REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Djilali Bounaama, Khemis Miliana
Faculté des Sciences de la Matière et d'informatique
Conseil Scientifique de la Faculté

جامعة جيلالي بونعامة خميس مليانة
كليّة علوم المادة والإعلام الآلي
المجلس العلمي للكلية

Ref : 01 /..../CSF/ 2026

EXTRAIT DU PV
DE LA REUNION ORDINAIRE DU CONSEIL SCIENTIFIQUE
Du 02/02/2026

Objet : : Expertise de polycopié pédagogique

En l'an deux mille vingt-six (2026), le lundi 02 février à 13 h 30, une réunion ordinaire du Conseil Scientifique de la Faculté des Sciences de la Matière et de l'Informatique s'est tenue dans la salle de réunion de la faculté (Bloc B).

Suite aux rapports favorables reçus de la part des experts cités ci-après concernant l'expertise du polycopié pédagogique, le CSF a prononcé favorablement pour la conformité du polycopié pédagogique en vue de préparer son professorat.

- **Auteur du polycopié** : Dr. KELLECHE Abdelkarim (MCA)
- **Intitulé du polycopié**: Introduction to probabilities and descriptive statistic: Lessons and Solved Exercises.
- **Destiné aux étudiants de** : 1ére année Licence Mathématiques.
- **Experts du polycopié** :

  - HOUASNI Mohamed — MCA — UDB-Khemis Miliana
  - ABDALLAOUI Athmane — MCA — ENS-Boussaâda.

Président du
Conseil Scientifique de la Faculté SMI
Dr. BOUDERBALA Mihoub

رئيس المجلس العلمي
إمضاء: بودربالة ميهوب

المجلس العلمي
كلية علوم المادة و الإعلام الآلي

University Djilali Bounaama-Khemis Miliana
Faculty of Matter Sciences and Computer Science
Department of Mathematics

Coursebook:
Introduction to Probabilities and Descriptive Statistic: Lessons and Solved Exercises

For

First-year graduate students in Mathematics and Computer Science

Presented by:

Abdelkarim Kelleche

2025–2026

Introduction

This course is designed as a first graduate-level introduction to statistics, combinatorial analysis, and probability theory. It provides both the conceptual foundations and the practical methods needed to understand, organize, and interpret data, as well as to reason about randomness and uncertainty. The material is intended for students in mathematics, engineering, economics, and the sciences who wish to develop a solid grounding in statistical thinking and probabilistic reasoning.

Statistics, as presented here, is not merely a collection of techniques but a systematic approach to studying data — from observation and measurement to analysis and decision-making. Combinatorics and probability, on the other hand, form the mathematical framework that allows us to model random phenomena and quantify uncertainty. Together, these topics build a coherent pathway from data description to predictive analysis.

The book is organized into four main chapters, each building upon the previous one:

Chapter 1: Preliminaries This introductory chapter lays the foundation of descriptive statistics. It explains the essential vocabulary of the field — population, individual, variable, and modality — and distinguishes between qualitative and quantitative data. Various forms of data representation are presented, including statistical tables, bar charts, pie charts, and histograms. The chapter concludes with a series of solved exercises to reinforce basic concepts and familiarize students with data summarization techniques.

Chapter 2: Numerical Representation of Data The second chapter deepens the study of data analysis by introducing numerical measures of position (mean, median, quartiles) and dispersion (variance, standard deviation, range). Both discrete and continuous variables are examined, and students learn how to construct frequency distributions, cumulative frequency curves, and histograms. The chapter combines theory with numerous applied exercises to develop analytical and computational proficiency.

Chapter 3 combines combinatorial analysis and probability theory in a unified framework designed for solving past exam problems from Khemis Miliana University with full solutions. It introduces counting methods such as permutations, arrangements, and combinations, with or without repetition, to determine the number of possible outcomes in structured situations. These techniques are then used as a foundation for probability theory, which studies random experiments and events, along with their operations such as union, intersection, and complement. The chapter also presents the axioms of probability, followed by conditional probability, the law of total probability, and Bayes' theorem, all illustrated through exam-style problems to develop systematic problem-solving skills.

Chapter 4 is reserved for past exam problems from Khemis Miliana University with full solutions. It provides a structured set of exercises covering all key concepts of probability theory, allowing students to practice random experiments, event operations, and probability laws in real exam contexts. Each problem is solved step by step to strengthen understanding, improve accuracy, and prepare students for final examinations.

Throughout the textbook, theoretical explanations are supported by illustrative examples and step-by-step exercises, ensuring that students not only understand the concepts but also gain the ability to apply them effectively. The text balances rigor and clarity, making it suitable both for classroom instruction and for independent study.

Ultimately, this book aims to provide readers with a clear and coherent understanding of the tools that underpin modern data analysis and decision science — from the organization of data to the mathematical modeling of uncertainty.

Contents

List of Figures

Chapter 1

Preliminaries

1.1 Introduction

Statistics is the study of the collection of data, their analysis, their processing, the interpretation of the results and their presentation in order to make the data understandable by everyone. It is at the same time a science, a method and a set of techniques. Data analysis is used to describe the phenomena studied, make predictions and make decisions about them. In this way, statistics is an essential tool for understanding and managing complex phenomena. Statistics are useful in all disciplinary fields, from economics to biology through psychology and of course engineering sciences. The statistics consist of:

- Collect data.
- Present and summarize this data.
- Draw conclusions about the population studied and assist in decision-making.
- In the presence of time-dependent data, we try to make predictions.

1.2 Vocabulary

Descriptive statistics aim to study the characteristics of a set of observations such as the measurements obtained during an experiment. The experiment is the preliminary step to any statistical study. Generally, the statistical method is based on the following concept.

1.2.1 Statistical test

Definition 1.1. The **statistical test** is an experience that we provoke.

Example 1.1. A manufacturer of electric bulbs having the choice between 4 types of filaments intends to study the influence of the nature of the filament on the lifespan of the bulbs manufactured. To do this, he will have 4 samples of identical bulbs made, except for the filament, burn the bulbs until they go out, then compare the results obtained.

1.2.2 Population

In statistics, the term **population** applies to any statistical object studied, whether students (of a university or a country), households or any other group on which we make statistical observations.

Definition 1.2. We call **population** the group on which our statistical study relates. This set is denoted Ω.

Example 1.2.

1. We consider all the students in section A. We are interested in the number of brothers and sisters of each student. In this case Ω = all students.

2. If we now look at automobile traffic in a city, the population is then all the vehicles likely to be circulating in this city on a given date. In this case: Ω = all vehicles.

1.2.3 Individual (statistical unit)

A population is composed of individuals. The individuals who compose a statistical population are called **statistical units**.

Definition 1.3. We call **individual** any element of the population Ω, it is denoted $\omega \in \Omega$.

Example 1.3. If we study the annual production of a factory of metal beverage cans. The population is all the boxes produced during the year and a box constitutes an individual.

1.2.4 Character (statistical variable)

"Descriptive" statistics, as its name suggests, seeks to describe a given population. We are interested in the characteristic of units which can take different values.

Definition 1.4. We call **character** (or **statistical variable**, denoted S.V) any application

$$X : \Omega \to C$$

where C is the set of values of the character X (this is what is measured or observed on individuals).

Example 1.4. Height, temperature, nationality, eye color, professions.

1.2.5 Modalities

The modalities of a statistical variable are the different values that can take the statistical variable.

Example 1.5. Variable: "family situation". Modalities: single, married, divorced.

