Ministry of Higher Education and Scientific Research

University of JIJEL

Faculty of Economics, Commercial and Management Sciences

Fundamental Education Department

STATISTICS II

Lecture notes

Asma CHAKER

2025-2026

Table of contents

Table of contents

Introduction

Probability is a branch of mathematics that studies and analyzes randomness and uncertainty. Since many real-world phenomena are unpredictable and random, this field provides essential tools for understanding, modeling, analyzing and anticipating future outcomes. Although its origins lie in games of chance, probability has become a highly valued discipline due to its broad applications—such as modelling natural uncertainties, optimizing industrial processes, and analyzing biological data—providing a rigorous framework for addressing everyday random events.

This textbook offers an introduction to key aspects of this fascinating science through carefully structured chapters. We will begin by exploring the fundamentals of set theory, events, and different probabilistic approaches. We will then apply these concepts through calculation rules, conditional probabilities, and the total probability theorem and Bayes' theorem.

Finally, we will cover random variables and their distributions, which form the cornerstone of probabilistic modeling. You will learn about the most important probability distributions—from the uniform distribution to the binomial, Poisson, and normal distributions. These concepts will help you understand expectation and variance and lay the groundwork for studying limit theorems that connect probability and inferential statistics. You will also become familiar with moments, which help characterize and summarize the shape and key properties of a distribution; Markov's and Chebyshev's inequalities, which are especially useful when the distribution of a random variable is unknown; and conclude with the Law of Large Numbers.

Chapter I: Introduction to Set Theory

Probability theory is based on set theory, or the algebra of sets. It contributes to defining and analysing mathematical structures. It is widely used in different fields, such as probability which makes extensive use of set operations.
To illustrate set theory, we use the Venn diagram, originally introduced by Venn in his book, *Symbolic Logic*, published in 1881, can be used to depict a sample point, a sample space, an event, and related concepts. Specifically, the Venn diagram is made up of two or more overlapping circles, where one circle denotes one set. It is often used in mathematics to show relationships between sets (Yongmiao Hong, 2013).

1. Definition of a set

A set is a collection of objects, called elements of the set. Sets are usually denoted by capital letters (e.g., A, B, C).
If A is a set and x is an element of A, we write $x \in A$.
If x is not an element of A, we write $x \notin A$.
The elements of a set can be letters, numbers or any objects.

2. Types of Sets

Sets can be defined in a variety of ways.

- A set with a finite or limited number of elements, is a **Finite Set**, and we write $x_1, x_2, \dots, x_n$ as a list of elements, in braces: A $= \{x_1, x_2, \dots, x_n\}$

For example, the set of possible outcomes of a die roll is $\{1, 2, 3, 4, 5, 6\}$, and the set of possible outcomes of a coin toss is $\{H, T\}$, (H: heads; and T tails).

- A set with an infinitely many elements, is an **Infinite Set** or **countably infinite**.

If A contains unlimited number of elements $x_1, x_2, \dots$, which can be enumerated in a list, we write

$$A = \{x_1, x_2, \dots\}$$

For example, the set of odd integers can be written as $\{0, 1, -1, 3, -3 \dots\}$ is **countably infinite**.

- A set can have no elements, in which case it is called an **Empty (Null) Set**, denoted by $\emptyset$. By convention, an empty set $\emptyset$ is a subset of any set.
- The set containing all elements under discussion is a **Universal Set**, denoted by Ω.
- If every element of set A is in set B, then A is a **Subset** of B, denoted by $A \subset B$. If every element of a set A is also an element of a set B, we say that A is a **Subset** of B, and we write $A \subset B$ or $B \supset A$. If $A \subset B$ and $B \subset A$, the two sets are **Equal**, and we write $A = B$. Certain set theory notions have special terminology when used in the context of probability, and a subset A of Ω is called an event. In this context, for any events A, B ∈ Ω, A ⊂ B means that occurrence of A implies occurrence of B.
- If A is a subset of B but $A \neq B$, then A is a **Proper Subset** of B, denoted by $A \subset B$.

3. Set operations

Set operations help define relationships between different sets.
Let A and B be two events in the sample space **Ω**. Then we have the following definitions:

3.1 Union of Sets

The union of A and B; $A \cup B$, is the set of all elements that belong to either A or B. The union of A and B occurs if and only if either A or B (or both) occurs. It is also called the logical sum. It can be represented by the Venn diagram in Figure 1.2 and is denoted by $A \cup B$.

$$A \cup B = \{x|\ x \in A\ or\ x \in B\}$$

In the context of probability, for any events A, $B \in \Omega$, $A\ U\ B$ is 'A or B'.
Example: $A = \{1,2,3\}$, $B = \{3,5,8\}$, then $A\ U\ B = \{1,2,3,5,8\}$.

3.2 Intersection of Sets

The **intersection** of two sets A and B, denoted $A \cap B$, is the set of all elements that belong to both A and B. The intersection occurs if and only if both events A and B occur. It is also called the logical product, and can be represented by the Venn Diagram in Figure 1. 1.
Thus,

$$A \cap B = \{x|\ x \in A\ and\ x \in B\}$$

In the context of probability, for any events A, B ∈ Ω, $A \cap B$ is 'A and B'.
Example: $A = \{1,2,3\}$, $B = \{3,5,8\}$, then $A \cap B = \{3\}$.

In some cases, we consider the union or the intersection of several, even an infinite number of sets, defined in the obvious way. For example, if for every positive integer n, we are given a set A_n, then

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \ldots = \{x|\ x \in A_n\ for\ some\ n\}$$

And

$$\bigcap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \ldots = \{x|\ x \in A_n\ for\ all\ n\}$$

3.3 Complement of a Set

The **complement** of a set A (relative to the universal set Ω), with respect to the universe Ω, is the set $\{x \in \Omega \mid x \notin \text{A}\}$ of all elements (basic outcomes of a random experiment) belonging to Ω but not to A, and is denoted as A^c, A' or $\bar{A}$.
The complement of event A is also called the negation of A. It can be represented by the Venn diagram in Figure 1.4.
In the context of probability, the complement event $\bar{A}$ is the event that A does not occur, or 'not A': $\bar{A} = \Omega - A$
Note that $\bar{\Omega} = \emptyset$.
Obviously, any event A and its complement $\bar{A}$ are mutually **exclusive** and collectively **exhaustive**. That is,
$\bar{A} = \Omega - A$ and $A = \Omega - \bar{A}$
$A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = \Omega$.
Example: If $\Omega = \{1,2,3,4,5\}$ and $A = \{1,2,3\}$, then $\bar{A} = \{4,5\}$.

3.4 Difference of Sets

The difference of sets A and B is the set of elements that belong to A but not to B, denoted as:
$A - B = \{x \mid x \in A, x \notin B\}$
$A - B = A \cap \bar{B}$
The difference $A - B$ can be represented by the Venn diagram in figure 1. 5.
Example: $A = \{1,2,3\}$, $B = \{3,4,5\}$, then $A - B = \{1,2\}$.

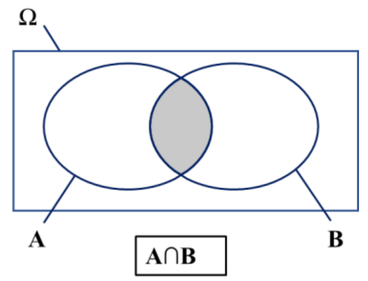


Figure 1. 1

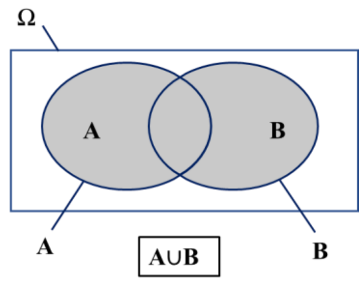


Figure 1. 2

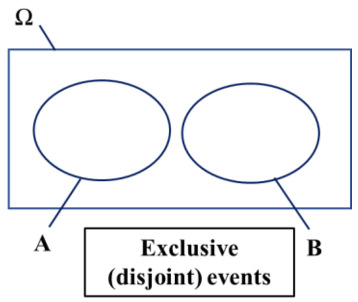


Figure 1. 3

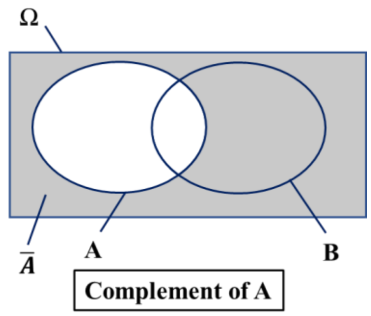


Figure 1. 4

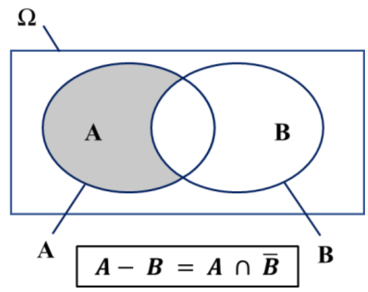


Figure 1. 5

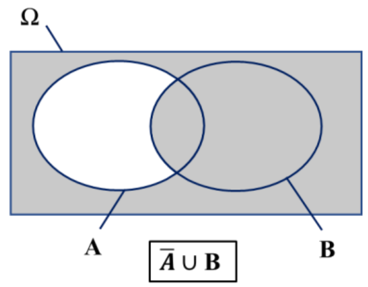


Figure 1. 6

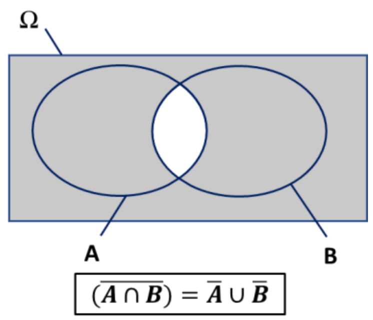


Figure 1. 7

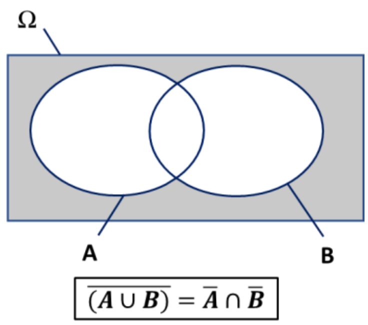


Figure 1. 8

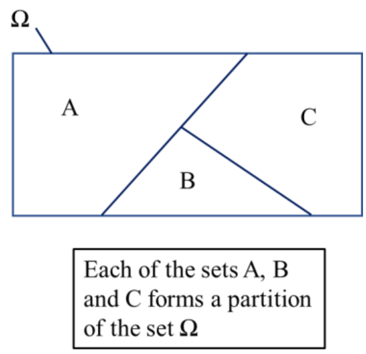


Figure 1. 9

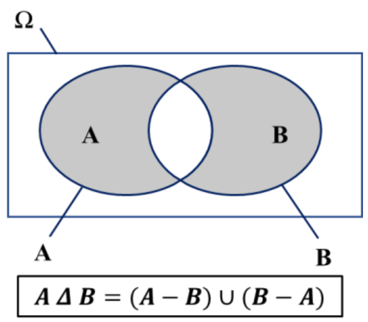


Figure 1. 10

Figure 1: Examples of Venn diagrams.

3.5 Symmetric Difference Definition

Symmetric difference between two sets A and B is denoted as $A \Delta B$ and is defined as the set of elements that are in either of set but not in their intersection. Mathematically, this can be expressed as:

$$A \Delta B = (A - B) \cup (B - A)$$

Symmetric difference between two sets A and B can be represented by the Venn Diagram in figure 1. 10.

3.6 Exclusiveness (disjoint events)

If A and B have no common basic outcomes, they are called mutually exclusive and their intersection is empty set $\emptyset$, i.e., $A \cap B = \emptyset$. By definition, mutually exclusive events cannot occur simultaneously. They are also called disjoint because they do not overlap when represented in the Venn diagram (example in Figure 1. 3). More generally, several sets are said to be disjoint if no two of them have a common element.

3.7 Collective Exhaustiveness (partitions)

A collection of sets is said to be a partition of a set Ω if the sets in the collection are disjoint and their union is Ω. It can be represented by the Venn diagram (figure 1. 9). Suppose $A_1, A_2, \dots, A_n$ are *n* events in the sample space Ω, where n is any positive integer. If $\bigcup_{i=1}^{n} A_i = \Omega$; then these *n* events are said to be collectively exhaustive or a partition of **Ω**.
In the context of probability, for any events A, B ∈ Ω, A ∩ B = Ø is A and B are **mutually exclusive** or **disjoint events**.

Intuitively, a sequence of collectively exhaustive and mutually exclusive events forms a partition **Ω**. In certain sense, it can be viewed as a set of orthogonal bases which can represent any event A in Ω; and $A_i \cap A$ could be viewed as the projection of event A on the base A_i.
Sets and the associated operations are easy to visualize in terms of Venn diagrams, as illustrated in Figure 1.

Example:
We roll a fair six-sided die.
Let:
Event A = "The number rolled is even"
Event B = "The number rolled is at least 4"
Find: $A, B, \bar{A}, \bar{B}, A - B, B - A, A \cap B, A \cup B, A \cap \bar{A}, A \cup \bar{A}$

Solution:
The sample space Ω = {1,2,3,4,5,6}.
$A = \{2,4,6\}$;
$B = \{4,5,6\}$;
$\bar{A} = \{1,3,5\}$;
$\bar{B} = \{1,2,3\}$;
$A - B = \{2\}$;
$B - A = \{5\}$;
$A \cap B = \{4,6\}$;
$A \cup B = \{2,4,5,6\}$;
$A \cap \bar{A} = \emptyset$;
$A \cup \bar{A} = \{1,2,3,4,5,6\} = \Omega$.

4. The Algebra of Sets

Set operations have several properties, which result from the definitions. Some examples for any three events A, B, C defined on a sample space Ω, are:
$A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = \Omega$.
$A \cup \Omega = \Omega$, $A \cap \Omega = A$.

4.1 Complementation
$\bar{\bar{A}} = A$,
$\bar{\emptyset} = \Omega$,
$\bar{\Omega} = \emptyset$.

4.2 Commutativity of union and intersection
$A \cup B = B \cup A$,
$A \cap B = B \cap A$.

4.3 Associativity of union and intersection
$(A \cup B) \cup C = A \cup (B \cup C)$,
$(A \cap B) \cap C = A \cap (B \cap C)$.

4.4 Distributivity laws
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
More generally, for any n ≥ 1,
$B \cap (\bigcup_{i=1}^{n} A_i) = \bigcup_{i=1}^{n} (B \cap A_i)$,
$B \cup (\bigcap_{i=1}^{n} A_i) = \bigcap_{i=1}^{n} (B \cup A_i)$.

4.5 De Morgan's law
De Morgan's Law is a law that deals with the union and intersection of sets with the intersection and union of individual sets respectively. There are two laws under it namely De Morgan's Law of Union represented in the Figure 1. 8 and De Morgan's Law of Intersection represented in the Figure 1. 7. Two particularly useful properties are given by De Morgan's laws which state that
$(\overline{A \cup B}) = \bar{A} \cap \bar{B}$,
$(\overline{A \cap B}) = \bar{A} \cup \bar{B}$.
More generally, for any $n \geq 1$,

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$$

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}$$

Set theory provides a rigorous foundation for probability. Understanding sets, their operations, and their applications is crucial to understand different fields.

Chapter I Exercises

Exercise 1:
We toss a coin three times. Let A and B be two subsets:
$$A = \{THT, TTT, HTH, HTT\}$$
$$B = \{TTT, HHH, TTH, HTT\}$$

Find:
$\bar{B}, \bar{A}$, $A \cap B, A \cup B, \overline{A \cap B}$, $\overline{A \cup B}, A \cap \bar{B}, \bar{A} \cap B, \bar{A} \cap \bar{B}$

Solution:
The sample space Ω for tossing a coin three times (with outcomes H (heads) and T (tails)) is:
$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

Given:
$$A = \{THT, TTT, HTH, HTT\}$$
$$B = \{TTT, HHH, TTH, HTT\}$$

Complement of B ($\bar{B}$): All outcomes in Ω not in B.
$$\bar{B} = \{THT, THH, HTH, HHT\}$$

Complement of A ($\bar{A}$): All outcomes in Ω not in A.
$$\bar{A} = \{TTH, THH, HHT, HHH\}$$

Intersection of A and B ($A \cap B$): Outcomes that are in both A and B.
$$A \cap B = \{TTT, HTT\}$$

Union of A and B ($A \cup B$): Outcomes that are in A or B or both.
$$A \cup B = \{TTT, THT, HTH, HTT, HHH, TTH\}$$

Complement of $\overline{A \cap B}$: All outcomes not in $A \cap B$.
$$\overline{A \cap B} = \{TTH, THT, THH, HTH, HHT, HHH\}$$

Complement of $\overline{A \cup B}$: All outcomes not in $A \cup B$.
$$\overline{A \cup B} = \{THH, HHT\}$$

Intersection of A and complement of B ($A \cap \bar{B}$): Outcomes in A that are not in B.
$$A \cap \bar{B} = \{THT, HTH\}$$

Intersection of complement of A and B ($\bar{A} \cap B$): Outcomes not in A but in B.
$$\bar{A} \cap B = \{HHH, TTH\}$$

Intersection of complement of A and complement of B ($\bar{A} \cap \bar{B}$): Outcomes not in A and not in B.
$$\bar{A} \cap \bar{B} = \{THH, HHT\}$$

Exercise 2:
We toss a fair coin four consecutive times where H represents "Head" and T represents "Tail".

1- Determine the number of possible outcomes.

2- What are all possible outcomes of the experiment?

Let the following events be:
A: Getting exactly three Heads
B: Getting exactly one Head
C: Getting at least one Head
D: Getting exactly two Tails
E: Getting at least two Tails

3- Express the previous events using subset notation:
4- Find: $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$, $\bar{E}$, $\bar{D} \cap \bar{E}$, $\bar{D} \cup \bar{E}$, $D \cap \bar{E}$, $A \cap C$, $\bar{A} \cap \bar{C}$, $A \cap \bar{C}$, $\bar{A} \cup \bar{C}$

Solution:

1- The number of possible outcomes

We toss a fair coin four times. The sample space Ω consists of all possible sequences of Heads (H) and Tails (T). Each toss has 2 outcomes, so the total number of possible outcomes is $2^4 = 16$.
(We will see this in more detail in Chapter III).

2- The sample space Ω is:

$\Omega = \{HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT\}$

3- Defining the events as subsets:

A: Exactly three Heads
A represents the outcomes with exactly three H's:

$$A = \{HHHT, HHTH, HTHH, THHH\}$$

B: Exactly one Head
B represents the outcomes with exactly one H:

$$B = \{HTTT, THTT, TTHT, TTTH\}$$

C: At least one Head
C represents the outcomes with one or more H's (i.e., all except all Tails):
C={HHHH,HHHT,HHTH,HTHH,THHH,HHTT,HTHT,HTTH,THHT,THTH,TTHH,HTTT, THTT,TTHT,TTTH}

D: Exactly two Tails
D represents the outcomes with exactly two T's (which implies exactly two H's):
D={HHTT,HTHT,HTTH,THHT,THTH,TTHH}

E: At least two Tails
E represents the outcomes with two or more T's:
E={HHTT,HTHT,HTTH,THHT,THTH,TTHH,HTTT,THTT,TTHT,TTTH,TTTT}

4- The events:

$\bar{A}$: Not exactly three Heads

$\bar{A} = \{\text{HHHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT}\}$

$\bar{B}$: Not exactly one Head

$$\bar{B} = \{\text{HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, TTTT}\}$$

$\bar{C}$: No Heads (i.e., all Tails)

$$\bar{C} = \{\text{TTTT}\}$$

$\bar{D}$: Not exactly two Tails

$$\bar{D} = \{\text{HHHH, HHHT, HHTH, HTHH, THHH, HTTT, THTT, TTHT, TTTH, TTTT}\}$$

$\bar{E}$: Fewer than two Tails (i.e., at least three Heads)

$$\bar{E} = \{\text{HHHH, HHHT, HHTH, HTHH, THHH}\}$$

$\bar{D} \cap \bar{E}$: Not exactly two Tails AND fewer than two Tails
$\bar{D}$ Not exactly two Tails (as above)
$\bar{E}$ Fewer than two Tails (at least three Heads)
So,

$$\bar{D} \cap \bar{E} = \{\text{HHHH, HHHT, HHTH, HTHH, THHH}\}$$

(since these are in both $\bar{D}$ and $\bar{E}$).

$\bar{D} \cup \bar{E}$ Not exactly two Tails OR fewer than two Tails
This is the union of $\bar{D}$ and $\bar{E}$:
Since D⊂E (every outcome with exactly two Tails has at least two Tails), $D \cap E = D$
So $\bar{D} \cup \bar{E} = \bar{D}$ (which we already have).

Alternatively, list all outcomes in $\bar{D}$ or $\bar{E}$:

$$\bar{D} \cup \bar{E} = \{\text{HHHH, HHHT, HHTH, HTHH, THHH, HTTT, THTT, TTHT, TTTH, TTTT}\}$$

$D \cap \bar{E}$: Exactly two Tails BUT NOT at least two Tails?
But note: If exactly two Tails, then it is automatically at least two Tails. So D⊂E, thus

$$D \cap \bar{E} = \emptyset$$

$A \cap C$: Exactly three Heads AND at least one Head
Since all outcomes in A have at least one Head,

$$A \cap C = A = \{HHHT, HHTH, HTHH, THHH\}$$

$\bar{A} \cap \bar{C}$: Not exactly three Heads AND no Heads
$\bar{C} = \{\text{TTTT}\}$ (no Heads), and this is also not exactly three Heads. So
$\bar{A} \cap \bar{C} = \{\text{TTTT}\}$

$A \cap \bar{C}$: Exactly three Heads BUT no Heads?
This is impossible because exactly three Heads requires at least one Head. So

$$A \cap \bar{C} = \emptyset$$

$\bar{A} \cup \bar{C}$: Not exactly three Heads OR no Heads
This is the union of $\bar{A}$ and $\bar{C}$
But $A \cap C = A$ (since A⊂C), so $\bar{A} \cup \bar{C} = \bar{A}$

Alternatively, list all outcomes in $\bar{A}$ or $\bar{C}$:

$\bar{A} \cup \bar{C}$={HHHH,HHTT,HTHT,HTTH,THHT,THTH,TTHH,HTTT,THTT,TTHT, TTTH,TTTT}

Exercise 3:

Let **Ω** be a sample space defined as follows:

$$\Omega = \{1,2,3,4,5,6,7,8,9\}$$

Let the events A, B, and C be three subset events.

$$A = \{2,7,6,8\}$$
$$B = \{1,2,3,7\}$$
$$C = \{3,7,5,6\}$$

Find:

$A \cap B \cap C, A \cap B \cap \bar{C}, A \cap \bar{B} \cap C, \bar{A} \cap B \cap C, A \cap \bar{B} \cap \bar{C}, \bar{A} \cap B \cap \bar{C}, \bar{A} \cap \bar{B} \cap C, \bar{A} \cap \bar{B} \cap \bar{C},$
$A \cup B \cup C, (A \cup B) \cap (A \cup B), A \cup (B \cap C), (\overline{A \cup B}), \overline{A \cap B}, (A \cap \bar{B}) \cup B, B \cap (\overline{A \cup C}).$

Solution:

Sample space $\Omega = \{1,2,3,4,5,6,7,8,9\}$

Events:

$$A = \{2,7,6,8\}$$
$$B = \{1,2,3,7\}$$
$$C = \{3,7,5,6\}$$

$A \cap B \cap C$:

$$A \cap B \cap C = \{7\}$$

$A \cap B \cap \bar{C}$:

$$A \cap B \cap \bar{C} = \{2\}$$

$A \cap \bar{B} \cap C$:

$$A \cap \bar{B} \cap C = \{6\}$$

$\bar{A} \cap B \cap C$:

$$\bar{A} \cap B \cap C = \{3\}$$

$A \cap \bar{B} \cap \bar{C}$:

$$A \cap \bar{B} \cap \bar{C} = \{8\}$$

$\bar{A} \cap B \cap \bar{C}$:

$$\bar{A} \cap B \cap \bar{C} = \{1\}$$

$\bar{A} \cap \bar{B} \cap C$:

$$\bar{A} \cap \bar{B} \cap C = \{5\}$$

$\bar{A} \cap \bar{B} \cap \bar{C}$:

$$\bar{A} \cap \bar{B} \cap \bar{C} = \{4,9\}$$

$A \cup B \cup C$:

$$A \cup B \cup C = \left(\overline{\bar{A} \cap \bar{B} \cap \bar{C}}\right) = \{1,2,3,5,6,7,8\}$$

$(A \cup B) \cap (A \cup B)$:

This is just $A \cup B$.

$$(A \cup B) \cap (A \cup B) = A \cup B = \{1,2,3,6,7,8\}$$

$A \cup (B \cap C)$:

$$B \cap C = \{3,7\}$$

So $A \cup (B \cap C) = \{2,6,7,8\} \cup \{3,7\} = \{2,3,6,7,8\}$

$(\overline{A \cup B})$:
Everything not in A∪B:

$$(\overline{A \cup B}) = \Omega - (A \cup \mathrm{B}) = \{4,5\}$$

$\overline{A \cap B}$:
Everything not in $A \cap B$.
$A \cap B = \{2,7\}$, so.

$$\overline{A \cap B} = \Omega - (A \cap \mathrm{B}) = \{1,3,4,5,6,8,9\}$$

$(A \cap \bar{B}) \cup B$:
$A \cap \bar{B} = \{6,8\}$ (from A not in B)

$$(A \cap \bar{B}) \cup B = \{6,8\} \cup \{1,2,3,7\} = \{1,2,3,6,7,8\}$$

$B \cap (\overline{A \cup C})$:
$A \cup C = \{2,5,6,7,8,3\}$ (all elements in A or C)
$\overline{A \cup C} = \{1,4,9\}$ (not in A or C)
Then $B \cap (\overline{A \cup C}) = \{1\}$ (since 1∈B and not in A∪C)
Shade only $B \cap (\overline{A \cup C}) = \bar{A} \cap B \cap \bar{C} = \{1\}$.

Exercise 4:
After the deliberations for the common core students, we found that 200 students failed, among whom:
84 did not pass the math module,
110 did not pass the statistics module,
60 passed both the first and the second module.
What is the number of students who will retake at least one of the two modules?
What is the number of students who will retake both modules?
What is the number of students who will retake only one of the two modules?

Solution:
M: the number of students who did not pass the math module.
M=84

S: the number of students who did not pass the statistics module.
S=110

$$\bar{M} \cap \bar{S} = 60$$

The number of students who will retake at least one of the two modules?

$$(M \cup S) = \Omega - (\bar{M} \cap \bar{S}) = 200 - 60 = 140$$

What is the number of students who will retake both modules?

$$M \cap S = M + S - M \cup S = 84 + 110 - 140 = 54$$

What is the number of students who will retake only one of the two modules?

$$(M \cap \bar{S}) \cup (\bar{M} \cap S) = [M - (M \cap S)] \cup [S - (M \cap S)] = (84 - 54) + (110 - 54) = 86$$

Chapter II: Basic Concepts of Probability

Probability is the mathematical framework for quantifying uncertainty and modelling random phenomena. It is essential across fields such as statistics, science, economics, and biology, as it provides tools to model scenarios with uncertain outcomes. The foundations of probability deal with core concepts such as random experiments, sample spaces, and events. To rigorously discuss probability theory, we must first establish fundamental definitions and terminology.

1. Random experiment

When we speak about probabilities, there is always an implied context, which we formally call the "random experiment". It is a process or an action with uncertain outcomes that is unpredictable, repeatable, and has well-defined possible outcomes. i.e. the outcome cannot be predicted with certainty when the experiment is repeated under identical conditions.
For example, before rolling a fair six-sided die, one cannot predict whether it will land on 1, 2, 3, 4, 5, or 6. The outcomes are $\{1, 2, 3, 4, 5, 6\}$.

2. Sample Space

The concept of sample space was first introduced by Ludwig von Mises, an Austrian mathematician and engineer, in 1931. It is important to note that for a random experiment, one knows the set of all possible basic outcomes, but one does not know which outcome will arise before performing the random experiment.
A Sample Space is the set of all possible outcomes of an experiment, it is denoted by **Ω**.
Examples:
Rolling a 6-sided die, The basic outcomes are the numbers 1, 2, 3, 4, 5, 6 : $\Omega = \{1, 2, 3, 4, 5, 6\}$
In the experiment of throwing a Coin, there are two possible outcomes: $\Omega = \{Head, Tail\}$
Throwing two coins: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$
The number of messages received in a day:

$$\Omega = \{x|\ x \in Z, x \geq 0\}$$

TikTok hours in a day:

$$\Omega = \{x|\ x \in \mathbb{R}, 0 \leq x \leq 24\}$$

A sample space **Ω** can be countable or uncountable. This dictates the way in which probabilities are assigned.
In the experiment of throwing two coins: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. The sample space **Ω** is **finite** and **countable**.
The number of messages you receive at a given interval of time will be $\{0,1,2, \dots\}$, and we write: $\Omega = \{x|\ x \in Z, x \geq 0\}$. This sample space is **countable** but **infinite**.
Suppose t_0 is the lowest level of rainfall in an area, and t_1 is the highest level of rainfall in the same area. Let *t* denotes the possible levels of rainfall in the area. Then the sample space of t is

$$\Omega = \{t \in \mathbb{R}: t_0 \leq t \leq t_1\}$$

The sample space in this example is **uncountable**

3. Basic Outcomes

The possible outcomes of a random experiment are called "basic outcomes", and the set of all basic outcomes constitutes the sample space **Ω**. The basic outcomes are the basic building blocks for a sample space. They cannot be divided (partitioned or separated) into more primitive or more elementary kinds of outcomes.

When an experiment is performed, the realization of the experiment is one (and only one) outcome in the sample space. If the experiment is performed a number of times, different outcomes may occur each time or some outcomes may repeat.
A sample space **Ω** is sometimes called an outcome space. Each outcome in **Ω** is called an element of **Ω**; or simply a sample point.

4. Event

In probability theory, an event A is a collection of basic outcomes from the sample space **Ω** that share certain common features or equivalently obey certain restrictions. An event E is a subset of outcomes in **Ω**. It represents outcomes or sets of outcomes from a random experiment.
The event A is said to occur if the random experiment gives rise to one (and only one) of the constituent basic outcomes in A. That is, an event occurs if any of its basic outcomes has occurred (or equivalently if the outcome of the random experiment is an element of event E).
reformulate
An Event is some subset of **Ω** that we ascribe meaning to.
Example:
Throwing a coin and getting a tail $E = \{Tail\}$
Throwing two coins and getting at least 1 Tail $E = \{(T,T),(T,H),(H,T)\}$
Rolling a die and getting more than 4: $E = \{5,6\}$
Receiving less than 20 messages in a day:

$$\Omega = \{x \mid x \in Z, 0 \leq x \leq 20\}$$

Considering it a wasted day if 3 hours or more are spent on TikTok:

$$\Omega = \{x \mid x \in R, 3 \leq x \leq 24\}$$

Mathematically speaking, an event is equivalent to a set. Thus, the words 'set' and 'event' are interchangeable.
For events A and B in the sample space **Ω**:
- If every element of A is an element of B, A is a subset of B, and we write A ⊂ B.
- The union $A \cup B$ includes all outcomes that are in either A or B (or both).
- The intersection $A \cap B$ includes the elements that are in both A and B.

