 2$$

8.6 Exercise 6

Assume that, on average, 2% of the population is left-handed. Given a sample of 100 individuals, use the Poisson approximation for binomial random variables to compute the probability that at least 3 individuals are left handed.
If X denotes the number of left-handed individuals among the sample, assume that, for a small p and a large N such that $Np \approx 1$, it is possible to perform the following approximation:

$$X \backsim Binomial(N, p) \approx Y \backsim Poisson(Np)$$

In this case, let $X \backsim (100, \frac{2}{100}) \approx Y \backsim Poisson(2)$, resulting in:

$$\mathbb{P}(X \geq 3) = \mathbb{P}(Y \geq 3) = 1 - \sum_{j=0}^{2} \mathbb{P}(Y = j) = 1 - \sum_{j=0}^{2} \frac{e^{-2}2^j}{j!} = 1 - 5e^{-2}$$

8.7 Exercise 7

A biased coin with head probability $p \in (0, 1)$ is tossed a random number of times (independently of the results of the coin tosses) with Poisson distribution of parameter $\lambda > 0$.
Find the probability distribution of the total number of heads and tails obtained, and prove that these two random variables are independent.
Assuming $N \backsim Poisson(\lambda)$ denotes the total number of coin tosses, let X denote the total number of heads and Y denote the total number of tails, and, by applying the law of total probability, let:

$$\mathbb{P}(X = k) = \sum_{n=0}^{\infty} \mathbb{P}(X = k | N = n)\mathbb{P}(N = n)$$

However, by definition of N and X, it must hold that $k \leq n$, which allows to consider that:

$$\mathbb{P}(X = k) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\lambda}\lambda^n}{n!} = \frac{e^{-\lambda}p^k}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} \cdot \lambda^{n-k+k}$$

So, by simplifying, this results in:

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}(\lambda p)^k}{k!} \cdot \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!}$$

If $n - k = n'$, the sum becomes a McLaurin expansion for $f(x) = e^x$, with $x = \lambda(1-p)$, resulting in:

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}(\lambda p)^k}{k!} \cdot e^{\lambda(1-p)} = \frac{e^{-\lambda p}(\lambda p)^k}{k!}$$

So, $X \backsim Poisson(\lambda p)$, and, similarly, $Y \backsim Poisson(\lambda(1-p))$.
This last statement can be used to show that X and Y are independent random variables: by definition, this requires that:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

To do so, pick two arbitrary values $k, j \geq 0$ and apply conditional probability in order to state that:

$$\mathbb{P}(X = k, Y = j) = \mathbb{P}(X = k, Y = j, N = k+j) = \mathbb{P}(X = k, Y = j | N = k+j)\mathbb{P}(N = k+j)$$

Exploit the fact that $N \backsim Poisson(\lambda)$ and state that:

$$\mathbb{P}(X = k, Y = j) = \binom{k+j}{k} p^k (1-p)^j \cdot \frac{e^{-\lambda}\lambda^{k+j}}{(k+j)!}$$

Rearrange the value to recover that:

$$\mathbb{P}(X = k, Y = j) = \frac{(\lambda p)^k}{k!} \cdot \frac{[\lambda(1-p)]^j}{j!} \cdot e^{-\lambda p - \lambda(1-p)} = \frac{e^{-\lambda}(\lambda p)^k}{k!} \cdot \frac{e^{-\lambda(1-p)}[\lambda(1-p)]^j}{j!}$$

However, since $X \backsim Poisson(\lambda p)$ and $Y \backsim Poisson(\lambda(1-p))$, it is possible to conclude that:

$$\mathbb{P}(X = k, Y = j) = \mathbb{P}(X = k)\mathbb{P}(Y = j)$$

Chapter 9

Exercise Sheet 9: Properties of random variables

9.1 Exercise 1

Let X and Y be random variables.

1. Prove that if X is a degenerate random variable, that is $X = c$ for some $c \in \mathbb{R}$, then X and Y are independent.
   By definition, X and Y are independent if and only if it holds that:

   $$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)\ \forall\ (x, y) \in Im(X) \times Im(Y)$$

   Most particularly, notice that $Im(X) = \{c\}$, implying that $\mathbb{P}(X = c) = 1$, which means that:

   $$\mathbb{P}(X = c, Y = y) = 1 \cdot \mathbb{P}(Y = y) = \mathbb{P}(X = c)\mathbb{P}(Y = y)\ \forall\ y \in Im(Y)$$

   Since the values factorise, X and Y are indeed independent random variables.

2. Prove that if X and Y are binary, that is $|Im(X)| = |Im(Y)| = 2$, then the random variables X and Y are independent if and only if $Cov(X, Y) = 0$.
   Consider the definition of covariance between two random variables X and Y:

   $$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

   If X and Y are independent, then $Cov(X, Y) = 0$ by factorisation, although the converse is not always true.
   However, if $X \in \{x_1, x_2\}$ and $Y \in \{y_1, y_2\}$, it is possible to notice that

the converse is also true by replacing $X - x_1 = \{0, x\}$ and $Y - y_1 = \{0, y\}$ for some $x, y \in \mathbb{R}$.
In fact, by computing the expectations, it is possible to recover:

$$\mathbb{E}(X) = \sum_{x \in \{0,x\}} x\mathbb{P}(X = x) = xp_x$$

$$\mathbb{E}(Y) = \sum_{y \in \{0,y\}} y\mathbb{P}(Y = y) = yp_y$$

$$\mathbb{E}(XY) = \sum_{xy \in \{0,x\} \times \{0,y\}} xy(XY = xy) = xyp_{xy}$$

In this case, it is possible to state that $Cov(X, Y) = 0$ if and only if the following equivalence holds:

$$xy\mathbb{P}(X = x, Y = y) = x\mathbb{P}(X = x)y\mathbb{P}(Y = y) \Leftrightarrow \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

Therefore, it is possible to conclude that, in this case, $Cov(X, Y) = 0$ if and only if X and Y are independent.

3. Give an example of two variables X and Y such that $Cov(X, Y) = 0$ but X and Y are not independent.
Pick X to have a symmetric distribution with respect to 0, meaning that:

$$\mathbb{P}(X = x) = \mathbb{P}(X = -x) \ \forall \ x \in Im(X) \Rightarrow \mathbb{E}(X) = 0$$

Now, let $Y = X^2$ and compute the covariance as:

$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X^3) = 0 \text{ by symmetry.}$$

However, notice that X and Y are not independent random variables because Y is expressed in terms of X: indeed, assuming $\mathbb{P}(X = x) = p_x$, it is possible to show that it is possible to violate independence by choosing a value x such that:

$$\mathbb{P}(X = x, Y = x^2) = p_x \neq 2p_x^2 = \mathbb{P}(X = x)\mathbb{P}(Y = x^2)$$

9.2 Exercise 2

For an event A, let 1_A denote the random variable which takes value 1 if $\omega \in A$ and value 0 if $\omega \in A^C$.

1. Let A and B be events. Check that $1_{A^C} = 1 - 1_A$ and $1_{A \cap B} = 1_A 1_B$.
Consider the indicator variable for the complement:

$$1_{A^C} = \begin{cases} 1 & \text{if } \omega \in A^C \\ 0 & \text{if } \omega \in (A^C)^C \end{cases} = \begin{cases} 1 & \text{if } \omega \in A^C \\ 0 & \text{if } \omega \in A \end{cases} = 1 - 1_A$$

Similarly, consider the indicator variable for the intersection:

$$1_{A\cap B} = \begin{cases} 1 & \text{if } \omega \in A \cap B \\ 0 & \text{if } \omega \in (A \cap B)^C \end{cases}$$

By analysing the combinations, it is possible to obtain $1_{A\cap B} = 1_A 1_B$.
N.B.: Alternatively, it is sufficient to just show the computation with the values that the indicator variables would take in every case.