1.3 Types of Characters

We distinguish two categories of characters: qualitative characters and quantitative characters.

1.3.1 Qualitative character

Qualitative characters are those whose modalities cannot be ordered, that is to say if we consider two characters taken at random, we cannot say of one that it is inferior or equal to the other.

Definition 1.5. The elements of C cannot be represented by numbers.

Example 1.6. The condition of a house: we can consider the following modalities: ancient, degraded, new, renovated.

1.3.2 Quantitative character

Quantitative characters are those whose modalities can be ordered. Thus, the age, life size or salary of an individual are quantitative characteristics.

Definition 1.6. All values are represented by numbers. Likewise, it is divided into two kinds of characters, discrete and continuous.

Example 1.7.

1. The salary of factory employees. Modalities: 10000da, 20000da, etc. Type: discrete.
2. The rigidity of the springs. Modalities: [10, 20] N/m. Type: continuous.

1.4 Statistical Tables and Graphical Representations

Statistical tables and graphs are used to summarize and illustrate the results of data collection. They allow us to visualize how the values of a variable are distributed in a given population. We distinguish between the case of **qualitative variables** and **quantitative variables**.

1.4.1 Case of Qualitative Variables

Statistical tables and graphs are used to summarize and illustrate the results of data collection. They allow us to visualize how the values of a variable are distributed in a given population. We distinguish between the case of **qualitative variables** and **quantitative variables**.

1.4.2 Definition of a Statistical Table

Definition 1.7. A **statistical table** is an organized arrangement of data that summarizes the observed values of a variable and their corresponding frequencies. It allows a clear and structured presentation of information before any graphical representation or calculation of indicators (mean, variance, etc.).

A statistical table generally contains the following components:

- **Modalities (or classes):** the different values or categories taken by the studied variable.

- **Absolute frequency (n_i):** the number of times a modality or value occurs in the dataset.

- **Relative frequency (f_i):** the ratio of n_i to the total number of observations N:

$$f_i = \frac{n_i}{N}, \quad \text{with } \sum_i f_i = 1$$

- **Percentage frequency:** obtained by multiplying the relative frequency by 100:

$$\text{Percentage} = f_i \times 100$$

- **Cumulative frequency:** for ordered variables, it represents the sum of the frequencies of all modalities less than or equal to a given value. It can be expressed in absolute, relative, or percentage form.

We now distinguish between the case of **qualitative** and **quantitative** variables.

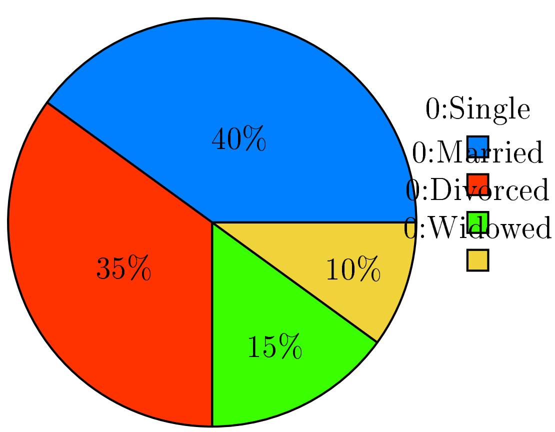
Example: Marital Status of 100 People

Modality	**Absolute Frequency**	**Relative Frequency**	**Percentage (%)**
Single	40	0.40	40%
Married	35	0.35	35%
Divorced	15	0.15	15%
Widowed	10	0.10	10%
Total	100	1.00	100%

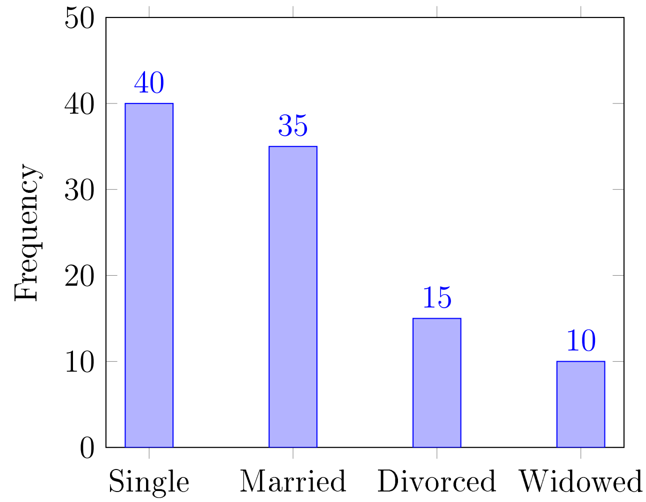
1. Circular Diagram (Pie Chart)

Each modality is represented by a circular sector whose angle is proportional to its frequency:

$$\text{Angle} = \frac{n_i}{N} \times 360^\circ$$



2. Bar Chart



1.4.3 Case of Quantitative Variables

Definition 1.8. A **qualitative variable** (or categorical variable) describes a quality, attribute, or category that cannot be measured numerically. Its possible values are called **modalities**.

Example: color of a car, marital status, nationality, type of housing, etc.

Definition 1.9. A **modality** is one of the possible categories or labels that a qualitative variable can take. Each individual in the population belongs to one and only one modality of the variable.

The statistical table for a qualitative variable usually includes:

- the list of modalities;
- the absolute frequencies (n_i);
- the relative frequencies (f_i);
- and optionally, cumulative or percentage frequencies.

Example: A survey of 50 people on their preferred drink gives the following results:

Drink	**Absolute frequency (n_i)**	**Relative frequency (f_i)**	**Percentage (%)**
Tea	20	0.40	40%
Coffee	15	0.30	30%
Juice	10	0.20	20%
Water	5	0.10	10%
Total	50	1.00	100%

Interpretation: Tea is the most preferred drink among respondents (40%).

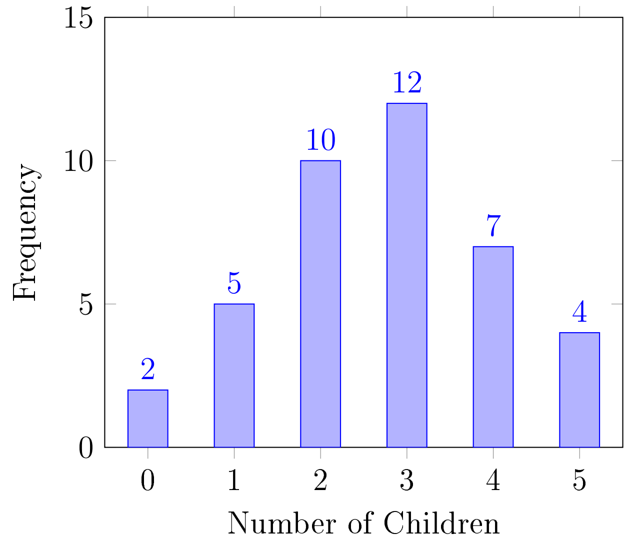
Graphical representations for qualitative variables include:

- Pie chart (circular diagram);
- Bar chart (column diagram);
- Band diagram (100% stacked bar).