5. Types of events

There are different types of events which are important for the definition and calculation of probabilities in various situations.

5.1 A simple Event (Elementary Event)

A simple event is an event that contains one single outcome (basic outcome) of a random experiment.
Example: Rolling a die and getting a number between 6 and 8, $A = \{6\}$.

5.2 A compound Event

A compound event is an event that contains two or more outcomes of a random experiment.
It includes more than one simple event.
Example: Rolling a die and getting an odd number $A = \{1, 3, 5\}$.

5.3 A certain Event (Sure Event)

A certain event contains all possible outcomes of the random experiment. It is an event that is sure to occur.
Example: Rolling a die and getting a number less than 7.

5.4 An impossible Event

An impossible event is an event that has no outcomes $\emptyset = \{\}$. It cannot occur regarding to the random experiment.
Example: Rolling a die and getting a negative number.

5.5 A complementary Event

The complementary event of an event A, denoted as A^c, A' or $\bar{A}$, consists of all outcomes in the sample space Ω which are not in A.

Example: If A is rolling an odd number, $\bar{A}$ is rolling an even number.

5.6 Independent Events

Two events are independent if the occurrence of one does not affect the occurrence of the other.
Example: Rolling a die and drawing a card from a deck are two independent events.

5.7 Dependent Events

Two events are dependent if the occurrence of one affects the occurrence of the other.
Example: Drawing two numbered balls (0,1,2,3,4,5,6,7,8,9) from a bag without replacement. The probability of the second number being 2 depends on the first ball drawn.

5.8 Mutually Exclusive Events (Disjoint Events)

Two events are mutually exclusive if they cannot occur at the same time. The events A and B are disjoint if they have no outcomes in common: $A \cap B = \emptyset$.
Example: Rolling a die and getting an even number or a number less than 2 (they cannot both occur in a single roll).

5.9 Exhaustive Events (partitions)

A set of events is exhaustive if at least one of the events must occur. The union of these events covers the entire sample space Ω. The events A_1, A_2 , ... are partitions of Ω if they are mutually disjoint and their union is Ω.
Example: drawing a red or a black card from a deck of cards (these two events cover all possible outcomes).

5.10 Equally Likely Events

Two events or more are equally likely if they have the same probability of occurring.
Example: Drawing a numbered balls (0,1,2,3,4,5,6,7,8,9) from a bag, each ball (1 through 10) has an equal probability of $\frac{1}{10}$.

5.11 Conditional Event

A conditional event is an event whose probability depends on the occurrence of another event.
Example: The probability of drawing a diamond from a deck of cards given that the card drawn is black.

6. The Probability of an Event: P(E)

A measure of how likely an event is to occur.

Chapter II Exercises

Exercise 1:
We roll a fair six-sided die twice.

1- Determine the sample space Ω.
2- Identify the following events:

A: “The sum of the two rolls is 7.”
B: “Getting a number greater than 1 on both rolls.”
C: “Getting the same number on both rolls.”
D: “The first roll is a 6.”
E: “The sum of the two rolls is an even number.”

3- Determine the complementary event $\bar{A}$, and the event $\bar{\bar{D}}$.
4- Find the intersection of events A and E (A∩E).
5- Find the union of events C and E (C∪E).

Solution:

1- The Sample Space Ω

The sample space consists of all ordered pairs (first roll, second roll):

$$\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

There are $6 \times 6 = 36$ equally likely outcomes.

2- Event A: “The sum of the two rolls is 7.”

A={(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)}

Event B: “Getting a number greater than 1 on both rolls.”

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

Event C: “Getting the same number on both rolls.”

$$C = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

Event D: “The first roll is a 6.”

$$D = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

Event E: “The sum of the two rolls is an even number.”
Both odd: (1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5) → 9 outcomes
Both even: (2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6) → 9 outcomes

$$E = \{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5),(2,2),(2,4),(2,6),\\ (4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$$

3- $\bar{A}$: “The sum is not 7.”

This event includes every possible outcome from rolling two dice except the six outcomes where the sum is exactly 7.

$$\bar{A} = \{(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),(2,3),(2,4),(2,6),\\ (3,1),(3,2),(3,3),(3,5),(3,6),(4,1),(4,2),(4,4),(4,5),(4,6),\\ (5,1),(5,3),(5,4),(5,5),(5,6),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

Event $\overline{\overline{D}}$

The complement of a complement returns to the original event. This is a fundamental rule of set theory.
D: “The first roll is a 6.”
$\overline{D}$: “The first roll is not a 6.”
$\overline{\overline{D}}$: “It is not true that the first roll is not a 6.” This double negative simplifies to “The first roll is a 6.”

Therefore:

$$\overline{\overline{D}} = D$$

Then, the event $\overline{\overline{D}}$ is identical to the original event D.

4- **Intersection $A \cap E$:**
We find outcomes where the sum is 7 AND even.
But 7 is an odd number.
Therefore, no outcome satisfies both conditions: $A \cap E = \emptyset$

5- **Union $C \cup E$: We find all outcomes that are in C or E or both.**
Notice that all elements of C have the same number, so they automatically have the same parity.
Therefore, $C \subset E$.
Thus:

$$C \cup E = E$$

Exercise 2:
An experiment consists of flipping two distinct coins (Coin 1 and Coin 2) and spinning a spinner divided into 3 equal sections labelled 1, 2, and 3.

1- What is the total number of possible outcomes?
2- Find the sample space Ω for this experiment.
3- Identify the following events:

A: “The two coins show different faces AND the spinner shows a 2.”
B: “The two coins show the same face AND the spinner shows an even number.”
C: “At most one coin shows Heads AND the spinner shows an odd number.”
D: “Both coins show Heads AND the spinner shows a 3.”

Solution:

1- **The Total Number of Possible Outcomes**

Each coin has 2 outcomes: Heads (H) or Tails (T).
The spinner has 3 outcomes: 1, 2, or 3.

By the fundamental counting principle:
Total Outcomes=2×2×3=12

2- **The Sample Space Ω**

The sample space consists of all ordered triples (Coin 1, Coin 2, Spinner):
$$\Omega = \{(H,H,1),(H,H,2),(H,H,3),(H,T,1),(H,T,2),(H,T,3),(T,H,1),(T,H,2),(T,H,3),(T,T,1),(T,T,2),(T,T,3)\}$$

3- **The Identification of Events**

Event A: "The two coins show different faces AND the spinner shows a 2."
Coins different: (H, T) and (T, H)
Spinner = 2
$$A = \{(H,T,2),(T,H,2)\}$$

Event B: "The two coins show the same face AND the spinner shows an even number."
Coins same: (H, H) and (T, T)
Spinner even: 2
$$B = \{(H,H,2),(T,T,2)\}$$

Event C: "At most one coin shows Heads AND the spinner shows an odd number."
At most one Head: 0 Heads (T, T) or 1 Head (H, T), (T, H)
Spinner odd: 1 or 3
$$C = \{(T,T,1),(T,T,3),(H,T,1),(H,T,3),(T,H,1),(T,H,3)\}$$

Event D: "Both coins show Heads AND the spinner shows a 3."
Coins = (H, H)
Spinner = 3
$$D = \{(H,H,3)\}$$

Exercise 3:
A quality control inspector tests two components from a production line. Each component can be classified as Defective (D) or Functional (F).

1- Determine the sample space Ω for this experiment.
2- Identify the following events:

A: "Exactly one component is defective."
B: "At least one component is functional."
C: "Both components are of the same type."
D: "The first component is defective."
E: "No more than one component is defective."

3- Find the intersection of events B and E (B∩E).
4- Find the union of events C and D (C∪D).

Solution:

1- **Sample Space Ω**

The sample space consists of all possible ordered pairs (Component 1, Component 2):
Ω={(F,F),(F,D),(D,F),(D,D)}

Where:
F: Functional
D: Defective

2- **The Identification of Events**

A: "Exactly one component is defective."
This means one is Defective (D) and the other is Functional (F).

$$A = \{(F, D), (D, F)\}$$

Event B: "At least one component is functional."
This means 1 or 2 functional components.

$$B = \{(F, F), (F, D), (D, F)\}$$

Event C: "Both components are of the same type."
Both Functional or both Defective.

$$C = \{(F, F), (D, D)\}$$

Event D: "The first component is defective."
The first component is D.

$$D = \{(D, F), (D, D)\}$$

Event E: "No more than one component is defective."
This means 0 or 1 defective components.

$$E = \{(F, F), (F, D), (D, F)\}$$

3- **The Intersection of B and E ($\boldsymbol{B \cap E}$)**

$$B = \{(F, F), (F, D), (D, F)\}$$
$$E = \{(F, F), (F, D), (D, F)\}$$

The common elements are all elements in B (or E):

$$B \cap E = \{(F, F), (F, D), (D, F)\}$$
$$B \cap E = B$$

(Since $E = B$ in this specific case)

4- **The Union of C and D ($\boldsymbol{C \cup D}$)**

$$C = \{(F, F), (D, D)\}$$
$$D = \{(D, F), (D, D)\}$$

Combining all elements from both sets:

$$C \cup D = \{(F, F), (D, D), (D, F)\}$$

Chapter III: Combinatorial analysis

Widely used in probability, and inextricably linked to set theory, combinatorial analysis is a branch of mathematics dealing with counting rules, permutation, arrangement, and combination of elements within a set.

1. Basic Counting principles

1.1 Addition Rule

If two events are mutually exclusive (cannot occur simultaneously), the total number of outcomes for either event is the sum of the number of outcomes for each event.
If event A has n_1 outcomes and event B has n_2 outcomes, and A and B are mutually exclusive, the total number of outcomes for A or B is:
If two events are mutually exclusive (cannot occur simultaneously), the total number of outcomes for either event is the sum of the number of outcomes for each event.
Suppose there are k events $A_1, A_2, A_3, \dots, A_k$, where each A_i have n_i possible outcomes, and all the events $A_1, A_2, A_3, \dots, A_k$ are mutually exclusive (no two events can occur at the same time, i.e., no overlaps), the total number of possibilities for A_1 or A_2 or A_3 or $\dots$ or A_k is the sum of the outcomes of each individual event:

$$N = n_1 + n_2 + n_3 + \cdots + n_k$$

Example: Mohamed can choose either the street A, B or C to go to work. He has to buy a newspaper. There are 3 shops in the street A, 5 in the street B, and 4 in the street C. how many choices has he to buy his newspaper?
And A, B, and C are mutually exclusive. Then the total number of shops for A, B, or C is:

$$N = 3 + 5 + 4 = 12$$

1.2 Multiplication Rule

If an experiment consists of multiple stages, and the number of outcomes at each stage is independent of the others, the total number of possible outcomes is the product of the number of outcomes at each stage.
Suppose there are k stages, there are n_1 possibilities for the first stage; then n_2 possibilities for the second stage; then n_3 possibilities for the third stage and n_i possible outcomes for the i^{th} stage, then the total number of possibilities for the set of outcomes is:

$$N = n_1 \times n_2 \times n_3 \times \dots \times n_k$$

Example: Mohamed has to go through 3 streets to go to work, he has to buy his newspaper. There are 3 shops on the first street, 4 on the second and 2 on the third.
How many choices has Mohamed to buy his newspaper?

$$N = 3 \times 4 \times 2 = 24$$

He has 24 choices.

2. Arrangements

Arrangements refer to the number of ways to arrange k objects or events out of a set of n objects or events, i.e. the different ways they can be ordered or organized.

2.1 Arrangements without Repetition

The different of ways to arrange k objects or events in a set of n distinct objects or events without repetition is given by:

$$A_n^k = \frac{n!}{(n-k)!}$$

Example: In how many ways can we arrange 4 books chosen from 8 on a shelf?

$$A_8^4 = \frac{8!}{(8-4)!} = 1680$$

There are 1680 different possible arrangements.

N.B: If the shelf can contain only 4 books, we are dealing with Permutations.

2.2 Arrangements with repetition

If objects can be repeated, the number of arrangements increases. The different of ways to arrange k objects or events in a set of n distinct objects or events with repetition allowed is given by:

$$A_n^k = n^k$$

Example:

A code consists of 2 letters (A-Z) followed by 3 digits (0-9). How many different codes can be created if letters and digits can be repeated?

For the letters:

$$n = 26$$
$$k = 2$$

$A_{26}^2 = 26^2 = 676.$

For the digits:

$$n = 10$$
$$k = 3$$

$A_{10}^3 = 10^3 = 1000.$

$$A_{26}^2 \times A_{10}^3 = 26^2 \times 10^3 = 676 \times 1000 = 676000$$

So, there are 676000 possible codes.

3. Permutations

A permutation is an arrangement of n objects or events in a specific order.

3.1 Permutations without Repetition

The number of ways to arrange n distinct objects or events is given by:

$$P_n = n!$$

where $n! = n \times (n-1) \times \ldots \times 2 \times 1$.

Example 1: we want to arrange 5 people on a 5-seater bench. The number of unique arrangements is:

$$P_5 = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

So, there are 120 different possibilities.

Example 2: How many ways can you arrange these 4 letters: T, R, P, N.

$$P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$$

We can illustrate all the possibilities through the following decision tree:

T	R	P	N	TRPN
		N	P	TRNP
	P	R	N	TPRN
		N	R	TPNR
	N	R	P	TNRP
		P	R	TNPR
R	T	P	N	RTPN
		N	P	RTNP
	P	T	N	RPTN
		N	T	RPNT
	N	T	P	RNTP
		P	T	RNPT
P	T	R	N	PTRN
		N	R	PTNR
	R	T	N	PRTN
		N	T	PRNT
	N	T	R	PNTR
		R	T	PNRT
N	T	R	P	NTRP
		P	R	NTPR
	R	T	P	NRTP
		P	T	NRPT
	P	T	R	NPTR
		R	T	NPRT

Figure 02: The probability tree diagram

Example 3: How many ways can 4 books be organized out of 4 on a shelf?

$$P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$$

$$A_4^4 = \frac{4!}{(4-4)!} = 24$$

There are 24 different possibilities.

3.2 Circular permutations

Unlike linear arrangements (e.g., arranging people or objects in a line), we fix one person or object to exclude equivalent rotations, reducing the number of possible arrangements.
The number of distinct ways to arrange n distinct objects around a circular table is:

$$P_n = (n-1)!$$

Example: we want to arrange 7 people around a circular table. The number of unique arrangements is:

$$(7-1)! = 6! = 720$$

There are 720 possibilities arrange 7 people around a table.

Special Case 1: If the table has a fixed reference point, then the number of arrangements is simply $n!$, as the fixed reference point influences the rotations, which are no longer considered identical.

Example: we want to arrange 7 people around a circular table, where one chair is bigger than the others. The number of unique arrangements is:

$$7! = 5040$$

There are 5040 possibilities arrange 7 people around a table where there is a distinct chair.

Special Case 2: If there are duplicates, i.e. the objects are not all distinct, the formula takes into account indistinguishable arrangements, dividing by the factorial of the number of identical objects.

Example: we want to arrange 7 canning jars storing homemade jam around a circular tray, but 3 of them are identical. The number of unique arrangements is:

$$\frac{(7-1)!}{3!} = \frac{720}{6} = 120$$

There are 120 distinct ways to arrange 7 canning jars on a circular tray when 3 of the jars are identical.

3.3 Permutations with Repetition

If there are *n* objects or events, with some repeated, the number of distinct ways to arrange them is:

$$P_n^{n_1,n_2,n_3,\ldots,n_k} = \frac{n!}{n_1! \times n_2! \times n_3! \times \ldots \times n_k!}$$

where $n_1, n_2, n_3, \ldots, n_k$ are the frequencies of identical objects or events.
k the number of objects or events without repetition.

Example 1: We want to calculate the number of distinct permutations of the letters in the word "TENNESSEE".

$$n = 9$$

$n_1 = 1$ (the letter T),
$n_2 = 4$ (the letter E),
$n_3 = 2$ (the letter N),
$n_4 = 2$ (the letter S).

$$P_9^{1,4,2,2} = \frac{9!}{1! \times 4! \times 2! \times 2!} = 3780$$

So, there are 3780 permutations of the letters in the word "TENNESSEE".

Example 2: Suppose you have 10 balls: 3 are red, 5 green, and 2 yellow. How many distinct ways can you arrange them in a row?

$$n = 10$$

$n_1 = 3$ (red balls),
$n_2 = 5$ (green balls),
$n_3 = 2$ (yellow balls).

$$P_{10}^{3,5,2} = \frac{10!}{3! \times 5! \times 2!} = 2520$$

So, there are 2520 distinct arrangements of these balls.

4. Combinations

A combination reflects the ways of choosing k objects or events out of a set of n objects or events, it is a selection of objects where order is insignificant.

4.1 Combinations without repetition

The number of ways to choose k objects from n objects where order does not matter and without repetition is:

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

Example: Suppose you have 12 people, and you want to choose 3 people to form a committee. How many different committees can you constitute?

$$n = 12$$
$$k = 3$$

$$C_{12}^3 = \frac{12!}{3!\,(12-3)!} = 220$$

So, there are 220 possible committees.

4.2 Combinations with Repetition

If repetition is allowed, the number of ways to choose k objects from n is:

$$C_{n+k-1}^k = \frac{(n+k-1)!}{k!\,(n-1)!}$$

Example: Suppose you are at candy shop that offers candies with 3 flavours (Banana, Lemon, Orange), and you want to buy 2 candies. How many different combinations can you choose if you can repeat flavours?

$$n = 3$$
$$k = 2$$

$$C_{3+2-1}^2 = \frac{(3+2-1)!}{2!\,(3-1)!} = \frac{4!}{2! \times 2!} = 6$$

So, there are 6 possible combinations:

Banana + Banana
Banana + Lemon
Banana + Orange
Lemon + Lemon
Lemon + Orange
Orange + Orange

PASCAL'S Triangle

These calculation rules provide a recursive computational approach for determining combinations, embodied in the structure known as Pascal's Triangle.

NEWTON'S BINOMIAL THEOREM

The Binomial Theorem is a key algebraic principle that provides a formula for expanding powers of a binomial expression, $(a+b)^n$, where n is a non-negative integer.

Let a and b be two real numbers. For any non-negative integer n, the expansion of $(a+b)^n$ is given by:

$$(a+b)^n = \sum_{k=0}^{n} C_n^K \, a^{n-k} b^k$$

Where
C_n^K is the binomial coefficient, calculated as:

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

$n!$ is the factorial of n ($n! = n \times (n-1) \times \cdots \times 1 n! = n \times (n-1) \times \cdots \times 1$).
The summation $\sum$ means that we sum the terms for all values of k from 0 to n.

BINOMIAL COEFFICIENTS AND PASCAL'S TRIANGLE

Binomial coefficients form Pascal's Triangle, that is arranged in a triangular pattern, where each number is the sum of the two numbers directly above it.
Let a and b be two real numbers. Then:

	0	1	2	3	4	....	
n=0	1						
n=1	1	1					
n=2	1	2	1				
n=3	1	3	3	1			
n=4	1	4	6	4	1		
n=6	1	5	10	10	5	1	
n=7	1	6	15	20	15	6	1
	....						
	....						
n-1	1	C_{n-1}^1	C_{n-1}^{p-1}	C_{n-1}^{p}	....		
n	1	C_n^1	C_n^{p-1}	C_n^{p}			

$$\boxed{C_n^p = C_{n-1}^{p-1} + C_{n-1}^p}$$

Figure 03: Pascal's Triangle

In Pascal's Triangle, the n^{th} row corresponds to the binomial coefficients in the expansion of $(a+b)^n$, where n is a non-negative integer.

Example: In the birthday paradox problem, how many people are needed for the probability of a birthday match to be greater than 1/2?

Solution:

We are looking for the smallest integer n such that

$$P(at\ least\ two\ people\ share\ a\ birthday) > \frac{1}{2}.$$

Ignoring leap years, there are 365 possible birthdays.
Also, birthdays are equally likely and independent.

We use complementary probability, since it is easier to first compute the probability of no birthday match:

$$P_{no\ match}(n) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365-n+1}{365}$$

That is:

$$P_{no\ match}(n) = \prod_{k=0}^{n-1}(1 - \frac{k}{365})$$

Probability of at least one match

$$P_{match}(n) = 1 - P_{no\ match}(n)$$

We want:

$$1 - \prod_{k=0}^{n-1}(1 - \frac{k}{365}) > 0.5$$

For n=22:

$$P_{match}(22) \approx 0.4757(< \ 0.5)$$

For n=23:

$$P_{match}(23) \approx 0.5073(> \ 0.5)$$

With 23 people, the probability that at least two share the same birthday is approximately 50.73%, which is greater than 0.5.

Chapter III Exercises

Exercise 1:
Answer the following:

1- In how many ways can 6 students line up in a ceremony?
2- In how many different ways can 6 students and their teacher line up in a ceremony if the teacher must always be placed at the right end?
3- In how many different ways can 6 students and their 2 teachers line up in a ceremony if one teacher must always be placed at the right end and the other one at the left end?
4- In how many ways can 6 students sit around a circular table for lunch?
5- How many words can be formed using all the letters of the words: Mississippi, Bookkeeper?

Solution:

1- This is a permutation of all 6 students, since the order matters. The number of ways to line up 6 students in a ceremony is given by:

$$P_n = n!$$
$$P_6 = 6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

There are 720 different ways for the 6 students to line up.

2- We have 6 students + 1 teacher = 7 people total. But only 6 positions (to the left) are available for the 6 students, since the teacher has a fixed position (at the right end).

The teacher's position is fixed (1 way) … $n_1 = 1$
The number of ways to arrange 6 distinct students in the 6 left positions…$n_2 = P_6 = 6! = 720$
We use the multiplication rule: $N = n_1 \times n_2 = 1 \times 720 = 720$

3- We have 6 students + 2 teachers = 8 people total. But only 6 positions (the middle positions) are available for the 6 students, since the teachers have fixed position (one at the right end and the other one at the left end).

Teacher A must be at the right end.
Teacher B must be at the left end.

The number of ways to arrange 6 distinct students in the 6 left positions… $n_1 = P_6 = 6! = 720$
Ther are 2 ways to assign the teachers to the ends (2 ends for 2 students)… $n_2 = P_2 = 2! = 2$

We use the multiplication rule: $N = n_1 \times n_2 = 720 \times 2 = 1440$

4- since there's no fixed starting point when arranging people around a circular table, rotations of the same arrangement are considered identical.

So, we fix one person's position to account for rotation symmetry. Then, the number of distinct ways to arrange *n* distinct people around a circle is:

$$(n-1)!$$

Here $n = 6$:

$$(6-1)! = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

There are 120 distinct seating arrangements around the circular table.

5- MISSISSIPPI, BOOKKEEPER

MISSISSIPPI has 11 letters with repetitions ($n = 11$), we use a permutation with repetitions:

$$P_n^{n_1,n_2,n_3,\dots,n_k} = \frac{n!}{n_1! \times n_2! \times n_3! \times \dots \times n_k!}$$

We count letters and repetitions:
I: $n_1 = 4$
S: $n_2 = 4$
P: $n_3 = 2$

$$P_{11}^{4,4,2} = \frac{11!}{4! \times 4! \times 2!} = \frac{39{,}916{,}800}{1152} = 34{,}650$$

BOOKKEEPER has 10 letters with repetitions ($n = 10$), we use a permutation with repetitions:

$$P_n^{n_1,n_2,n_3,\dots,n_k} = \frac{n!}{n_1! \times n_2! \times n_3! \times \dots \times n_k!}$$

We count letters and repetitions:
O: $n_1 = 2$
K: $n_2 = 2$
E: $n_3 = 3$

$$P_{10}^{2,2,3} = \frac{10!}{2! \times 2! \times 3!} = \frac{3{,}628{,}800}{24} = 151{,}200$$

Exercise 2:
What is the number of license plates that can be obtained by using 3 letters from the alphabet (26 letters) and 2 digits to the right of the letters, in the following cases:
Case 1: Without repeating letters or digits.
Case 2: With repetition of letters allowed, but without repetition of digits.
Case 3: With repetition of both letters and digits allowed.
Case 4: Without repetition of letters, but with repetition of digits allowed, provided that the first digit is odd.

Solution:
Format: LLL DD (3 letters then 2 digits)

Case 1: Without repeating letters or digits
Letters (3 positions, no repetition):
The first letter: $n_1 = 26$
The second letter: $n_2 = 25$
The third letter: $n_3 = 24$

$$N_1 = n_1 \times n_2 \times n_3 = 26 \times 25 \times 24 = 15{,}600$$

Digits (2 positions, no repetition):
The first digit: $n_1 = 10$
The second digit: $n_2 = 9$

$$N_2 = n_1 \times n_2 = 10 \times 9 = 90$$

Total:

$$N = N_1 \times N_2 = 15{,}600 \times 90 = 1{,}404{,}000$$

Case 2: Repetition of letters allowed, but no repetition of digits
Letters (repetition allowed):
The first letter: $n_1 = 26$
The second letter: $n_2 = 26$
The third letter: $n_3 = 26$

$$N_1 = n_1 \times n_2 \times n_3 = 26^3 = 17{,}576$$

Digits (no repetition):
The first digit: $n_1 = 10$
The second digit: $n_2 = 9$

$$N_2 = n_1 \times n_2 = 10 \times 9 = 90$$

Total:

$$N = N_1 \times N_2 = 17{,}576 \times 90 = 1{,}581{,}840$$

Case 3: Repetition of both letters and digits allowed
Letters:
Letters (repetition allowed):
The first letter: $n_1 = 26$
The second letter: $n_2 = 26$
The third letter: $n_3 = 26$

$$N_1 = n_1 \times n_2 \times n_3 = 26^3 = 17{,}576$$

Digits:
The first digit: $n_1 = 10$
The second digit: $n_2 = 10$

$$N_2 = n_1 \times n_2 = 10^2 = 100$$

Total:

$$N = N_1 \times N_2 = 17{,}576 \times 100 = 1{,}757{,}600$$

Case 4: No repetition of letters, repetition of digits allowed, first digit is odd
Letters (3 positions, no repetition):
The first letter: $n_1 = 26$
The second letter: $n_2 = 25$
The third letter: $n_3 = 24$

$$N_1 = n_1 \times n_2 \times n_3 = 26 \times 25 \times 24 = 15{,}600$$

Digits:
First digit: odd → {1,3,5,7,9} → $n_1 = 5$
Second digit: any digit 0–9 → $n_2 = 10$

$$N_2 = n_1 \times n_2 = 5 \times 10 = 50$$

Total:

$$N = N_1 \times N_2 = 15{,}600 \times 50 = 780{,}000$$

Exercise 3:

1- Given the following set:
A={1,1,2,3,3,3}
How many 6-digit numbers can be formed from A?

2- Given the following set:
B={2,3,4,5,6}
How many even 3-digit numbers can be formed from B?
(a) With repetition allowed
(b) Without repetition

3- How many 3-digit numbers that are odd or multiples of 2 can be formed from B, using each digit only once?

4- How many 5-digit numbers with distinct digits can be formed from the set:
$C = \{1,2,3,4,5,6,7,8,9\}$?

5- How many 8-digit numbers can be formed from the set:

$$D = \{1,2,4,5,6,7,8,9\}$$

(a) With repetition allowed
(b) Without repetition

6- How many 4-digit numbers can be formed from the set E:
E={0,1,2,3,4,5,6,7}, such that:
(a) Digits can be repeated
(b) All digits are distinct (without repetition)
(c) The number is a multiple of 10 (with repetition)

Solution:

1- We want the number of distinct permutations of the multiset $A = \{1,1,2,3,3,3\}$
We count digits and frequencies:
Total digits $n = 6$

repetitions:
The number 1 appears 2 times: $n_1 = 2$
The number 2 appears 1 time: $n_2 = 1$
The number 3 appears 3 times: $n_3 = 3$

$$P_6^{2,1,3} = \frac{6!}{2! \times 1! \times 3!} = \frac{720}{12} = 60$$

2- B = {1, 2, 3, 4, 5}
Even digits in $B = \{2, 4\}$ → 2 possibilities for the units place.

(a) Repetition allowed
Units digit (U): 2 choices (2, 4); $n_1 = 2$
Tens digit (T): 5 choices; $n_2 = 5$

Hundreds digit (H): 5 choices; $n_3 = 5$

$$N = n_1 \times n_2 \times n_3 = \ 2 \times 5 \times 5 = 50$$

(b) Without repetition
Units (U): 2 choices (even digits); $n_1 = 2$
Hundreds (H): cannot repeat U → 4 choices left from B; $n_2 = 4$
Tens (T): cannot repeat H or U → 3 choices left; $n_3 = 3$

$$N = n_1 \times n_2 \times n_3 = 4 \times 3 \times 2 = 24$$

3- odd or multiples of 2 digits from B, without repetition
Odd numbers without repetition
Odd digits in B: {1, 3, 5} → 3 choices for units place.
Units: 3 choices; $n_1 = 3$
Hundreds: from remaining 4 digits (since 0 not in B) → 4 choices; $n_2 = 4$
Tens: from remaining 3 digits → 3 choices; $n_3 = 3$

Odd numbers: $N = n_1 \times n_2 \times n_3 = 3 \times 4 \times 3 = 36$

Even numbers without repetition
Even digits in B: {2, 4} → 2 choices for units place.
Units: 2 choices; $n_1 = 2$
Hundreds: from remaining 4 digits → 4 choices; $n_2 = 4$
Tens: from remaining 3 digits → 3 choices; $n_3 = 3$

Even count $N = n_1 \times n_2 \times n_3 = 2 \times 4 \times 3 = 24$

Since a number is either odd or even (mutually exclusive):
Odd or Even: $36 + 24 = 60$

This matches the total 60, which makes sense when we calculate all 3-digit numbers from B without repetition (either odd or even numbers)
Total arrangements of 3 digits from 5 digits:

$$A_5^3 = \frac{5!}{(5-3)!} = 60$$

4- We have C={1,2,3,4,5,6,7,8,9}— 9 digits, none is 0, so $n = 9$
We want 5-digit numbers with distinct digits, so $k = 5$
Since all digits are nonzero, any arrangement of 5 distinct digits from these 9 will form a valid 5-digit number.

Number of ways = Arrangement of 5 out of 9:

$$A_9^5 = \frac{9!}{(9-5)!} = 15120$$

5- We have D = {1, 2, 4, 5, 6, 7, 8, 9} — 8 digits, no 0. $n = 8$

We want 8-digit numbers with distinct digits.
Since all digits are nonzero, any arrangement of 8 distinct digits from these 8 will form a valid 8-digit number.

(a) With repetition allowed
We form 8-digit numbers: each of the 8 positions can be filled with any of the 8 digits.

$$8^8 = 16{,}777{,}216$$

(b) Without repetition
We are arranging all 8 digits of D in a sequence (since we need an 8-digit number from 8 distinct digits).
We calculate the number of permutations of 8 distinct digits:

$$P_n = n!$$
$$P_8 = 8! = 40{,}320$$

6- E = {0, 1, 2, 3, 4, 5, 6, 7} — 8 digits.