2. Let $a_1, b_1, \dots, a_n, b_n \in \mathbb{R}$. Discuss the binomial identity:

$$\prod_{i=1}^{n}(a_i + b_i) = \sum_{I\subset\{1,\dots,n\}} \prod_{i\in I} a_i \prod_{j\in I^C} b_j$$

Basically, for each couple a_i, b_i, either a_i or b_i can be chosen for the summation: assuming I denotes the set of indices for the couples where a was chosen over b, consider a simple example for $n = 3$:

$(a_1+b_1)(a_2+b_2)(a_3+b_3) = a_1a_2a_3+a_1b_2b_3+\dots$ and so on and so forth $\forall\ I' \subseteq I$

Notice that this procedure can be generalized for any n, meaning that the identity is indeed true.

3. Use the previous points together with the properties of the expectation to prove the inclusion/exclusion principle.
Recall the inclusion/exclusion formula for a sequence $A_1, \dots, A_n$ of events:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1\le i_1<\dots<i_k\le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

Consider this value in terms of its complement:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - \mathbb{P}((\bigcup_{k=1}^{n} A_k)^C) = 1 - \mathbb{P}(\bigcap_{k=1}^{n} A_k^C)$$

To simplify the proof, consider the following identity linking the probability of an event to the expectation of its indicator variable:

$$\mathbb{P}(A) = \mathbb{E}(1_A) \text{ because } \mathbb{E}(1_A) = 0\mathbb{P}(A^C) + 1\mathbb{P}(A) = \mathbb{P}(A)$$

This allows to rewrite:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - \mathbb{P}(\bigcap_{k=1}^{n} A_k^C) = 1 - \mathbb{E}(\prod_{k=1}^{n} 1_{A_k^C}) = 1 - \mathbb{E}(\prod_{k=1}^{n}(1 - 1_{A_k}))$$

Assuming $a_i = -1_{A_i}$ and $b_i = 1$, it is possible to exploit the binomial identity to simplify:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - \mathbb{E}(\prod_{k=1}^{n}(1 - 1_{A_k})) = 1 - \mathbb{E}(\sum_{I\subseteq\{1,\dots,n\}} (\prod_{i\in I}(-1_A))(\prod_{j\in I^C} 1))$$

Notice that, for the subset $I = \emptyset$, the summation is equal to 1, allowing to rewrite:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - 1 + \mathbb{E}(\sum_{\emptyset \subset I \subseteq \{1,\dots,n\}} (-1)^{|I|-1} \prod_{i \in I} 1_{A_i}) = \mathbb{E}(\sum_{\emptyset \subset I \subseteq \{1,\dots,n\}} (-1)^{|I|-1} \prod_{i \in I} 1_{A_i})$$

By exploiting the identity $\mathbb{P}(A) = \mathbb{E}(1_A)$, it is possible to conclude that, for any set I with cardinality k, the equation does indeed to simplify to the inclusion/exclusion formula:

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{\emptyset \subset I \subseteq \{1,\dots,n\}} (-1)^{|I|-1} \mathbb{P}(\bigcap_{i \in I} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\dots,n\}:|I|=k} \mathbb{P}(\bigcap_{i \in I} A_i)$$

9.3 Exercise 3

Consider a biased coin with head probability p unknown. In order to estimate p, the coin is tossed n times and p is estimated via $\frac{S_n}{n}$, where S_n is the number of heads in the n coin tosses. Given $\delta > 0$, determine how large should n be in order for the probability that $|\frac{S_n}{n} - p| < \delta$ to be at least 95%.
Notice that $S_n \backsim Binomial(n, p)$ can also be expressed as the sum of identically distributed and independent random variables $X_i \backsim Bernoulli(p)$, meaning that, by linearity of the expectation, it is possible to state that:

$$\mathbb{E}(\frac{S_n}{n}) = \frac{1}{n} \mathbb{E}(\sum_{i=1}^{n} X_i) = \frac{np}{n} = p$$

Therefore, apply Chebyshev's inequality and reason in terms of the complement:

$$\mathbb{P}(|\frac{S_n}{n} - p| \geq \delta) \leq \frac{\mathbb{V}(\frac{S_n}{n})}{\delta^2} \leq \frac{5}{100}$$

Knowing that $S_n \backsim Binomial(n.p)$, the variance can be computed as the following:

$$\mathbb{V}(\frac{S_n}{n}) = \frac{1}{n^2} \mathbb{V}(\sum_{i=1}^{n} X_i) = \frac{n(p - p^2)}{n^2} = \frac{p(1-p)}{n}$$

Therefore, it must hold that:

$$\mathbb{P}(|\frac{S_n}{n} - p| \geq \delta) \leq \frac{p(1-p)}{n\delta^2}$$

Notice, however, that, since p is not known a priori, the strategy to follow is to bound the probability by picking a value $p \in [0, 1]$ such that the upper bound is maximum, meaning that:

$$\mathbb{P}(|\frac{S_n}{n} - p| \geq \delta) \leq \max_{p \in [0,1]} (\frac{p(1-p)}{n\delta^2})$$

By applying differentiation, it is possible to discover that this upper bound is maximised for $p = \frac{1}{2}$, meaning that:

$$\frac{1}{4n\delta^2} \leq \frac{5}{100} \Leftrightarrow n \geq \frac{100}{20\delta^2} = \frac{5}{\delta^2}$$

9.4 Exercise 4

Consider an album with n coupons. In order to complete the album, you buy one coupon per day (uniformly chosen among all possible coupons, independently of the other days).

1. Show that the expected value of days needed to complete the album is given by:

$$K_n = n(1 + \frac{1}{2} + \cdots + \frac{1}{n})$$

If X denotes the total number of days taken to complete the album, express it in the following way:

$$X = \sum_{i=1}^{n} X_i \text{ and } X_i \text{ counts the days between the } (i-1)^{\text{th}} \text{ and the } i^{\text{th}} \text{ coupons.}$$

Most particularly, notice that, while $X_1 = 1$ is degenerate, it is possible to generalize $X_i \backsim Geometric(p_i)$ with $p = \frac{n-i+1}{n}$.
Therefore, by applying the linearity of expectation, it is possible to conclude that:

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{n-n+1} = n(\sum_{i=1}^{n} \frac{1}{i})$$

N.B.: It is possible to apply a change of variable and let $i' = n - (i-1)$ and rewrite the expectation as:

$$\mathbb{E}(X) = \sum_{i'=1}^{n} \frac{n}{i'}$$

2. Prove the asymptotic relation:

$$K_n = n(\ln n + o(\ln n))$$

Apply the integral test in the interval $[1, n]$:

$$\sum_{i=1}^{n} \frac{1}{i} \approx \int_{1}^{n} \frac{1}{i}\, di = [\ln i]_1^n = \ln n \text{ (not 100\% accurate due to error estimation)}$$

Therefore, it is possible to estimate the expectation as:

$$\sum_{i=1}^{n} \frac{n}{i} \approx n \ln n = n(\ln n + o(\ln n))$$

N.B.: It is possible to find a lower bound as well by applying the integral test on $f(i) = \frac{1}{i+1}$ over the interval $[2, n]$.