Example: Number of Children per Family (Discrete Variable)

Number of Children (x_i)	**Frequency (n_i)**	**Relative Freq. (f_i)**	**Cumulative Freq.**
0	2	0.05	0.05
1	5	0.13	0.18
2	10	0.25	0.43
3	12	0.30	0.73
4	7	0.18	0.91
5	4	0.09	1.00
Total	40	1.00	

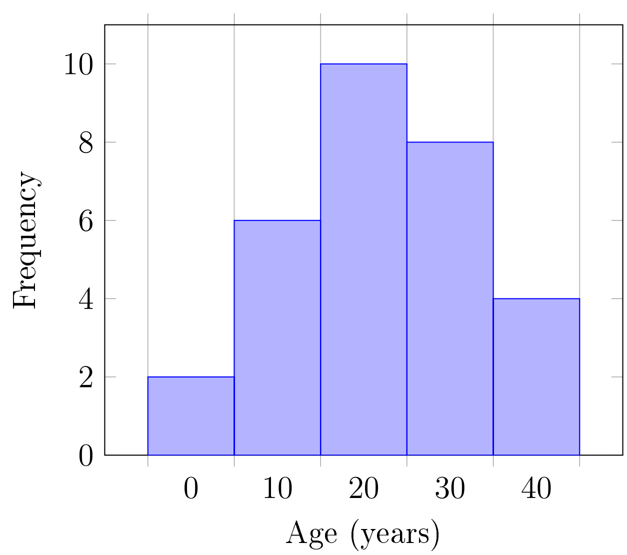
1. Bar Diagram



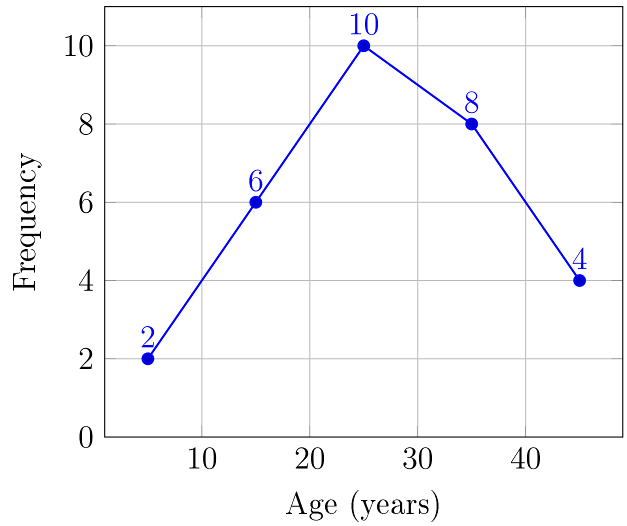
Example: Age Distribution (Continuous Variable)

Interval	Frequency	Midpoint	Cumulative
$[0, 10[$	2	5	2
$[10, 20[$	6	15	8
$[20, 30[$	10	25	18
	etc.		

2. Histogram



3. Frequency Polygon



1.5 Exercises with Solutions

1.5.1 Exercises on Basic Statistical Concepts

Exercise 1.1. School Survey

A study is carried out in a high school to determine:

- the number of siblings of each student,
- the favorite sport of each student,
- and the weight of each student.

1. Identify the **population**.
2. What is an **individual** in this study?
3. For each of the three characteristics above, give:
   (a) the **name of the variable (character)**;
   (b) the possible **modalities**;

(c) the **type of character** (qualitative or quantitative; if quantitative, discrete or continuous).

Solution.

- **Population:** all students of the high school.
- **Individual:** one student from the high school.
- **Variables:**

1. Number of siblings: modalities $0, 1, 2, 3, \dots \to$ **quantitative discrete**.
2. Favorite sport: modalities (football, basketball, swimming, etc.) $\to$ **qualitative**.
3. Weight: modalities expressed in kg (e.g., 55 kg, 62 kg, ...) $\to$ **quantitative continuous**.

Exercise 1.2. Car Factory

In a car factory, we record for each car:

- the color of the bodywork;
- the engine power (in horsepower);
- the country of destination of the sale.

1. Define the **population** and the **individual**.
2. For each variable, give:

(a) its set of **modalities**;

(b) its **type of character**.

Solution.

- **Population:** all cars produced in the factory during a given period.
- **Individual:** one car produced by the factory.
- **Variables:**
  1. Color of bodywork → modalities: red, black, white, gray, etc. → **qualitative nominal**.
  2. Engine power (HP) → modalities: numerical values (e.g., 80, 90, 120, 150) → **quantitative discrete**.
  3. Country of destination → modalities: France, Germany, Italy, etc. → **qualitative nominal**.

Exercise 1.3. Hospital Patients

A hospital collects data on its patients including:

- sex,
- age,
- blood pressure,
- and the department of hospitalization.

1. Identify the population and the individuals.
2. Determine the type of each variable.

Solution.

- **Population:** all patients admitted to the hospital during the study period.
- **Individual:** one patient.
- **Variables:**

1. Sex → modalities: male, female → **qualitative**.
2. Age → modalities: numerical values (years) → **quantitative continuous**.
3. Blood pressure → modalities: numerical values (mmHg) → **quantitative continuous**.
4. Department → modalities: cardiology, pediatrics, surgery, etc. → **qualitative**.

Exercise 1.4. University Grades

A teacher records the exam results of students in a statistics course:

- Student's name,
- Student's major (mathematics, computer science, physics),
- Final exam grade (on 20),
- and attendance rate (in %).

1. Identify the population and individuals.
2. For each variable, specify:
   (a) whether it is qualitative or quantitative;
   (b) and, if quantitative, whether it is discrete or continuous.

Solution.

- **Population:** all students enrolled in the statistics course.
- **Individual:** one student.
- **Variables:**
  1. Name → qualitative (identifier, not measurable).
  2. Major → qualitative (categorical).

3. Grade (on 20) → quantitative continuous.
4. Attendance rate (%) → quantitative continuous.

Exercise 1.5. Interpretation Practice

For each of the following studies, identify the **population**, one possible **individual**, the **variable studied**, and its **type**.

1. Measuring the temperature of each day in January.
2. Recording the color of cars passing through a toll gate.
3. Measuring the number of visits each patient makes to a clinic.
4. Measuring the height of 100 randomly chosen trees in a forest.

Solution.

1. Population: all days in January; individual: one day; variable: temperature; type: quantitative continuous.
2. Population: all cars passing the toll gate; individual: one car; variable: color; type: qualitative.
3. Population: all patients of the clinic; individual: one patient; variable: number of visits; type: quantitative discrete.
4. Population: all trees in the forest; individual: one tree; variable: height; type: quantitative continuous.

Exercise 1.6. Qualitative Variable

A survey was conducted among 60 students to find out their preferred mode of transportation to university. The following results were obtained:

Mode of Transport	Bus	Car	Bicycle	On foot
Number of students	25	15	10	10

1. Construct the frequency table (absolute, relative, and percentage frequencies).

2. Represent the data using:

   (a) a pie chart,

   (b) a bar chart.