(a) Digits can be repeated
First digit cannot be 0 → 7 choices (1-7)
Each of the remaining 3 digits → 8 choices each

$$7 \times \mathrm{r}A_8^3 = 7 \times 8^3 = 7 \times 512 = 3584$$

(b) All digits are distinct
First digit: 7 choices (1-7)
Remaining 3 digits chosen from the remaining 7 digits (including 0) and arranged:

$$7 \times A_7^3 = 7 \times \frac{7!}{(7-3)!} = 7 \times (7 \times 6 \times 5) = 7 \times 210 = 1470$$

(c) Multiple of 10 (with repetition)
A multiple of 10 must end in 0.
Last digit: 1 choice (0)
First digit: cannot be 0 → 7 choices (1-7)
Middle 2 digits: 8 choices each

$$7 \times \mathrm{r}A_8^3 \times 1 = 7 \times 8^2 \times 1 = 7 \times 64 = 448$$

Exercise 4:
A box contains 7 white balls (W) and 3 black balls (B). We draw 4 balls at once.
What is the number of ways to draw:

1- 4 balls (any color)
2- 4 white balls
3- 4 black balls
4- 1 white and 3 black balls

Solution:
We have:
White balls (B) = 7
Black balls (N) = 3
Total balls $n = 10$

We draw 4 balls at once ($k = 4$)→ order doesn't matter → **combinations**.

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

1- 4 balls (any color)

$$C_{10}^4 = \frac{10!}{4!\,(10-4)!} = 210$$

2- 4 white balls

Choose 4 white from 7 white, 0 black from 3 black:

$$C_7^4 = \frac{7!}{4!\,(7-4)!} = 35$$

3- 4 black balls

There are only 3 black balls available → it is an impossible event.

4- 1 white and 3 black balls

$$C_7^1 \times C_3^3 = 7 \times 1 = 7$$

Exercise 5:

In a round-robin tennis tournament, each of 7 players plays exactly one match against every other player. How many matches are played in total?

Solution:

In a round-robin tournament with ($n = 7$) players, each match corresponds to choosing 2 distinct players to play against each other ($k = 2$).

$$C_7^2 = \frac{7!}{2!\,(7-2)!} = 21$$

There is 21 matches are played in total.

Chapter IV: Introduction to Probability Theory

Probability theory provides essential tools to analyze random events. Understanding its core concepts and theorems proves indispensable across diverse fields including statistics, economics, finance, and empirical sciences.
This chapter focuses on applied conditional probability, beginning with foundational probability computations. We examine the addition and multiplication rules - indispensable tools for determining probabilities of interrelated events. These principles enable us to quantify relationships between dependent and independent events systematically.
As we progress, and building on these fundamentals, we introduce two fundamental theorems; the Total Probability Theorem, which facilitates probability computation when conditional probabilities are known or readily derivable, and Bayes' Theorem, which provides a mechanism for updating probability estimations based on new evidences.

1. Probability of an event

Consider the sample space **Ω** associated with an experiment. To fully establish the probabilistic model, we must specify a probability measure. Intuitively, this determines how likely any particular outcome or collection of outcomes may be. Formally, the probability measure assigns to each event A a real number $P(A)$, known as the probability of A.
The probability of A $P(A)$ is a function P which assigns a numerical value to the event A, meaning that P is a function from the space of events to the real line.

$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total possible outcomes}}$ (if outcomes are equally likely)

The probability of A satisfies the following axioms (Kolmogorov axioms):

1- **Non-Negativity**: $P(E) \geq 0$ for any event E.
2- **Normalization**: $P(\Omega) = 1$, meaning that the probability of the entire sample space **Ω** is equal to 1.

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset)$$

Therefore, the probability of an event quantifies the likelihood of its occurrence, with values ranging between impossible event ($P(\emptyset) = 0$) and certain event ($P(\Omega) = 1$).

3- **Additivity**: If A and B are two disjoint events (mutually exclusive), then the probability of their union satisfies

$$P(A \cup B) = P(A) + P(B)$$

If the sample space has an infinite number of elements, and $E_1, E_2, \ldots$ is a sequence of mutually exclusive events, then:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = P(E_1 \cup E_2 \cup \ldots) = P(E_1) + P(E_2) + \ldots = \sum_{i=1}^{\infty} P(E_i)$$

Thus, the probability of the union of finitely many disjoint events is always equal to the sum of the probabilities of these events.

Properties of the Probability Function
The following properties of probability functions can be derived from the axioms.
For events A and B,

1.The probability that an event does not occur equals one minus the probability that the event occurs.

$$P(\bar{A}) = 1 - P(A)$$

2. $P(\emptyset) = 0.$
"nothing happens" occurs with zero probability.

3. By Kolmogorov's second axiom, the probability of any event cannot exceed 1.

$$P(A) \leq 1$$

4. Monotonicity of probability measures guarantees means that larger sets necessarily have larger probability. For any events A and B
If $A \subset B$, then $P(A) \leq P(B)$.

5. The inclusion-Exclusion Principle is a useful decomposition of the probability of the union of two events, where:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

6. Boole's inequality establishes that for any countable collection of events, the probability measure of their union is dominated by the sum of their individual probabilities.
For events A and B

$$P(A \cup B) \leq P(A) + P(B)$$

For an infinite number of events A_i

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

7. Bonferroni's inequality establishes a lower bound for intersection probabilities through the inclusion-exclusion principle.
For events A and B

$$P(A \cap B) \geq P(A) + P(B) - 1$$

For an infinite number of events A_i

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) \leq 1 - \sum_{i=1}^{\infty} (1 - P(A_i))$$

2. Fundamental Theorems

2.1 Addition Rule (Union)

The addition rule calculates the probability that at least one of the events *A* or *B* occurs.

2.1.1 Case of mutually exclusive events: the basic version

A and B are two disjoint events of the sample space Ω ($A \cap B = \emptyset$)

$$P(A \cup B) = P(A) + P(B)$$

Example: rolling one fair die one time, let *A* be the event of rolling a 2, and *B* the event of rolling a number ≥ 6

$$P(A) = 1/6$$
$$P(B) = 1/6.$$

A and B are mutually exclusive events, then the probability that at least *A* or *B* occurs is

$$P(A \cup B) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

2.1.2 Case of non-mutually exclusive events: the general version

A and B are two events of the sample space Ω, if $A \cap B \neq \emptyset$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example 1: rolling one fair die one time, let *A* be the event of rolling an odd number, and *B* the event of rolling a number ≥ 3,

$A = \{1,3,5\}, P(A) = \frac{3}{6}$

$B = \{3,4,5,6\}, P(B) = \frac{4}{6}$

$A \cap B = \{3,5\} \Rightarrow (P(A \cap B) = \frac{2}{6}$

$$A \cup B = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$$

Example 2: In a 52 deck of cards, calculate the probability of drawing a Spade (♠) or a Queen (Q)?

$$P(♠ \cup \mathrm{Q}) = P(♠) + P(Q) - P(♠ \cap Q)$$
$$P(♠ \cup \mathrm{Q}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$$

2.2 Multiplication Rule, Intersection Rule (Conjunction)

The conjunction rule calculates the probability that **both *A* and *B*** occur.

2.2.1 Conditional Probability

Consider two events A and B, the conditional probability of A given B (or A assuming that B occurs) : $P\ (A \setminus B)$, stands for the fraction of the time that *A* occurs once we know that *B* occurs. If the events A and B are independent (unrelated), the probability that an event is unaffected by the outcome of another event is

$$P(A \setminus B) = P(A)$$
$$P(B \setminus A) = P(B)$$

If the events are dependent

$$P(A \setminus B) \neq P(A)$$
$$P(B \setminus A) \neq P(B)$$

2.2.2 Dependence of two events

The definition of $P(A \mid B)$ is immediately derived from the multiplication formula

$P(A \cap B) = P(A \mid B)P(B)$ where $P(A \mid B)$ is the probability of *A* knowing that *B* has occured

$P(A \cap B) = P(B \mid A)P(A)$ where $P(B \mid A)$ is the probability of *B* knowing that *A* has occured

Note that

$$P(A \cap B) = P(A \setminus B)P(B) = P(B \setminus A)P(A)$$

It is computed as the ratio of the probability that *A* and *B* both occur, divided by the probability that *B* occurs, with $P(B) > 0$, as follows

$$P(A \mid B) = \frac{P(A \cap B)}{\mathrm{P(B)}}$$

i.e. once we know that *B* has occurred, the only way for *A* to occur is if the outcome is in the intersection $A \cap B$.

This is used to compute the joint probability of *A* and *B* when we are given the probability of *A* and the conditional probability of *B* given *A*.

The probability of *B* given *A*, with $P(A) > 0$ is

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

If *A* and *B* are independent then

$$P(A \setminus B) = \frac{P(A \cap B)}{\mathrm{P(B)}} = \frac{P(A)P(B)}{\mathrm{P(B)}} = \mathrm{P(A)}$$

Also $P(A\setminus \bar{B}) = P(A)$.

$$P(A \setminus \bar{B}) = \frac{P(A \cap \bar{B})}{\mathrm{P}(\bar{B})} = \frac{P(A)P(\bar{B})}{\mathrm{P}(\bar{B})} = \mathrm{P(A)}$$

So knowing whether *B* occurs or not does not influence the probability that *A* occurs.

Example: Drawing 2 cards without replacement from a 52 deck of cards. Let *A* be the 1st card is a Spade, and *B* the 2nd card is a Spade

$$P(A) = \frac{13}{52}$$

$$P(B|A) = \frac{12}{51}$$

$$P(A \cap B) = \frac{13}{52} \times \frac{12}{51} = \frac{156}{2652} = \frac{13}{221}$$

N.B. $P(A \setminus B) > P(A) \Leftrightarrow P(B \setminus A) > P(B)$.

2.2.3 Independence of two events

Events are independent if their occurrence is unrelated, i.e. knowing that one event occurs does not affect the conditional probability of the other event. For example, consider flipping a coin and rolling a die. If there is no mechanism connecting the two we would typically expect that neither rolling the die nor throwing the coin is affected by the outcome of the other; these two events are independent. As another example, consider the birth rate in Algeria and the price of coffee in China. There is no reason to expect one of these two events affecting the other; these two events are independent.

Thus, two unrelated (independent) events A and B will satisfy the properties

$$P(A\setminus B) = P(A)$$
$$P(B\setminus A) = P(B)$$

When events are independent then joint probabilities can be calculated by multiplying individual probabilities. In words, when the occurrence of B provides no information and does not alter the probability that A has occurred. From the definition of conditional probability this implies

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

Independence is a symmetric property; i.e. if *A* is independent of *B*, then *B* is independent of *A*, and we can unambiguously say that *A* and *B* are independent events.
Independence is often easy to grasp intuitively, but on the other hand, it is not easily visualized in terms of the sample space. A common first thought is that two events are independent if they are mutually exclusive, but in fact the opposite is true: two mutually exclusive events *A* and *B* with $P(A) > 0$ and $P(B) > 0$ are never independent, since their intersection is empty ($A \cap B = \emptyset$) and has a ull probability.

Example 1: Consider the case of rolling a die and flipping a coin at the same time, if *A* is the event of rolling a number ≥ 5, and *B* is the event that the coin is Head, are the events *A* and *B* independent?

$\Omega = \{(1,H),(2,H),(3,H),(4,H),(5,H),(6,H),(1,T),(2,T),(3,T),(4,T),(5,T),(6,T)\}$

$P(A) = P(\{(5,H),(6,H),(5,T),(6,T)\}) = \frac{4}{12} = \frac{1}{3}$, and

$$P(B) = P(\{(1,H),(2,H),(3,H),(4,H),(5,H),(6,H)\}) = \frac{6}{12} = \frac{1}{2}$$

Also, $P(A \cap B) = P(\{(5,H),(6,H)\}) = \frac{2}{12} = \frac{1}{6}$, which is indeed equal to $\frac{1/3}{1/2} = \frac{1}{6}$. So, *A* and *B* are independent in this example.

Example 2: Consider rolling a fair six-sided die, with the events:

$$A = \{1,2,3,4\}$$
$$B = \{4,5,6\}$$

The intersection is $A \cap B = \{4\}$, which has probability:

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

The probability of *B* is:

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

The conditional probability $P(A \mid B)$ is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = 1/3$$

This can also be reasoned directly: given *B*, the outcomes {4},{5},{6} are equally likely, each with probability $\frac{1}{3}$. Event *A* occurs given *B* only if the outcome is {4}, so $P(A \mid B) = \frac{1}{3}$.

2.2.4 Properties of conditional probability

A and B are independent, then considering set theory
$P(A \cap \bar{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$; so A and $\bar{B}$ are independent
$\bar{A}$ and $\bar{B}$ are independent, and
$\bar{A}$ and B are independent.
$P(A \cap B) = P(A)P(B \setminus A) = P(B)P(A \setminus B)$,
$(A \cap B \cap C) = P(A \setminus B \cap C)\, P(B \setminus C)\, P(C)$,

$P\,(A \setminus B\, \cap\, C) = \frac{P\,(A \cap B \setminus C)}{P\,(B \setminus C)}$,
Conditional probability satisfies its own law of total probability.
If A and B are independent, and $P(A) > 0$ and $P\,(B) > 0$, then $P\,(A \setminus B) = P\,(A)$ and $P\,(B \setminus A) = P\,(B)$.
If in addition, $P(B) > 0$, independence is equivalent to the condition $P(A \mid B) = P(A)$.
Two events A and B are said to be conditionally independent, given another event C with $P(C) > 0$, if
$P(A \cap B|C) = P(A|C)P(B|\,C)$.
If in addition, $P(B \cap C) > 0$, conditional independence is equivalent to the condition
$P(A|B \cap C) = P(A|C)$.
Independence does not imply conditional independence, and vice versa.

Example 1: What is the probability of getting a tail on the first coin flip and a tail on the second flip when flipping a coin two times?

$$\Omega = \{(H,H),(H,t),(T,T),(T,H)\}$$

A: getting a tail on the 1st coin flip, A $= \{(T,T),(T,H)\}$
$P(A) = 1/2$
B: getting a tail on the 2nd flip, B $= \{(H,t),(T,T)\}$
$P(B) = 1/2$.

$$P(A \cap B) = P(A) \times P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Example 2: What is the probability of rolling at least 10 from a pair of fair die.
Answer the same question supposing that the dice are still independent but are not fair, such as the probability of a "5" is 2/6 and the probability of a "6" is 0.

The sample space for two independent dice

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

Let A be the event of rolling at least 10 from a pair of die.

$$A = \{(4,6),(5,5),(6,4),(5,6),(6,5),(6,6)\}.$$

The outcomes are disjoint. Thus the probability of rolling at least 10 is the sum

$$P(A) = P(4,6) + P(5,5) + P(6,4) + P(5,6) + P(6,5) + P(6,6)$$

Assuming that the two dice are independent of one another, applying the multiplication rule

$$P(A) = P(4) \times P(6) + P(5) \times P(5) + P(6) \times P(4) + P(5) \times P(6) + P(6) \times P(5) + P(6) \times P(6)$$

$$P(A) = \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6}$$

$$P(A) = \frac{1}{6}$$

Now suppose the dice are still independent but are not fair. Such as the the probability of a "5" is 2/6 and the probability of a "6" is 0.

$$P(A) = P(4) \times P(6) + P(5) \times P(5) + P(6) \times P(4) + P(5) \times P(6) + P(6) \times P(5) + P(6) \times P(6)$$

$$P(A) = \frac{1}{6} \times 0 + \frac{2}{6} \times \frac{2}{6} + 0 \times \frac{1}{6} + \frac{2}{6} \times 0 + 0 \times \frac{2}{6} + 0 \times 0$$

$$P(A) = \frac{4}{36} = \frac{1}{9}$$

Example 3: Two fair dice are thrown. Let:
C1: The first die shows a prime number $\{2, 3, 5\}$.
C2: The second die shows an even number $\{2, 4, 6\}$.
C3: The product of the two numbers is even.

1- Are C1 and C2 independent?
2- Are C1 and C3 independent?

Solution:
First, we determine Individual Probabilities
Sample Space: 36 equally likely outcomes.
Probability of C1:
Prime numbers on a die: {2, 3, 5} → 3 outcomes.

$$P(C1) = \frac{3}{6} = \frac{1}{2}$$

Probability of C2:
Even numbers on a die: {2, 4, 6} → 3 outcomes.

$$P(C2) = \frac{3}{6} = \frac{1}{2}$$

Probability of C3:
The product is even if at least one die shows an even number.
It's easier to compute the complement: product is odd only if both dice are odd.
Odd numbers: {1, 3, 5} → 3 outcomes per die.

$$P(product\ odd) = \frac{3 \times 3}{36} = \frac{9}{36} = \frac{1}{4}$$

$$P(C3) = 1 - \frac{1}{4} = \frac{3}{4}$$

Independence of C1 and C2
We need to calculate P(C1∩C2):
C1: First die is 2, 3, or 5.
C2: Second die is 2, 4, or 6.
Number of favorable outcomes: 3×3=9

$$P(C1 \cap C2) = \frac{9}{36} = \frac{1}{4}$$

Check condition for independence:

$$P(C1) \cdot P(C2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$P(C1 \cap C2) = \frac{1}{4} = P(C1) \cdot P(C2)$$

Therefore, C1 and C2 are independent.

Independence of C1 and C3
We need to calculate $P(C1 \cap C3)$:
C3: Product is even → At least one die is even.
It's easier to find P($C1 \cap \overline{C3}$): Product is odd AND first die is prime.
Product is odd only if both dice are odd.
First die prime AND odd: {3, 5} → 2 outcomes.
Second die must be odd: {1, 3, 5} → 3 outcomes.

$$P(C1 \cap \overline{C3}) = \frac{2 \times 3}{36} = \frac{6}{36} = \frac{1}{6}$$

Therefore:

$$P(C1 \cap C3) = P(C1) - P(C1 \cap \overline{C3}) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$

Check condition for independence:

$$P(C1) \cdot P(C3) = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$$

$$P(C1 \cap C3) = \frac{1}{3} \approx 0.333 \; and \; P(C1) \cdot P(C3) = \frac{3}{8} = 0.375$$

$$P(C1 \cap C3) \neq P(C1) \cdot P(C3)$$

Therefore, C1 and C3 are NOT independent.

2.2.5 Independence of multiple events

Events A_1, A_2 , ... are independent (or 'mutually independent') if for any finite subcollection of distinct events A_{i1}, A_{i2} , ... , A_{ij}

$$P(A_{i1} \cap A_{i2} \cap \cdots \cap A_{ir}) = P(A_{i1})P(A_{i2}) \cdots P(A_{ir})$$

For multiple events, the independence of events is more involved to some extent, they can be pairwise independent without being (mutually) independent.

If two events are not independent we say that they are dependent. In this case the joint event $A \cap B$ occurs at a different rate than predicted if the events were independent.

2.3 Law of Total Probability

Conditional probability leads to the law of total probability, in a different and sometimes more helpful way.
The law of total probability allows us to calculate the probability $P(A)$ by partitioning the sample space into mutually exclusive events B_1, B_2 , ... , B_n. The probability that A occurs is a weighted average of its conditional probability considering each event B_i, where each B_i is weighted according to its (unconditional) probability. The figure bellow provides a graphical representation of the total probability theorem.

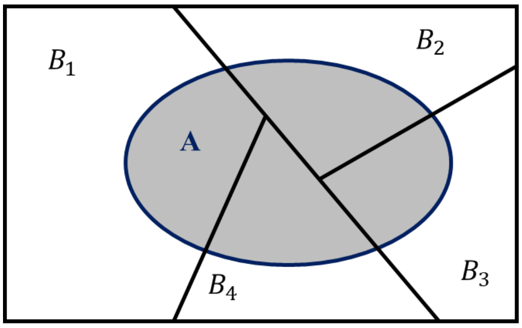


Figure 04: Visualization of the Bayesian and Total Probability Theorems

Consider B_i, a finite or countable collection of disjoint events such that $\bigcup_i B_i = \Omega$ is partition of the sample space Ω (each possible outcome is included in one and only one of the events B_i), and assume that $P(B_i) > 0$. For any event A,

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n) = \sum_i P(A \cap B_i)$$

Since the events $(A \backslash B_i\ )$ are disjoint, the application of the intersection rule and the definition of conditional probability, with $P(B_i) > 0$ for all i gives that

$$P(A) = P(B_1)P(\,A \mid B_1\,) + P(B_2)P(\,A \mid B_2\,) + \cdots + P(B_n)P(A \mid B_n) = \sum_i P(B_i)P(A \mid B_i)$$

This is called the Law of Total Probability, where the second summation extends only over those B_i with $P(B_i)\ >\ 0$.
B_i are mutually exclusive, $B_i \cup B_j = \emptyset$
Their union covers Ω, $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$, i.e. $\bigcup_i B_i = \Omega$
This theorem allows to compute the probability of various events A for which the conditional probabilities $P(A \mid B_i)$ are given or easy to derive. The idea is to choose appropriately the partition $B_1, B_2, \ldots, B_n$, and this choice depends on the problem structure.

Example 1: Suppose a factory with two machines: M_1 (65% of production) and M_2 (53% of production).
M_1 has 4% defect rate, M_2 has 1%. An item is chosen uniformly at random from this factory. What is the probability that the chosen item will be defective?
Let D be the set of all defective items.
$P(M_1) = 0.65, P(M_2) = 0.35, P(\,D \mid M_1\,) = 0.04, P(\,D \mid M_2\,) = 0.01$
We compute the probability of a defective item by Total Probability Theorem

$$P(D) = P(M_1)P(\,D \mid M_1\,) + P(M_2)P(\,D \mid M_2\,) = 0.65 \times 0.04 + 0.35 \times 0.01 = 0.0295$$

so the probability that the randomly chosen Item be defective is 2.95%.

Example 2: Suppose a firm with 55% women and 45% men. Suppose that 7% of the girls are left handed, and 5% of the boys are left handed. An employee is chosen uniformly at random from this firm. What is the probability that the chosen employee will be left handed?
Let B_1 be the set of women and B_2 be the set of men. Then $\{B_1, B_2\}$ is a partition of the firm.
Let A be the set of all left handed employees.
$P(B_1) = 0.55, P(B_2) = 0.45, P(\,A \mid B_1\,) = 0.07, P(\,A \mid B_2\,) = 0.05$
We have to compute P(A), using Total Probability Theorem

$$P(A) = P(B_1)P(\,A \mid B_1\,) + P(B_2)P(\,A \mid B_2\,) = 0.55 \times 0.07 + 0.45 \times 0.05 = 0.061$$

so there is a 6.1% chance that the randomly chosen employee be left handed.

Example 3: During a darts competition, probability of loosing a game for a player is 0.3 against half the players (B_1), 0.4 against a quarter of the players (B_2), and 0.5 against the remaining quarter of the players (B_3). He plays a game against a randomly chosen opponent. What is the probability of winning?
Let B_i be the event of playing with an opponent of type i. We have
P(B_1) = 0.5, P(B_2) = 0.25, P(B_3) = 0.25.
Let also A be the event of loosing. So
$P(A \mid B_1) = 0.3, P(A \mid B_2) = 0.4, P(A \mid B_3) = 0.5$
Then
$P(\bar{A} \mid B_1) = 0.7, P(\bar{A} \mid B_2) = 0.6, P(\bar{A} \mid B_3) = 0.5$

Thus, by the total probability theorem,

$$\begin{aligned} P(\bar{A}) &= P(B_1)P(\bar{A} \mid B_1) + P(B_2)P(\bar{A} \mid B_2) + P(B_3)P(\bar{A} \mid B_3) \\ &= 0.5 \times 0.7 + 0.25 \times 0.6 + 0.25 \times 0.5 = 0.625 \end{aligned}$$

The probability of winning is 62.5%.

2.4 Bayes' Theorem (Bayes' formula)

Bayes' rule is often used for inference. There are some causes that may lead to a certain result. When the latter is observed, it is possible to infer the cause. The disjoint events $B_1, B_2, \dots, B_n$ that form a partition of the sample space are associated with the causes and the event A represents the result. The probability $P(A \mid B_i)$ that the result will be observed when the cause B_i occurs is associated to a probabilistic causal model. Given that the result A has been observed, we wish to evaluate the (conditional) probability $P(B_i \mid A)$ that the cause Bi is present, assuming that $P(B_i) > 0$.

The definition of conditional probability implies

$$P(A \cap B) = P(A \setminus B)P(B) = P(B \setminus A)P(A)$$

Then

$$P(A \cap B_i) = P(A \setminus B_i)P(B_i) = P(B_i \setminus A)P(A)$$

If $P(A) > 0$ and $P(B_i) > 0$, the probability of an event B_i after observing evidence A is

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$

Using the total probability theorem to develop P(A) using the partition B_i we obtain

$$P(B_i \mid R) = \frac{P(R \mid B_i)P(B_i)}{\sum_j P(R \mid B_i)P(B_i)}$$

This famous result is credited to Reverend Thomas Bayes. Bayes Rule is terrifically useful in different contexts.

Standard applications of the multiplication formula, the law of total probabilities, and Bayes' theorem occur with two steps or stages. Typically, we are given the probabilities for the first stage and the conditional probabilities for the second stage, then it is possible to calculate:

- The joint probabilities for what happens at both stages applying the multiplication formula ;
- The probabilities for what happens at the second stage applying the of total probability theorem; and
- The conditional probabilities for the first stage, given what has occurred at the second stage applying Bayes' theorem.

Example 1: A screening test has 99% accuracy while the disease prevalence is 1%. what is the probability that a person has the disease given a positive test?
Let D be the event of having the disease, and + be the event of testing positive.
$P(\mathrm{D}) = 0.01, P(\overline{\mathrm{D}}) = 0.99$
$P(+|\ \mathrm{D}\ ) = 0.99, P(+|\ \overline{\mathrm{D}}\ ) = 0.01$
Applying Bayes' rule

$$P(\ \mathrm{D} \mid +\ ) = \frac{P(\mathrm{D}) \times P(+|\ \mathrm{D}\ )}{P(\mathrm{D}) \times P(+|\ \mathrm{D}\ ) + P(\overline{\mathrm{D}}) \times P(+|\ \overline{\mathrm{D}}\ )} = 0.5$$

We obtain

$$P(\ \mathrm{D} \mid +\ ) = \frac{0.01 \times 0.99}{0.01 \times 0.99 + 0.99 \times 0.01} = 0.5$$

Probability of having the disease given a positive test is 50%.

Example 2: Let us return to the darts competition of Example 00. Here B_i is the event of playing with an opponent of type i, and
P(B_1) = 0.5, P(B_2) = 0.25, P(B_3) = 0.25.
Also, A is the event of loosing, and
$P(\ A \mid B_1\ ) = 0.3, P(\ A \mid B_2\ ) = 0.4, P(\ A \mid B_3\ ) = 0.5$
Suppose that the player wins. What is the probability that he had an opponent of type 1?
Using Bayes' rule, we have

$$P(Bi \mid \bar{A}) = \frac{P(\bar{A} \mid B_1)P(B_1)}{P(\bar{A})}$$

$$P(\ Bi \mid \bar{A}\ ) = \frac{0.5 \times 0.7}{0.625} = 0.56$$

The probability that the player who wins had an opponent of type 1 is 56%.

Example 3: A teacher estimates that some of his students cheat on exams. The plagiarism detector comes up with the following accuracy:
Correctly flags 48% of cheating students (true positive rate).
Mistakenly flags 4% of honest students (false positive rate).
If 20% of the students are flagged by the detector, what is the probability that a student is mistakenly flagged (did not cheat)?
$P(B_1) = 0.2, P(B_2) = 0.8, P(\ A \mid B_1\ ) = 0.48, P(\ A \mid B_2\ ) = 0.04$
Applying Bayes' Theorem,

$$P(Bi \mid A) = \frac{P(A \mid Bi)P(Bi)}{\sum_j P(A \mid B_i)P(B_i)}$$

$$P(Bi \mid A) = \frac{0.2 \times 0.48}{0.2 \times 0.48 + 0.8 \times 0.04} = 0.128$$

The probability that a student is mistakenly flagged is 12.8%.

Chapter IV Exercises

Exercise 1:
A school club has 5 members: (Paul P, Quentin Q, Rachel R, Sophie S, and Thomas T). A committee of 2 members is chosen randomly to meet the principal. What is the probability of:
1- Selecting member Paul P.
2- Selecting both members Paul P and Sophie S.
3- Selecting one of the two members, Paul P or Sophie S.
4- Not selecting member Paul P.

Solution:
We are choosing 2 members out of a total of 5. The order in which they are chosen does not matter. Therefore, we calculate the number of combinations.

Total number of ways to choose the committee:

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

$$n = 5$$
$$k = 2$$

$$C_5^2 = \frac{5!}{2!\,(5-2)!} = 10$$

1- Probability of Selecting Member Paul P

If Paul P is selected, we need to choose the 1 remaining member from the other 4 (Quentin, Rachel, Sophie, Thomas).

Number of favorable outcomes (committees that include Paul):

$$C_1^1 \times C_4^1 = 1 \times 4 = 4$$

So,

$$P(\text{Paul}) = \frac{\text{Number of favorable outcomes}}{\text{Total possible outcomes}} = \frac{4}{10} = \frac{2}{5} = 0.4$$

2- Probability of Selecting Both Paul P and Sophie S
If both Paul and Sophie are selected, the committee is complete. There is only 1 way to choose this specific pair.

Number of favorable outcomes: 1

$$P(\text{Paul and Sophie}) = \frac{1}{10} = 0.1$$

3- Probability of Selecting One of the Two (Paul P or Sophie S)
This means exactly one of them is in the committee, but not both.

Let’s break it down:

Case 1: Paul is selected, Sophie is not.
We choose Paul, and the other member must be chosen from Quentin, Rachel, or Thomas (3 people).
Number of ways:

$$C_1^1 \times C_3^1 = 1 \times 3 = 3$$

Case 2: Sophie is selected, Paul is not.
Similarly, we choose Sophie, and the other member is chosen from Quentin, Rachel, or Thomas.
Number of ways:

$$C_1^1 \times C_3^1 = 1 \times 3 = 3$$

Total favorable outcomes: $3 + 3 = 6$

So,

$$P(\text{Paul or Sophie, but not both}) = \frac{6}{10} = \frac{3}{5} = 0.6$$

(Note: Here we face an "exclusive OR". If the question was "at least one of them", the probability would be higher, but the phrasing "one of the two" typically implies exactly one).

4- Probability of Not Selecting Member Paul P

If Paul is not selected, we must choose both members from the other 4 people (Quentin, Rachel, Sophie, Thomas).