9.5 Exercise 5

Show that if a random variable $X \geq 0$ takes integer values, then:

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

Apply the definition of expectation:

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} k\mathbb{P}(X = k)$$

Rewrite n as a summation of many ones, resulting in:

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} (\sum_{k=1}^{n} 1)\mathbb{P}(X = k)$$

It is possible to interchange the summations for $1 \leq k \leq n$ and $n \geq 1$, allowing to conclude that:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{P}(X = n) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

N.B.: A useful application of this proof is for the computation of the expectation of a geometric random variable by means of a survival function.

9.6 Exercise 6

A server can process two types of requests: A and B. Assume that the total number of requests is described by a Poisson random variable with parameter 5 and each request, independently from the others, is of type A with probability $\frac{2}{3}$ and of type B with probability $\frac{1}{3}$.

1. Compute the probability that the server receives 5 requests of type A.
   Assuming $X \backsim Poisson(5)$ denotes the total number of requests processed by the server, if $X_A \backsim Poisson(\lambda_A)$ denotes the number of requests of type A and $X_B \backsim Poisson(\lambda_B)$ denotes the number of requests of type B, it is

possible to consider:

$$X = X_A + X_B, \text{ with } \lambda_A = \frac{10}{3} \text{ and } \lambda_B = \frac{5}{3}.$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(X_A = 5) = \frac{e^{-\lambda_A}\lambda_A^5}{5!} = \frac{e^{-\frac{10}{3}}(\frac{10}{3})^5}{5!}$$

2. Compute the expectation values of the requests of type A and B.
   Knowing that $X_A \backsim Poisson(\frac{10}{3})$ and $X_B \backsim Poisson(\frac{5}{3})$, it is possible to directly conclude that:

$$\mathbb{E}(X_A) = \lambda_A = \frac{10}{3} \text{ and } \mathbb{E}(X_B) = \lambda_B = \frac{5}{3}$$

3. Knowing that the total number of requests is 9, compute the probability that 6 are of type A.
   Apply the definition of conditional probability:

$$\mathbb{P}(X_A = 6|X = 9) = \frac{\mathbb{P}(X_A = 6, X = 9)}{\mathbb{P}(X = 9)}$$

   Most particularly, notice that having 6 out of 9 requests of type A implies that the remaining 3 will be of type B, allowing to conclude that:

$$\mathbb{P}(X_A = 6|X = 9) = \frac{\mathbb{P}(X_A = 6, X_B = 3)}{\mathbb{P}(X = 9)}, \text{ resulting in:}$$

$$\mathbb{P}(X_A = 6|X = 9) = \frac{e^{-\frac{10}{3}}(\frac{10}{3})^6}{6!} \cdot \frac{e^{-\frac{5}{3}}(\frac{5}{3})^3}{3!} \cdot \frac{9!}{e^{-(\frac{10}{3}+\frac{5}{3})}(\frac{10}{3}+\frac{5}{3})^{6+3}} = \binom{9}{6}(\frac{2}{3})^6(\frac{1}{3})^3$$

   N.B.: In this case, when conditioned by X, X_A behaves like a binomial random variable.

4. Knowing that the total number of requests is at least three, compute the probability that the first three are of type A.
   Apply the definition of conditional probability:

$$\mathbb{P}(\{\text{first three} = A|X \geq 3\}) = \frac{\mathbb{P}(\{\text{first three} = A, X \geq 3\})}{\mathbb{P}(X \geq 3)}$$

   Observe that, by independence, it is possible to factorise the joint probability, allowing to conclude that:

$$\mathbb{P}(\{\text{first three} = A|X \geq 3\}) = \mathbb{P}(\{\text{first three} = A\}) = (\frac{2}{3})^3$$

9.7 Exercise 7

Alice and Bob participate to a prize game. They have a fair die (common to Alice and Bob) and two independent coins (one for Alice and one for Bob) which give head with probability $p \in (0,1)$ and tails with probability $1-p$. The rules of the game are the following: First, the die is tossed, and, if it gives 1, there are no prizes, otherwise both Alice and Bob toss their coin until it returns head, and they win 1 euro for each toss they did.
Let X and Y denote the euros won by Alice and Bob, respectively.

1. Are X and Y independent random variables?
   The random variables are not independent because of the die roll.
   in fact, by considering the definition of the experiment, it is possible to discover the following counterexample by assuming that Alice and Bob win 0 euros each:

   $$\mathbb{P}(X=0, Y=0) = \mathbb{P}(\{\text{the die gives } 1\}) = \frac{1}{6} \neq \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(X=0)\mathbb{P}(Y=0)$$

2. Compute the probability distribution of X and Y.
   By definition of the experiment, X and Y are symmetric random variables.
   Most particularly, notice that Alice/Bob wins 0 euros if the die rolls 1, otherwise, Alice/Bob wins k euros if the die does not roll 1 and the first head is seen at the k^{th} toss, meaning that:

   $$\mathbb{P}(X=k) = \mathbb{P}(Y=k) = \begin{cases} \frac{1}{6} & \text{if } k=0 \\ \frac{5}{6}(1-p)^{k-1}p & \text{if } k \geq 1 \end{cases}$$

3. Compute the expectation value of X and Y.
   Knowing that X and Y are symmetric, apply the definition of expectation:

   $$\mathbb{E}(X) = \mathbb{E}(Y) = \sum_{k=0}^{\infty} k\mathbb{P}(X=k) = \sum_{k=1}^{\infty} k \cdot \frac{5}{6}(1-p)^{k-1}p = \frac{5}{6p}$$

4. Compute the covariance between X and Y.
   Consider the definition of covariance:

   $$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

   Most particularly, since $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are already known, the goal is to find $\mathbb{E}(XY)$:

   $$\mathbb{E}(XY) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} kj\mathbb{P}(XY=kj) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} kj \cdot \frac{5}{6}(1-p)^{k-1}p(1-p)^{j-1}p, \text{ resulting in:}$$

   $$\mathbb{E}(XY) = \frac{5}{6}\sum_{k=1}^{\infty}(k(1-p)^{k-1}p)\sum_{j=1}^{\infty}(j(1-p)^{j-1}p) = \frac{5}{6p^2}$$

Therefore, it is possible to conclude that:

$$Cov(X,Y) = \frac{5}{6p^2} - (\frac{5}{6p})^2 = \frac{5}{36p^2}$$

N.B.: Since $Cov(X,Y) > 0$, X and Y are said to be positively correlated.

Chapter 10

Exercise Sheet 10: Joint distributions and continuous probability

10.1 Exercise 1

From a collection of 7 batteries, of which 3 are new, 2 are used and 2 are broken, 3 batteries are chosen at random. Let X and Y denote respectively the number of new batteries and used working batteries, among the ones chosen.