Solution. **(1) Frequency table**

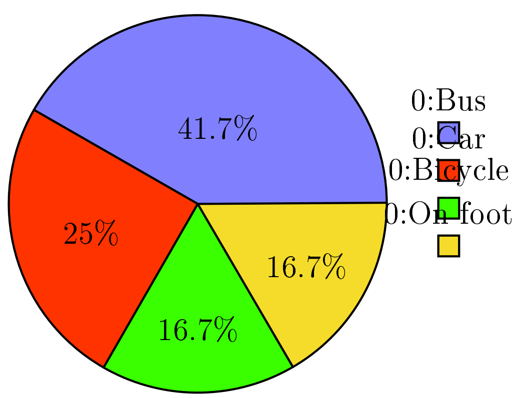
Mode of Transport	**Absolute**	**Relative**	**Percentage (%)**
Bus	25	$25/60 = 0.4167$	41.7%
Car	15	$15/60 = 0.25$	25%
Bicycle	10	$10/60 = 0.1667$	16.7%
On foot	10	$10/60 = 0.1667$	16.7%
Total	60	1.00	100%

(2a) Pie chart:

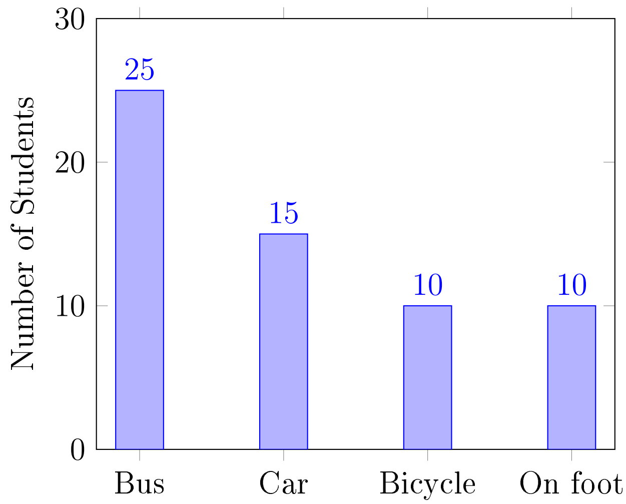
$$\text{Angle} = \frac{n_i}{N} \times 360^\circ$$

So the corresponding angles are:

$$\text{Bus: } 150^\circ, \quad \text{Car: } 90^\circ, \quad \text{Bicycle: } 60^\circ, \quad \text{On foot: } 60^\circ$$



(2b) Bar chart:



Interpretation: Most students use the bus to reach university, while the smallest groups walk or ride a bicycle.

Exercise 1.7. Quantitative Discrete Variable

The table below shows the number of children in 20 families.

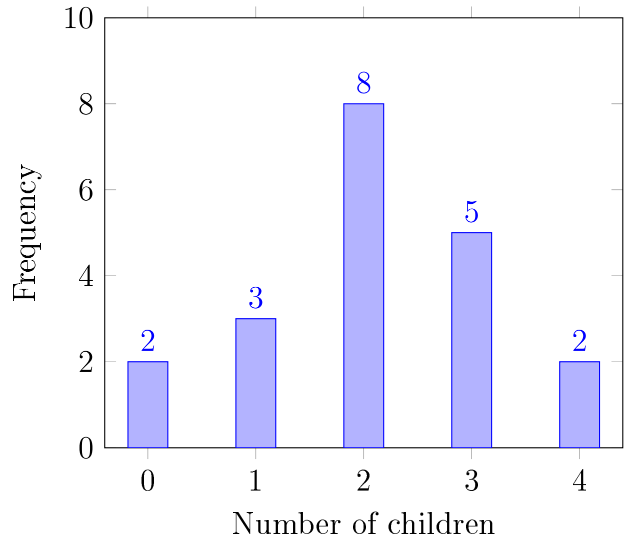
Children per family	0	1	2	3	4
Frequency	2	3	8	5	2

1. Compute the relative and cumulative frequencies.
2. Draw the corresponding bar chart.

Solution. **(1) Frequency table**

x_i	n_i	$f_i = \frac{n_i}{N}$	Cumulative Frequency
0	2	0.10	0.10
1	3	0.15	0.25
2	8	0.40	0.65
3	5	0.25	0.90
4	2	0.10	1.00

(2) Bar chart



Interpretation: Most families have two children. Few families have none or four.

The following data represent the ages (in years) of 40 workers in a factory:

$$[20, 30],\ [30, 40],\ [40, 50],\ [50, 60],\ [60, 70]$$

with respective frequencies:

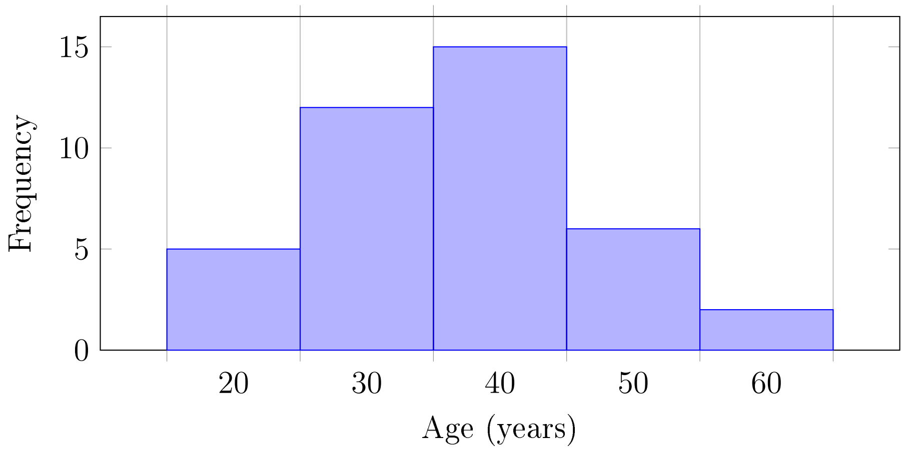
$$5,\ 12,\ 15,\ 6,\ 2$$

1. Build the frequency table (with midpoints and cumulative frequencies).
2. Draw the histogram and the frequency polygon.

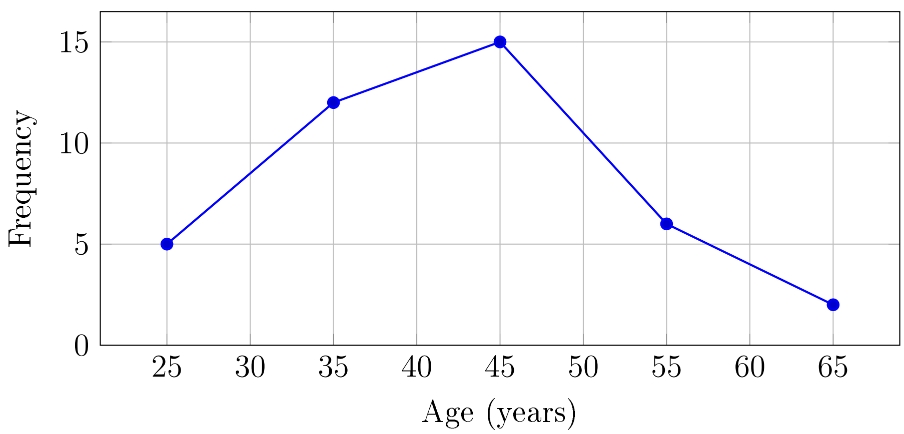
Solution. **(1) Frequency table**

Class Interval	Midpoint (x_i)	Frequency (n_i)	Cumulative Frequency
$[20, 30[$	25	5	5
$[30, 40[$	35	12	17
$[40, 50[$	45	15	32
$[50, 60[$	55	6	38
$[60, 70[$	65	2	40

(2a) Histogram



(2b) Frequency Polygon



Interpretation: Most workers are between 40 and 50 years old, while few are over 60.

Chapter 2

Numerical representation of data

2.1 Introduction

The statistical character can take a finite number of values (score, number of children, number of rooms, etc.). In this case, the statistical character studied is called a **discrete character**.

I. Discrete character: Throughout this part, we consider the situation:

$$X : \Omega \to \{x_1, x_2, \ldots, x_N\},$$

with $\text{Card}(\Omega) = N$ the number of individuals in the study.