Number of favorable outcomes:

$$C_4^2 = \frac{4!}{2!\,(4-2)!} = 6$$

So,

$$P(\text{Not Paul}) = \frac{6}{10} = \frac{3}{5} = 0.6$$

We can also use the complement rule from Part 1:

$$P(\text{Not Paul}) = 1 - \text{P(Paul)} = 1 - \frac{2}{5} = \frac{3}{5}$$

Exercise 2:

A telecommunications company offers two types of mobile plans: a "Data Plus" plan and a "Talk & Text" plan. A survey of the company's customers shows that:
60% have the "Data Plus" plan.
55% have the "Talk & Text" plan.
30% have both plans.

A customer is selected at random.

1- What is the probability that the customer has neither plan?
2- What is the probability that the customer has exactly one plan?
3- What is the probability that the customer has at most one plan?
4- What is the probability that the customer has at least one plan?

Solution:

Let D be the event that a customer has the "Data Plus" plan. $P(D) = 0.60$
Let T be the event that a customer has the "Talk & Text" plan. $P(T) = 0.55$
The event that a customer has both plans is $D \cap T$. $P(D \cap T) = 0.30$

A Venn diagram is very helpful for visualizing this.
The probability of having only the Data Plus plan:

$$P(D\ only) = P(D) - P(D \cap T) = 0.60 - 0.30 = 0.30$$

The probability of having only the Talk & Text plan:

$$P(T\ only) = P(T) - P(D \cap T) = 0.55 - 0.30 = 0.25$$

The probability of having at least one plan (the union of D and T) is:

$$P(D \cup T) = P(D) + P(T) - P(D \cap T) = 0.60 + 0.55 - 0.30 = 0.85$$

1- Probability that the customer has neither plan

This is the complement of having at least one plan.

$$P(Neither) = 1 - P(D \cup T) = 1 - 0.85 = 0.15$$

2- Probability that the customer has exactly one plan

This means the customer has "Data Plus only" OR "Talk & Text only". These are mutually exclusive events.

$$P(Exactly\ One) = P(D\ only) + P(T\ only) = 0.30 + 0.25 = 0.55$$

3- Probability that the customer has at most one plan

"At most one" means the customer has neither plan OR exactly one plan.
We can calculate this as 1 minus the probability of having both plans.

$$P(At\ most\ one) = 1 - P(D \cap T) = 1 - 0.30 = 0.70$$

Alternatively, we can add the probabilities we already found:

$$P(At\ most\ one) = P(Neither) + P(Exactly\ One) = 0.15 + 0.55 = 0.70$$

4- Probability that the customer has at least one plan

This is the union of events D and T, which we calculated at the beginning.

$$P(At\ least\ one) = P(D \cup T) = 0.85$$

Exercise 3:

A box contains 6 balls numbered from 1 to 6. Two balls are drawn simultaneously from the box. Let:
Event A: The sum of the numbers on the two balls equals 7
Event B: The absolute difference between the numbers on the two balls equals 3

- Find the conditional probability: $P(Sum\ =\ 7 \mid Difference\ =\ 3)$

Solution:

This is a conditional probability problem. We need to find:

$$P(Sum = 7 \mid Difference = 3)$$

By definition, this is:

$$\mathrm{P(Sum = 7 \mid Diff = 3)} = \frac{\mathrm{P(Sum = 7 \cap Diff = 3)}}{\mathrm{P(Diff = 3)}}$$

$$P(\text{Sum} = 7 \mid \text{Diff} = 3) = \frac{\text{Number of outcomes where Sum} = 7 \text{ AND Diff} = 3}{\text{Number of outcomes where Diff} = 3}$$

Since the balls are drawn simultaneously, the order of selection does not matter. We are dealing with combinations.

The total number of ways to draw 2 balls from 6 is given by the combination formula:

$$C_6^2 = \frac{6!}{2!\,(6-2)!} = 15$$

$Total\ Outcomes = (62) = 6 \times 52 \times 1 = 15?$

We can list all possible unordered pairs:

$\Omega = \{(1{,}2), (1{,}3), (1{,}4), (1{,}5), (1{,}6), (2{,}3), (2{,}4), (2{,}5), (2{,}6), (3{,}4), (3{,}5), (3{,}6), (4{,}5), (4{,}6), (5{,}6)\}$.

We identify outcomes where the difference is 3, i. e. we find all pairs (a, b) where |a - b| = 3.

$$|4-1| = 3 \rightarrow (1,4)$$
$$|5-2| = 3 \rightarrow (2,5)$$
$$|6-3| = 3 \rightarrow (3,6)$$

So, the number of outcomes where the difference is 3 is:

$$n(Diff = 3) = 3$$

And the successful outcomes for the condition are: $\{(1{,}4), (2{,}5), (3{,}6)\}$

Now, we identify outcomes where both conditions are met (sum=7 and diff=3). So, from the list above, we check which pairs also have a sum of 7.

$$(1,4){:}\ Sum = 1 + 4 = 5$$
$$\boldsymbol{(2,5){:}\ Sum = 2 + 5 = 7}$$
$$(3,6){:}\ Sum = 3 + 6 = 9$$

Only one pair satisfies both conditions: (2, 5).

So, the number of outcomes where both events occur is:

$$n(Sum = 7 \cap Diff = 3) = 1$$

Step 5: Calculate the Conditional Probability

We apply the conditional probability formula:

$$P(Sum = 7 \mid Diff = 3) = \frac{n(Sum = 7 \cap Diff = 3)}{n(Diff = 3)} = \frac{1}{3}$$

Exercise 4:

We have 4 boxes numbered from 1 to 4.

Box 1 contains 5 balls: 2 red and 3 blue.

Box 2 contains 5 balls: 4 red and 1 blue.

Box 3 contains 5 balls: 3 red and 2 blue.

Box 4 contains 5 balls: all 5 are red.

A box is chosen at random, and then a ball is drawn at random from that box. The drawn ball is red.

What is the probability that:

1- The ball came from Box 1?

2- The ball came from Box 2?
3- The ball came from Box 3?
4- The ball came from Box 4?

Solution:
We start by defining the events
Let B_1, B_2, B_3, and B_4 be the events that Box 1, Box 2, Box 3, and Box 4 are chosen, respectively.
Since a box is chosen at random: $P(B_1) = P(B_2) = P(B_3) = P(B_4) = \frac{1}{4}$

Let R be the event that a red ball is drawn.

Then, we find the probability of drawing red from each box. These are conditional probabilities.
From Box 1: $P(R \mid B1) = \frac{2}{5}$
From Box 2: $P(R \mid B2) = \frac{4}{5}$
From Box 3: $P(R \mid B3) = \frac{3}{5}$
From Box 4: $P(R \mid B4) = \frac{5}{5} = 1$

We Calculate the total probability of drawing a red ball, P(R)
We use the Law of Total Probability:

$$P(A) = \sum_i P(B_i)P(A \mid B_i)$$

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + P(B_3)P(A \mid B_3) + P(B_4)P(A \mid B_4)$$

$$P(A) = (\frac{1}{4} \times \frac{2}{5}) + (\frac{1}{4} \times \frac{4}{5}) + (\frac{1}{4} \times \frac{3}{5}) + (\frac{1}{4} \times 1)$$

$$P(A) = \frac{2}{20} + \frac{4}{20} + \frac{3}{20} + \frac{5}{20} = \frac{14}{20} = \frac{7}{10}$$

Finally, we apply Bayes' Theorem for each question

Bayes' Theorem states:

$$P(B_i \mid R) = \frac{P(R \mid B_i)P(B_i)}{P(R)}$$

1- Probability the ball came from Box 1

$$P(B_1 \mid R) = \frac{P(R \mid B_1)P(B_1)}{P(R)}$$

$$P(B_1 \mid R) = \frac{\frac{1}{4} \times \frac{2}{5}}{\frac{7}{10}} = \frac{\frac{2}{20}}{\frac{7}{10}} = \frac{1}{7}$$

2- Probability the ball came from Box 2

$$P(B_2 \mid R) = \frac{P(R \mid B_2)P(B_2)}{P(R)}$$

$$P(B_2 \mid R) = \frac{\frac{1}{4} \times \frac{4}{5}}{\frac{7}{10}} = \frac{\frac{4}{20}}{\frac{7}{10}} = \frac{2}{7}$$

3- Probability the ball came from Box 3

$$P(B_3 \mid R) = \frac{P(R \mid B_3)P(B_3)}{P(R)}$$

$$P(B_3 \mid R) = \frac{\frac{1}{4} \times \frac{3}{5}}{\frac{7}{10}} = \frac{\frac{3}{20}}{\frac{7}{10}} = \frac{3}{14}$$

4- Probability the ball came from Box 4

$$P(B_4 \mid R) = \frac{P(R \mid B_4)P(B_4)}{P(R)}$$

$$P(B_4 \mid R) = \frac{\frac{1}{4} \times 1}{\frac{7}{10}} = \frac{\frac{1}{4}}{\frac{7}{10}} = \frac{5}{14}$$

Exercise 5:
Suppose the probability of rain is 20%, and the probability of a traffic accident is 10%. Suppose further that the conditional probability of an accident, given that it rains, is 40%. What is the conditional probability that it rains, given that there is an accident?

Solution:
We start by defining the events
Let R be the event that it rains. $P(R) = 0.20$
Let A be the event that there is a traffic accident. $P(A) = 0.10$
Conditional probability of an accident given that it rains $P(A \mid R) = 0.40$

We are asked to find $P(R \mid A)$, the conditional probability that it rained, given that there was an accident.
We apply Bayes' Theorem, which provides the formula for reversing conditional probabilities:

$$P(R \mid A) = \frac{P(A \mid R)P(R)}{P(A)}$$

Plug the known values into the formula:

$$P(R \mid A) = \frac{(0.40) \cdot (0.20)}{0.1} = 0.8$$

Chapter V: Discrete Random Variables and Their Probability Distributions

In Chapter IV, we established the probability model as the foundational framework of probability theory, centered on defining a probability measure P over subsets of the sample space. However, practical applications often require more efficient methods for characterizing probability assignments than working directly with P.
This chapter introduces powerful analytical tools including Random Variables and their associated Distribution Functions, Probability Mass Functions (for discrete cases), Probability Density Functions (for continuous cases)
While previous chapters covered the construction of probability models (sample space S and probability measure P), we now build upon this foundation by defining random variables that map these abstract models to concrete, analyzable quantities. This transition from abstract measures to functional representations marks a crucial step in applying probability theory to practical problems.

1. Random Variables

1.1 Definition of a Random Variable

The basic outcomes of an experiment may be nonnumerical (e.g., "heads" or "tails") or quantitative (e.g., "rolling a die"). A random variable assigns a unique numerical value to each outcome (sample point) of a random experiment. Its value is determined by the result of the trial, translating outcomes into a measurable numerical form.
A random variable is a real-valued function that assigns a real numerical value to each outcome in a sample space of a random experiment.
In practice, it is often useful to convert random outcomes numerically. When an outcome is both numerical and one-dimensional, we classify it as a random variable. The latter are known as a cornerstone of probability theory.
Examples:
- Nonnumerical outcome: “Coin flip = Tails”
- Numerical (quantitative) outcome: “six-sided die roll = 3”

Consider a coin toss with outcomes H (heads) or T (tails). We can represent this numerically by defining a random variable X such that:
$X = 1$ if the outcome is H,
$X = 0$ if the outcome is T.
Since the value of X depends on the random outcome of the coin flip, X is itself random. In the coin flip example the function is

$$X = \begin{cases} 1 & if\ H \\ 0 & if\ T \end{cases}$$

A simple example is a coin toss: we can record heads as 1 and tails as 0. Since the outcome can only be 0 or 1, this represents a basic case of a discrete distribution.

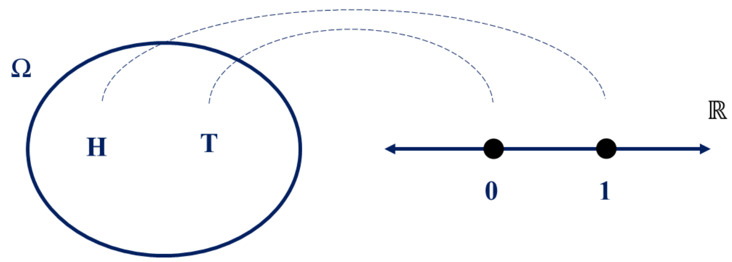


Figure 05: A Random Variable is a Function

By convention, it is useful to distinguish between random variables and their realizations. In probability theory and statistics, upper case letters (e.g. X, Y…) are used to indicate a random variable and lower case letters (e. g. x, y…) are used to indicate its realization or specific value.

Example:
X: Number of heads in 3 coin tosses (discrete).
Y: Time until a light bulb fails (continuous).

1.2 Types of Random Variables (RV)
There are two main types of random variables:

Discrete Random Variable (DRV)

Takes on a countable number of distinct values (e.g., rolling a die, integers).

Continuous Random Variable (CRV)

Takes on uncountable values i.e. an infinite number of values within a given interval (e.g., measuring temperature, real numbers).

Random variables are either discrete (sample points can be listed in order) or continuous (sample points cannot be fully enumerated due to their infinite density). This distinction is central to their probability analysis.
For a continuous random variable, the number of possible sample points is always infinite, and these points are uncountable—between any two points, there exists an infinite number of additional points, making it impossible to list them sequentially.
In contrast, a discrete random variable may have either a finite or infinite number of sample points. Even if infinite, these points can be enumerated in order (though listing them all would take infinite time). This distinguishes discrete variables from continuous ones, where such ordered enumeration is impossible.

2. Discrete Random Variables

2.1 Definition
A discrete random variable is a random variable that can take on a finite or countably infinite set of distinct values, resulting from counting processes. (e. g. the number of tails in 50 coin tosses).

2.2 Probability Mass Function (PMF)
The probability distribution for a discrete random variable X associates with each of the distinct outcomes x_i, ($i = 1,2,3,\dots,k$) a probability $P(X = x_i)$. It is called the Probability Mass Function (PMF) or the probability function. It is a function that gives the probability that X takes the value xi. Below is the probability distribution table for the discrete random variable.

X_i	x_1	x_2	x_3	…	x_n	$\sum$
$P(X = x_i)$	P_1	P_2	P_3	…	P_n	1

Where
$0 \leq P(X = x_i) \leq 1$ for all x_i
All probabilities must sum to 1: $\sum_x P(X = xi) = 1$

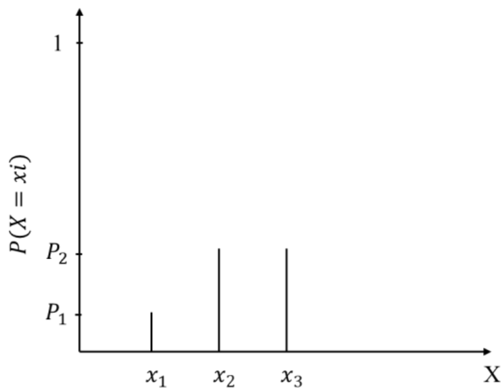
Graphical Representation of a Probability Mass Function (PMF)

A Probability Mass Function describes the probability distribution of a Discrete Random Variable (DRV), and assigns probabilities to specific individual values. The graphical representation typically uses bars (like a histogram or a stick diagram) to show probabilities at each possible value and emphasize discrete jumps, it can also use dots with lines (especially for infinite DRVs like Poisson).

Fundamental components of a PMF Graph

The X-axis: shows each possible outcome x_i of the random variable *X* (e. g. the number of heads in 3 coin flips: $x = \{0,1,2,3\}$, rolling a die: $x = \{1,2,3,4,5,6\}$).

The Y-axis: the probabilities corresponding to each possible outcome *X*, where the height of the bar at xi represents $P(X = x_i)$.
Ensure that all probabilities must sum to 1: $\sum_x P(X = x_i) = 1$
Each possible value of *X* is represented by a vertical bar whose height equals the probability $P(X = x_i)$, as follows:



2.3 Cumulative Distribution Function (CDF) or Cumulative Probability Function
The cumulative probability distribution or cumulative probability function for a discrete random variable *X*, denoted *F(X)*, and provides the probability that *X* will be at or below any given value of x_i, i.e. $P\ (X \leq x_i)$ for all x_i.
The CDF of a discrete random variable *X* is defined as:
$F(x) = P\ (X \leq x_i) = \sum_{t \leq x} P(X = t)$. It is a non-decreasing step function.
It is a non-decreasing function, which is displayed in a table like the one below.

X_i	x_1	x_2	x_3	…	x_n	$\sum$
$P(X = x_i)$	P_1	P_2	P_3	…	P_n	1
$F(x)$	P_1	$P_1 + P_2$	$P_1 + P_2 + P_3$	…	$P_1 + P_2 + P_3 + \ldots + P_n = 1$	

Or takes the following form

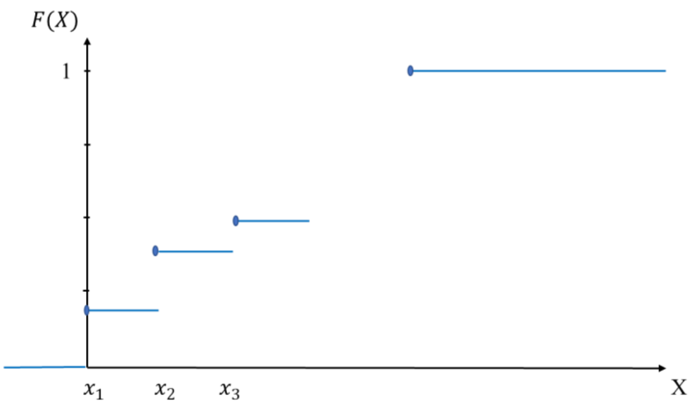
$$F(x) = \begin{cases} 0 & if\ x < x_1 \\ P_1 & if\ x_1 \leq x < x_2 \\ P_1 + P_2 & if\ x_2 \leq x < x_3 \\ P_1 + P_2 + P_3 & if\ x_3 \leq x < x_4 \\ \quad \cdots & \\ P_1 + P_2 + P_3 + \ldots + P_{n-1} & if\ x_{n-1} \leq x < x_n \\ 1 & if\ x \geq x_n \end{cases}$$

Properties

The CDF is always bounded between 0 and 1, i.e., $0 \leq F(X) \leq 1$.
The CDF is a non-decreasing function.
The CDF $F(X)$ approaches 0 as X tends to $-\infty$ and approaches 1 as X tends to $+\infty$.

Graphical Representation of a Cumulative Distribution Function (CDF)

The graph of a CDF provides a visual way to understand how probabilities accumulate across the range of possible values.
For discrete random variables, the cumulative distribution function (CDF) graph exhibits a step-like or staircase pattern. It jumps at every possible value of X, where the magnitude of each jump corresponds to the probability $P(X = x_i)$. Additionally, the CDF is right-continuous, which is visually represented by a solid dot at the top of each jump.



Properties of the graph

The function "jumps" at each possible value of x, and the size of the jump at x is equal to $P(X = x_i)$.
The function is constant between possible values.
The filled circles (●) indicate the value of the function at the point x.
The function is continuous from the right, it starts at 0 and ends at 1.

2.4 Expected Value, Variance and standard deviation

2.4.1 The Expected Value (Mean)

Definition of the Expected Value

The expected value of a discrete random variable X is also known as its mean value, and denoted by $E(X)$ and defined

$$E(X) = \sum_{i=1}^{k} \mathrm{x}_i \mathrm{P}(\mathrm{X} = \mathrm{x}_i)$$

$E(X)$ is a weighted average of the possible outcomes with the probability values as weights. That is why, it is called the mean of the probability distribution of X.
The mean or expected value is a number that does not correspond to any particular outcome.

Properties of Mathematical Expectation

The expectation of a constant a is the constant itself: $E(a) = a$.
The expectation of an expectation is the expectation itself:
$E(E(X)) = E(X)$.
If X is a random variable and a, b are constants, then:
$E(aX) = a \times E(X)$,
$E(aX \pm b) = a \times E(X) \pm b$.
If X and Y are random variables, then
The expectation of their sum is the sum of their expectations:
$E(X + Y) = E(X) + E(Y)$.
The expectation of their difference is the difference of their expectations:
$E(X - Y) = E(X) - E(Y)$.
Then the expectation of their product is the product of their expectations:
$E(X \times Y) = E(X) \times E(Y)$.

2.4.2 The Variance

Definition of the variance

The variance is calculated as a weighted average of the squared deviations of X's outcomes from their expected value, with each deviation weighted by its probability of occurrence. In this way, it quantifies how much the outcomes of X diverge from their expected value—similar to how the variance of quantitative variables measures the spread of values around their mean.
The variance of a discrete random variable X is denoted by $Var(X)$ and defined as

$$Var(X) = \sum_{i=1}^{k} (\mathrm{x}_i - \mathrm{E(X)})^2 \mathrm{P}(\mathrm{x}_i) = E\left[\left(\mathrm{x}_i - E(X)\right)^2\right] = E(X^2) - [E(X)]^2$$

Properties of Variance

The variance of a constant a is zero:

$$V(a) = 0$$

If X is a random variable and a is a constant, then:

$$V(aX) = a^2 V(X)$$

If X is a random variable and a and b are constants, then:

$$V(aX \pm b) = a^2 V(X)$$

If X and Y are random variables, then
The variance of their sum is the sum of their variances:

$$V(X + Y) = V(X) + V(Y)$$

The variance of their difference is the difference of their variances:

$$V(X - Y) = V(X) + V(Y)$$

2.4.3 The standard deviation

Since the variance is simply the expected value of the squared deviations of the values from their mean. The standard deviation, denoted by σ, is defined as the square root of the variance. The standard deviation is denoted by σ and defined as

$$\sigma = \sqrt{Var(X)}$$

2.5 Probability calculation for a discrete random variable

Calculating probabilities for a discrete random variable involves using the following

$P(x \leq a) = F(a)$,
$P(X<a) = P(x \leq a-1) = F(a-1)$,
$P(X>a) = 1 - P(x \leq a) = 1 - F(a)$,
$P(X \geq a) = 1 - P(X<a) = 1 - F(a-1)$,
$P(a<X<b) = P(X<b) - P(X \leq a) = F(b-1) - F(a)$,
$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-1)$,
$P(a \leq X < b) = P(X < b) - P(X < a) = F(b-1) - F(a-1)$,
$P(a<X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.

Example: Let the discrete random variable *X* represent the number of heads obtained when a fair coin is tossed twice.
Probability Mass Function (PMF)

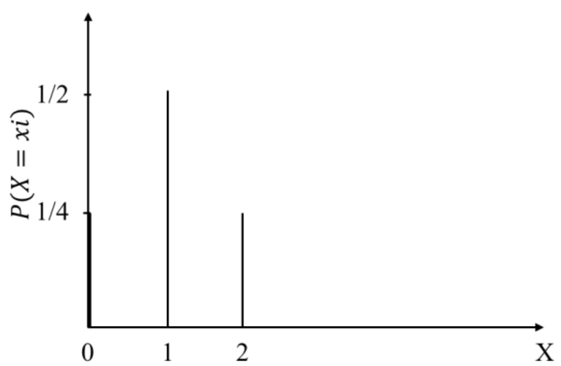
The possible values of *X* are {0,1,2}. The PMF is:

$$P(X=0) = P(TT) = \frac{1}{4}$$
$$P(X=1) = P(HT, TH) = \frac{2}{4} = \frac{1}{2}$$
$$P(X=2) = P(HH) = \frac{1}{4}$$

The table takes the form:

xi	0	1	2	$\sum$
$P(X=x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

Graphical Representation of the PMF
The PMF can be graphically represented using a probability histogram or stick diagram:

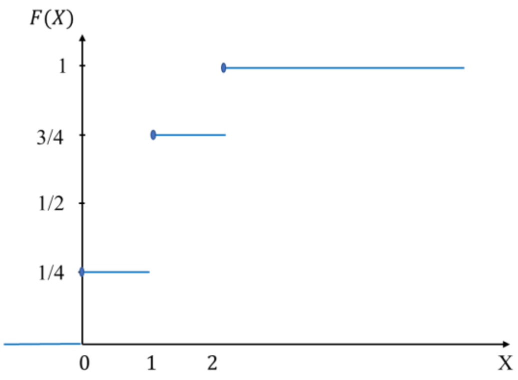


The CDF

xi	0	1	2	$\sum$
$P(X=x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1
$F(x)$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{4}{4}=1$	

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ \frac{1}{4} & if\ 0 \leq x < 1 \\ \frac{3}{4} & if\ 1 \leq x < 2 \\ 1 & if\ x \geq 2 \end{cases}$$

The CDF representation:



The Expected Value (Mean):

$$E(X) = \sum_{i=1}^{k} \mathrm{x}_i \mathrm{P}(\mathrm{X} = \mathrm{x}_i)$$

xi	0	1	2	$\sum$
$P(X=x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1
$E(X)$	0	$\frac{2}{4}$	$\frac{2}{4}$	1

$$E(X) = \sum_{i=1}^{k} \mathrm{x}_i \mathrm{P}(\mathrm{X} = \mathrm{x}_i)$$

$$E(X) = 1$$

The Variance

$$Var(X) = E(X^2) - [E(X)]^2$$

xi	0	1	2	$\sum$
$P(X=x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1
X^2	0	1	4	

$E(X^2)$	0	$\frac{2}{4}$	$\frac{4}{4}$	$\frac{6}{4}$	$=\frac{3}{2}=1.5$

$$Var(X) = 1.5 - [1]^2 = 0.5$$

The standard deviation

$$\sigma = \sqrt{Var(X)}$$
$$\sigma = \sqrt{0.5} \approx 0.707$$

3. Common Discrete Distributions

3.1 Uniform Probability Distributions

3.1.1 Introduction

The Uniform probability distribution or rectangular probability distribution is one of the simplest probability distributions, resulting when the probabilities of each outcome in the sample space are equal. Such probability distributions may be either discrete or continuous. The discrete uniform distribution describes experiments with a finite n number of outcomes $(x_1, x_2, \dots, x_n)$ and the same probability for each outcome $\frac{1}{n}$. (e. g. Rolling a fair six-sided die ($n = 6$), Drawing a card from a well-shuffled deck ($n = 52$)).

3.1.2 Probability Mass Function (PMF)

The PMF of a discrete uniform random variable *X* is given by:

$$P(X = x_i) = \begin{cases} \frac{1}{n} & if\ x \in \{x_1, x_2, \dots, x_n\} \\ 0 & otherwise \end{cases}$$

Meaning that $P(X = 1) = P(X = 2) = \cdots = P(X = n)$

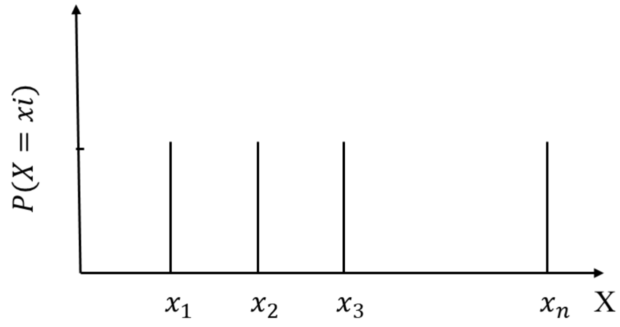
The probability distribution for a uniform discrete random variable is defined by its possible values and their corresponding probabilities, typically presented in a table format as shown below.

x_i	x_1	x_2	x_3	…	x_n	$\sum$
$P(X = x_i)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	…	$\frac{1}{n}$	1

The discrete uniform distribution is defined for a finite number of outcomes, each with equal probability. It has two fundamental properties:
The probability for each outcome is non-negative: $Pi \geq 0$.
The sum of all probabilities equals 1: $\sum Pi = 1$.

Discrete Uniform Distribution PMF Representation

A graph of a discrete uniform probability distribution is shown in the Figure after.



3.1.3 Cumulative Distribution Function (CDF)

The CDF of a discrete uniform probability distribution gives $P(X \leq x_i)$:

$$F(x) = \begin{cases} 0 & if\ x < x_1 \\ \frac{k}{n} & if\ x_k \leq x < x_{k+1} \\ 1 & if\ x \geq x_n \end{cases}$$

3.1.4 Expected Value (Mean) and Variance

Expected Value (Mean)

$$E(X) = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{n+1}{2}$$

Variance

$$Var(X) = \frac{n^2 - 1}{12}$$

Example: Consider a random number generator that produces integers uniformly distributed between 1 and 10. For the random variable *X*, perform the following:

1- State the distribution law and provide its graphical representation.
2- Calculate the expected value $E(X)$, the variance $V(X)$, and the standard deviation $\sigma(X)$.

Solution:

1- The probability distribution and its graphical representation

The random variable *X* follows a Discrete Uniform Distribution.
The generator produces integers from 1 to 10, and each integer is equally likely to be selected.
The probability that any one of the 10 numbers will be turned up is $1/10$ or 0.1.

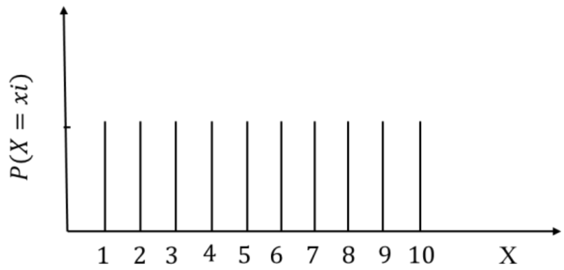
$$P(X = x_i) = \frac{1}{10} \qquad for\ x = 1,2,3,\ldots,10$$

The probability distribution for this process is therefore

x_i	1	2	3	4	5	6	7	8	9	10	$\sum$
$P(X = x_i)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1

Graphical Representation

It can be represented by a probability histogram where each value from 1 to 10 has a bar of equal height.



$E(X) = \frac{10+1}{2} = 5.5.$

$Var(X) = \frac{10^2-1}{12} = \frac{99}{12} = 8.25$

$\sigma = \sqrt{8.25} \approx 2.872.$

In statistical modelling, the discrete uniform distribution is indispensable for representing equally likely events. Its theoretical properties make it vital for more complex distributions, and designing unbiased sampling protocols across research domains.

3.2 Bernoulli Distribution

3.2.1 Introduction

The Bernoulli distribution is the simplest discrete probability distribution, modeling only a single trial (a single experiment) with exactly two possible binary outcomes (yes/no, true/false, heads/tails):

The first one representing the success (usually coded as 1) with probability p.

And the second one representing the failure (usually coded as 0) with probability $q = 1 - p$.

(e. g. Coin flip (Heads = 1, Tails = 0); Medical test (Positive = 1, Negative = 0); Quality control (Defective = 1, Non-defective = 0)).

3.2.2 Probability Mass Function (PMF)

The Probability Mass Function of a Bernoulli random variable *X* is given by:

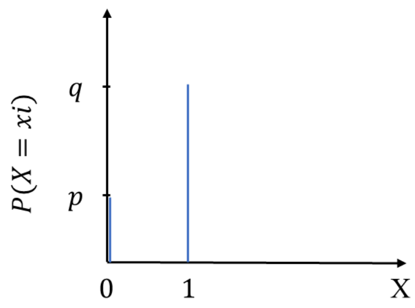
$$P(X = x_i) = \begin{cases} p & if\ x = 1(success) \\ 1-p & if\ x = 0(failure) \end{cases}$$

Compact Form

$$P(X = x_i) = p^x(1-p)^{1-x}, where\ x \in \{0, 1\}$$

x_i	0	1	$\sum$
$P(X = x_i)$	q	p	1

The probability distribution for a Bernoulli random variable *X* is shown in the Figure after.