1. Determine the joint distribution of (X, Y) and the marginal distributions of X and Y.
   Start by finding the dimension of the sample space, which represents all the possible ways such that the three batteries can be chosen among the group of 7:

$$|\Omega| = \binom{7}{3} = \frac{7!}{(7-3)! \cdot 3!} = 35$$

   At this point, compute the probability that, among the three chosen batteries, x are new and y are used but working:

$$\mathbb{P}(X = x, Y = y) = \frac{\binom{3}{x} \cdot \binom{2}{y} \cdot \binom{2}{3-(x+y)}}{35}, \text{ provided: } \begin{cases} x \leq 3 \wedge y \leq 2 \\ x + y \leq 3 \\ 3 - (x+y) \leq 2 \end{cases}$$

   At this point, it is possible to exploit the statement to build the table for the joint distribution of (X, Y):

	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	0	$\frac{3}{35}$	$\frac{6}{35}$	$\frac{1}{35}$
$Y=1$	$\frac{2}{35}$	$\frac{12}{35}$	$\frac{6}{35}$	0
$Y=2$	$\frac{2}{35}$	$\frac{3}{35}$	0	0

Therefore it is possible to recover the marginal distributions of both X and Y through the law of total probability:

- Marginal probability distribution of X:

$$\mathbb{P}(X=x)=\sum_{j=0}^{2}\mathbb{P}(X=x,Y=j)=\begin{cases}\frac{4}{35} & \text{if } X=0\\ \frac{18}{35} & \text{if } X=1\\ \frac{12}{35} & \text{if } X=2\\ \frac{1}{35} & \text{if } X=3\end{cases}$$

- Marginal probability distribution of Y:

$$\mathbb{P}(Y=y)=\sum_{i=0}^{3}\mathbb{P}(X=i,Y=y)=\begin{cases}\frac{10}{35} & \text{if } Y=0\\ \frac{20}{35} & \text{if } Y=1\\ \frac{5}{35} & \text{if } Y=2\end{cases}$$

2. Compute $Cov(X,Y)$. Are the random variables X and Y independent? Apply the definition of covariance:

$$Cov(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$$

At this point, derive the various expectations by applying the definition:

$$\mathbb{E}(X)=\sum_{x\in Im(X)}x\mathbb{P}(X=x)=\frac{9}{7}$$

$$\mathbb{E}(Y)=\sum_{y\in Im(Y)}y\mathbb{P}(Y=y)=\frac{6}{7}$$

$$\mathbb{E}(XY)=\sum_{x\in Im(X)}\sum_{y\in Im(Y)}xy\mathbb{P}(XY=xy)=\frac{6}{7}$$

Therefore, it is possible to conclude that:

$$Cov(X,Y)=\frac{6}{7}-\frac{9}{7}\cdot\frac{6}{7}=\frac{6}{7}-\frac{54}{49}=-\frac{12}{49}$$

Notice that, since $Cov(X,Y)\neq 0$, X and Y will for sure not be independent random variables: instead, $Cov(X,Y)<0$ implies that X and Y are negatively correlated.

3. The three chosen batteries are inserted in a machine which works if none of them is broken. Determine the probability that the machine works.
It is possible to make sure that the machine works by just choosing among the new or working batteries, meaning that, since two batteries are broken, the machine will work if and only if no battery is chosen among the broken ones, resulting in:

$$\mathbb{P}(\{\text{the machine works}\}) = \frac{\binom{7-2}{3}}{\binom{7}{3}} = \frac{5!}{2!\cdot 3!} \cdot \frac{4!\cdot 3!}{7!} = \frac{4\cdot 3}{2!\cdot 7\cdot 6} = \frac{2}{7}$$

10.2 Exercise 2

A die has a blue face, two red faces and three green faces. The die is rolled twice. Let R denote the number of times a red face is obtained and G denote the number of times a green face is seen.

1. Construct the table of joint distribution of (R, G).
Notice that, by definition of the experiment, R and G are independent random variables, meaning that the table will be:

	$R=0$	$R=1$	$R=2$
$G=0$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{4}{36}$
$G=1$	$\frac{6}{36}$	$\frac{12}{36}$	0
$G=2$	$\frac{9}{36}$	0	0

2. Determine the distribution of $Z = \max\{R, G\}$ and compute $\mathbb{E}(Z)$ and $\mathbb{V}(Z)$.
By definition of the experiment, let $Im(Z) = \{0, 1, 2\}$ and define the probability distribution of Z in the following way:

$$\mathbb{P}(Z = z) = \begin{cases} \frac{1}{36} & \text{if } Z = 0 \\ \frac{22}{36} & \text{if } Z = 1 \\ \frac{13}{36} & \text{if } Z = 2 \end{cases}$$

At this point, apply the definition of expectation and compute:

$$\mathbb{E}(Z) = \sum_{z\in Im(Z)} z\mathbb{P}(Z = z) = 0 \cdot \frac{1}{36} + 1 \cdot \frac{22}{36} + 2 \cdot \frac{13}{36} = \frac{4}{3}$$

Similarly, apply the definition of variance and compute:

$$\mathbb{V}(Z) = \mathbb{E}(Z^2) - \mathbb{E}^2(Z) = \frac{37}{18} - (\frac{4}{3})^2 = \frac{5}{18}$$

10.3 Exercise 3

The electrical components produced in a factory are defective, independently of each other, with probability $p \in (0,1)$ and they are functioning with probability $1-p$. The components are subject to a quality check with the following procedure: each component, independently of the others, is checked with probability $\alpha \in (0,1)$ and it is not checked with probability $1-\alpha$.
A component which is found defective is discarded, while the others are put on sale.
Suppose that the factory produced n components.

1. Compute the distribution of the number of components which are discarded after the quality check procedure.
   Consider the outcomes in which a component may get discarded:

	Defective	Not defective
Checked	$p\alpha$	$(1-p)\alpha$
Not checked	$p(1-\alpha)$	$(1-p)(1-\alpha)$

   Assuming X denotes the total number of discarded components, it is possible to express it as the sum of independent Bernoulli random variables of type X_i, which denote whether component i has been discarded or not, meaning that:

$$X = \sum_{i=1}^{n} X_i, \text{ with } X_i = \begin{cases} 1 & \text{with probability } p\alpha \\ 0 & \text{with probability } 1-p\alpha \end{cases}$$

   This is equivalent to letting $X \backsim Binomial(n, p\alpha)$, resulting in:

$$\mathbb{P}(X=k) = \binom{n}{k}(p\alpha)^k(1-p\alpha)^{n-k}, \text{ for } k \in \{0,\dots,n\}.$$

   N.B.: An alternative way to recover $X \backsim Binomial(n, p\alpha)$ is via indicator random variable:

$$X = \sum_{i=1}^{n} 1_{A=\{\text{component } i \text{ is discarded}\}}, \text{ with } 1_A \backsim Bernoulli(p\alpha)$$

2. Knowing that k components have been discarded after the quality check procedure, compute the distribution of the number of defective components among the $n-k$ components which are put on sale.
   Assuming Y denotes the total number of defective components that are put on sale among the ones that got through the quality check procedure, it is possible to express it as the sum of independent Bernoulli random variables of type Y_i, which denote whether component i was defective and

put on sale or not, meaning that:

$$Y = \sum_{i=1}^{n-k} Y_i, \text{ with } Y_i = \begin{cases} 1 & \text{with probability } q \\ 0 & \text{with probability } 1-q \end{cases}$$

Most particularly, let q denote the probability that the component was defective given that it managed to get through the quality check procedure, resulting in:

$$q = \mathbb{P}(\{\text{defective}|\text{passed check}\}) = \frac{\mathbb{P}(\{\text{defective and passed check}\})}{\mathbb{P}(\{\text{passed check}\})} = \frac{p(1-\alpha)}{1-p\alpha}$$