Example 2.1. A survey was conducted in a village concerning the number of children per family. We denote X as the number of children. The results are:

x_i	0	1	2	3	4	5	6
n_i	18	32	66	41	32	9	2

- Ω: all families,
- w: a family,
- X: number of children per family,

- x_i: different values of X (modalities).

$$X : w \mapsto X(w)$$

For each family w, we associate $X(w)$ = number of children in that family.

2.2 Frequency and Cumulative Frequency

For each value x_i, we define:

$$n_i = \text{Card}\{w \in \Omega : X(w) = x_i\}.$$

n_i is called the **partial number** of x_i.

Example 2.2. In the previous example, 66 is the number of families that have 2 children:

$$n_3 = \text{Card}\{w \in \Omega : X(w) = x_3 = 2\} = 66.$$

2.2.1 Cumulative Frequencies

We define:

$$N_i^+ = n_1 + n_2 + \cdots + n_i, \quad N_i^- = n_i + n_{i+1} + \cdots + n_k.$$

Remark 2.1.

$$N_1^+ = n_1, \quad N_k^+ = N, \quad N_i^+ = N_{i-1}^+ + n_i,$$

$$N_1^- = N, \quad N_k^- = n_k, \quad N_i^- = N_{i-1}^- - n_{i-1}.$$

Example 2.3.

x_i	0	1	2	3	4	5	6
n_i	18	32	66	41	32	9	2
N_i^+	18	50	116	157	189	198	200
N_i^-	200	182	150	84	43	11	2

Hence the total number is:

$$N = \sum_{i=1}^{k} n_i = 200.$$

2.3 Relative and Cumulative Relative Frequencies

Definition.

$$f_i = \frac{n_i}{N}, \quad \text{(Relative frequency of } x_i\text{)}$$

Cumulative relative frequencies:

$$F_i^+ = f_1 + f_2 + \cdots + f_i, \quad F_i^- = f_i + f_{i+1} + \cdots + f_k.$$

Example 2.4.

x_i	0	1	2	3	4	5	6	Total
n_i	18	32	66	41	32	9	2	200
f_i	0.09	0.16	0.33	0.205	0.16	0.045	0.01	1
F_i^+	0.09	0.25	0.58	0.785	0.945	0.99	1	
F_i^-	1	0.91	0.75	0.42	0.215	0.055	0.01	

2.4 Graphical Representation of Discrete Variables

Two representations are possible: **bar graph** and **cumulative graph**.

2.4.1 Bar Graph

Each modality corresponds to a bar; heights are proportional to n_i.

2.4.2 Distribution Function

Let the function $F_X : \mathbb{N} \to [0, 1]$ be defined by:

$$F_X(x) = \text{percentage of individuals whose character value is} \leq x.$$

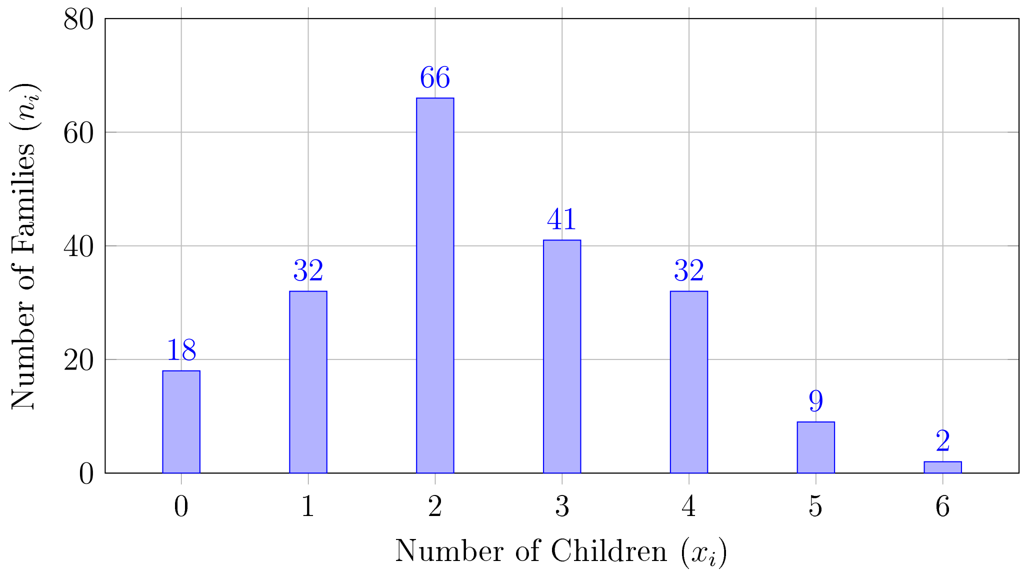


Figure 2.1: Distribution of families according to their number of children

Then:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.09, & 0 \leq x < 1 \\ 0.25, & 1 \leq x < 2 \\ 0.58, & 2 \leq x < 3 \\ 0.785, & 3 \leq x < 4 \\ 0.945, & 4 \leq x < 5 \\ 0.99, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases}$$

2.5 Position Parameters

2.5.1 Mode

The **mode** of a statistical variable is the value x_i having the greatest frequency:

$$Mo = \arg\max_i (f_i)$$

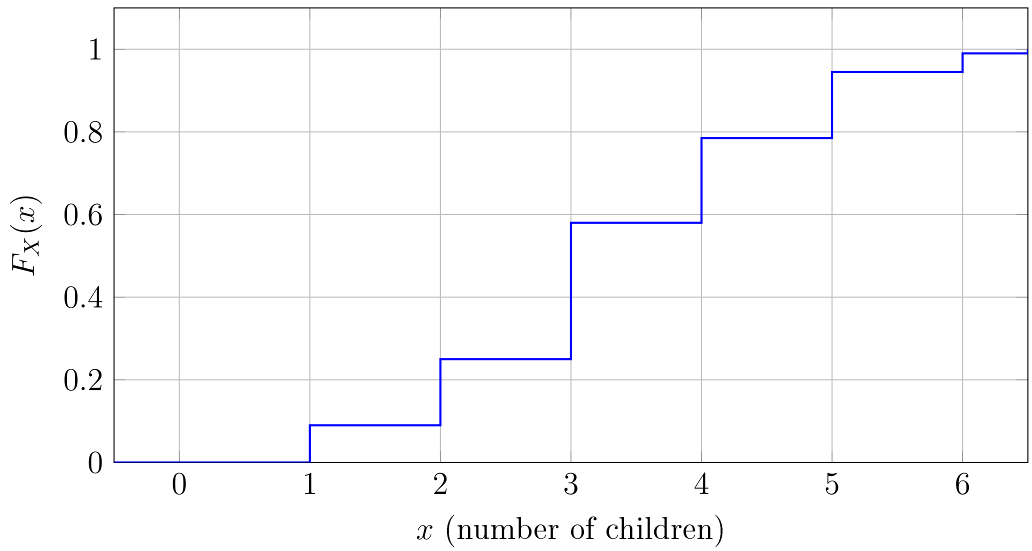


Figure 2.2: Distribution function of the variable: number of children per family

Example 2.5. Determine the mode for each of the following samples:

1. (2,3,3,3,4,5,6) $\rightarrow Mo = 3$
2. (2,3,4,5,6,6,6) $\rightarrow Mo = 6$
3. (2,3,3,4,4,4,5,5,6) $\rightarrow Mo = 4$
4. (2,3,3,3,4,4,4,5) $\rightarrow Mo = 3.5$
5. (2,3,3,3,4,5,6,6,6,8,2,2,3,3,4,4,5,5,6,6) $\rightarrow Mo = 3$ and 5

2.5.2 Median

The median (Me) divides the population into two equal parts.