3.2.3 Cumulative Distribution Function (CDF)

The Cumulative Distribution Function of a Bernoulli random variable X gives $P(X \leq x_i)$:

$$P(X = x_i) = \begin{cases} 0 & if\ x < 0 \\ 1 - p & if\ 0 \leq x < 1 \\ 1 & if\ x \geq 1 \end{cases}$$

3.2.4 Expected Value (Mean) and Variance

Expected Value (Mean)

$$E(X) = \sum_x xP(X = x_i) = (1.p) + \big(0.(1 - p)\big) = p$$

Variance

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

Standard Deviation

$$\sigma_x = \sqrt{Var(X)} = \sqrt{p(1 - p)}$$

Example 1:
A biased coin has $P(Heads) = 0.6$. Let X be the random variable representing the outcome ($Heads = 1, Tails = 0$).

- Calculate E(X) and Var(X).

Solution:
This example demonstrates a Bernoulli distribution with success probability $p = 0.6$, where the random variable X takes value 1 with probability 0.6 and value 0 with probability 0.4.

$$X \sim Bernoulli(p)$$

Probability Mass Function:

$$P(X = 1) = 0.6, P(X = 0) = 0.4$$

Expected Value:

$$E(X) = \sum x \cdot P(X = x) = (1)(0.6) + (0)(0.4) = 0.6$$

Variance:

$$Var(X) = p(1-p) = 0.6 \times 0.4 = 0.24$$

Example 2:

In a medical test, the probability of a false positive is 0.05. Let $X = 1$ if the test is positive, and $X = 0$ if the test is negative.

1- Compute $P(X = 0)$.

2- What is the standard deviation?

Solution:

1- Computing $P(X = 0)$

Probability of false positive: $P(positive \mid actually\ negative) = 0.05$

But we need $P(X = 0)$, the probability the test is negative

We assume the test can give either positive or negative results, and the total probability must sum to 1.

Let $p = P(X = 1)$ be the probability of a positive test.

Since we're only given the false positive rate but not the disease prevalence, we'll assume this refers to:

$P(X = 1) = 0.05$ (probability of positive test)

Therefore: $P(X = 0) = 1 - P(X = 1) = 1 - 0.05 = 0.95$

2- The standard deviation

Probability Distribution:

$$X \sim Bernoulli(p = 0.05)$$

$$P(X = 1) = 0.05, P(X = 0) = 0.95$$

Variance for Bernoulli distribution:

$$Var(X) = p(1-p) = 0.05 \times 0.95 = 0.0475$$

Standard Deviation:

$$\sigma = \sqrt{Var(X)} = \sqrt{0.0475} \approx 0.2179$$

Bernoulli distribution is fundamental in probability theory, it provides the theoretical underpinning for derived complex distributions such as the Binomial and Geometric distributions.

3.3 Binomial Probability Distributions

3.3.1 Definition

The Bernoulli random trials are experiments in which there are two basic outcomes of a qualitative nature (e.g. the coin comes up either heads or tails, the experiment either confirms the hypothesis or it doesn't, the train is either on time or it's late …).

The binomial probability distribution is a discrete probability distribution since X can only take the discrete values $0, 1, \ldots, n$. It assigns arbitrarily one outcome the value 0 and the other the value 1. This random variable, $X_i = \{0, 1\}$ is called a Bernoulli random variable. Usually we are interested in a fixed number (n) of independent trials, where each trial has two possible outcomes: success (probability p) or failure (probability $q = 1 - p$), that remain constant for all trials.

Notation: $X \sim Binomial(n, p)$

In practical applications, the binomial model requires specifying both the trial count (n) and success probability (p), generating an entire class of distributions where each (n, p) combination produces different probability patterns for count outcomes (e. g. Number of heads in 20 coin

flips; Number of defective items in a batch of 100; Number of patients responding to a treatment in a clinical trial).
The process satisfies independence when all X_{i} are probabilistically unrelated, and stationarity when each binary outcome $X_{\mathrm{i}} \in \{0,1\}$ follows the same fixed Bernoulli distribution across all trials.

3.3.2 Probability Mass Function (PMF)

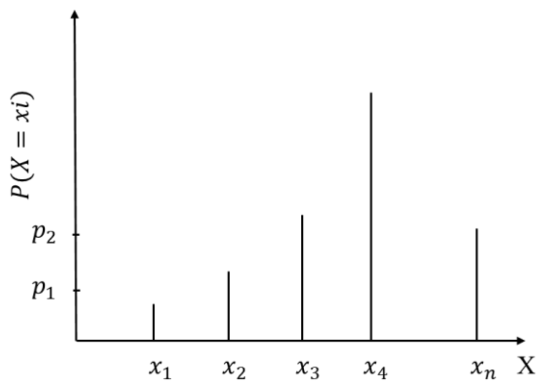
The PMF gives the probability of getting exactly x_i successes in n trials:
The Binomial Probability Mass Function, which gives the probabilities that *X* will get exactly x_i successes in n trials, is

$$P(x_i) = P(X = x_i) = \binom{n}{x} p^x (1-p)^{n-x} = C_n^x p^x (1-p)^{n-x}$$

where $P(x_i) = P(X = x_i)$
$x = 0, 1, 2, \ldots, n$, and
$0 \leq p \leq 1$. The parameter p is the probability that $X_i = 1$. It is the same for all *i* because the Bernoulli random variables X_i are identically distributed.
$C_n^x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient

The probability distribution for a Binomial random variable $X \sim Binomial(n,p)$ can be presented in a table. The table lists the possible values *x* (from 0 to *n*) and their corresponding probabilities, calculated using the formula $P(X = x_i) = C_n^x p^x (1-p)^{n-x}$. The sum of all probabilities in the table equals 1.

x_i	0	1	…	n	$\sum$
$P(X = x_i)$	$C_n^0 p^0 (1-p)^n$	$C_n^1 p^1 (1-p)^{n-1}$	…	$C_n^n p^n (1-p)^{n-n}$	1



The shape of the distribution depends on parameters *n* and *p*:
When $p < 0.5$: the distribution is skewed right
When $p > 0.5$: the distribution is skewed left
When $p = 0.5$: the distribution is symmetrical
As n increases, it approaches normal distribution (Central Limit Theorem)

3.3.3 Cumulative Distribution Function (CDF)

The CDF gives the probability of getting up to xi successes:

$$\mathrm{F(X)} = \mathrm{P(X} \leq \mathrm{x}_i) = C_n^i p^i (1-p)^{n-i}$$

Properties

$$F(n) = 1$$

Can be calculated using statistical tables or software

3.3.4 Expected Value and Variance

Mean (Expected Value)

The mean of the binomial distribution is

$$E(X) = n \times p$$

Variance

$$Var(X) = n \times p \times (1 - p) = n \times p \times q$$

Standard Deviation

$$\sigma = \sqrt{Var(X)} = \sqrt{n \times p \times (1 - p)} = \sqrt{n \times p \times q}$$

Example:

A salesperson has a 30% chance of making a sale per call. They make 10 calls.

1- Probability of exactly 4 sales?
2- Probability of at most 2 sales?
3- Expected number of sales?

Solution:

$$X \sim Binomial(n = 10, p = 0.3)$$
$$P(X = k) = C_{10}^k (0.3)^k \, (0.7)^{10-k}$$

1- Probability of exactly 4 sales

$$P(X = 4) = C_{10}^4 (0.3)^4 \, (0.7)^6 = 210 \times 0.0081 \times 0.117649 \approx 0.2001$$

2- Probability of at most 2 sales

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$P(X = 0) = C_{10}^0 (0.3)^0 \, (0.7)^{10} = 1 \times 1 \times 0.028248 \approx 0.0282$$

$$P(X = 1) = C_{10}^1 (0.3)^1 \, (0.7)^9 = 10 \times 0.3 \times 0.040354 \approx 0.1211$$

$$P(X = 2) = C_{10}^2 (0.3)^2 \, (0.7)^8 = 45 \times 0.09 \times 0.057648 \approx 0.2335$$

$$P(X \leq 2) = 0.0282 + 0.1211 + 0.2335 \approx 0.3828$$

3- Expected number of sales

$$E(X) = n \times p = 10 \times 0.3 = 3$$

The binomial distribution essential when analyzing processes with binary outcomes, making it indispensable in statistics. This discrete probability distribution models scenarios characterized by a fixed number of independent trials, each with an identical probability of success, and a discrete number of successes. Fully specified by its two parameters: n -the number of trials- and p -the probability of success in each trial-.

3.4 Geometric Distribution

3.4.1 Definition

The Geometric Distribution is a discrete probability distribution that models the number of trials needed to achieve the first success or the number of failures before the first success in a sequence of independent Bernoulli trials (experiments with two outcomes: success or failure).

$$x \sim Geometric(p)$$

3.4.2 Probability Mass Function (PMF):

$$P(X = k) = (1 - p)^{k-1}p$$

where:
k is the trial number of the first success ($k = 1,2,3, \ldots$)
p is the probability of success on each trial (on a single trial), i. e. the probability that the K^{th} trial is a success.
$(1 - p)^{k-1}$ is the probability that the first $k - 1$ trials are failures.

3.4.3 Expected value, variance and standard deviation

Mean (Expected Value)

$$E(X) = \frac{1}{p}$$

Variance

$$Var(X) = \frac{1 - p}{p^2}$$

Standard deviation

$$\sigma(X) = \frac{\sqrt{1 - p}}{p^2}$$

Example:
We roll a fair six-sided die repeatedly until we get a 6. Let X be the number of rolls needed to get the first 6.

1- What's the probability we get our first 6 on roll 3?
2- What's the probability we get our first 6 on roll 1?
3- What's the probability we need more than 4 rolls?

Solution:
This follows a geometric distribution because:
We count the number of trials until the first success
The probability of success is constant: $p = \frac{1}{6}$
The trials are independent

Notation: $X \sim Geometric(p = \frac{1}{6})$
For geometric distribution:

$$P(X = k) = (1 - p)^{k-1}p$$

where:
$k = 1,2,3, \ldots$ (number of trials until first success)
$p = \frac{1}{6}$ (probability of success on each trial)

The probability we get our first 6 on roll 3

$$P(X = 3) = \left(\frac{5}{6}\right)^2 \times \frac{1}{6} = \frac{25}{36} \times \frac{1}{6} = \frac{25}{216} \approx 0.1157$$

The probability we get our first 6 on roll 1

$$P(X = 1) = \left(\frac{5}{6}\right)^0 \cdot \frac{1}{6} = \frac{1}{6} \approx 0.1667$$

The probability we need more than 4 rolls to get 6

$$P(X > 4) = (1 - p)^4 = \left(\frac{5}{6}\right)^4 = \frac{625}{1296} \approx 0.4823$$

4- Expected value

For geometric distribution:

$$E(X) = \frac{1}{p} = \frac{1}{\frac{1}{6}} = 6$$

On average, we need 6 rolls to get our first 6.

3.5 Poisson Distribution

3.5.1 Definition

The Poisson Distribution is a discrete probability distribution that models the number of rare events occurring in a fixed interval of time or space, given a constant average rate λ.

$$X \sim Poisson(\lambda)$$

3.5.2 Probability Mass Function (PMF)

$$\mathrm{P}(\mathrm{X} = \mathrm{x}_i) = \frac{\lambda^x e^{-\lambda}}{\mathrm{x}!}$$

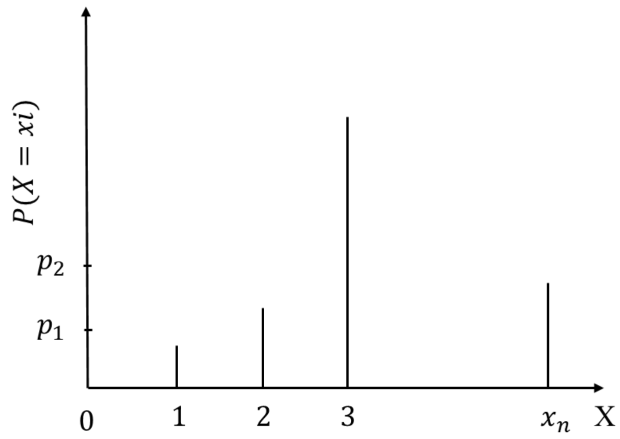
λ is the average number of events per interval, i. e. the average rate (mean and variance).
$e^{-\lambda}$ ensures the total probability sums to 1.

$x!$ accounts for the number of ways to arrange k events.
x the number of occurrences.

The probability distribution for a Poisson random variable $X \sim Poisson(\lambda)$ is given by the formula $P(X = x_i) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, 3, \ldots$. The sum of all these probabilities equals 1. It is illustrated in the following table.

x_i	0	1	…	∞	$\sum$
$P(X = x_i)$	$\frac{\lambda^0 e^{-\lambda}}{0!}$	$\frac{\lambda^1 e^{-\lambda}}{1!}$	…	$\frac{\lambda^\infty e^{-\lambda}}{\infty!}$	1

PMF Graphical representation



3.5.3 Expected value, variance and standard deviation

Mean and Variance

$$E(X) = Var\ (X) = \lambda$$

Standard deviation

$$\sigma(X) = \sqrt{Var\ (X)} = \sqrt{\lambda}$$

Example:
Cars enter a parking garage at an average rate of 4 cars per hour. Let X be the random variable representing the number of cars entering the garage each hour.

1- Find the probability distribution of X
2- Calculate the variance

Solution:

1- The probability distribution of X

This scenario follows a Poisson distribution because:
We're counting the number of events (cars arriving) in a fixed time interval (1 hour)
Events occur independently at a constant average rate
The probability of more than one car arriving in an infinitesimal time interval is negligible

Parameters:
Average rate: $\lambda = 4$ cars per hour

Probability Mass Function (Poisson):

$$P(X = x_i) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^x e^{-4}}{x!}$$

for $x = 0,1,2,3, ...$

X (cars)	0	1	2	3	4	5	6	…	$\sum P(X = x)$
$P(X = x_i)$	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	…	$\approx 1 \; as \; x \to \infty$

2- Calculating the variance

For a Poisson distribution, the variance equals the mean:

$$Var(X) = \lambda = 4$$

3.6 Hypergeometric distribution

3.6.1 Definition

The hypergeometric distribution is a discrete probability distribution that describes the probability of k successes in n draws, without replacement, from a finite population of size N that contains exactly K success states.

$$X \sim Hypergeometric(N, K, n)$$

3.6.2 Probability Mass Function (PMF)

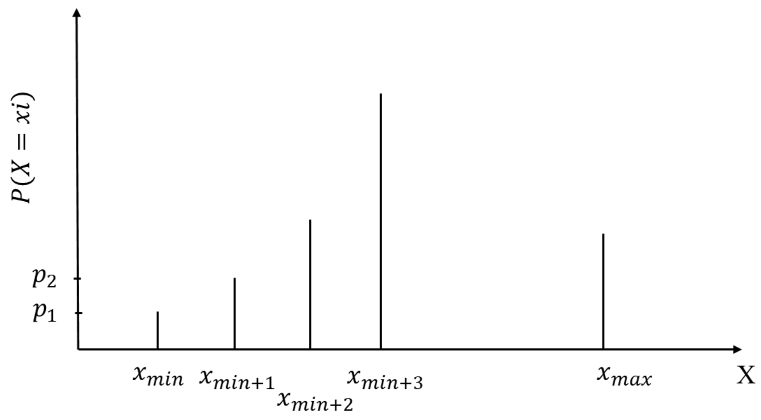
$$P(X = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}$$

Where:
N is the population size
K is the number of success states in the population
n is the number of draws (sample size)
k is the number of observed successes

The hypergeometric distribution can be summarized in the following table:

k (number of successes)	k_{min}	k_{min+1}	…	k_{max}	$\sum$
$P(X = k)$	$\frac{C_K^k C_{N-K}^{n-k}}{C_N^n}$	$\frac{C_K^k C_{N-K}^{n-k}}{C_N^n}$		$\frac{C_K^k C_{N-K}^{n-k}}{C_N^n}$	1

PMF Graphical representation



3.6.3 Expected value, variance and standard deviation

Mean

$$E(X) = n \times p = n \times \frac{K}{N}$$

Variance

$$Var(X) = n \times p \times q \frac{N-n}{N-1} = n \times \frac{K}{N} \times \frac{N-K}{N} \times \frac{N-n}{N-1}$$

Standard Deviation

$$\sigma = \sqrt{n \times p \times q \frac{N-n}{N-1}} = \sqrt{n \times \frac{K}{N} \times \frac{N-K}{N} \times \frac{N-n}{N-1}}$$

Range of k: $max(0, n + K - N) \leq k \leq min(n, K)$
So, $k \in [max(0, n + K - N), min(n, K)]$

$\frac{N-n}{N-1}$: is called the Finite Population Correction (FPC) factor. When the sample size n is very small compared to the population size *N*, this factor approaches 1.

Example:
A box contains 10 good items and 5 defective items. We randomly draw 7 items without replacement.
Let X be the random variable representing the number of good items obtained.

1- Find the probability distribution of X
2- Calculate the expected value of good items obtained

Solution:

1- The probability distribution of X

This follows a hypergeometric distribution with:
Population size: $N = 15$
Number of successes in population: $K = 10$ (good items)
Sample size: $n = 7$

Number of observed successes: k (good items in sample)

Probability Mass Function:

$$P(X = k) = \frac{C_{10}^{k} C_{5}^{7-k}}{C_{15}^{7}}$$

for $k = \{2,3,4,5,6,7\}$

The minimum k=2 because we're drawing 7 items from 15 total, and there are only 5 defective items available

For $k = 2$:

$$P(X = 2) = \frac{C_{10}^{2} C_{5}^{7-2}}{C_{15}^{7}} = \frac{C_{10}^{2} C_{5}^{5}}{C_{15}^{7}} = \frac{45 \times 1}{6435} = \frac{45}{6435} \approx 0.0070$$

For $k = 3$:

$$P(X = 3) = \frac{C_{10}^{3} C_{5}^{7-3}}{C_{15}^{7}} = \frac{C_{10}^{3} C_{5}^{4}}{C_{15}^{7}} = \frac{120 \times 5}{6435} = \frac{600}{6435} \approx 0.0932$$

For $k = 4$:

$$P(X = 4) = \frac{C_{10}^{4} C_{5}^{7-4}}{C_{15}^{7}} = \frac{C_{10}^{4} C_{5}^{3}}{C_{15}^{7}} = \frac{210 \times 10}{6435} = \frac{2100}{6435} \approx 0.3263$$

For $k = 5$:

$$P(X = 5) = \frac{C_{10}^{5} C_{5}^{7-5}}{C_{15}^{7}} = \frac{C_{10}^{5} C_{5}^{2}}{C_{15}^{7}} = \frac{252 \times 10}{6435} = \frac{2520}{6435} \approx 0.3916$$

For $k = 6$:

$$P(X = 6) = \frac{C_{10}^{6} C_{5}^{7-6}}{C_{15}^{7}} = \frac{C_{10}^{6} C_{5}^{1}}{C_{15}^{7}} = \frac{210 \times 5}{6435} = \frac{1050}{6435} \approx 0.1632$$

For $k = 7$:

$$P(X = 7) = \frac{C_{10}^{7} C_{5}^{7-7}}{C_{15}^{7}} = \frac{C_{10}^{7} C_{5}^{0}}{C_{15}^{7}} = \frac{120 \times 1}{6435} = \frac{120}{6435} \approx 0.0187$$

Probability Distribution Table:

k	2	3	4	5	6	7	$\sum$
$P(X = k)$	0.0070	0.0932	0.3263	0.3916	0.1632	0.0187	1

2- Calculating the expected value of good items obtained

Mean of hypergeometric distribution:

$$E(X) = n \times \frac{K}{N} = 7 \times \frac{10}{15} =\approx 4.667$$

On average, we expect to obtain about 4.667 good items in our sample of 7 items.

Chapter V Exercises

Exercise 1:
Three fair coins are tossed simultaneously. If all three coins show the same face (all heads or all tails), you win 80 DZD. Otherwise, you lose 20 DZD.

1- What is the expected gain?
2- Is the game fair? Why or why not?

Solution:

1- We start by defining the events and probabilities

When three fair coins are tossed, the total number of possible outcomes is:

$$2^3 = 8$$

The winning outcomes (all three coins the same):{H H H, T T T }
So, the number of winning outcomes is 2.
The probability of winning (all same): $P(win) = \frac{2}{8} = \frac{1}{4}$
The probability of losing (not all same): $P(lose) = 1 - \frac{1}{4} = \frac{3}{4}$

Then, we define the monetary values
Gain if you win: +80 DZD
Gain if you lose: −20 DZD (a loss is a negative gain)

Finally, we calculate the expected gain
The formula for expected value E is:

$$E = [P(win) \times Gain\ from\ win] + [P(lose) \times Gain\ from\ lose]$$

Substitute the values:

$$E = (\frac{1}{4} \times 80) + (\frac{3}{4} \times (-20))$$
$$E = (20) + (-15) = 20 - 15 = 5$$

The expected gain is 5 DZD.

2- A game is considered fair if the expected value is 0. This means that, on average, a player neither wins nor loses money over many plays.

In this game, the expected value is 5 DZD, which is greater than 0.
So, the game is not fair.

Justification: the expected value is positive (E=5 DZD), meaning that, on average, a player can expect to win 5 DZD per game. Therefore, the game is favorable to the player and unfavorable to the person running the game.

Exercise 2:
Two fair six-sided dice are rolled once in a random experiment.
The random variable X represents the maximum of the two numbers that appear.

1- Find the probability distribution of the random variable X.
2- Write the probability density function and the cumulative distribution function F(X).
3- Calculate the mathematical expectation E(X) and the variance V(X).

Solution:

1- The Probability Distribution of X

The random variable X is the maximum of the two numbers on the dice. The possible values for X are the integers from 1 to 6.

$$\Omega_X = \{1,2,3,4,5,6\}$$

We roll two fair dices, so the total number of outcomes is $6 \times 6 = 36$, all equally likely.

To find $P(X = x_i)$, we count the number of outcomes where the maximum of the two dice is exactly x_i.

If $X = 1$: The only way the maximum is 1 is if both dice show 1.
Outcomes: (1,1)
Number of outcomes: 1

$$P(X = 1) = \frac{1}{36}$$

$X = 2$: The maximum is 2. This means both dices are 2 or less, but they are not both 1.
Favorable outcomes: (1,2), (2,1), (2,2)
Number of outcomes: 3

$$P(X = 2) = \frac{3}{36} = \frac{1}{12}$$

$X = 3$: The maximum is 3. This means both dice are 3 or less, but not both 2 or less.
Number of outcomes with both dice ≤ 3: $3 \times 3 = 9$
Number of outcomes with both dice ≤ 2: $2 \times 2 = 4$
Favorable outcomes: $9 - 4 = 5$
List: (1,3), (2,3), (3,1), (3,2), (3,3)

$$P(X = 3) = \frac{5}{36}$$

$X = 4$: Both dice ≤ 4, but not both ≤ 3.
Both ≤ 4: $4^2 = 16$
Both ≤ 3: $3^2 = 9$
Favorable outcomes: $16 - 9 = 7$

$P(X = 4) = \frac{7}{36}$

$X = 5$: Both dice ≤ 5, but not both ≤ 4.
Both ≤ 5: $5^2 = 25$
Both ≤ 4: $4^2 = 16$
Favorable outcomes: $25 - 16 = 9$

$$P(X = 5) = \frac{9}{36} = \frac{1}{4}$$

$X = 6$: Both dice ≤ 6, but not both ≤ 5.
Both ≤ 6: $6^2 = 36$

Both ≤ 5: $5^2 = 25$
Favorable outcomes: $36 - 25 = 11$
$P(X = 6) = \frac{11}{36}$

Let's verify the probabilities sum to 1:
$\frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36} + \frac{11}{36} = \frac{36}{36} = 1$. Correct.

Probability Density Function (PDF) or Probability Mass Function (PMF):

X	1	2	3	4	5	6	∑
P(X=x)	1/36	3/36	5/36	7/36	9/36	11/36	1

2- The Cumulative Distribution Function (CDF), $F(x)$:
The CDF is defined as $F(x) = P(X \leq x)$.
For $x < 1$: F(x)=0
For $1 \leq x < 2$: $F(x) = P(X \leq 1) = P(X = 1) = \frac{1}{36}$
For $2 \leq x < 3$: $F(x) = P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{36} + \frac{3}{36} = \frac{4}{36}$
For $3 \leq x < 4$: $F(x) = P(X \leq 3) = \frac{4}{36} + \frac{5}{36} = \frac{9}{36}$
For $4 \leq x < 5$: $F(x) = P(X \leq 4) = \frac{9}{36} + \frac{7}{36} = \frac{16}{36}$
For $5 \leq x < 6$: $F(x) = P(X \leq 5) = \frac{16}{36} + \frac{9}{36} = \frac{25}{36}$
For $x \geq 6$: $F(x) = P(X \leq 6) = \frac{25}{36} + \frac{11}{36} = \frac{36}{36} = 1$

X	1	2	3	4	5	6	∑
P(X=x)	1/36	3/36	5/36	7/36	9/36	11/36	1
F(x)	1/36	4/36	9/36	16/36	25/36	36/36	

$$F(x) = \begin{cases} 0 & if\ x < 1 \\ \frac{1}{36} & if\ 1 \leq x < 2 \\ \frac{4}{36} & if\ 2 \leq x < 3 \\ \frac{9}{36} & if\ 3 \leq x < 4 \\ \frac{25}{36} & if\ 5 \leq x < 6 \\ \frac{25}{36} & if\ 5 \leq x < 6 \\ 1 & if\ x \geq 6 \end{cases}$$

3- Calculate the Mathematical Expectation E(X) and the Variance V(X)
(a) Expectation E(X):

$$E(X) = \sum_{i=1}^{6} x_i P(X = x_i)$$

X	1	2	3	4	5	6	∑

P(X=x)	1/36	3/36	5/36	7/36	9/36	11/36	1
E(x)	1/36	6/36	15/36	28/36	45/36	66/36	161/36

$$E(X) = \frac{161}{36} \approx 4.472$$

(b) Variance V(X):

We use the formula:

$$V(X) = E(X^2) - [E(X)]^2$$

First, calculate $E(X^2)$:

$$E(X^2) = \sum_{i=1}^{6} \mathrm{x}_i{}^2 \mathrm{P}(\mathrm{X} = \mathrm{x}_i)$$

X	1	2	3	4	5	6	$\sum$
P(X=x)	1/36	3/36	5/36	7/36	9/36	11/36	1
$E(X^2)$	1/36	12/36	45/36	112/36	225/36	396/36	791/36

$$V(X) = (X^2) - [E(X)]^2$$

$$V(X) = \frac{791}{36} - \left[\frac{161}{36}\right]^2 = \frac{791}{36} - \frac{25921}{1296} = \frac{2555}{1296}$$

Exercise 3:

Let X be a discrete random variable with the following probability mass function:

$$P(X = x_i) = \begin{cases} \frac{k}{x^3} & if\ x \in \{1, 2, 3\} \\ 0 & otherwise \end{cases}$$

1- Find the value of the constant k that makes this a valid probability distribution.
2- Find the expected value E(X).

Solution:

1- The value of the constant k

For this to be a valid probability mass function, the sum of all probabilities must equal 1.

$$\sum_{i=1}^{3} \mathrm{P}(\mathrm{X} = \mathrm{x}_i) = 1$$

$$P(X = 1) + P(X = 2) + \mathrm{P}(\mathrm{X} = 3) = 1$$

$$\frac{k}{1^3} + \frac{k}{2^3} + \frac{k}{3^3} = 1$$

$$\frac{(216 + 27 + 8)k}{216} = 1$$

$$\frac{251k}{216} = 1$$

$$k = \frac{216}{251}$$

2- The expected value E(X)

The expected value for a discrete random variable is given by:

$$E(X) = \sum_{i=1}^{n} \mathrm{x}_i \mathrm{P}(\mathrm{X} = \mathrm{x}_i)$$

$$E(X) = \sum_{i=1}^{3} \mathrm{x}_i \times \frac{k}{x^3} = \frac{k}{x^2} = k \times (\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2}) = k \times \frac{49}{36}$$

$$E(X) = \frac{216}{251} \times \frac{49}{36} = \frac{294}{251}$$

Exercise 4 :

We toss a biased coin three times. The probability of getting Heads (H) is $\frac{2}{3}$.

We define the random variable X as the number of times Heads (H) appears.

1- Write the probability distribution.
2- Calculate the mathematical expectation (mean), variance, and standard deviation.
3- Calculate the probabilities:

$P(X < 0), \mathrm{P}(\mathrm{X} = 2), \mathrm{P}(\mathrm{X} < 2), \mathrm{P}(\mathrm{X} \leq 2), \mathrm{P}(\mathrm{X} \geq 2), \mathrm{P}(0 < \mathrm{X} < 2).$

4- Calculate E(2X) and V(2X).

Solution:

1- We toss a biased coin three times. So the total number of outcomes is $2^3 = 8$

$$\Omega = \{HHH, HHT, HTH, HTT, TTT, TTH, THT, THH\}$$

X is a random variable that represents the number of times Heads (F) appears.

$$\Omega_X = \{0,1,2,3\}$$

We have a biased coin where:

$$P(Heads) = P(H) = \frac{2}{3}$$

$$P(Tails) = P(T) = 1 - P(H) = \frac{1}{3}$$

We calculate $P(X = x_i)$

If $X = 0$: There is only one outcome with no Heads.
Favorable outcomes: (TTT).
Number of outcomes: 1

$$P(X = 1) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$$

$X = 1$: There is three outcomes with one Heads.
Favorable outcomes: (TTH), (THT), (HTT)
Number of outcomes: 3

$$P(X = 2) = (\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3}) + (\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3}) + (\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}) = \frac{6}{27}$$

$X = 2$: There is three outcomes with two Heads.
Favorable outcomes: (THH), (HHT), (HTH)

Number of outcomes: 3

$$P(X=2)=\left(\frac{1}{3}\times\frac{2}{3}\times\frac{2}{3}\right)+\left(\frac{2}{3}\times\frac{2}{3}\times\frac{1}{3}\right)+\left(\frac{2}{3}\times\frac{1}{3}\times\frac{2}{3}\right)=\frac{12}{27}$$

If $X=3$: There is only one outcome with three Heads.
Favorable outcomes: (HHH).
Number of outcomes: 1

$$P(X=2)=\left(\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}\right)=\frac{8}{27}$$

Let's verify the probabilities sum to 1:
$\frac{1}{27}+\frac{6}{27}+\frac{12}{27}+\frac{8}{27}=\frac{27}{27}=1$. Correct.