Therefore, it is possible to conclude that $Y \backsim Binomial(n-k, \frac{p(1-\alpha)}{1-p\alpha})$, meaning that:

$$\mathbb{P}(Y=h|X=k) = \binom{n-k}{h}(\frac{p(1-\alpha)}{1-p\alpha})^h(1-\frac{p(1-\alpha)}{1-p\alpha})^{n-k-h}, \text{ for } h \in \{0,\dots,n-k\}.$$

N.B.: An alternative way to compute the probability of the event is by considering:

$$\mathbb{P}(Y=h|X=k) = \frac{\mathbb{P}(X=k, Y=h)}{\mathbb{P}(X=k)} = \frac{\binom{n}{k\ h\ n-(k+h)}(p\alpha)^k(p(1-\alpha))^h(1-p)^{n-(k+h)}}{\binom{n}{k}(p\alpha)^k(1-p\alpha)^{n-k}}$$

This will eventually simplify to a binomial distribution:

$$\mathbb{P}(Y=h|X=k) = \binom{n-k}{h}(\frac{p(1-\alpha)}{1-p\alpha})^h(1-\frac{p(1-\alpha)}{1-p\alpha})^{n-(k+j)}$$

10.4 Exercise 4

Let X, Y be two independent Bernoulli random variables with parameter p. Define:

$$Z = X(1-Y) \text{ and } W = 1 - XY$$

1. What is the joint distribution of (Z, W)?
   Start by considering the joint distribution of (X, Y):

	$X=0$	$X=1$
$Y=0$	$(1-p)^2$	$p(1-p)$
$Y=1$	$(1-p)p$	p^2

   For this reason the joint distribution of (Z, W) can be written as:

	$Z=0$	$Z=1$
$W=0$	p^2	0
$W=1$	$1-p$	$p(1-p)$

2. What are the marginal distributions of Z and W?
   It is possible to recover the marginal distributions of both Z and W through the law of total probability:

   - Marginal probability distribution of Z:

   $$\mathbb{P}(Z=z)=\begin{cases} p^2+1-p & \text{if } Z=0 \\ p-p^2 & \text{if } Z=1 \end{cases}$$

     Notice that $Z \backsim Bernoulli(p(1-p))$.
   - Marginal probability distribution of W:

   $$\mathbb{P}(W=w)=\begin{cases} p^2 & \text{if } W=0 \\ 1-p^2 & \text{if } W=1 \end{cases}$$

     Notice that $W \backsim Bernoulli(1-p^2)$.

3. For which values of p are the random variables Z and W independent?
   The goal is to find p such that the joint distribution factorises, meaning that:

   $$\mathbb{P}(Z=z,W=w)=\mathbb{P}(Z=z)\mathbb{P}(W=w)$$

   Observe that, since Z and W are binary random variables, it is possible to exploit the following remark:

   $$Z \text{ and } W \text{ are independent} \Leftrightarrow Cov(Z,W)=0$$

   Therefore, the goal is to find p such that $Cov(Z,W)=0$:

   $$Cov(Z,W)=\mathbb{E}(ZW)-\mathbb{E}(Z)\mathbb{E}(W)=0$$

   However, since Z and W are Bernoulli random variables, it is possible to simplify computation and directly derive that:

   $$Cov(Z,W)=p(1-p)-p(1-p)(1-p^2)=p^3(1-p)=0 \Leftrightarrow p=0 \vee p=1$$

   Therefore, Z and W will be independent if and only if they are both degenerate random variables.
   N.B.: An alternative, but longer approach, is to check the independence criterion for all possible cases.

10.5 Exercise 5

Let X be a continuous random variable with $X \geq 0$. Show that:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x)\ dx$$

Generally speaking, if a continuous random variable X has probability density function $f_X(x)$, then its expectation can be expressed as:

$$\mathbb{E}(X) = \int_{-\infty}^\infty x f_X(x)\ dx$$

In this case, it is possible to exploit this information and reason in terms of the complement, resulting in:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x)\ dx = \int_0^\infty 1 - \mathbb{P}(X \leq x)\ dx = \int_0^\infty 1 - F_X(x)\ dx$$

Now, knowing that, by definition, $f_X(x) = F'_X(x)$, it is possible to compute the integral by applying integration by parts:

$$\mathbb{E}(X) = \int_0^\infty x f_X(x)\ dx = [x(F_X(x)-1)]_0^\infty - \int_0^\infty F_X(x) - 1\ dx = \int_0^\infty 1 - F_X(x)\ dx$$

Therefore, it is possible to conclude that:

$$\mathbb{E}(X) = \int_0^\infty 1 - \mathbb{P}(X \leq x) = \int_0^\infty \mathbb{P}(X > x)\ dx$$

N.B.: For simplicity, assume $x f_X(x) = o(\frac{1}{x})$.

10.6 Exercise 6

A and B play the following game: A writes 1 or 2 on a piece of paper, and B guesses the number written by A. If A wrote $i \in \{1, 2\}$ and B guesses right, then A pays i euros to B. If, on the other hand, B guesses wrong, then B pays 0.75 euros to A.

1. Assume that B uses a random strategy, guessing 1 with probability p and 2 with probability $1-p$: Knowing that A wrote 1, determine the expected gain of B.
   Let X denote the gain of B after a single guess.
   Assuming A wrote 1, there are only two possible outcomes: either B guesses 1 correctly (probability p) and wins 1 euro or B guesses 2 incorrectly (probability $1-p$) and loses 0.75 euros.
   This means that X can be defined in the following way:

$$\mathbb{P}(X = x | \{\text{A wrote } 1\}) = \begin{cases} 1 & \text{with probability } p \\ -0.75 & \text{with probability } 1-p \end{cases}$$

Therefore, it is possible to find the expectation as:

$$\mathbb{E}(X|\{\text{A wrote 1}\}) = \sum_{x \in Im(X)} x\mathbb{P}(X = x|\{\text{A wrote 1}\}) = \frac{7p-3}{4}$$

2. Knowing that A wrote 2, determine the expected gain of B.
   Again, let X denote the gain of B after a single guess, and apply the same logic as the previous point.
   Assuming A wrote 2, there are only two possible outcomes: either B guesses 2 correctly (probability $1-p$) and wins 2 euros or B guesses 1 incorrectly (probability p) and loses 0.75 euros.
   This means that X can be defined in the following way:

$$\mathbb{P}(X = x|\{\text{A wrote 1}\}) = \begin{cases} 2 & \text{with probability } 1-p \\ -0.75 & \text{with probability } p \end{cases}$$

   Therefore, it is possible to find the expectation as:

$$\mathbb{E}(X|\{\text{A wrote 2}\}) = \sum_{x \in Im(X)} x\mathbb{P}(X = x|\{\text{A wrote 2}\}) = \frac{8-11p}{4}$$

3. Determine the value of p which maximises the minimum between the two expected gains.
   Maximising the minimum gain is equivalent to minimising the maximum gain, so consider:

$$\mathbb{E}(X|\{\text{A wrote 1}\}) = \mathbb{E}(X|\{\text{A wrote 2}\}\}, \text{ resulting in:}$$
$$\frac{7p-3}{4} = \frac{8-11p}{4} \Leftrightarrow 18p = 11 \Leftrightarrow p = \frac{11}{18}$$