Method 1:

$$\begin{cases} \text{If } N \text{ odd: } Me = y_{\frac{N+1}{2}} \\ \text{If } N \text{ even: } Me = \frac{y_{\frac{N}{2}} + y_{\frac{N}{2}+1}}{2} \end{cases}$$

Method 2: Compute N_i^+ or F_i^+ and find the smallest value $\geq N/2$ or ≥ 0.5 respectively.

2.5.3 Quartiles

There exist three quartiles:

$$Q_1 : \text{smallest } F_i^+ \geq 0.25,$$
$$Q_2 = Me,$$
$$Q_3 : \text{smallest } F_i^+ \geq 0.75.$$

Example 2.6. From the previous data:

x_i	0	1	2	3	4	5	6
N_i^+	18	50	116	157	189	198	200
F_i^+	0.09	0.25	0.58	0.785	0.945	0.99	1

$$Q_1 = 1, \quad Q_2 = 2, \quad Q_3 = 3.$$

2.5.4 Mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i.$$

Example: $\bar{x} = 2.46$, meaning on average a family has 2.46 children.

2.6 Dispersion Parameters

Range:

$$e = x_{\max} - x_{\min}.$$

2.6.1 Variance

$$\mathrm{Var}(X) = \sum_i f_i (x_i - \bar{x})^2 = \sum_i f_i x_i^2 - \bar{x}^2.$$

2.6.2 Standard Deviation

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

When σ_X is small, data are concentrated around the mean (homogeneous). When σ_X is large, data are dispersed (heterogeneous).

II. Continuous character:

Definition 2.1. We call a continuous statistical variable (or continuous character) any application from Ω to real values that takes a "large" number of values (Continuous characters are those that have an infinite number of modalities).

Example 2.7. Let Ω be the set of all newborns born at the university hospital of a city during the first 3 months of 2023. We designate by X the weight of the newborns. We suppose that

$$e_{\min} = 2.701 \quad \text{and} \quad e_{\max} = 5.001.$$

Remark 2.2. To study this character, we distribute the values taken by X into classes.

2.6.3 Classes of Values

Definition 2.2. We call a class of values of X an interval of type $[a, b]$ such that $X \in [a, b[$ if and only if $a \leq X(w) < b$, that is, the character values are in the class $[a, b]$.

When a character is identified as continuous, its modalities $C_k = [e_i, e_{i+1}[$ are intervals with:

- e_{i-1}: lower bound.
- e_i: upper bound.
- $a_i = e_i - e_{i-1}$: its amplitude, step, or length.
- $x_i = (e_{i-1} + e_i)/2$: its center.

2.6.4 Number of Classes

In how many classes do we distribute values? The answer is not unique. Let N be the total number. In this course we can consider three responses as examples:

1. One answer: $\sqrt{N}$, $[\sqrt{N}]$ (integer part) or $[\sqrt{N}]+1$. Therefore, the number of classes
$$k = \sqrt{N}.$$

   Example 2.8. Consider 30 values between 56.5 cm and 97.8 cm. In this case, $k = [\sqrt{N}]+1$ and we take $k = 6$.

2. One answer: Sturge formula
$$k = 1 + 3.3\log_{10}(N).$$

3. One answer: the Yule formula
$$k = 2.5\sqrt[4]{N}.$$

   Example 2.9. If we take $N = 30$, then the number of classes is given, for example, by:

   - either the Sturge formula $k = 1 + 3.3\log_{10}(30) = 6$,
   - or the Yule formula $k = 2.5\sqrt[4]{30} = 6$.

   We mention that the two formulas are almost the same if $N \ll 200$.

Definition 2.3. The number
$$e = x_{\max} - x_{\min}$$
is called the range of X. In this case, we can define the length by
$$a_i = \frac{e}{k} = \frac{x_{\max} - x_{\min}}{k}.$$

Exercise

We measure the height of 50 students in centimeters:

152 152 152 153 153 154 154 154 155 155 156 156 156 156 156 157 157 158 158 159 159 160 160 160 160 161 161 162 162 162 163 164 164 164 164 165 166 167 168 168 168 169 169 170 171 171 171 171.

1. Determine by different ways the number of classes k.
2. Determine the length a_i and the center x_k.
3. Determine the statistical table according to the frequency n_i.

2.7 Graphical Representation

2.7.1 Histogram and Polygon

1. The length a_i of classes is equal: n_i or f_i.
2. The length a_i of classes is different: we evaluate $n_i' = \frac{n_i}{e_i}$ or $f_i' = \frac{f_i}{a_i}$.
3. **Polygon**: it is obtained by joining the midpoints of the upper sides of each rectangle of the histogram using segments.

2.7.2 Graphical Representation of the Distribution Function

2.7.3 $\mathbf{N}_i^+, \mathbf{N}_i^-, \mathbf{F}_i^+, \mathbf{F}_i^-$

The curves of N_i^+, N_i^-, F_i^+, F_i^- are obtained by joining the following points:

1. N_i^+ or F_i^+: $(e_0, 0), (e_1, N_1^+), \ldots, (e_i, N_i^+), \ldots, (e_k, N)$, or $(e_0, 0), (e_1, F_1^+), \ldots, (e_i, F_i^+), \ldots, (e_k, 1)$.
2. N_i^- or F_i^-: $(e_0, N), (e_1, N_2^-), \ldots, (e_{i-1}, N_i^-), \ldots, (e_k, 0)$, or $(e_0, 1), (e_1, F_2^-), \ldots, (e_{i-1}, F_i^-), \ldots, (e_k,$

Remark 2.3. The point of intersection of the curves of N_1^+, N_1^- is the point $(Me, \frac{N}{2})$ and the point of intersection of the curves of F_1^+, F_1^- is the point $(Me, 0.5)$.

Example 2.10.

c_i	$[0,2[$	$[2,4[$	$[4,6[$	$[6,8[$	$[8,10[$	Total
x_i	1	3	5	7	9	
n_i	2	4	5	1	8	20
f_i	0.1	0.2	0.25	0.05	0.4	1
N_i^+	2	6	11	12	20	
N_i^- (ICRF)	20	18	14	9	8	
F_i^+	0.1	0.3	0.55	0.6	1	
F_i^-	1	0.9	0.7	0.45	0.4	

Table 2.1: Statistical table for continuous distribution

2.8 Distribution Function

Definition 2.4. The distribution function is defined by

$$\begin{aligned} F_X &: \mathbb{R} \to [0,1] \\ x &\to F_X(x) \end{aligned}$$

such that

$$F_X(x) = \begin{cases} 0, & x < e_0 \\ 0 + (x - e_0)\frac{f_i}{a_1}, & e_0 \le x < e_1 \\ \dots & \\ F_{i-1}^+ + (x - e_{i-1})\frac{f_i}{a_i}, & e_{i-1} \le x < e_i \\ \dots & \\ 1, & x \ge e_k. \end{cases}$$