The probability mass function is:

X	0	1	2	3	$\sum$
P(X=x)	$^{1}/_{27}$	$^{6}/_{27}$	$^{12}/_{27}$	$^{8}/_{27}$	$\frac{27}{27}$

2- Calculate Expectation, Variance, and Standard Deviation

Expectation E(x)

$$E(X)=\sum_{i=0}^{3}\mathrm{x}_i\mathrm{P}(\mathrm{X}=\mathrm{x}_i)$$

X	0	1	2	3	$\sum$
P(X=x)	$^{1}/_{27}$	$^{6}/_{27}$	$^{12}/_{27}$	$^{8}/_{27}$	$\frac{27}{27}$
E(x)	0	$^{6}/_{27}$	$^{24}/_{27}$	$^{24}/_{27}$	$^{54}/_{27}=2$

$$E(X)=2$$

Variance $Var(x)$
We use the formula:

$$V(X)=E(X^2)-[E(X)]^2$$

First, calculate $E(X^2)$:

$$E(X^2)=\sum_{i=0}^{3}\mathrm{x}_i{}^2\mathrm{P}(\mathrm{X}=\mathrm{x}_i)$$

X	0	1	2	3	$\sum$
P(X=x)	$^{1}/_{27}$	$^{6}/_{27}$	$^{12}/_{27}$	$^{8}/_{27}$	$^{27}/_{27}=1$
E(x²)	0	$^{10}/_{27}$	$^{32}/_{27}$	$^{72}/_{27}$	$^{114}/_{27}$

$$V(X)=\frac{114}{27}-[2]^2=\frac{2}{3}\approx 0.667$$

Standard Deviation σ

$$\sigma=\sqrt{Var(X)}$$

$$\sigma = \sqrt{\frac{2}{3}} \approx 0.816$$

3- Calculate the probabilities:
$P(X < 0), \mathrm{P(X = 2)}, \mathrm{P(X < 2)}, \mathrm{P(X \leq 2)}, \mathrm{P(X \geq 2)}, \mathrm{P(0 < X < 2)}.$

$$P(X < 0) = 0$$
$$\mathrm{P(X = 2)} = {}^{12}/_{27}$$
$$\mathrm{P(X < 2) = P(X = 0) + P(X = 1)} = {}^{1}/_{27} + {}^{6}/_{27} = {}^{7}/_{27}$$
$$\mathrm{P(X \leq 2) = P(X = 0) + P(X = 1)} + P(X = 2) = {}^{1}/_{27} + {}^{6}/_{27} + {}^{12}/_{27} = {}^{19}/_{27}$$
$$\mathrm{P(X \geq 2)} = P(X = 2) + P(X = 3) = {}^{12}/_{27} + {}^{8}/_{27} = {}^{20}/_{27}$$
$$\mathrm{P(0 < X < 2) = P(X = 1)} = {}^{6}/_{27}$$

4- Calculate E(2X) and V(2X).
To calculate E(2X) and V(2X), we use the properties of expectation and variance:

$$E(aX) = a \cdot E(X)$$
$$V(aX) = a^2 \cdot V(X)$$

Here, $a = 2$.
$E(2X)$:

$$E(2X) = 2 \cdot E(X) = 2 \times 2 = 4$$

$V(2X)$:

$$V(2X) = 2^2 \cdot V(X) = 4 \times \frac{2}{3} = \frac{8}{3}$$

Exercise 5:
A quality control check at a factory found that, on average, 2% of the items produced are defective. If a random sample of 10 items is taken from the production line, calculate the following probabilities:

1- All selected items are non-defective.
2- At most one item is defective.
3- At least two items are defective.

Solution:
We are dealing with a Binomial distribution because we have a fixed number of independent trials (selecting 10 items), each with the same probability of being defective.

Sample size: $n = 10$
Probability of a defective item: $p = 0.02$
Probability of a non-defective item: $q = 1 - p = 0.98$

Let X be the random variable representing the number of defective items in the sample.

$$X \sim Bin(n = 10, p = 0.02)$$

The Binomial probability formula is:

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

1- Probability that all selected items are non-defective

This is the probability of finding zero defective items, P(X=0)

$$P(X = 0) = C_{10}^0 (0.02)^0 (0.98)^{10}$$
$$P(X = 0) = 1 \times 1 \times (0.98)^{10}$$
$$P(X = 0) \approx (0.98)^{10} \approx 0.8171$$

2- Probability that at most one item is defective

This is the probability of finding zero or one defective item,

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

We already have $P(X = 0) \approx 0.8171$.
Now, calculate P(X=1):

$$P(X = 1) = C_{10}^1 (0.02)^1 (0.98)^9$$
$$P(X = 1) = 10 \times 0.02 \times (0.98)^9$$
$$P(X = 1) \approx 0.2 \times 0.8337 \approx 0.1667$$

So,

$$P(X \leq 1) = P(X = 0) + P(X = 1) \approx 0.8171 + 0.1667 \approx 0.9838$$

3- Probability that at least two items are defective

This is the probability of finding two or more defective items. This is the complement of the event "at most one item is defective".

$$P(X \geq 2) = 1 - P(X \leq 1)$$
$$P(X \geq 2) \approx 1 - 0.9838 \approx 0.0162$$

Exercise 6:
The number of typographical errors on a page of a manuscript follows a Poisson distribution with an average of 1.2 errors per 2 pages.

1- Find the probability of having at most three errors in a chapter that is 5 pages long.
2- Find the length of the manuscript (in pages) if the probability of having no errors is 0.7.
3- Find the standard deviation of the number of errors in a section that is 2 pages long.

Solution:
We are given that the number of errors follows a Poisson distribution.
Average rate: 1.2 errors per 2 pages.
This means the rate per page is $\lambda_{page} = \frac{1.2}{2} = 0.6\ errors/page$.

Let X be the number of errors in a given number of pages.
If t is the number of pages, then $X \sim Poisson(\lambda t)$, where $\lambda = 0.6\ errors/page$.

1- Probability of at most three errors in 5 pages

First, find the mean for $t = 5$ pages:

$$\mu_1 = \lambda t = 0.6 \times 5 = 3$$

So, $X \sim Poisson(3)$.
We want $P(X \leq 3)$.
Using the Poisson probability formula

$$P(X = k) = \frac{e^{-\mu}\mu^{k}}{k!}$$

$$P(X = 0) = \frac{e^{-3}3^{0}}{0!} = e^{-3}$$
$$P(X = 1) = \frac{e^{-3}3^{1}}{1!} = 3e^{-3}$$
$$P(X = 2) = \frac{e^{-3}3^{2}}{2!} = \frac{9}{2}e^{-3} = 4.5e^{-3}$$
$$P(X = 3) = \frac{e^{-3}3^{3}}{3!} = \frac{27}{6}e^{-3} = 4.5e^{-3}$$

So,

$$P(X \leq 3) = e^{-3} + 3e^{-3} + 4.5e^{-3} + 4.5e^{-3} = e^{-3}(1 + 3 + 4.5 + 4.5)$$
$$P(X \leq 3) = e^{-3} \times 13$$
$$P(X \leq 3) \approx 0.049787 \times 13 \approx 0.6472$$

2- Find the length (in pages) if the probability of no errors is 0.7

Let t be the number of pages. The mean for tt pages is μ=0.6t.

The probability of no errors is:

$$P(X = 0) = e^{-0.6t}$$

We are given $P(X = 0) = 0.7$, so:

$$e^{-0.6t} = 0.7$$

Take the natural logarithm of both sides:

$$-0.6t = ln(0.7)$$
$$-0.6t \approx -0.356675$$
$$t \approx \frac{0.356675}{0.6} \approx 0.5945$$

3- Standard deviation for a section of 2 pages

For $t = 2$ pages, the mean is:

$$\mu_2 = \lambda t = 0.6 \times 2 = 1.2$$

For a Poisson distribution, the variance equals the mean:

$$\sigma^2 = \mu_2 = 1.2$$

The standard deviation is:

$$\sigma = \sqrt{1.2} \approx 1.0954$$

Exercise 7:

Consider a fair 20-sided die numbered from 1 to 20. How many rolls must a player make to be at least 95% certain of rolling the number 15 at least once?

Solution:
First, we define the probability of success and failure

We have a fair 20-sided die. The probability of rolling the number 15 (a "success") on a single roll is:

$$p = \frac{1}{20} = 0.05$$

The probability of not rolling a 15 on a single roll (a "failure") is:

$$q = 1 - p = 1 - 0.05 = 0.95$$

Then, we set up the inequality for "at least one success"

Let n be the number of rolls. The probability of getting at least one 15 in n rolls is equal to 1 minus the probability of getting no 15s in n rolls.

$$P(at\ least\ one\ 15) = 1 - P(no\ 15s) = 1 - q^n = 1 - (0.95)^n$$

We want this probability to be at least 95% (or 0.95).

$$1 - (0.95)^n \geq 0.95$$
$$-(0.95)^n \geq 0.95 - 1$$
$$-(0.95)^n \geq -0.05$$
$$(0.95)^n \leq 0.05$$

Now, take the natural logarithm (ln) of both sides:

$$ln\ ((0.95)^n) \leq ln\ (0.05)$$
$$n \cdot ln(0.95) \leq ln(0.05)$$

N.B. Since $ln(0.95)$ is a negative number, dividing both sides by it will reverse the inequality sign again.

$$n \geq \frac{ln(0.05)}{ln(0.95)}$$

Calculate the logarithms:

$$ln(0.05) \approx -2.9957$$
$$ln(0.95) \approx -0.05129$$

Now compute n:

$$n \geq \frac{-2.9957}{-0.05129} \approx 58.4$$

Since n must be an integer (number of rolls), we round up to the nearest whole number to ensure the probability is at least 0.95.

$$n = 59$$

The player must roll the die 59 times to be at least 95% certain of rolling the number 15 at least once.

Chapter VI: Continuous Random Variables and Their Probability Distributions

So far, we have focused on the probabilities of events within a random experiment. However, this approach has limitations when it comes to quantifying and numerically manipulating the outcomes of these experiments.
The concept of random variables will allow us to move beyond these limitations and step forward. By assigning a numerical value to every possible outcome of a random experiment, it opens the door to mathematical analysis while providing a powerful tool for characterizing uncertain phenomena.

1. Continuous Random Variables

1.1 Definition

A continuous random variable assumes values on a continuum, i. e. it can take an uncountable number of possible values, typically over an interval in ℝ. Since there are an infinity of values between any two points on a continuum, assigning a probability value with a point on that continuum is not meaningful. Instead, probabilities are assigned to intervals along the continuum.
The continuous uniform distribution is fundamental when modeling complete randomness in intervals. It serves as a basis for more complex distributions, provides theoretical foundation for random sampling, and has an important role in simulation and statistical theory (e. g. the time taken to complete a test; the temperature on a given day).

1.2 Probability Density Function (PDF)

The probability density function (PDF) of a continuous random variable *X*, denoted by $f(x)$, is a mathematical function such that the area under its curve over any interval equals the probability that *X* will take on a value in that interval, describing the likelihood of *X* taking a value within an infinitesimal range. It corresponds to the probability density at *x*.
The mathematical formula takes the following form:

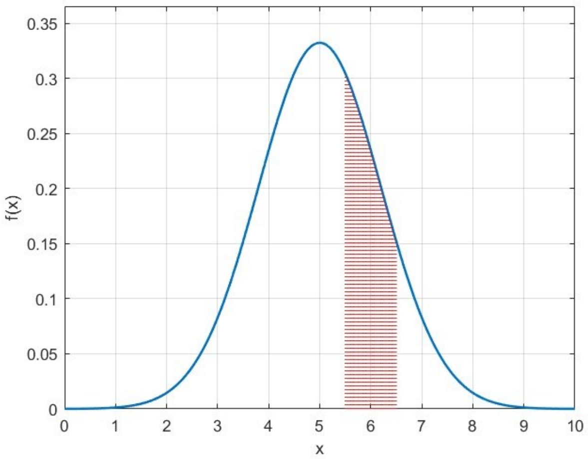
$$f(x) = \begin{cases} f(x) & if\ a \le x \le b \\ 0 & Otherwise \end{cases}$$

1.2.1 Properties of PDF:

- The PDF is non-negative: $f(x) \geq 0, \forall\, x$
- The total area under the curve is equal to 1: $\int_{-\infty}^{+\infty} f(x)dx = 1$

1.2.2 Graphical Representation of a Probability Density Function (PDF)

The Probability Density Function (PDF) of a continuous random variable (CRV) provides a graphical representation of its probability distribution. Unlike discrete distributions modeled by probability mass functions with discrete bars, the PDF is characterized by a smooth curve where probabilities correspond to areas under the curve over specific intervals.



1.2.3 Key Properties of a PDF Graph

1- The shape of the Curve

The curve $y = f(x)$ is continuous (no jumps or breaks).
The height at any point x is not the probability $P(X = x_i)$ (which is always 0 for a CRV), but rather the density.
The total area under the curve is 1 (since $\int_{-\infty}^{+\infty} f(x)dx = 1$).

2- Probability as an Area

The probability that *X* falls in $[a, b]$ is the area under $f(x)$ between a and b:
$P(a \leq X \leq b) = \int_a^b f(x)dx$
An example is given in the PDF graph with the shaded area being the probability that *X* will take a value between *a* and *b*.

3- The possible Shapes

It can be:
Unimodal: Single peak (e.g., normal distribution).
Bimodal: Two peaks.
Uniform: Flat line (equal probability over an interval).
Skewed: Longer tail on one side (e.g., exponential distribution).

1.3 the probability over an interval [a, b]

In mathematical terms we can express the probability function over an interval as

$$P(a \leq x \leq b) = \int_a^a f(x)dx$$

Note: For a single point, $P(X = c) = 0$ since the integral over a point is zero.

1.4 The Cumulative Distribution Function (CDF)

The CDF of a continuous random variable *X* is denoted F(x), and is defined

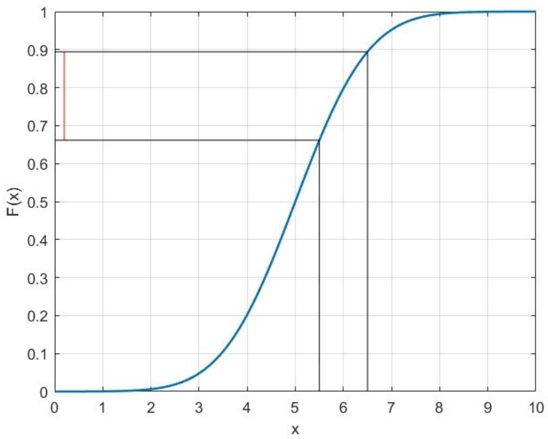
$$F(x) = P(X \leq x_i) = \int_{-\infty}^{x} f(x)dx$$

where $-\infty \leq x \leq +\infty$.

$$F(x) = \begin{cases} 0 & if\ x \leq 0 \\ x & if\ 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Graphical Representation of a Cumulative Distribution Function (CDF)

The Cumulative Distribution Function corresponds to the area under the probability density function to the left of x, and the probability that the outcome of X in a random trial will be less than or equal to any specified value x.



In the cumulative distribution function (CDF) graph, the vertical separation between the points at $X \leq 5.5$ and $X \leq 6.5$ corresponds to the shaded region under the probability density function (PDF) curve between those same points (p. 84). Furthermore, the CDF's value at $X = 5.5$ itself represents the total area under the PDF curve (p. 84) to the left of $X = 5.5$.

Properties of CDF

1- As X increases $F(X)$ never decreases (Monotonic increasing).
2- The Limits:

$$\lim_{x \to -\infty} F(x) = 0$$
$$\lim_{x \to +\infty} F(x) = 1$$

3- The Relation with the PDF:

$$f(x) = \frac{d}{dx} F(x)$$

For continuous random variables, the cumulative distribution function's graph is a smooth, monotonically increasing curve that ranges from 0 (as $x \to -\infty$) to 1 (as $x \to +\infty$). The derivative of this CDF at any point x yields the probability density function $f(x)$.

Graphic

1.5 Expected Value (Mean) and Variance

Expected Value (Mean)

The continuous uniform distribution is characterized by a constant probability density between a and b, a linear cumulative distribution, and a simple mean and variance formulas.
The mean or the expected value of a continuous random variable X is denoted as

$$E(x) = \mu = \int_{-\infty}^{+\infty} xf(x)dx$$

The integral used in the case of continuous random variable performs the same role as the summation does for a discrete one.

Variance

Similarly to the expected value, the variance of a continuous random variable is defined as

$$Var(X) = E\left[\left(x_i - E(X)\right)^2\right] = \int_{-\infty}^{+\infty} [(x_i - E(X)]^2 f(x)dx$$

$$Var(x) = \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left[\int_{-\infty}^{+\infty} xf(x)dx\right]^2 = \int_{-\infty}^{+\infty} (X-\mu)^2 f(x)dx$$

Alternatively, it can be computed as:

$$V(x) = E(X^2) - [E(X)]^2$$

The variance measures the spread of *X* around its mean.

1.6 Probabilities calculations

$P(X = a) = \int_a^a f(X)dx = 0,$

$P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f(X)dx = F(a),$

$P(X > a) = P(X \geq a) = \int_a^{+\infty} f(X)dx = 1 - \int_{-\infty}^{a} f(X)dx = 1 - F(a),$

$P(a < X < b) = \int_a^b f(X)dx = \int_{-\infty}^{b} f(X)dx - \int_{-\infty}^{a} f(X)dx = F(b) - F(a),$

$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b).$

Exercise 1:

Let *X* be a continuous random variable that follows a uniform distribution on the interval [1,4]. Its probability density function (PDF) is given by:

$$f(x) = \begin{cases} c & for\ 1 \leq x \leq 4 \\ 0 & otherwise \end{cases}$$

1- Find the value of the constant c.
2- Find the cumulative distribution function (CDF), $F(x)$.
3- Calculate the following probabilities: $P(X < 2.5), P(1.5 \leq X \leq 3.5), P(X > 3)$

Solution:

1- The value of the constant c

For any continuous probability density function, the total area under the curve must equal 1.

The PDF is a constant *c* over the interval [1,4]. The length of this interval is $4 - 1 = 3$.
The area under the curve is the area of a rectangle: $Base \times Height = (4-1) \times c = 3c$.

We set this equal to 1:

$$3c = 1$$

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

Therefore, the PDF is:

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

2- The cumulative distribution function (CDF), $F(x)$

The CDF, $F(x)$, is defined as $F(x) = P(X \leq x)$. For a continuous uniform distribution on [a,b], it is found by integrating the PDF from a to *x*.

Here, $a = 1$, $b = 4$, and $f(x) = \frac{1}{3}$.

$$F(x) = \int_1^x f(x)\,dt = \int_1^x \frac{1}{3}\,dt$$

We consider different intervals for *x*:

For $x < 1$: $F(x) = 0$

For $1 \leq x \leq 4$:

$$F(x) = \int_1^x \frac{1}{3}\,dt = \frac{1}{3}(x-1)$$

For $x > 4$: $F(x) = 1$

Therefore, the CDF is:

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

3- Calculating the Probabilities

We can calculate these probabilities using the CDF F(x) or by finding areas under the PDF curve. Using the CDF is often more straightforward.

a) $P(X < 2.5)$

For a continuous variable, $P(X < k) = P(X \leq k) = F(k)$.

$$P(X < 2.5) = F(2.5) = \frac{2.5-1}{3} = \frac{1.5}{3} = 0.5$$

b) $P(1.5 \leq X \leq 3.5)$

$$P(1.5 \leq X \leq 3.5) = F(3.5) - F(1.5)$$

$$F(3.5) = \frac{3.5-1}{3} = \frac{2.5}{3} \approx 0.8333$$

$$F(1.5) = \frac{1.5-1}{3} = \frac{0.5}{3} \approx 0.1667$$

$$P(1.5 \leq X \leq 3.5) = F(3.5) - F(1.5) \approx 0.8333 - 0.1667 = 0.6666$$

c) $P(X > 3)$

This is the complement of $P(X \leq 3)$.

$$P(X > 3) = 1 - P(X \leq 3) = 1 - F(3)$$

$$F(3) = \frac{3-1}{3} = \frac{2}{3} \approx 0.6667$$

$$P(X > 3) = 1 - \frac{2}{3} = \frac{1}{3} = 0.3333$$

2. Common Continuous Distributions

2.1 Uniform Distribution

2.1.1 Definition

The continuous uniform random variable is the simplest continuous probability distribution, it has uniform probability density over an interval, i. e. all outcomes in an interval are equally likely (e. g. Random number generation; time of arrival when equally likely at any moment). The continuous uniform probability density function has the notation:

$$X \sim Uniform(a, b)$$

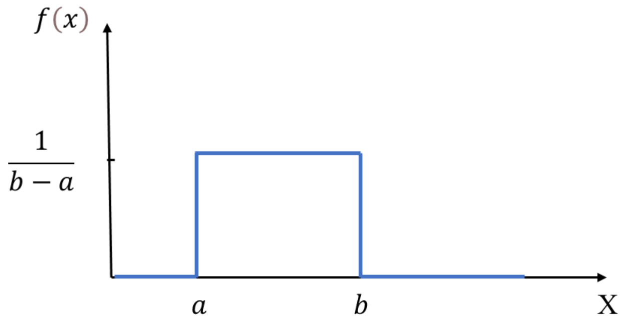
2.1.2 Probability Density Function (PDF) of a Continuous Uniform Distribution

The Probability Density Function of a continuous uniform random variable is constant over the interval [a, b]:

$$f(x) = \begin{cases} \frac{1}{b-a} & if\ a \leq x \leq b \\ 0 & Otherwise \end{cases}$$

where
a: Lower bound of the interval
b: Upper bound of the interval ($b > a$)

Graphical Representation of PDF of a Continuous Uniform Distribution



Properties of PDF

The total area under the curve is equal to 1.
Height is constant (uniform) between a and b.
The probability outside $[a, b]$ equals zero.

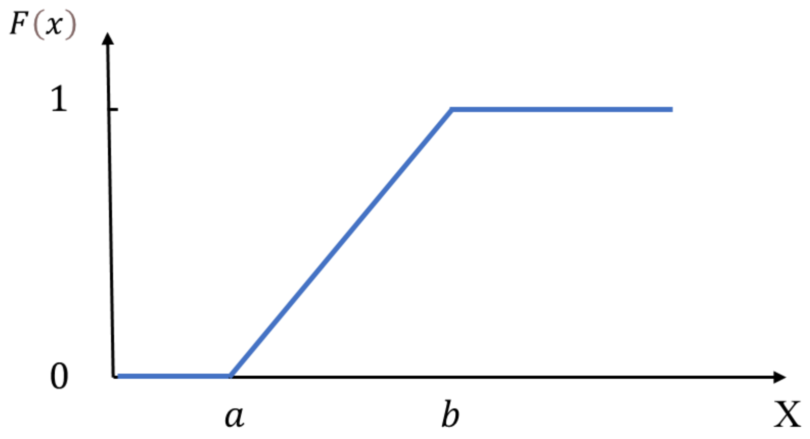
2.1.3 Cumulative Distribution Function (CDF) of a Continuous Uniform Distribution

The cumulative probability function of a continuous uniform random variable gives

$$F(x) = P(X \le x) = \begin{cases} 0 & for\ x < a \\ \dfrac{x-a}{b-a} & for\ a \le x \le b \\ 1 & for\ x > b \end{cases}$$

The CDF has a linear increase from 0 to 1 over [a, b]

Graphical Representation of CDF



2.1.4 Expected Value and Variance

Mean (Expected Value)

$$E(X) = \frac{a+b}{2}$$

Variance

$$Var(X) = \frac{(b-a)^2}{12}$$

Standard Deviation

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{(b-a)^2}{12}}$$

Example:
A bus arrives every 10 minutes, and arrival times are uniformly distributed.

1- Probability a passenger waits ≤3 minutes?
2- Expected waiting time?
3- Probability of waiting between 2 and 7 minutes?

Solution:
Let X be the waiting time (in minutes)

$$X \sim Uniform(0,10)$$

$$f(x) = \frac{1}{10}, when\ 0 \leq x \leq 10$$
$$F(x) = \frac{x}{10}, when\ 0 \leq x \leq 10$$

1- Probability a passenger waits ≤ 3 minutes?

$$P(X \leq 3) = F(3) = \frac{3}{10} = 0.3$$

There is 30% chance of waiting 3 minutes or less.

2- Expected waiting time?

For a uniform distribution $Uniform(a, b)$:

$$E(X) = \frac{a+b}{2} = \frac{0+10}{2} = 5 \text{ minutes}$$

The average waiting time is 5 minutes

3- Probability of waiting between 2 and 7 minutes?

$$P(2 \leq X \leq 7) = F(7) - F(2) = \frac{7}{10} - \frac{2}{10} = \frac{5}{10} = 0.5$$

There is 50% chance of waiting between 2 and 7 minutes.

2.2 Normal (Gaussian) Probability Distributions

2.2.1 Introduction to the Normal Distribution

The normal probability distributions is the most important continuous probability distribution of all for our purposes. It is a continuous distribution defined over the interval $]-\infty, +\infty[$. In addition to be unbounded, it provides an accurate model for a wide variety of real-world phenomena.

Its importance is due to its symmetricity about the mean (bell-shaped curve).

Normal distribution is important due to its symmetricity about the mean (bell-shaped curve), and is often referred to using the compact notation

$$X \sim N(\mu, \sigma^2)$$

2.2.2 Probability Density Function (PDF)

The probability density function of the famous "bell curve" is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

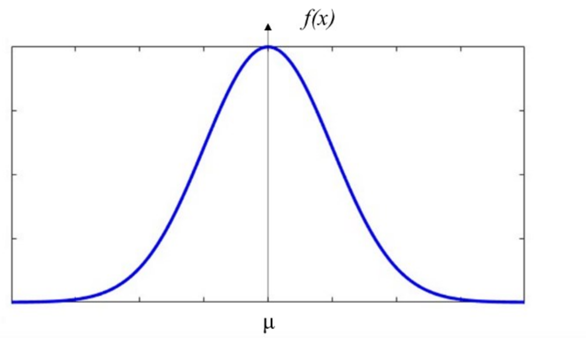
Where

$-\infty \leq x \leq +\infty$,
$-\infty \leq \mu \leq +\infty$,
$\sigma > 0$,
$\pi = 3.14159$, and
$e = 2.71828$.

Properties

The normal distribution is symmetric about the mean.
The total area under curve is equal to 1.
Its points of inflection are at $x = \mu \pm \sigma$
For a Normal Distribution:
$\approx$ 68% of values fall within ± 1 standard deviation of the mean $(\mu \pm \sigma)$
$\approx$ 95% of values fall within ± 2 standard deviations of the mean $(\mu \pm 2\sigma)$
$\approx$ 99.7% of values fall within ± 3 standard deviations of the mean $(\mu \pm 3\sigma)$
i.e. it follows 68-95-99.7 Rule.

Graphical Representation



2.2.3 Cumulative Distribution Function (CDF)

No closed-form solution, but defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

2.2.4 Expected Value and Variance

The mean, variance and standard distribution of a normal probability distribution are
Mean (Expected Value)

$$E(X) = \int_{-\infty}^{+\infty} x \times f(x)dx = \int_{-\infty}^{+\infty} x \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$$

With the change of variable $z = \frac{x-\mu}{\sigma}$

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\mu + \sigma z) e^{-\frac{1}{2}z^2} dz$$

$$E(X) = \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot e^{-\frac{1}{2}z^2} dz$$

$$E(X) = \mu \cdot 1 + \sigma \cdot 0 = \mu$$

Variance

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma^2$$

$$Var(X) = \sigma^2$$

The variance of a normal distribution $N(\mu, \sigma^2)$ is indeed σ^2, which confirms that the parameter σ^2 in the notation represents the variance.

Standard Deviation

$$\sigma = \sqrt{Var(X)}$$

The normal distributions is characterized by two parameters—the mean (μ), which is a location parameter and standard deviation (σ), which is a spread parameter. Each uniques combination (μ, σ) defines a distinct member of the normal distribution family. All normal distributions are symmetric and bell-shaped, centerd at the mean (μ) with a spread determined by the standard deviation (σ).

2.3 Standard Normal Distribution

2.3.1 Definition

The standardissed normal distribution is the most important member of the family of normal probability distributions—the one with $\mu = 0$ and $\sigma = 1$. The normal random variable distributed according to the standard normal distribution is called the standard normal variable and is denoted by Z.
Standardised Normal Distributions are often referreed to using the compact notation

$$Z \sim N(0,1)$$

Any normal distribution $Z \sim N(0,1)$ can be expressed as

$$Z = \frac{X - \mu}{\sigma}$$

2.3.2 PDF for a standard normal distribution

The PDF $\varphi(z)$ of a standard normal distribution is given by:

$$\varphi(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} & if \ -\infty \leq z \leq +\infty \\ 0 & Otherwise \end{cases}$$

2.3.3 Cumulative Distribution Function: Φ(z) (tabulated values)

The CDF (Cumulative Distribution Function) $\Phi(z)$ is defined as:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Or in piecewise form:

$$\Phi(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} & for \ all \ z \in R \\ 0 & as \ z \to -\infty \\ 1 & as \ z \to +\infty \end{cases}$$

The use of the standard normal distribution table:
For the calculation of the critical value t, we associate it with a value α, which can be interpreted in terms of probability, proportion, or percentage.
Knowing that the total area under the curve is equal to 1, this probability is equal to the area between the curve and the horizontal axis of values.
α represents the probability located at the two tails of the distribution, and we write t_{α}. If we want only one tail, we divide α by 2.

For example:

t_{α}		0.06	
1.9		0.975	

for $t_{\alpha} = 1.96, \alpha = 0.975$

Example:
The lifespan of an industrial pump has a normal distribution with a mean of 8000 hours and a standard deviation of 500 hours.
Find the probability that the pump fails before 7500 hours.
Find the probability that the pump's lifespan exceeds 8500 hours.
If a factory uses 4 of these pumps in its production line, find the probability that exactly 2 of them are still operational after 8200 hours.

Solution:

1- Probability that the pump fails before 7500 hours

Let X be the lifespan of a pump. We are given $X \sim N(\mu = 8000, \sigma = 500)$.