4. Suppose that A uses a random strategy, writing 1 with probability q and 2 with probability $1-q$: knowing that B guessed 1, determine the expected loss of A.
   Let Y denote the loss of A after a single guess.
   Assuming B guessed 1, there are only two possible outcomes: either A wrote 1 (probability q) and loses 1 euro or A wrote 2 (probability $1-q$) and gain 0.75 euros.
   This means that Y can be defined in the following way:

$$\mathbb{P}(Y = y|\{\text{B guessed 1}\}) = \begin{cases} 1 & \text{with probability } q \\ -0.75 & \text{with probability } 1-q \end{cases}$$

   Therefore, it is possible to find the expectation as:

$$\mathbb{E}(Y|\{\text{B guessed 1}\}) = \sum_{y \in Im(Y)} y\mathbb{P}(Y = y|\{\text{B guessed 1}\}) = \frac{7q-3}{4}$$

5. Knowing that B guessed 2, determine the expected loss of A.
   Again, let Y denote the loss of A after a single guess, and apply the same logic as the previous point.
   Assuming B guessed 2, there are only two possible outcomes: either A wrote 1 (probability q) and gains 0.75 euros or A wrote 2 (probability $1-q$) and loses 2 euros.
   This means that Y can be defined in the following way:

$$\mathbb{P}(Y=y|\{\text{B guessed 2}\}) = \begin{cases} 2 & \text{with probability } 1-q \\ -0.75 & \text{with probability } q \end{cases}$$

   Therefore, it is possible to find the expectation as:

$$\mathbb{E}(Y|\{\text{B guessed 2}\}) = \sum_{y \in Im(Y)} y\mathbb{P}(Y=y|\{\text{B guessed 2}\}) = \frac{8-11q}{4}$$

6. Determine the value of q which minimises the maximum between the two expected gains above.
   Minimising the maximum gain is equivalent to maximising the minimum gain, so consider:

$$\mathbb{E}(Y|\{\text{B guessed 1}\}) = \mathbb{E}(\{\text{B guessed 2}\}), \text{ resulting in:}$$
$$\frac{7q-3}{4} = \frac{8-11q}{4} \Leftrightarrow 18q = 11 \Leftrightarrow q = \frac{11}{18}$$

10.7 Exercise 7

Let X, Y be two random variables. For $y \in Im(Y)$, let $\mathbb{E}(X|y)$ denote the expected value of X conditional on the event $\{Y=y\}$. Prove that:

$$\mathbb{E}(X) = \sum_{y \in Im(Y)} \mathbb{P}(Y=y)\mathbb{E}(X|y)$$

Consider the definition of conditional expectation:

$$\mathbb{E}(X|y) = \sum_{x \in Im(X)} x\mathbb{P}(X=x|Y=y) = \sum_{x \in Im(X)} x\frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$$

By applying the law of total probability, it is possible to rewrite the expectation of X in the following way:

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x\mathbb{P}(X=x) = \sum_{x \in Im(X)} \sum_{y \in Im(Y)} x\mathbb{P}(X=x|Y=y)\mathbb{P}(Y=y)$$

Now, it is possible to switch the summation order to conclude that:

$$\mathbb{E}(X) = \sum_{y \in Im(Y)} \mathbb{P}(Y=y) \sum_{x \in Im(X)} x\mathbb{P}(X=x|Y=y) = \sum_{y \in Im(Y)} \mathbb{P}(Y=y)\mathbb{E}(X|y)$$

10.8 Exercise 8

Let X be a continuous random variable with probability density function:

$$f(x) = \begin{cases} \frac{1}{6}(x+k) & \text{if } x \in [0,k] \\ 0 & \text{otherwise} \end{cases}$$

1. Compute the value of $k \geq 0$.
   The goal is to find $k \geq 0$ such that $f(x)$ is a legit probability density function:

   $$\int_{-\infty}^{\infty} f(x)\ dx = \int_0^k f(x)\ dx = 1$$

   Therefore, solve the integral:

   $$\int_0^k \frac{x+k}{6}\ dx = [\frac{(x+k)^2}{12}]_0^k = \frac{4k^2 - k^2}{12} = \frac{k^2}{4} = 1 \Leftrightarrow k = \pm 2$$

   However, since $k \geq 0$ by definition, $k = -2$ is not an acceptable solution, meaning that $f(x)$ will be a legit probability density function if and only if $k = 2$.

   $$\text{Therefore, } f(x) = \begin{cases} \frac{x+2}{6} & \text{if } x \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

2. Compute $\mathbb{P}(1 \leq X \leq 2)$.

   $$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 \frac{x+2}{6}\ dx = [\frac{(x+2)^2}{12}]_1^2 = \frac{16-9}{12} = \frac{7}{12}$$

Chapter 11

Exercise Sheet 11: Continuous probability

11.1 Exercise 1

Let X_1, X_2 be independent uniform random variables in $[0, 1]$.

1. Compute the probability density function of $X_1 + X_2$.
   Start by considering the probability density functions of $X_1, X_2 \backsim Uniform(0, 1)$:

$$f_{X_1}(x) = f_{X_2}(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

   By applying the fundamental theorem of integral calculus, it is possible to recover the cumulative distribution functions of X_1 and X_2 as the antiderivative of their probability density functions, meaning that:

$$F_{X_1}(x) = F_{X_2}(x) = \mathbb{P}(X_1 \leq x) = \int_{-\infty}^{x} f_{X_1}(y)\ dy = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \in [0,1] \\ 1 & \text{if } x > 1 \end{cases}$$

   Notice that, if $X = X_1 + X_2$, then $Im(X) = [0, 2]$, meaning that it is possible to start denoting its cumulative distribution function in the following way:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 + X_2 \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ ??? & \text{if } x \in [0,2] \\ 1 & \text{if } x > 2 \end{cases}$$

   Observe that, if $X = X_1 + X_2 = x$, then $\mathbb{P}(X \leq x)$ for $x \in [0, 2]$ is basically the ratio between the area of the region D of points (x_1, x_2) such that $x_1 + x_2 \leq x$ and the total area $[0,1] \times [0,1]$, meaning that:

$$F_X(x) = \int\int_D f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 = \int\int_D f_{X_1}(x) f_{X_2}(x)\, dx_1 dx_2 \text{ by independence.}$$

Most particularly, by plotting these regions graphically, it is possible to notice that, given a line of equation $x_1 + x_2 = x$, then $x_1 + x_2 \leq x$ is represented by the area below the line, meaning that it is possible to define two main cases according to the value of x:

- If $x \in [0,1]$, then the region is the triangle with legs equal to x, meaning that:

$$\mathbb{P}(X \leq x, x \in [0,1]) = \frac{x^2}{2}$$

- If $x \in [1,2]$, then the region is the portion of the unit square but without the upper region, whose area is $\frac{(2-x)^2}{2}$, meaning that:

$$\mathbb{P}(X \leq x, x \in [1,2] = 1 - \frac{(2-x)^2}{2}$$

Therefore, it is possible to define the cumulative distribution function of $X = X_1 + X_2$ in the following way:

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{2} & \text{if } x \in [0,1] \\ 1 - \frac{(2-x)^2}{2} & \text{if } x \in [1,2] \\ 1 & \text{if } x > 2 \end{cases}$$

For this reason, by applying differentiation, it is possible to recover the probability density function as:

$$f_X(x) = \frac{d(F_X(x))}{dx} = \begin{cases} x & \text{if } x \in [0,1] \\ 2 - x & \text{if } x \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$