We need to calculate $P(X < 7500)$

First, calculate the Z-score:

$$Z = \frac{X - \mu}{\sigma} = \frac{7500 - 8000}{500} = \frac{-500}{500} = -1.00$$

Now, find the probability:

$$P(X < 7500) = P(Z < -1.00)$$

Using the standard normal distribution table or properties:

$$P(Z < -1.00) = 0.1587$$

2- Probability that the pump's lifespan exceeds 8500 hours

We need to calculate $P(X > 8500)$

Calculate the Z-score:

$$Z = \frac{8500 - 8000}{500} = \frac{500}{500} = 1.00$$

Now, find the probability:

$$P(X > 8500) = P(Z > 1.00) = 1 - P(Z < 1.00)$$
$$P(Z < 1.00) = 0.8413 \Rightarrow P(Z > 1.00) = 1 - 0.8413 = 0.1587$$

3- Probability that exactly 2 out of 4 pumps are still operational after 8200 hours

This is a binomial probability problem. First, we need the probability p that a single pump is still operational after 8200 hours.
We need to calculate $p = P(X > 8200)$

Calculation of the Z-score:

$$Z = \frac{8200 - 8000}{500} = \frac{200}{500} = 0.40$$
$$P(X > 8200) = P(Z > 0.40) = 1 - P(Z < 0.40)$$
$$P(Z < 0.40) = 0.6554 \Rightarrow P(Z > 0.40) = 1 - 0.6554 = 0.3446$$

Now, we use the Binomial Distribution:
Let Y be the number of pumps still operational after 8200 hours.

$$Y \sim Bin(n = 4, p = 0.3446)$$

We need to calculate $P(Y = 2)$

The binomial probability formula is:

$$P(Y = k) = C_n^k p^k (1 - p)^{n-k}$$
$$P(Y = 2) = C_4^2 (0.3446)^2 (1 - 0.3446)^{4-2}$$
$$P(Y = 2) = 6 \times (0.3446)^2 \times (0.6554)^2$$
$$P(Y = 2) = 6 \times 0.1187 \times 0.4295$$
$$P(Y = 2) = 6 \times 0.0510 = 0.3060$$

Chapter VI Exercises

Exercise 1:
Let X be a random variable with the following cumulative distribution function:

$$F(x)=\begin{cases} 0 & \text{for } x<1 \\ \dfrac{x^2-1}{8} & \text{for } 1\le x<2 \\ \dfrac{5x-7}{8} & \text{for } 2\le x<3 \\ 1 & \text{for } x\ge 3 \end{cases}$$

1- Find the probability density function (PDF), $f(x)$.
2- Calculate $P(1.5 < X \leq 2.5)$
3- Find the expected value $E(X)$.
4- Find the value xi such that $P(X \leq xi) = 0.7$.

Solution:
1- Finding the probability density function (PDF), $f(x)$.

$$f(x).=\begin{cases} 0 & \text{for } x<1 \\ \dfrac{d}{dx}\left(\dfrac{x^2-1}{8}\right)=\dfrac{2x}{8}=\dfrac{x}{4} & \text{for } 1\le x<2 \\ \dfrac{d}{dx}\left(\dfrac{5x-7}{8}\right)=\dfrac{5}{8} & \text{for } 2\le x<3 \\ 0 & \text{for } x\ge 3 \end{cases}$$

$$f(x).=\begin{cases} 0 & \text{for } x<1 \\ \dfrac{x}{4} & \text{for } 1\le x<2 \\ \dfrac{5}{8} & \text{for } 2\le x<3 \\ 0 & \text{for } x\ge 3 \end{cases}$$

2- Calculating $P(1.5 < X \leq 2.5)$

$$P(1.5 < X \leq 2.5) = F(2.5) - F(1.5)$$

$$P(1.5 < X \leq 2.5) = \frac{5(2.5)-7}{8} - \frac{(1.5)^2-1}{8}$$

$$P(1.5 < X \leq 2.5) = \frac{5.5}{8} - \frac{1.25}{8} = \frac{4.25}{8} = 0.53125$$

3- The expected value E(X)

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

$$E(x) = \int_1^2 x.\frac{x}{4}dx + \int_2^3 x.\frac{5}{8}dx$$

$$E(x)=\frac{1}{4}\int_1^2 x^2\,dx+\frac{5}{8}\int_2^3 x\,dx$$

$$E(x)=\frac{1}{4}\left[\frac{x^3}{3}\right]_1^2+\frac{5}{8}\left[\frac{x^2}{2}\right]_2^3$$

$$E(x)=\frac{1}{4}\left(\frac{8}{3}-\frac{1}{3}\right)+\frac{5}{8}\left(\frac{9}{2}-\frac{4}{2}\right)=\frac{1}{4}\times\frac{7}{3}+\frac{5}{8}\times\frac{5}{2}=\frac{7}{12}+\frac{25}{16}$$

$$E(x)=\frac{103}{48}\approx 2.146$$

4- Finding x_i such that $P(X\leq xi)=0.7$

We need to solve $F(xi)=0.7$

Check the range of the second piece at $x=2$: $F(2)=\frac{3}{8}=0.375$

Check the third piece at $x=3$: $F(3)=1$

Since $0.375<0.7<1$, xi is in the third interval [2,3[.

Use the third piece of the CDF:

$$\frac{5xi}{8}-\frac{7}{8}=0.7$$
$$\frac{5xi-7}{8}=0.7$$
$$5xi-7=5.6$$
$$5xi=12.6$$
$$xi=\frac{12.6}{5}=2.52$$

Exercise 2:

Consider the following function:

$$f(x)=\begin{cases}\frac{1}{9}x^2 & for\ 0<x<3\\ 0 & otherwise\end{cases}$$

1- Verify that the given function is a probability density function (PDF) and sketch its graph.
2- Find the probability $P(1<X<2)$.

Solution:

1- Sketch the graph of the PDF

The graph of $f(x)=\frac{1}{9}x^2$ on the interval (0, 3) has the following characteristics:

It is a parabola opening upwards.

It passes through the origin: $f(0)=0$.

At $x=3$, $f(3)=\frac{1}{9}(9)=1$.

The function is zero for $x\leq 0$ and $x\geq 3$.

Description of the Sketch:
The graph starts at the point (0, 0). It curves upwards smoothly, passing through the point (3, 1). The area under this parabolic curve between x=0 and x=3 is exactly 1. The graph is drawn only for x between 0 and 3, with the line falling to zero outside this interval.

2- The probability $P(1 < X < 2)$

The probability is found by integrating the PDF over the desired interval.

$$P(1 < X < 2) = \int_1^2 f(x)\,dx = \int_1^2 \frac{1}{9}\,x^2\,dx$$

$$P(1 < X < 2) = \frac{1}{9}\int_1^2 x^2\,dx$$

$$P(1 < X < 2) = \frac{1}{9}\left[\frac{x^3}{3}\right]_1^2$$

$$P(1 < X < 2) = \frac{1}{9}\left(\frac{8}{3} - \frac{1}{3}\right) = \frac{7}{27}$$

Exercise 3:
The length of a component produced by a machine has a uniform distribution between 9.8 cm and 10.2 cm. Suppose the required specifications state that the length must be between 9.85 cm and 10.15 cm.

1- Write the probability density function.
2- What is the probability that the length of a component meets the required specifications?
3- Determine the length that is exceeded by 80% of the produced components.

Solution:

1- The Probability Density Function (PDF)

For a continuous uniform distribution on the interval [a,b], the PDF is constant.
Lower limit: $a = 9.8$
Upper limit: $b = 10.2$

The PDF is given by:

$$f(x) = \frac{1}{b - a} \quad for\ a \leq x \leq b$$

First, calculate the range:

$$b - a = 10.2 - 9.8 = 0.4$$

Therefore, the PDF is:

$$f(x) = \begin{cases} \dfrac{1}{0.4} & for\ 9.8 \leq \mathrm{x} \leq 10.2 \\ 0 & otherwise \end{cases}$$

2- Probability that the length meets specifications

The specifications require the length X to be between 9.85 cm and 10.15 cm. We need $P(9.85 \leq X \leq 10.15)$

For a uniform distribution, this probability is the area of the rectangle under the PDF over the interval [9.85,10.15]

$$P(9.85 \leq X \leq 10.15) = (Base) \times (Height)$$
$$P(9.85 \leq X \leq 10.15) = (10.15 - 9.85) \times 2.5$$
$$P(9.85 \leq X \leq 10.15) = (0.3) \times (2.5) = 0.75$$

Interpretation: There is a 75% chance that a randomly selected component meets the specifications.

3- Length exceeded by 80% of components

If a length L is exceeded by 80% of components, this means that only 20% of components have a length greater than L. Equivalently, 80% of components have a length less than or equal to L. So, L is the 80th percentile, and we need to find L such that $P(X \leq L) = 0.8$

The Cumulative Distribution Function (CDF) for a uniform distribution is:

$$F(x) = P(X \leq x) = \frac{x - a}{b - a} \qquad for\ a \leq x \leq b$$

We set F(L)=0.8 and solve for L:

$$\frac{L - 9.8}{0.4} = 0.8$$
$$L - 9.8 = 0.32$$
$$L = 9.8 + 0.32 = 10.12$$

Interpretation: 80% of the produced components have a length of 10.12 cm or less, meaning 20% have a length greater than 10.12 cm. Therefore, the length 10.12 cm is exceeded by 20% of the components.

(Note: The original question asks for the length exceeded by 80%. If taken literally, this would be the 20th percentile, which would be $L = 9.8 + 0.2 \times 0.4 = 9.88$ cm. However, the standard interpretation in quality control for "exceeded by 80%" is that 80% are below the value, making it the 80th percentile. The solution above follows this standard interpretation.)

Chapter VII: Moments and Moment Generating Functions

1. Moments

1.1 Definition

Moments are quantitative indicators that describe key features of a random variable's probability distribution, including its central tendency (mean), spread (variance), asymmetry (skewness), and tail behavior (kurtosis). While used to define measures of dispersion and shape in statistics, they are linked to the characteristic function in probability theory.

1.2 Types of Moments

For a real random variable *X* with expected value μ and an integer $r \geq 1$, the moments of *X* are defined by the following expected values:
The centered (at 0) moment of order *r* of a random variable *X* is defined as: $m_r(X) = E(X^r)$
The centered moment of order *r* of a random variable *X* is defined as: $\mu_r(X) = E((X-\mu)^r)$
The absolute moment of order *r* of a random variable *X* is defined as: $M_r(X) = E(|X|^r)$

PARTICULAR case

The mathematical expectation (mean) μ is the first raw moment (m_1), while the variance σ^2 is the second central moment (μ_2). Furthermore, by the very definition of μ, the first central moment is always zero, since $E(X-\mu) = \mu - \mu = 0$.

If the random variable is discrete (defined by its possible values $\{xi\}$ and their corresponding probabilities $P_i = P(X = x_i)$), the moments are computed using the following sums:

$$\sum_{x_i} x_i^r P_i, \sum_{x_i} (x_i - \mu)^r P_i, \sum_{x_i} |x_i|^r P_i$$

If the random variable is absolutely continuous (defined by its probability density function $f(x)$), the moments are calculated using the following integrals:

$$\int_{-\infty}^{+\infty} x^r f(x)dx, \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx, \int_{-\infty}^{+\infty} |x|^r f(x)dx,$$

It should be noted that there is no guarantee that the sums will converge when dealing with an infinite set of possible values, or that the integrals will converge, unless explicitly verified. Some random variables may not possess moments of order r beyond a certain threshold. Finally, all probabilistic moments (as defined above in probability theory) have their statistical counterparts, which can be computed by straightforward adaptation of the summation formulas.

1.2.1 Raw Moments

The *rth* raw moment of a random variable *X* is:

$$m_r(X) = E(X^r) = \begin{cases} \sum_{x_i} x_i^r P_i = \sum_{x_i} x_i^r P(X = x_i) & (Discrete) \\ \int_{-\infty}^{\infty} x^r f(x)dx & (Continuous) \end{cases}$$

Key RAW Moments

$$m_0 = E(X^0) = E(1) = 1$$
$$m_1 = E(X^1) = E(X) = \mu$$
$$m_2 = E(X^2)$$

1.2.2 Central Moments

The r^{th} central moment (about the mean μ):

$$M_r(X) = E((X-\mu)^r) = \begin{cases} \sum_{x_i} (x_i - \mu)^r P_i & (Discrete) \\ \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx & (Continuous) \end{cases}$$

Key** CENTRAL **Moments

$$M_0 = 1$$
$$M_1 = 0$$
$$M_2 = m_2 - m_1^2 = \sigma^2$$

1.2.3 Absolute Moments

The r^{th} absolute moment of a random variable X is:

$$M_r(X) = E(|X|^r) = \begin{cases} \sum_{x_i} |x_i|^r P_i & (Discrete) \\ \int_{-\infty}^{+\infty} |x|^r f(x)dx & (Continuous) \end{cases}$$

Example:
Let X be a discrete random variable with the following probability distribution:

X	1	2	3	4
$P(X = x_i)$	0.1	0.4	0.3	0.2

1. Calculate the raw moments (moments about the origin) of orders 0, 1, 2, and 3.
2. Calculate the central moments (moments about the mean) of orders 0, 1, 2, and 3.

Solution:

1- Calculating the Raw Moments

The r^{th} raw moment of a discrete random variable is given by:

$$m_r(X) = E(X^r) = \sum_{x_i} x_i^r P_i = \sum_{x_i} x_i^r P(X = x_i)$$

Order 0:

$$m_0 = E[X^0] = E[1] = 1$$

Order 1:

$$m_1 = E(X) = (1)(0.1) + (2)(0.4) + (3)(0.3) + (4)(0.2) = 0.1 + 0.8 + 0.9 + 0.8 = 2.6$$

Order 2:

$$m_2 = E(X^2) = (1^2)(0.1) + (2^2)(0.4) + (3^2)(0.3) + (4^2)(0.2) = 0.1 + 1.6 + 2.7 + 3.2$$
$$= 7.6$$

Order 3:

$$m_3 = E(X^3) = (1^3)(0.1) + (2^3)(0.4) + (3^3)(0.3) + (4^3)(0.2) = 0.1 + 3.2 + 8.1 + 12.8$$
$$= 24.2$$

2- Calculating the Central Moments

The r^{th} central moment of a discrete random variable is given by:

$$M_r(X) = E[(X-\mu)^r] = \sum_{x_i} (x_i - \mu)^r P_i$$

where $\mu = E(X) = 2.6$

Order 0:

$$M_0 = E[(X-\mu)^0] = E[1] = 1$$

Order 1:

$$M_1 = E(X-\mu) = E(X) - \mu = 2.6 - 2.6 = 0$$

Order 2 (Variance):

$$M_2 = E[(X-\mu)^2] = \sum (x-2.6)^2 P(X=x)$$
$$= (1-2.6)^2(0.1) + (2-2.6)^2(0.4) + (3-2.6)^2(0.3) + (4-2.6)^2(0.2)$$
$$= (2.56)(0.1) + (0.36)(0.4) + (0.16)(0.3) + (1.96)(0.2)$$
$$= 0.256 + 0.144 + 0.048 + 0.392 = 0.84$$

Order 3:

$$M_3 = E[(X-\mu)^3] = \sum (x-2.6)^3 P(X=x)$$
$$= (1-2.6)^3(0.1) + (2-2.6)^3(0.4) + (3-2.6)^3(0.3) + (4-2.6)^3(0.2)$$
$$= (-4.096)(0.1) + (-0.216)(0.4) + (0.064)(0.3) + (2.744)(0.2)$$
$$= -0.4096 - 0.0864 + 0.0192 + 0.5488 = 0.072$$

1.3 The relationship between central moments and raw moments

While raw moments provide insights into the distribution's structure relative to the origin, central moments (computed about the mean) reveal key properties such as variance, skewness, and kurtosis. There is a relationship between central moments and raw moments. Specifically, we can express the r^{th} central moment (centered around the mean $\mu = E(X)$) in terms of raw moments using the following expansion:

$$M_r(X) = E([X - E(X)]^r) = E\left[\sum_{j=0}^{r} C_r^j . X^j . (-E(X))^{r-j}\right]; r \in \mathbb{N}$$

Knowing that E[X]=μ is a constant, we can express the r^{th} central moment in terms of raw moments as follows:

$$M_r(X) = E([X - E(X)]^r) = \sum_{j=0}^{r} (-1)^{r-j} C_r^j . E(X^j)[E(X)]^{r-j}; r \in \mathbb{N}$$

Then

$$M_r(X) = E([X - E(X)]^r) = \sum_{j=0}^{r} (-1)^{r-j} C_r^j . m_j m_1^{r-j}; r \in \mathbb{N}$$

2. Moment Generating Functions (MGFs)

2.1 Definition

The moment-generating function is a real-valued function of a real variable that can be associated with certain real random variables and is directly related to the moments.

Let *X* be a real random variable. The moment-generating function of *X*, denoted m_X, is the function of the real variable *t* defined by:

$$M_x(t) = E[e^{tX}]$$

on the set of values of *t* for which this mathematical expectation exists.

– **Expression in the discrete case:** If *X* takes values x_k with probabilities p_k:

$$M_x(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tX} p(X = x)$$

– **Expression in the absolutely continuous case:** If X has the density *f(x)*:

$$M_x(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tX} f(x) dx$$

When the random variable is bounded, the moment-generating function exists and is continuous for all *t*.

If the random variable X has moments of all orders, and we have:

$$M_x(t) = 1 + E(X)\frac{u}{1!} + E(X^2)\frac{u}{2!} + \ldots + E(X^k)\frac{u}{k!} + \ldots$$

This implies in particular that for all $r \geq 1$, we have $E(X^r) = M_X^{(r)}(0)$.

Behavior under affine transformation:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Behavior under addition of independent random variables:

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

Despite the direct relationship between the moment-generating function and the moments, mathematicians prefer working with the characteristic function $E(e^{itX})$, which can be defined for any random variable *X*, is bijective, and exists for all values of *t*.

2.2 Key Properties

Uniqueness Theorem: If MGFs exist and match, distributions are identical
Derivatives Give Moments:

$$M_r = \frac{d^r}{dt^r} M(t)|_{t=0} = E(X^r) = M_X^{(r)}(0)$$

2.3 MGF under Bernoulli distribution

$$X \sim \mathcal{B}ernoulli(1, p) \Rightarrow X = \{0,1\}$$

$$P(X = x_i) = \begin{cases} p & if x = 1(success) \\ q = 1 - p & if x = 0(failure) \end{cases}$$
$$E(X) = V(X) = p$$
$$m(t) = q + p.e^t$$

MGF

$$m(t) = E(e^{tX}) = \sum_{x_i} e^{tX} P(X = x_i)$$
$$m(t) = e^{tX} P(X = 0) + e^{tX} P(X = 1)$$
$$m(t) = e^{t0} q + e^{t1} p$$
$$m(t) = q + p.e^t$$
$$E(X) = V(X) = p$$

2.4 MGF under Binomial distribution

$$X \sim \mathcal{B}inomial(n, p) \Rightarrow X = \{0,1, \dots, n\}$$
$$P(X = x_i) = \binom{n}{x} p^x q^{n-x} = C_n^x p^x q^{n-x}$$
$$P + q = 1$$

$$E(X) = np$$
$$V(X) = npq$$
$$m(t) = (q + p.e^t)^n$$

MGF

$$m(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tX} . C_n^x p^x q^{n-x}$$
$$m(t) = \sum_{x=0}^{n} C_n^x . (pe^t)^x . q^{n-x} = (q + pe^t)^n$$
$$E(X) = np$$
$$V(X) = npq$$

Example:

Let X be a continuous random variable following a uniform distribution on the interval [0,1], denoted as $X \sim Uniform(0,1)$.

1- Find the first two raw moments of X.
2- Compute the variance of X using the moments.
3- Calculate the moment generating function (MGF) of X.

Solution:

The probability density function (PDF) of X is:

$$f(x) = \begin{cases} 1 & for\ 0 \leq x \leq 1 \\ 0 & Otherwise \end{cases}$$

3- Calculating the First Two Raw Moments

The *r*

The r^{th} raw moment is $m_r = E(X^r) = \int_{-\infty}^{+\infty} x^r f(x) dx$

The First Raw Moment ($r = 1$):

$$m_1 = E(X) = \int_0^1 x \cdot 1 dx = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

The Second Raw Moment ($r = 2$):

$$m_2 = E(X^2) = \int_0^1 x^2 \, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

2- Computing the Variance Using Moments
The variance can be computed using the formula:

$$\mathrm{Var(X)} = E(X^2) - [E(X)]^2$$

Substitute the moments:

$$Var(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{4-3}{12} = \frac{1}{12}$$

4- Calculating the Moment Generating Function (MGF)
The moment generating function is defined as:

$$M_x(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tX} f(x) dx$$

For the uniform distribution on [0,1]:

$$M_x(t) = \int_0^1 e^{tX} 1 \, dx = \int_0^1 e^{tX} \, dx$$

Case 1: $t \neq 0$

$$M_x(t) = \left[\frac{e^{tX}}{t}\right]_0^1 = \frac{\mathrm{e}^t - 1}{t}$$

Case 2: $t = 0$

$$M_x(0) = \int_0^1 e^0 \, dx = \int_0^1 1 \, dx = 1$$

Therefore, the MGF is:

$$M_x(t) = \begin{cases} \dfrac{\mathrm{e}^t - 1}{t} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0 \end{cases}$$

Chapter VII Exercises

Exercise 1:
Consider the following probability distribution for a discrete random variable X:

X	0	1	2	3
$P(X=x)$	0.2	0.3	0.4	0.1

1- Verify that this is a valid probability distribution.
2- Calculate the expected value $E(X)$ and the variance $V(X)$.
3- If $Y=4X-2$, find $E(Y)$ and $V(Y)$.
4- Find the moment generating function (MGF) of X, $M_X(t)$.
5- Using the MGF, verify your calculated values for $E(X)$ and $V(X)$.

Solution:
1- Verifying that this is a valid probability distribution
A valid probability distribution must satisfy two conditions:
- Each probability $P(X=x)$ must be between 0 and 1:
- The sum of all probabilities must equal 1.

Checking the probabilities: $0.2, 0.3, 0.4, 0.1$ are all between 0 and 1.
The sum of probabilities: $0.2+0.3+0.4+0.1=1$
So, this is a valid probability distribution.

2- Calculating E(X) and V(X)

The Expected Value E(X):
We use the formula:

$$E(X)=\sum_{i=1}^{n} \mathrm{x}_i \mathrm{P}(\mathrm{X}=\mathrm{x}_i)$$

X	0	1	2	3	$\sum$
$P(X=x)$	0.2	0.3	0.4	0.1	1
$E(X)$	0	0.3	0.8	0.3	1.4

$$E(X)=1.4$$

The Variance V(X):
We use the formula:

$$V(X)=E(X^2)-[E(X)]^2$$

$$E(X^2)=\sum \mathrm{x}_i{}^2 \mathrm{P}(\mathrm{X}=\mathrm{x}_i)$$

X	0	1	2	3	$\sum$
$P(X=x)$	0.2	0.3	0.4	0.1	1
$E(X^2)$	0	0.3	1.6	0.9	2.8

$$V(X)=E(X^2)-[E(X)]^2=2.8-(1.4)^2=2.8-1.96=0.84$$

3- Finding E(Y) and V(Y) if $Y=4X-2$

Using the linearity properties of expectation and variance:

$$E(aX + b) = aE(X) + b$$
$$V(aX + b) = a^2V(X)$$

Here, $a = 4$, $b = -2$.

Expected Value E(Y):

$$E(Y) = E(4X - 2) = 4E(X) - 2 = 4(1.4) - 2 = 5.6 - 2 = 3.6$$
$$E(Y) = 3.6$$

Variance V(Y):

$$V(Y) = V(4X - 2) = 4^2V(X) = 16 \times 0.84 = 13.44$$

4- Finding the moment generating function $M_X(t)$

The MGF is defined as $M_X(t) = \mathrm{E}(\mathrm{e}^{\mathrm{tx}})$

$$M_X(t) = \sum \mathrm{e}^{\mathrm{tx}} P(X = x)$$
$$M_X(t) = e^{t\cdot 0}(0.2) + e^{t\cdot 1}(0.3) + e^{t\cdot 2}(0.4) + e^{t\cdot 3}(0.1)$$
$$M_X(t) = 0.2 + 0.3e^t + 0.4e^{2t} + 0.1e^{3t}$$

5- Verifying E(X) and V(X) using the MGF

The first and second moments can be found from the MGF:

$$E(X) = M_X'(0)$$
$$E(X2) = M_X''(0)$$

First Derivative $M_X'(t)$:

$$M_X'(t) = \frac{d}{dt}(0.2 + 0.3e^t + 0.4e^{2t} + 0.1e^{3t}) = 0 + 0.3e^t + 0.8e^{2t} + 0.3e^{3t}$$

Evaluate at $t = 0$:

$$MX'(0) = 0.3(1) + 0.8(1) + 0.3(1) = 0.3 + 0.8 + 0.3 = 1.4$$

$E(X) = 1.4$ Matches the answer 2

Second Derivative $M_X''(t)$:

$$M_X''(t) = \frac{d}{dt}(0.3e^t + 0.8e^{2t} + 0.3e^{3t}) = 0.3e^t + 1.6e^{2t} + 0.9e^{3t}$$

Evaluate at $t = 0$:

$$M_X''(0) = 0.3(1) + 1.6(1) + 0.9(1) = 0.3 + 1.6 + 0.9 = 2.8$$

$\Rightarrow E(X^2) = 2.8$ Matches the answer 2

Variance Verification:

$V(X) = E(X^2) - [E(X)]^2 = 2.8 - (1.4)^2 = 2.8 - 1.96 = 0.84$ Matches the answer 2.

Exercise 2:

Let the probability density function of a continuous random variable X be given by:

$$f(x) = \begin{cases} \dfrac{2x}{9} & for\ 0 < x < 3 \\ 0 & otherwise \end{cases}$$

1- Calculate the raw moments of order 0, 1, 2, and 3.
2- Calculate the central moments of order 0, 1, 2, and 3.

Solution:

1- Calculating the Raw Moments

The r^{th} raw moment is defined as $E(X^r) = \int_{-\infty}^{+\infty} x^r f(x)dx$

Since $f(x)$ is non-zero only on]0,3[, we have:

$$E(X^r) = \int_0^3 x^r \frac{2x}{9} dx = \frac{2}{9}\int_0^3 x^{k+1} dx$$

Order 0 ($r = 0$):

$$E(X^0) = E(1) = \frac{2}{9}\int_0^3 x^1 dx = \frac{2}{9}\left[\frac{x^2}{2}\right]_0^3 = \frac{2}{9}\cdot\frac{9}{2} = 1$$

(This verifies the PDF integrates to 1.)

Order 1 ($r = 1$):

$$E(X^1) = \frac{2}{9}\int_0^3 x^2 dx = \frac{2}{9}\left[\frac{x^3}{3}\right]_0^3 = \frac{2}{9}\cdot\frac{27}{3} = \frac{2}{9}\cdot 9 = 2$$

Order 2 ($r = 2$):

$$E(X^2) = \frac{2}{9}\int_0^3 x^3 dx = \frac{2}{9}\left[\frac{x^4}{4}\right]_0^3 = \frac{2}{9}\cdot\frac{81}{4} = \frac{162}{36} = \frac{9}{2} = 4.5$$

Order 3 ($r = 3$):

$$E(X^3) = \frac{2}{9}\int_0^3 x^4 dx = \frac{2}{9}\left[\frac{x^5}{5}\right]_0^3 = \frac{2}{9}\cdot\frac{243}{5} = \frac{486}{45} = \frac{54}{5} = 10.8$$

2- Calculating the Central Moments

The r^{th} central moment is $M_r = E[(X-\mu)^r]$, where $\mu = E(X) = 2$

Order 0 ($r = 0$):

$$M_r = E[(X-2)^0] = E(1) = 1$$

Order 1 ($r = 1$):

$$M_1 = E[(X-2)^1] = E(X) - 2 = 2 - 2 = 0$$

(The first central moment equals always 0.)

Order 2 ($r = 2$): Variance

We can use the formula: $M_2 = E(X^2) - [E(X)]^2$

$$M_2 = 4.5 - (2)^2 = 4.5 - 4 = 0.5$$

Order 3 ($r = 3$):

$$M_3 = E[(X-2)^3] = E[X3 - 6X^2 + 12X - 8]$$

Using linearity of expectation:

$$\mu_3 = E(X^3) - 6E(X^2) + 12E(X) - 8$$
$$M_3 = 10.8 - 6(4.5) + 12(2) - 8 = 10.8 - 27 + 24 - 8 = -0.2$$

Exercise 3:

Let X be a random variable. Find the moment generating function (MGF) in the following cases:

1- X follows a Bernoulli distribution where $P(X = x) = p^x(1-p)^{1-x}$ for $x = 0,1$.

2- X follows a Poisson distribution where $P(X = x) = \frac{e^{-2}2^x}{x!}$ for $x = 0,1,2,...$

Solution:

1- MGF for the Bernoulli Distribution

The probability mass function (PMF) is:

$P(X = 0) = 1 - p$ and $P(X = 1) = p$

The moment generating function is defined as $M_X(t) = E(e^{t.x})$.

$$M_X(t) = \sum_x \sum e^{t.x} P(X = x) = e^{t.0}P(X = 0) + e^{t.1}P(X = 1)$$
$$M_X(t) = (1)(1-p) + e^t(p)$$
$$M_X(t) = 1 - p + pe^t$$

2- MGF for the Poisson Distribution

The probability mass function (PMF) is:

$$P(X = x) = \frac{e^{-2}2^x}{x!}, x = 0,1,2,...$$

The moment generating function is:

$$M_X(t) = E(e^{t.x}) = \sum_{x=0}^{\infty} e^{tX}\, P(X = x) = \sum_{x=0}^{\infty} e^{tX}\, \frac{e^{-2}2^x}{x!}$$

Combine the exponential terms:

$$M_X(t) = e^{-2} \sum_{x=0}^{\infty} \frac{(2e^t)^x}{x!}$$

Recognize that the sum is the power series expansion of the exponential function

$$e^a = \sum_{x=0}^{\infty} \frac{a^x}{x!}, where\ a = 2e^t$$
$$M_X(t) = e^{-2} \cdot e^{2e^t} = e^{2(e^t - 1)}$$

Exercise 4:

Let X be a continuous random variable with the following probability density function:

$$f(x) = \begin{cases} 2 & for\ 0 \le x \le 0.5 \\ 0 & otherwise \end{cases}$$

1- Find the moment generating function (MGF) of the random variable X.

2- Calculate the raw moment m_k for $k = 1$.

3- Calculate the central moments M_k for $k = 0,1$.

Solution:

1- Finding the Moment Generating Function (MGF)

The moment generating function is defined as

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tX} f(x)dx$$

Since $f(x) = 2$ for $0 \leq x \leq 0.5$ and 0 otherwise:

$$M_X(t) = \int_0^{0.5} e^{tX}.2dx = 2\int_0^{0.5} e^{tX}dx$$

Case 1: $t \neq 0$

$$M_X(t) = 2\left[\frac{e^{tx}}{t}\right]_0^{0.5} = \frac{2}{t}(e^{0.5t} - e^0) = \frac{2}{t}(e^{0.5t} - 1)$$

Case 2: $t = 0$

$$M_X(0) = \int_0^{0.5} e^0.2dx = 2\int_0^{0.5} 1dx = 2[x]_0^{0.5} = 2(0.5) = 1$$

$$M_X(t) = \begin{cases} \frac{2}{t}(e^{0.5t} - 1) & for\ t \neq 0 \\ 1 & for\ t = 0 \end{cases}$$

2- Calculating the raw moment m_r for $k = 1$

The r^th raw moment is $m_r = E(X^r) = \int_{-\infty}^{+\infty} x^r f(x)dx$

For $r = 1$:

$$m1 = E(X) = \int_0^{0.5} x \cdot 2dx = 2\int_0^{0.5} xdx = 2\left[\frac{x^2}{2}\right]_0^{0.5} = [x^2]_0^{0.5} = (0.5)^2 = 0.25$$

3- Calculating the central moments M_r for $k = 0{,}1$

The r^{th} central moment is $M_r = E[(X-\mu)^r]$, where $\mu = E(X) = 0.25$.