2. Compute the probability density function of $\max\{X_1, X_2\}$.
Let $X = \max\{X_1, X_2\}$, and notice that $X = \max\{X_1, X_2\} \leq x$ if and only if $X_1 \leq x$ and $X_2 \leq x$.
For this reason, by exploiting independence, it is possible to derive the cumulative distribution function of X by applying the following reasoning:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x) = \mathbb{P}(X_1 \leq x)\mathbb{P}(X_2 \leq x) = F_{X_1}(x) F_{X_2}(x)$$

At this point, by remembering the definition of $F_{X_1}(x)$ and $F_{X_2}(x)$, it is possible to provide the following analytical expression for the cumulative distribution function of X:

$$F_X(x) = \begin{cases} x^2 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Therefore, by applying differentiation, it is possible to recover the following probability density function:

$$f_X(x) = \frac{d(F_X(x))}{dx} = \begin{cases} 2x & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

3. Compute the probability density function of $\min\{X_1, X_2\}$.
   Since, in this case, it is easier to reason in terms of the complement, let $X = \min\{X_1, X_2\}$, and notice that $X = \min\{X_1, X_2\} > x$ if and only if $X_1 > x$ and $X_2 > x$ as well.
   For this reason, by reasoning in terms of the complement, it is possible to derive the cumulative distribution function of X in the following way:

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - \mathbb{P}(X_1 > x, X_2 > x) = 1 - \mathbb{P}(X_1 > x)\mathbb{P}(X_2 > x)$$

   Most particularly, by applying the definition of cumulative distribution function, it is possible to simplify this result as:

$$F_X(x) = 1 - ((1 - \mathbb{P}(X_1 \leq x))(1 - \mathbb{P}(X_2 \leq x))) = 1 - (1 - F_{X_1}(x))(1 - F_{X_2}(x))$$

   At this point, by remembering the definition of $F_{X_1}(x)$ and $F_{X_2}(x)$, it is possible to provide the following analytical expression for the cumulative distribution function of X:

$$F_X(x) = \begin{cases} 1 - (1-x)^2 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

   Therefore, by applying differentiation, it is possible to recover the probability density function as:

$$f_X(x) = \frac{d(F_X(x))}{dx} = \begin{cases} 2 - 2x & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

11.2 Exercise 2

Let U be a uniform random variable in $[0,1]$ and let V be a random variable, independent of U, uniformly distributed in $[-1,1]$.

1. Compute the probability density function of V^2.
   Start by considering the probability density functions of U and V:

$$f_U(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

$$f_V(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

At this point, consider the cumulative distribution function of V^2:

$$F_{V^2}(x) = \mathbb{P}(V^2 \leq x) = \begin{cases} 0 & \text{if } x \leq 0 \\ ??? & \text{if } x \in (0,1) \\ 1 & \text{if } x \geq 1 \end{cases}$$

Observe that it is possible to rewrite the non-trivial case as:

$$F_{V^2}(x) = \mathbb{P}(V^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq V \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} f_V(V)\, dV = [\frac{V}{2}]_{-\sqrt{x}}^{\sqrt{x}} = \sqrt{x}$$

Therefore, it is possible to recover the probability density function of V^2 by applying differentiation:

$$f_{V^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

2. Compute the probability density function of $\ln(\frac{1}{U})$.
   Start by considering the cumulative distribution function of $\ln(\frac{1}{U})$:

$$F_{\ln(\frac{1}{U})}(x) = \mathbb{P}(\ln(\frac{1}{U}) \leq x) = \mathbb{P}(-\ln U \leq x) = \mathbb{P}(U \geq e^{-x}) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{if } x \geq 0 \end{cases}$$

   Therefore, it is possible to recover the probability density function of $\ln(\frac{1}{U})$ by applying differentiation:

$$f_{\ln(\frac{1}{U})}(x) = \frac{d(F_{\ln(\frac{1}{U})}(x))}{dx} = \begin{cases} 0 & \text{if } x < 0 \\ e^{-x} & \text{if } x \geq 0 \end{cases}$$

3. Compute $\mathbb{P}(U \leq V)$.
   By definition of $U \backsim Uniform(0,1)$ and $V \backsim Uniform(-1,1)$, if the region A denotes the set of all points $(x,y) \in [0,1] \times [-1,1]$ such that $x \leq y$, allowing to rewrite the event as:

$$\mathbb{P}(U \leq V) = \mathbb{P}(U - V \leq 0) = \mathbb{P}((U,V) \in A)$$

   Graphically, notice that this probability can be computed through the ratio between the area of A, which is represented by a triangle whose hypotenuse is the line $y = x$, and the area of $[0,1] \times [-1,1]$, meaning that:

$$\mathbb{P}(U \leq V) = \frac{Area(A)}{Area([0,1] \times [-1,1])} = \frac{\frac{1}{2}}{2} = \frac{1}{4}$$

11.3 Exercise 3

Let X be a Gaussian random variable with mean 2 and variance 25.
Provide answers to the following questions by using Gaussian integral tables.

1. Compute $\mathbb{P}(|X-2| \geq 7)$.
   In order to use the Gaussian integral tables for a Gaussian random variable $X \backsim \mathcal{N}(\mu, \sigma^2)$, start by applying a linear transformation to X in order to recover a standard Gaussian random variable $Z = \frac{X-\mu}{\sigma} \backsim \mathcal{N}(0,1)$.
   Therefore, let $Z = \frac{X-2}{5} \backsim \mathcal{N}(0,1)$ and rewrite the event in the following way:
   $$\mathbb{P}(|X-2| \geq 7) = \mathbb{P}(|Z| \geq \frac{7}{5}) = \mathbb{P}(\{Z \leq -\frac{7}{5}\} \cup \{Z \geq \frac{7}{5}\})$$
   By applying additivity and symmetry, this result can be simplified to:
   $$\mathbb{P}(|X-2| \geq 7) = (1-\Phi(\frac{7}{5}))-(1-\Phi(\frac{7}{5})) = 2(1-\Phi(\frac{7}{5})) = 2(1-0.9192) = 0.1616$$
2. Compute $\mathbb{P}(0 \leq X \leq 7)$.
   Again, standardize X into $Z \backsim \mathcal{N}(0,1)$, and rewrite the event in the following way:
   $$\mathbb{P}(0 \leq X \leq 7) = \mathbb{P}(-\frac{2}{5} \leq Z \leq 1) = \mathbb{P}(\{Z \geq -\frac{2}{5}\} \cap \{Z \leq 1\})$$
   By applying symmetry, it is possible to conclude that:
   $$\mathbb{P}(0 \leq X \leq 7) = \Phi(1)-\Phi(-\frac{2}{5}) = \Phi(1)-(1-\Phi(\frac{2}{5})) = 0.8413-(1-0.6554) = 0.4967$$
3. Determine α such that $\mathbb{P}(X \geq \alpha) \leq 0.1$.
   Again, standardize X into $Z \backsim \mathcal{N}(0,1)$, and rewrite the event in the following way:
   $$\mathbb{P}(X \geq \alpha) = \mathbb{P}(Z \geq \frac{\alpha-2}{5}) \leq 0.1$$
   At this point, it is possible to reason in terms of the complement, resulting in:
   $$\mathbb{P}(Z \geq \frac{\alpha-2}{5}) = 1-\Phi(\frac{\alpha-2}{5}) \leq 0.1 \Rightarrow \Phi(\frac{\alpha-2}{5}) \geq 0.9 \Rightarrow \frac{\alpha-2}{5} \geq 1.29 \Rightarrow \alpha \geq 8.45$$

11.4 Exercise 4

In order to transmit a bit from a source A to a receiver B via a pair of electrical wires, one applies a potential difference of +2V for the value 1 and -2V for the value 0 . Due to electromagnetic disturbance, if A applies $\mu = \pm 2$V, B reads $X = \mu + Z$, where Z represents the noise, described by a Gaussian random variable with mean 0 and variance 1. After reading X, B registers the message with the following rule: if $X \geq 0.5$, then B registers 1, while if $X < 0.5$ then B registers 0.