For $r = 0$:

$$M_0 = E[(X - 0.25)^0] = E(1) = 1$$

For $r = 1$:

$$M_1 = E[X - 0.25] = E(X) - 0.25 = 0.25 - 0.25 = 0$$

Chapter VIII: Markov's Inequality - Bienaymé-Chebyshev's Inequality - Law of Large Numbers

1. Markov's Inequality

Expectation and variance are closely related to the underlying distributions of random variables. This relationship allows us to prove certain inequalities that are often very useful, such as, Markov's inequality, Chebychev's inequality, Cauchy–Schwartz inequality…
Markov's Inequality is a fundamental result in probability theory that provides an upper bound for the probability that a non-negative random variable exceeds a certain value. It's one of the simplest and most widely used probability inequalities.

1.1 Statement of Markov's Inequality

Theorem: Let X be a non-negative random variable ($P(X \geq 0) = 1$) with finite expectation $E(X)$. Then for any $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Markov's inequality is very useful when the mean is known, for non-negative random variables, and allows for obtaining rough bounds quickly. However, this inequality is not very precise because it does not allow for the calculation of lower bounds, the bound can be very loose, and it is only useful for $a > E(X)$ (otherwise the bound is greater than or equal to 1).

The Markov's inequality only makes sense if the random variables are non-negative because negative values would make the bound meaningless.

Example 1:
Let X be the height of a tree with $E(X) = 10$ meters. What's the probability a tree exceeds 25 meters?

Solution:
$E(X) = 10$ meters (average tree height)
We want $P(X \geq 25)$
$a = 25$
Markov's Inequality states that for a non-negative random variable X and any $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

$$P(X \geq 25) \leq \frac{10}{25} = 0.4$$

This means that at most 40% of trees exceed 25 meters in height.
This is an upper bound, not the exact probability
The actual probability could be much lower than 40%
Markov's inequality gives the worst-case scenario
This bound is valid for any distribution of tree heights, as long as trees are non-negative and have mean height 10 meters

Example 2:
A factory produces components with average lifetime $E(X) = 1000\ hours$. What fraction fails before 200 hours?

Solution:

We want the probability that a component fails before 200 hours:

$$P(X \leq 200)$$

where X is the component's lifetime.

Markov's Inequality states that for a non-negative random variable X and any $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

This provides an upper bound for the probability that X is large.
We want $P(X \leq 200)$, but Markov gives $P(X \geq a)$. Use complement:

$$P(X \leq 200) = 1 - P(X > 200) \geq 1 - \frac{1000}{200} = 1 - 5 = -4$$

This shows Markov is not useful for lower bounds - it gives a negative probability, which is meaningless.
Markov's inequality cannot provide useful information for this question
Markov's inequality is designed to bound the probability that a random variable is large, not small. For lower tail probabilities (like early failure), we need different tools like Chebyshev's inequality or distribution-specific bounds.

2. Bienaymé-Chebyshev's Inequality: Theory and Applications

Among different inequalities (Markov's inequality, Chebychev's inequality, Cauchy–Schwartz inequality), Chebyshev's inequality stands out as an essential tool, linking variance to the concentration of a random variable around its mean. In this section, we will explore this inequality and its implications for bounding deviations from the expected value.
Chebyshev's Inequality is a key probability theorem that offers a conservative estimate of how far a random variable may deviate from its mean, applicable to any distribution (even unknown distributions).

2.1 Statement of Chebyshev's Inequality

The Bienaymé-Chebyshev theorem enables to calculate the probabilities for significant deviations of the random variable X from its mean.
Let X be a real random variable with mean μ ($\mu = E(X)$) and standard deviation σ ($\sigma^2 = Var(X)$). Then, for all $t > 0$, we have:

Chebyshev's Inequality states:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$

An equivalent alternative form (*Standard Form*), for all $a > 0$:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

(Where $\varepsilon = k\sigma$)

Another Complementary Form:

$$P(|X - \mu| \geq k\sigma) \geq 1 - \frac{1}{k^2}$$

This inequality allows us to prove the weak law of large numbers, and its theoretical relvance is therefore substantial.

However, if we want to use it numerically, its usefulness is limited to random variables X about which we know nothing (other than the existence and values of μ and σ. In such cases, this inequality is the only available information about the decay of $P(|X - \mu| \geq t\sigma)$ as t tends to infinity.

2.2 Interpretation and Implications

Chebyshev's Inequality guarantees that the probability of a data point being far from the mean is small, regardless of the underlying distribution.
For $k = 1$: The bound is 1 (100%), which is a useless, "trivial" guarantee. It doesn't tell us anything new, as all probabilities are ≤1 anyway.
For $k = 2$: No more than 25% of the data can lie beyond 2 standard deviations from the mean. This is a powerful and non-obvious fact for any dataset.
For $k = 3$: No more than 11% of the data can lie beyond 3 standard deviations. This is often much more conservative than the 0.3% predicted by a Normal distribution, showing it's a "worst-case" bound.
For $k = \sqrt{2}$: This is a special case showing that at least 50% of the data (the median) always lies within about 1.414 standard deviations of the mean.
These statements demonstrates that the further we go from the mean, the proportion of data found there must drop off quickly, no matter how strange or skewed the data distribution is.

This shows that Chebyshev's bound is very conservative when additional information (e.g., normality) is available.

2.3 Proof of Chebyshev's Inequality

Using Markov's Inequality on the non-negative random variable $(X - \mu)^2$:

1. Markov's Inequality: For $Y \geq 0$, $P(Y \geq a) \leq \frac{E(Y)}{a}$
2. Apply to $Y = (X - \mu)^2$ with $a = (k\sigma)^2$:

$$P((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{E[(X - \mu)^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

3. Result follows since $(X - \mu)^2 \geq k^2\sigma^2 \Leftrightarrow |X - \mu| \geq k\sigma$

2.4 Improved Versions and Related Inequalities

2.4.1 One-Sided Chebyshev Inequality (Cantelli's Inequality)

$P(X - \mu \geq k\sigma) \leq \frac{1}{1+k^2}$

2.4.2 Multivariate Chebyshev Inequality

For random vectors with covariance matrix Σ:

$$P((X-\mu)^t\Sigma^{-1}(X-\mu) \geq k^2) \leq \frac{n}{k^2}$$

(Where n = dimension)

Example: A manufacturing process produces items where a key quality characteristic has a mean (μ) of 50 units and a standard deviation (σ) of 5 units. The exact distribution of this characteristic is unknown.
Using probability theory, determine upper bounds for the following probabilities:

1- A randomly selected item has a measurement outside the range of 40 to 60 units.
2- A randomly selected item has a measurement exceeding 65 units.

Solution:
Chebyshev's Inequality states that for any random variable X with mean μ and standard deviation σ, the probability that X is more than k standard deviations away from the mean is at most $\frac{1}{k^2}$.

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

1- Probability a measurement is outside 40-60

The interval is $[40, 60]$. The mean is $\mu = 50$.
Finding k:
The distance from the mean to either bound is 10.
So, $k\sigma = 10$
Then, $k \times 5 = 10 \Rightarrow k = 2$

The interval $[40, 60]$ is exactly $\mu \pm 2\sigma$.
We Apply Chebyshev's Inequality:
We want $P(|X-50| \geq 10)$, which is the probability of being outside the interval.

$$P(|X-50| \geq 10) = P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = \frac{1}{4} = 0.25$$

The probability a measurement is outside 40-60 is at most 0.25 (or 25%).

2- Probability a measurement is above 65

The value is 65. The mean is $\mu = 50$.
Finding k:
The distance from the mean to 65 is 15.

$$k\sigma = 15$$
$$k \times 5 = 15 \Rightarrow k = 3$$

So, 65 is 3 standard deviations above the mean.
We apply Chebyshev's Inequality:
The standard Chebyshev inequality gives a bound for both tails (above and below the mean). We can derive a one-sided bound.

$$P(X-\mu \geq k\sigma) \leq \frac{1}{1+k^2}$$

Alternatively, the standard two-sided inequality also gives a valid, though less precise, bound:

$$P(X \geq 65) = P(X-\mu \geq 15) \leq P(|X-\mu| \geq 15) \leq \frac{1}{3^2} = \frac{1}{9} \approx 0.111$$

Using the one-sided inequality:

$$P(X \geq 65) \leq \frac{1}{1+3^2} = \frac{1}{10} = 0.1$$

Using the more precise one-sided bound, the probability a measurement is above 65 is at most 0.1 (or 10%).

3. Law of Large Numbers

This section explores convergence in probability and its major implications, especially the laws of large numbers. Historically, the law of large numbers refers to the observation that the frequency of an event converges toward its probability as the number of independent trials increases indefinitely. Initially viewed as a natural law, it became a rigorous mathematical theorem.

3.1 Definition of the Law of Large Numbers (LLN)

The Law of Large Numbers is a fundamental theorem in probability that describes how the average of a large number of trials converges to the expected value as more trials are performed.

$$\bar{X}_n \to \mu$$

in probability as $n \to \infty$.

In this section, we focus on the behavior of arithmetic means. We consider a sequence (X_n) of independent and identically distributed (i.i.d.) real-valued random variables with mathematical expectation μ. The sample average, denoted by $\bar{X}_n$, is computed as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

As the sample size *n* grows indefinitely, $\bar{X}_n$ converges to the expected value μ (by the Law of Large Numbers). The results related to this problem are called the Laws of Large Numbers (LLN).

- Expected Value (μ): Theoretical mean of the random variable $E(X)$

This constitutes the formal definition of almost sure convergence of the sequence (X_n) toward the constant random variable μ as n tends to ∞, i. e. as the sample size n tends to infinity ($n \to \infty$), the sample mean $\bar{X}_n$ converges to the population mean μ ($\bar{X}_n \to \mu$). While this result is deeply technical—chiefly employed by mathematicians in theoretical contexts—it parallels the weak law's evolution: throughout the 20th century, mathematicians derived progressively refined versions of the strong law under weaker assumptions.

3.2 Types of Law of Large Numbers

3.2.1 The Weak Law of Large Numbers (WLLN)

Let $\{Xn; n \geq 1\}$ be a sequence of real-valued, independent, identically distributed (i.i.d.), and square-integrable random variables. Let $\mu = E(X_1)$ (finite mean); then,

$$\bar{X}_n \to \mu \text{ as } n \to \infty$$

Then, for every $\varepsilon > 0$, we have:

$$P(|\bar{X}_n - \mu| > \varepsilon) \to 0 \; when \; n \to \infty$$

Convergence in probability:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad for \; any \; \varepsilon > 0$$

This is precisely the definition of the convergence in probability of the sequence (X_n) to the (constant) random variable μ as n approaches infinity. The first proof was provided by Bernoulli for the special case where the X_n are Bernoulli random variables (though he did not call them that!), and their sum is a binomial random variable: if the Bernoulli random variable is the indicator of an event, this corresponds to the convergence of the frequency to the event's probability. Several versions of the weak law of large numbers with relaxed conditions were proven in the 20th century (e.g., without assuming the X_n are independent, but only that their pairwise covariances are zero, or by dropping the requirement of identical distributions).

3.2.2 The Strong Law of Large Numbers (SLLN)

For the second family of the laws of large numbers, we have Kolmogorov's strong law of large numbers:
We consider a sequence (X_n) of independent and identically distributed (i.i.d.) real-valued random variables with mathematical expectation μ. We define the sample means:

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Then
$\bar{X}_n \xrightarrow{a.s.} \mu$ when $n \to \infty$
The sequence of random variables X_n converges almost surely to μ as $n \to \infty$, denoted:

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$$

The SLLN presents a stronger condition than WLLN.
- Requirements:
  - i.i.d. random variables
  - Finite mean $E(|X|) < \infty$ (same as WLLN)

3.2.3 Comparison between WLLN and SLLN

The Weak Law of Large Numbers guarantees that for a large sample, the sample average will be probably close to the expected value, making it very useful for most practical statistics. The Strong Law provides a stricter, more powerful guarantee that the sample average will almost surely converge permanently to the expected value, which is essential for advanced theoretical proofs but is often more than is needed for everyday applications.

3.3 Connection of Chebyshev's Inequality to the Law of Large Numbers

Chebyshev's Inequality directly proves the Weak Law of Large Numbers:
For sample mean $\bar{X}_n$ of i.i.d. variables with mean μ.
Using Chebyshev's Inequality, if $Var(X) = \sigma^2 < \infty$:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \; as \; n \to \infty$$

proving WLLN.

Example 1:
You have a fair coin. The theoretical probability of getting Heads is $p = 0.5$.

1- You flip the coin 10 times and get 7 Heads.

2- Later, you flip the same coin 1,000 times and get 505 Heads.

Solution:

1- After 10 flips:

Experimental probability of $Heads = \frac{7}{10} = 0.7$This is quite different from the true probability (0.5).

2- After 1,000 flips:

Experimental probability of $Heads = \frac{505}{1000} = 0.505$This is very close to the true probability (0.5).

This demonstrates the Law of Large Numbers. With a small number of trials, the experimental result can be far from the theoretical probability. But as the number of trials increases, the experimental average converges to the true expected value.

Example 2:

Let Xi represent the outcome of rolling a fair six-sided die, with theoretical mean $\mu = 3.5$. Consider two experiments:

Experiment A: 100 rolls are recorded, with a sample mean of $\bar{X}_{100} = 3.4$

Experiment B: 1,000,000 rolls are recorded, with a sample mean of $\bar{X}_{1,000,000} = 3.50001$

1- Calculate the absolute error between the sample mean and theoretical mean for each experiment.
2- Which experiment provides a more accurate estimate of the true population mean? Explain why using the appropriate statistical law.
3- If you conducted a third experiment with only 10 rolls, would you expect the sample mean to be closer to 3.5 than Experiment B's result? Justify your answer.

Solution:

1- Absolute error calculation:

Experiment A: $| 3.5 - 3.45 |= 0.05$

Experiment B: $| 3.5 - 3.50001 |= 0.00001$

2- More accurate estimate:

Experiment B provides a more accurate estimate of the true population mean.

This demonstrates the Law of Large Numbers, which states that as the sample size increases, the sample mean converges to the theoretical population mean.

3- Third experiment prediction:

No, with only 10 rolls, the sample mean would likely be farther from 3.5 than Experiment B's result.

The small sample size of 10 rolls would be more susceptible to random variation and sampling error, making it less reliable than the result from 1,000,000 rolls.

Example 3:

A statistics student is conducting a probability experiment by rolling a fair six-sided die 10,000 times. The theoretical expected value (mean) of a single die roll is 3.5.

Using the Law of Large Numbers, estimate the probability that the calculated average of all 10,000 rolls will fall between 3.4 and 3.6.

Explain your reasoning based on the principles of the Law of Large Numbers, rather than performing detailed probability calculations.

Solution:
Let X_i be the outcome of the i^{th} roll of a fair die.
The possible outcomes are $X_i \in \{1,2,3,4,5,6\}$.
The theoretical mean (expected value) for a fair die is:

$$\mu = E[X_i] = \frac{1+2+3+4+5+6}{6} = 3.5$$

The sample mean after n rolls is:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The sample size is $n = 10{,}000$.

The Law of Large Numbers states that as the number of trials n becomes large, the sample mean $\bar{X}_n$ converges to the population mean μ.

For a very large n (like $n = 10{,}000$), the LLN implies that $\bar{X}_n$ will be very close to $\mu = 3.5$ with a very high probability.

We are asked for $P(3.4 \leq \bar{X}_n \leq 3.6)$.

The interval [3.4, 3.6] is symmetric around the mean $\mu = 3.5$. The deviation from the mean is $3.6 - 3.5 = 0.1$.

According to the LLN:
For a small number of rolls, the average could easily fall outside this interval.
For a very large number of rolls like $n = 10{,}000$, the average is extremely likely to be very close to 3.5.
Therefore, the probability that the average is within 0.1 units of the true mean is very high.
Based on the direct application of the Law of Large Numbers:
The probability is very close to 1.

The Law of Large Numbers does not give us the exact probability, but it assures us that for a large sample size of $n = 10{,}000$, it is almost certain that the sample average will be very close to the theoretical average of 3.5. The interval $[3.4, 3.6]$ is quite narrow and centered on 3.5, so we can be highly confident the result will fall within it.

Chapter VIII Exercises

Exercise 1:
In a large statistics class, the average score on the final exam is 75%. The exam scores are all non-negative values.
Using probability theory, determine the maximum possible proportion of students who could have scored above 90% on this exam.

Solution:
The average test score is 75%: $E(X) = 75\%$
We want $P(X > 90\%)$
So $a = 90$
We apply Markov's Inequality
Markov's Inequality states that for a non-negative random variable X and any $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

$$P(X \geq 90) \leq \frac{75}{90} = \frac{5}{6} \approx 0.833$$

This is an upper bound, not the exact proportion
The actual proportion could be much lower than 83.3%
Markov's inequality gives the worst-case scenario
This bound is valid for any distribution of test scores, as long as scores are non-negative and have mean 75%
The bound is quite loose here since typically we'd expect a much smaller proportion of students to score above 90% when the average is 75%

Exercise 2:
At a customer service center, the average waiting time is 15 minutes. Waiting times are always non-negative.
Using probability theory, determine an upper bound for the probability that a randomly selected customer waits more than 1 hour (60 minutes).

Solution:
The average waiting time : $E(X) = 15\ minutes.$
We want $P(X > 60\ minutes)$
So, $a = 60$
We apply Markov's Inequality
Markov's Inequality states that for a non-negative random variable X and any $a > 0$:

$$P(X \geq 60) \leq \frac{15}{60} = \frac{1}{4} = 0.25$$

This is an upper bound, not the exact probability
The actual probability could be much lower than 25%
Markov's inequality gives the worst-case scenario
This bound is valid for any distribution of waiting times, as long as waiting times are non-negative and have mean 15 minutes
In most realistic queuing systems, the actual probability of waiting more than 1 hour would be much smaller than 25% when the average wait is only 15 minutes.

Exercise 3:

A factory produces electrical resistors. The resistance of these resistors (in ohms, Ω) is a random variable X with a known mean $\mu = 100\,\Omega$ and a standard deviation $\sigma = 5\,\Omega$. However, the exact probability distribution of X is unknown.

1- Using Chebyshev's Inequality, find the minimum probability that a randomly selected resistor has a resistance between 90 Ω and 110 Ω.

2- The quality control department requires that at least 96% of the resistors must lie between $\mu - k$ and $\mu + k$. Use Chebyshev's Inequality to find the smallest value of k (in ohms) that guarantees this, regardless of the actual distribution.

3- If it were later discovered that the resistance is, in fact, normally distributed, would the actual probability for part (1) be higher or lower than the bound you calculated? Explain why in one sentence.

Solution:

1- Chebyshev's Inequality states that for any random variable X with mean μ and standard deviation σ, and for any $k > 0$:

$$P(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2}$$

Equivalently, the probability that X lies within k standard deviations of the mean is:

$$P(\mid X - \mu \mid < k\sigma) \geq 1 - \frac{1}{k^2}$$

We are given the interval $90 \leq X \leq 110$. The mean is $\mu = 100$.
The distance from the mean to either bound is $100 - 90 = 10$, or $110 - 100 = 10$.
So, $\mid X - \mu \mid \leq 10$.

We find k such that $k\sigma = 10$. Given $\sigma = 5$:

$$k \times 5 = 10 \Rightarrow k = 2$$

Now apply Chebyshev's Inequality:

$$P(\mid X - 100 \mid < 10) = P(\mid X - 100 \mid < 2\sigma) \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 0.75$$

2- We require that at least 96% of resistors lie within the interval $[\mu - k, \mu + k]$. This means:

$$P(\mid X - \mu \mid \leq k) \geq 0.96$$

We relate k to the standard deviation: let $k = m\sigma$, where mm is the number of standard deviations.

From Chebyshev's Inequality:

$$P(\mid X - \mu \mid < m\sigma) \geq 1 - 1m2$$

We want this probability to be at least 0.96:

$$1 - 1m2 \geq 0.96$$

Solving for m:

$$1 - 0.96 \geq \frac{1}{m^2}$$

$$0.04 \geq \frac{1}{m^2}$$
$$m^2 \geq \frac{1}{0.04} = 25$$
$$m \geq 5$$

Now find k:

$$k = m\sigma = 5 \times 5 = 25\,\Omega$$

3- The smallest value of k that guarantees at least 96% of resistors are within $\mu \pm k$ is 25 Ω.

If the resistance is normally distributed, the actual probability for question (1) would be higher. For a normal distribution, the probability of being within 2 standard deviations of the mean is about 95%, which is much higher than the 75% lower bound provided by Chebyshev's Inequality, which is designed for any distribution and is thus more conservative.

Because the normal distribution concentrates more probability near the mean than the worst-case distribution assumed by Chebyshev's theorem.

Exercise 4:

The heights of a group of students follow a distribution with a mean of $\mu = 165\,cm$ and a standard deviation of $\sigma = 6\,cm$. Use Chebyshev's inequality to find an upper bound for the probability that a randomly selected student's height differs from the mean by at least 12 cm.

Solution:

We want the probability that the height *X* is outside the interval $[\mu - 12, \mu + 12]$.

That is, $P(|\,X - \mu\,| \geq 12)$.

Chebyshev's inequality states that for any random variable *X* with mean μ and standard deviation σ, and for any $k > 0$:

$$P(|\,X - \mu\,| \geq k\sigma) \leq \frac{1}{k^2}$$

We are given:

$$\mu = 165\,cm$$
$$\sigma = 6\,cm$$

The deviation from the mean is 12 cm

Find k such that:

$$k\sigma = 12$$
$$k \times 6 = 12 \Rightarrow k = \frac{12}{6} = 2$$

Substitute $k = 2$ into Chebyshev's inequality:

$$P(|\,X - \mu\,| \geq 12) \leq \frac{1}{2^2} = \frac{1}{4} = 0.25$$

The upper bound for the probability that a student's height differs from the mean by at least 12 cm is 0.25 (or 25%).

Note: Chebyshev's inequality provides a conservative upper bound that works for any distribution with a known mean and standard deviation. The actual probability is likely to be lower, especially if the distribution is normal.

Exercise 5:
Assume that a factory's daily production of a medical product is a random variable *X* with an expected value (mean) of $\mu = 950$ boxes and a variance of $\sigma^2 = 400$.

- Calculate the probability that the factory's daily production is between 900 and 1000 boxes.

Solution:
We know the mean $\mu = 950$ and variance $\sigma^2 = 400$, so the standard deviation is

$$\sigma = \sqrt{400} = 20$$

We want $P(900 \leq X \leq 1000)$.
Notice that this interval is symmetric around the mean:
Lower bound: $950 - 50 = 900$
Upper bound: $950 + 50 = 1000$
So, the interval is $\mu - 50 \leq X \leq \mu + 50$

The deviation from the mean is 50 boxes.
We express this in terms of standard deviations:

$$k\sigma = 50$$
$$k \times 20 = 50 \Rightarrow k = 5020 = 2.5$$

So, the interval [900,1000] is the same as $[\mu - 2.5\sigma, \mu + 2.5\sigma]$

We want $P(900 \leq X \leq 1000) = P(|\, X - \mu \,| \leq 50)$

Chebyshev's inequality gives a bound for the complement:

$$P(|\, X - \mu \,| \geq k\sigma) \leq \frac{1}{k^2}$$

Therefore, the probability of being within *k* standard deviations is:

$$P(|\, X - \mu \,| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Substitute $k = 2.5$:

$$P(|\, X - 950 \,| < 50) \geq 1 - \frac{1}{(2.5)^2} = 1 - \frac{1}{6.25} = 1 - 0.16 = 0.84$$

The probability that the daily production is between 900 and 1000 boxes is at least 0.84 (or 84%).

This is a lower bound using Chebyshev's inequality, which works for any distribution. The actual probability could be higher, especially if the production follows a distribution that concentrates more mass around the mean (like the normal distribution).

Exercise 6:

Suppose the number of products manufactured in a factory during a week is a random variable with a mean of 120 and a standard deviation of 15.

- What is the probability that the production for a given week differs from the mean by more than 30 units?

Solution:
We are looking for $P(|X - \mu| > 30)$, where:
$$\mu = 120$$
$$\sigma = 15$$
The deviation from the mean is 30

Chebyshev's inequality states that for any random variable X with mean μ and standard deviation σ, and for any $k > 0$:
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

First, find k such that $k\sigma = 30$:
$$k \times 15 = 30 \Rightarrow k = 2$$

We Calculate the Probability Bound
Substitute $k = 2$ into Chebyshev's inequality:
$$P(|X - \mu| \geq 30) \leq \frac{1}{2^2} = \frac{1}{4} = 0.25$$

The probability that the weekly production differs from the mean by more than 30 units is at most 0.25 (or 25%).
Note: Chebyshev's inequality provides an upper bound that holds for any probability distribution with a finite mean and variance. The actual probability may be lower.

Exercise 7:
The average daily energy consumption of a household in a neighborhood is 30 kWh, with a standard deviation of 5 kWh.

- Calculate the minimum probability that the daily energy consumption for a randomly selected household is between 20 and 40 kWh.

Solution:
We are given:
Mean $\mu = 30\ kWh$
Standard deviation $\sigma = 5\ kWh$
Interval: $[20, 40]\ kWh$

Notice that the interval is symmetric around the mean:
Lower bound: $30 - 10 = 20$
Upper bound: $30 + 10 = 40$
So, the interval is $\mu - 10 \leq X \leq \mu + 10$

The Number of Standard Deviations k
We express the deviation from the mean in terms of standard deviations:
$$k\sigma = 10$$
$$k \times 5 = 10 \Rightarrow k = 2$$

So, the interval $[20,40]$ corresponds to $[\mu - 2\sigma, \mu + 2\sigma]$.

- Calculating the Minimum Probability

We use Chebyshev's Inequality, which gives a lower bound for the probability of being within k standard deviations of the mean:

$$P(| X - \mu |< k\sigma) \geq 1 - \frac{1}{k^2}$$

Substitute $k = 2$:

$$P(| X - 30 |< 10) \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 0.75$$

The minimum probability is 0.75 (or 75%).

Exercise 8:

You have a fair coin. When you flip it, the probability of getting Heads is 0.5.

1- What is the experimental probability of getting Heads after 10 coin flips, and how does it compare to the true probability?

2- What is the experimental probability of getting Heads after 1000 coin flips, and how does it demonstrate the Law of Large Numbers?

Solution:

Case 1: Small Number of Flips

You flip the coin 10 times.

You get 7 Heads and 3 Tails.

The experimental probability of Heads is:

$$\hat{p}_{10} = \frac{7}{10} = 0.7$$

This is quite different from the true probability $p = 0.5$.

Case 2: Large Number of Flips

Now you flip the coin 1,000 times.

You get 510 Heads and 490 Tails.

The experimental probability of Heads is:

$$\hat{p}_{1000} = \frac{510}{1000} = 0.51$$

This is very close to the true probability $p = 0.5$.

The Law of Large Numbers tells us that as the number of trials increases, the experimental probability $\hat{p}_n$ gets closer to the true probability p.

Exercise 9:

You roll a fair six-sided die. The probability of rolling a 6 is $\frac{1}{6} \approx 0.167$.

1- What is the experimental probability of rolling a 6 after 30 rolls, and how does it compare to the theoretical probability of $\frac{1}{6} \approx 0.167$?

2- What is the experimental probability of rolling a 6 after 3000 rolls, and how does this result demonstrate the Law of Large Numbers?

Solution:
Case 1: Small Number of Rolls
You roll the die 30 times.
You get 2 sixes.
The experimental probability of rolling a 6 is:

$$\hat{p}_{30} = \frac{2}{30} \approx 0.067$$

This is quite different from the true probability $p \approx 0.167$.

Case 2: Large Number of Rolls
Now you roll the die 3,000 times.
You get 505 sixes.
The experimental probability of rolling a 6 is:

$$\hat{p}_{3000} = \frac{505}{3000} \approx 0.168$$

This is very close to the true probability $p \approx 0.167$.

The Law of Large Numbers tells us that as the number of trials increases, the experimental probability $\hat{p}_n$ gets closer and closer to the true theoretical probability p.

Exercise 10:
A bag contains 4 red marbles and 6 blue marbles. The probability of drawing a red marble is 0.4.

1- What is the experimental probability of drawing a red marble after 10 draws (with replacement), and how does it compare to the true probability of 0.4?
2- What is the experimental probability of drawing a red marble after 1000 draws (with replacement), and how does this result illustrate the Law of Large Numbers?

Solution:
Case 1: Small Number of Draws
You draw 10 marbles (with replacement)
You get 2 red marbles
Experimental probability:

$$\hat{p}_{10} = \frac{2}{10} = 0.2$$

This is quite different from the true probability $p = 0.4$

Case 2: Large Number of Draws
You draw 1,000 marbles (with replacement)
You get 398 red marbles
Experimental probability:

$$\hat{p}_{1000} = \frac{398}{1000} = 0.398$$

This is very close to the true probability $p = 0.4$
Thus, as the number of trials increases, the experimental results get closer to the theoretical probability, demonstrating the Law of Large Numbers.

Bibliography

Anderson, D. R., Sweeney, D. J., Camm, J. D., Williams, T. A., & Cochran, J. J. (2015). Statistiques pour l'économie et la gestion. De Boeck supérieur.

Bertsekas, D. P., & Tsitsiklis, J. N. (2000). Lecture notes for course 6.041/6.431 : Introduction to probability (Fall 2000). Massachusetts Institute of Technology.

Cantoni, E., Ronchetti, E., & Huber, P. (2006). Maîtriser l'aléatoire: Exercices résolus de probabilités et statistique. Paris: Springer Paris.

Carrieu, H. (2008). Probabilités : Exercices corrigés. EDP Sciences.

Delmas, J. F. (2013). Introduction au calcul des probabilités et à la statistique. Les Presses de l'ENSTA.

Dress, F. (2007). Les probabilités et la statistique de A à Z : 500 définitions, formules et tests d'hypothèse. Dunod.

Evans, M. J., & Rosenthal, J. S. (2024). Probability and statistics: The science of uncertainty (2nd ed.).

Floyd, J. E. (2010). Statistics for economists: A beginning. University of Toronto.

Hansen, B. E. (2022). Probability and statistics for economists. Princeton University Press.

Hong, Y. (2013). Probability and statistics for economists (Lecture notes). Cornell University.

Hurlin, C., & Mignon, V. (2022). Statistique et probabilités en économie-gestion (2e éd.). Dunod.

Kaji, T. (2023). Probability and Statistics Lecture notes. University of Chicago Booth School of Business.

Lecoutre, J. P. (2015). *TD Statistique et probabilités*: 6e édition. Dunod.

Lejeune Michel, *Statistique : La théorie et ses applications*, Springer, 2010.

Ross, S. M. (2020). *A first course in probability*. Harlow, UK: Pearson.

Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.

Scalas, E. (2008). Introduction to probability theory for economists (Lecture notes). Università del Piemonte Orientale Amedeo Avogadro.

Veysseyre, R. (2006). Aide-mémoire : Statistique et probabilités pour l'ingénieur (2e éd.). Dunod.

Weber, R. R. (2014). Probability (Lecture notes). University of Cambridge.