1. If A sends 0, compute the probability that B registers 1.
   By definition of the experiment, it must hold, in this case, that:

   $$X = \mu + Z \Rightarrow -2 + Z \geq 0.5 \Rightarrow Z \geq 2.5$$

   Most particularly, since $Z \backsim \mathcal{N}(0,1)$, it is possible to reason in terms of the complement in order to exploit Gaussian integral tables:

   $$\mathbb{P}(B = 1|A = 0) = \mathbb{P}(Z \geq 2.5) = 1 - \mathbb{P}(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

2. If A sends 1, compute the probability that B registers 0.
   Similarly to the previous case, by definition of the experiment it must hold, in this case, that:

   $$X = \mu + Z \Rightarrow 2 + Z < 0.5 \Rightarrow Z < -1.5$$

   Most particularly, since $Z \backsim \mathcal{N}(0,1)$, it is possible to exploit symmetry of the Gaussian integral tables:

   $$\mathbb{P}(B = 0|A = 1) = \mathbb{P}(Z < -1.5) = \Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$$

3. Suppose now that A sends 0 or 1 with equal probability: Compute the probability that B registers 1.
   In this case, B registers 1 either if A sends 1 and the message is correctly sent or if A sends 0 but the message is flipped due to the electromagnetic noise, meaning that, by applying the law of total probability, it is possible to state that:

   $$\mathbb{P}(B = 1) = \mathbb{P}(B = 1|A = 1)\mathbb{P}(A = 1) + \mathbb{P}(B = 1|A = 0)\mathbb{P}(A = 0)$$

   Most particularly, consider each of the two cases separately:

   - If B reads 1 and A indeed sent 1, then the following must be true:

     $$X = \mu + Z \Rightarrow 2 + Z \geq 0.5 \Rightarrow Z \geq -1.5$$

     Therefore, it is possible to state that:

     $$\mathbb{P}(B = 1|A = 1) = \mathbb{P}(Z \geq -1.5) = 1 - \mathbb{P}(Z \leq -1.5) = 1 - \Phi(-1.5)$$

     By symmetry, this value becomes:

     $$\mathbb{P}(B = 1|A = 1) = 1 - (1 - \Phi(1.5)) = \Phi(1.5) = 0.9332$$

- If B reads 1 but A actually sent 0, then the following must be true:

$$X = \mu + Z \Rightarrow -2 + Z \geq 0.5 \Rightarrow Z \geq 2.5$$

Therefore, it is possible to state that:

$$\mathbb{P}(B = 1|A = 0) = \mathbb{P}(Z \geq 2.5) = 1-\mathbb{P}(Z \leq 2.5) = 1-\Phi(2.5) = 1-0.9938 = 0.0062$$

Therefore, it is possible to conclude that:

$$\mathbb{P}(B = 1) = \frac{\Phi(1.5)}{2} + \frac{1 - \Phi(2.5)}{2} = \frac{0.9332}{2} + \frac{0.0062}{2} = 0.4697$$

4. If B has registered 1, compute the probability that the message registered coincides with the message sent.
By definition of the experiment, this statement means that B registered 1 and A indeed sent 1.
Therefore, assuming that A sends 0 or 1 with equal probability, it is possible to apply Bayes' theorem to revert the condition and obtain that:

$$\mathbb{P}(A = 1|B = 1) = \frac{\mathbb{P}(A = 1, B = 1)}{\mathbb{P}(B = 1)} = \frac{\mathbb{P}(B = 1|A = 1)\mathbb{P}(A = 1)}{\mathbb{P}(B = 1)}$$

Most particularly, by exploiting Gaussian integral tables, it is possible to compute this value in the following way:

$$\mathbb{P}(A = 1|B = 1) = \frac{\Phi(1.5) \cdot \frac{1}{2}}{\frac{1}{2}(\Phi(1.5) + 1 - \Phi(2.5))} = \frac{0.9332 \cdot 0.5}{0.4697} = 0.9934$$

11.5 Exercise 5

Two fair dice are rolled 300 times. Let X denote the number of rolls at which pair $(1, 1)$ is obtained.

1. Compute $\mathbb{E}(X)$ and $\mathbb{V}(X)$.
Notice that, by definition, $X \backsim Binomial(n, p)$, with $n = 300$ and $p = \frac{1}{36}$, meaning that it is possible to directly compute its expectation and variance in the following way:

$$\mathbb{E}(X) = np = \frac{300}{36}$$
$$\mathbb{V}(X) = np(1 - p) = \frac{300}{36}(1 - \frac{1}{36}) = \frac{300}{36} \cdot \frac{35}{36}$$

N.B.: An alternative way to recover $X \backsim Binomial(n, p)$ is by seeing X as the sum of identically distributed and independent random variables of type $X_i \backsim Bernoulli(p)$.
In this case, it is possible to compute the expectation and variance of X through linearity and independence of the Bernoulli variables.

2. Using the Gaussian approximation, compute the probability of obtaining $(1, 1)$ more than 10 times.
   Start by approximating X into $Z \backsim \mathcal{N}(0, 1)$ using the central limit theorem:

$$\mathbb{P}(X \geq 10) = \mathbb{P}(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{10 - np}{\sqrt{np(1-p)}}) = \mathbb{P}(Z \geq \frac{10 - \frac{300}{36}}{\sqrt{\frac{300}{36} \cdot \frac{35}{36}}}) = \mathbb{P}(Z \geq 0.54)$$

   At this point, it is possible to reason in terms of the complement in order to conclude that:

$$\mathbb{P}(X \geq 10) = \mathbb{P}(Z \geq 0.54) = 1 - \mathbb{P}(Z \leq 0.54) = 1 - \Phi(0.54) = 1 - 0.7054 = 0.2946$$

3. Now consider the case where the two dice are rolled n times: Using the Gaussian approximation, determine how large n should be so that the probability of obtaining $(1, 1)$ at least 10 times exceeds $\frac{1}{2}$.
   Generalize the previous point:

$$\mathbb{P}(X \geq 10) = \mathbb{P}(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{10 - np}{\sqrt{np(1-p)}}) = \mathbb{P}(Z \geq \frac{360 - n}{\sqrt{35n}}) > \frac{1}{2}$$

   At this point, it is possible to reason in terms of the complement, allowing to state that:

$$\mathbb{P}(Z \leq \frac{360 - n}{\sqrt{35n}}) = 1 - \mathbb{P}(Z \geq \frac{360 - n}{\sqrt{35n}}) = 1 - \Phi(\frac{360 - n}{\sqrt{35n}}) < \frac{1}{2} \Rightarrow \Phi(\frac{360 - n}{\sqrt{35n}}) > \frac{1}{2}$$

   Therefore, by using the Gaussian integral table, it is possible to conclude that:

$$\frac{360 - n}{\sqrt{35n}} \leq 0 \Rightarrow 360 - n \leq 0 \Rightarrow n \geq 360$$