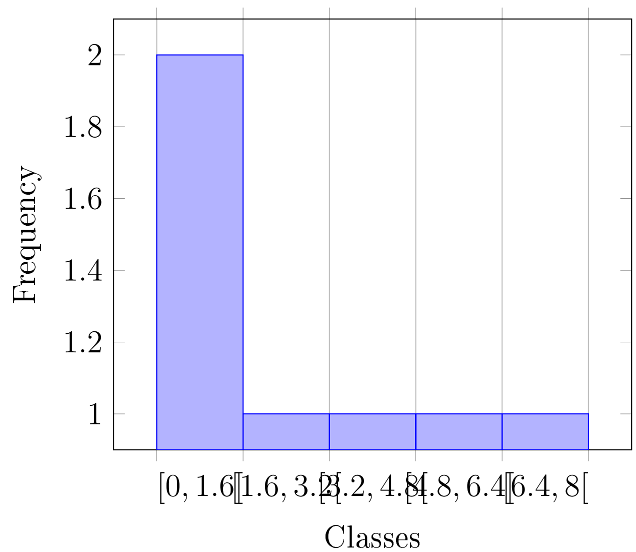


Figure 2.3: Histogram of continuous distribution

Example 2.11. For the previous example, we get

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{0.1}{2}x, & 0 \leq x < 2 \\ 0.1 + \frac{0.2}{2}(x-2), & 2 \leq x < 4 \\ 0.3 + \frac{0.25}{2}(x-4), & 4 \leq x < 6 \\ 0.55 + \frac{0.09}{2}(x-6), & 6 \leq x < 8 \\ 0.6 + \frac{0.4}{2}(x-8), & 8 \leq x < 10 \\ 1, & x \geq 10. \end{cases}$$

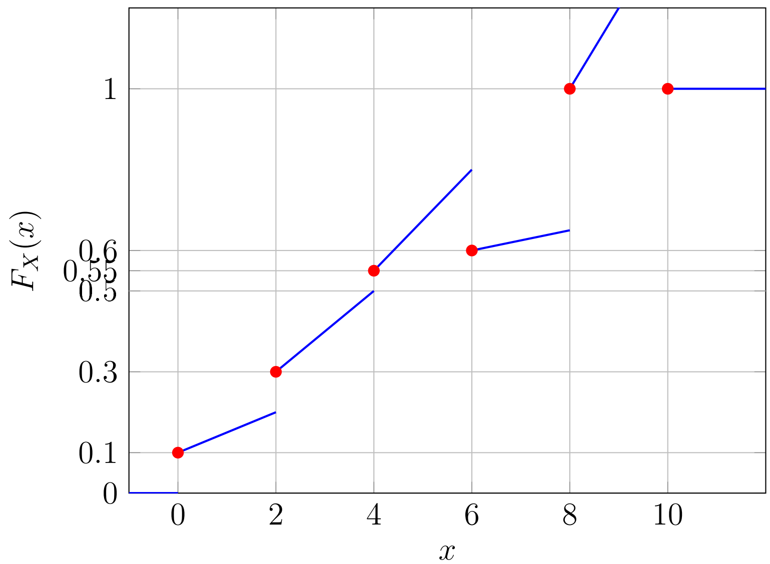


Figure 3.2: Cumulative curve or distribution function graph of the continuous variable

2.9 Position Parameters

2.9.1 Mean (or the Average)

Definition 2.5. The mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i.$$

2.9.2 Mode

Definition 2.6. We define the modal class as the class of values of X that has the greatest n_i or f_i. The mode is given by

$$Mo = e_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i$$

where:

- e_i: the lower bound of the modal class.
- a_i: the length of the modal class.

- $\Delta_1 = n_0 - n_1$, $\Delta_2 = n_0 - n_2$ or $\Delta_1 = f_0 - f_1$, $\Delta_2 = f_0 - f_2$.

- n_0 and f_0 are the frequency and relative frequency associated with the modal class.

- n_1 and f_1 are the frequency and relative frequency of the class that precedes the modal class.

- n_2 and f_2 are the frequency and relative frequency of the class that follows the modal class.

Remark 2.4. In the case of unequal length a_i, we use n'_i and f'_i instead of n_i and f_i.

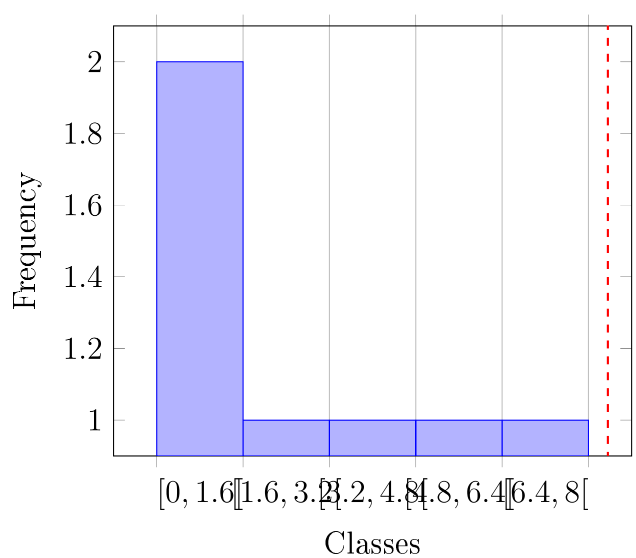


Figure 2.4: Graphical determination of the mode

Example 2.12. For the previous example: Modal class is $MC = [8, 10[$, $e_i = 8$, $a_i = 2$, $\Delta_1 = 9 - 7 = 2$, $\Delta_2 = 9 - 0 = 9$.

$$Mo = 8 + 2\frac{2}{9+2} = 8.36.$$

2.9.3 Quartiles

1. The first quartile (Q1): The class of Q1 corresponds to the cumulative frequency (or the cumulative relative frequency) that satisfies

$$N_{i-1}^{+} \leq \frac{N}{4} \leq N_{i}^{+} \quad \text{or} \quad F_{i-1}^{+} \leq 0.25 \leq F_{i}^{+}.$$

This class is denoted by $CQ1 = [e_{i-1}, e_i[$. The expression of $Q1$ is given by

$$Q1 = e_{i-1} + a_i \left(\frac{\frac{N}{4} - N_{i-1}^{+}}{n_i} \right),$$

or

$$Q1 = e_{i-1} + a_i \left(\frac{0.25 - F_{i-1}^{+}}{f_i} \right).$$

2. The second quartile (Median): The class of Me corresponds to the cumulative frequency (or the cumulative relative frequency) that satisfies

$$N_{i-1}^{+} \leq \frac{N}{2} \leq N_{i}^{+} \quad \text{or} \quad F_{i-1}^{+} \leq 0.5 \leq F_{i}^{+}.$$

This class is denoted by $CMe = [e_{i-1}, e_i[$. The expression of Me is given by

$$Me = e_{i-1} + a_i \left(\frac{\frac{N}{2} - N_{i-1}^{+}}{n_i} \right),$$

or

$$Me = e_{i-1} + a_i \left(\frac{0.5 - F_{i-1}^{+}}{f_i} \right).$$

3. The third quartile (Q3): The class of Q3 corresponds to the cumulative frequency (or the cumulative relative frequency) that satisfies

$$N_{i-1}^{+} \leq \frac{3N}{4} \leq N_{i}^{+} \quad \text{or} \quad F_{i-1}^{+} \leq 0.75 \leq F_{i}^{+}.$$

This class is denoted by $CQ3 = [e_{i-1}, e_i[$. The expression of $Q3$ is given by

$$Q3 = e_{i-1} + a_i \left(\frac{\frac{3N}{4} - N_{i-1}^{+}}{n_i} \right),$$

or

$$Q3 = e_{i-1} + a_i \left(\frac{0.75 - F_{i-1}^{+}}{f_i} \right).$$

Example 2.13. For the previous example, we see that:

1. The class of $Q1$ corresponds to the cumulative frequency (or the cumulative relative frequency) that satisfies

$$F_1^+ = 0.1 \leq 0.25 \leq 0.3 = F_2^+, \quad F_{i-1}^+ \leq 0.25 \leq F_i^+.$$

By identification, $i-1=1$ then $i=2$ and $CQ1 = [e_1, e_2] = [2, 4[$. The value of $Q1$ is calculated as

$$Q1 = 2 + 2\left(\frac{0.25 - 0.1}{0.2}\right) = 2.15.$$

2. The class of Me corresponds to

$$F_2^+ = 0.3 \leq 0.5 \leq 0.55 = F_3^+, \quad F_{i-1}^+ \leq 0.5 \leq F_i^+.$$

By identification, $i-1=2$ then $i=3$ and $CMe = [e_2, e_3] = [4, 6[$. The value of Me is calculated as

$$Me = 4 + 2\left(\frac{0.5 - 0.3}{0.25}\right) = 5.6.$$

3. The class of $Q3$ corresponds to

$$F_4^+ = 0.6 \leq 0.75 \leq 1 = F_5^+, \quad F_{i-1}^+ \leq 0.75 \leq F_i^+.$$

By identification, $i-1=4$ then $i=5$ and $CQ3 = [e_4, e_5] = [8, 10[$. The value of $Q3$ is calculated as

$$Q3 = 8 + 2\left(\frac{0.75 - 0.6}{0.4}\right) = 8.75.$$

2.10 Dispersion Parameters

2.10.1 Range

Definition 2.7. The range of a statistical array is the difference between the greatest value and the smallest value of the character, given by the quantity

$$e = e_k - e_0.$$

Example 2.14. For the previous example: $e = 10 - 0 = 10$.

2.10.2 Variance

Definition 2.8. We call the variance of a statistical array the number

$$Var(X) = \sum_{i=1}^{k} f_i(\bar{x} - x_i)^2 = \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2.$$

2.10.3 Standard Deviation

Definition 2.9. The standard deviation is given by

$$\sigma_X = \sqrt{Var(X)}.$$

2.11 Exercises

Exercise 2.1. Classify the following characteristics in the table below:

1. The different car brands in Algeria.
2. The average age of the Algerian population.
3. Infant mortality in Europe.
4. The destination countries of African migrants.
5. The number of floors in the tallest skyscrapers in the world.
6. The height of players of the national handball team.
7. The means of transport used by students.
8. The monthly salary of civil servants in an institution.
9. The number of municipalities per wilaya in Algeria.
10. The evolution of seasonal temperatures in Algeria.
11. Citrus fruit production in Mostaganem.

12. Fiscal expenditures.

Qualitative variable	Quantitative discrete variable	Quantitative continuous variable

Exercise 2.2. The results of observing the sequence of a DNA strand are:

G, G, A, T, A, G, C, T, A, G, G, A, T, G, C, C, T, G, C, T, A, G, T, A, G, A, T,
C, G, A, G, C, T, G, C, T, A, C, C, T, C, C, G, A, T, C, G, C, T, C.

A: Adenine, G: Guanine, C: Cytosine, T: Thymine.

1. Identify the statistical series presented above (the population, its size, the studied characteristic and its type).
2. Give the distribution of partial frequencies.
3. Draw the bar diagram of this statistical series.
4. Compute the frequencies.

Exercise 2.3. The means of transport used by SEI faculty students to attend classes are recorded as follows:

Transport	Bus	Taxi	On foot
Number	300	165	45
Frequency			
Angle			

1. What is the studied population? What is the studied variable? What is its type?
2. Represent the data of the table using a pie chart.

Exercise 2.4. We have recorded the statistical series:

$$23, 23, 23, 34, 39, 39, 39, 39, 39, 39, 39, 39, 45, 45, 45, 45, 48, 48, 52, 52, 52, 52, 52, 55, 62, 62, 62, 62, 62, 62.$$

1. Give the distribution of partial frequencies.
2. Compute the frequencies.
3. Compute the cumulative frequencies.
4. Determine the mode, median, and quartiles of this series.
5. Compute D_3 and C_{90}.

Exercise 2.5. The table below shows the speed of 200 vehicles recorded by a radar during a road inspection:

Speed (Km/h)	[75,80[	[80,90[	[90,95[	[95,100[	[100,105[	[105,110[	[110,115[
Number of vehicles	10	38	12	96	17	22	5

1. Identify the statistical series.
2. Determine the range and class width of the series.
3. Draw the histogram of the frequencies.
4. Calculate the class centers and draw the cumulative frequency curves.

Exercise 2.6. The blood glucose level (in g/l) measured in 32 persons is given as:

$$0.79, 0.83, 0.84, 0.87, 0.90, 0.90, 0.92, 0.94, 0.95, 0.97, 0.97, 0.97, 1.00, 1.01, 1.01, 1.03, 1.03, 1.03, 1.03, 1.03,$$

$$1.08, 1.09, 1.11, 1.11, 1.11, 1.12, 1.14, 1.18, 1.20.$$

1. What is the type of the studied variable?
2. Classify this series into classes of equal width 0.06.

3. Represent graphically this statistical series.

4. Compute the frequencies.

5. Determine the modal class and compute the mode.

6. Compute the median.

7. Compute the arithmetic ($\bar{x}$), geometric ($\bar{x}_G$) and harmonic ($\bar{x}_H$) means of this series.

Exercise 2.7. In an Algérie Télécom agency, the telephone bill amounts issued in a given day are summarized below:

Bill amount (DA)	[500,800[	[800,1100[	[1100,1400[	[1400,1700[	[1700,2000[	[2000,2300[
Number of bills	3	10	27	12	26	22

1. Identify the statistical series x (population, size, studied variable and its type).

2. Determine the modal class and compute the mode.

3. Find the median of this series.

4. Compute the 75th percentile C_{75}, the arithmetic mean $\bar{x}$, the variance $Var(x)$ and the coefficient of variation C_v.

Exercise 2.8. A quality control was carried out during 100 working hours on two machines producing mechanical parts. Some parts are defective and therefore unusable. The number of defective parts per hour was recorded as follows:

Machine A	0	1	2	3	4	5	6	7
Number of hours	13	42	38	2	2	1	1	1

Machine B	0	1	2	3	4	5
Number of hours	35	40	1	1	10	13

1. Compare the two distributions (machines A and B) using measures of central tendency (mean and median).

2. Compare the two distributions using measures of dispersion (range, mean absolute deviation, standard deviation).

3. Which parameter seems most relevant to compare the two machines?

Exercise 2.9. The weight of olives harvested in 150 farms is expressed in tons as follows:

Weight (tons) $[x_i, x_{i+1}[$	Frequency n_i
[5.00, 5.01[	4
[5.01, 5.02[	18
[5.02, 5.03[	25
[5.03, 5.04[	36
[5.04, 5.05[	30
[5.05, 5.06[	22
[5.06, 5.07[	11
[5.07, 5.08[	3
[5.08, 5.09[	1

1. Identify the statistical series x (population, studied variable, and its type).

2. Let $y = \frac{x-5.045}{0.01}$, compute the mean and variance of y.

3. Deduce the mean and variance of x.

4. Compute the coefficient of variation of y, then deduce that of x.

1.7 Solutions

Exercise 1

Qualitative variables	**Quantitative variables (discrete / continuous)**
1, 4, 7	3, 5, 9
	2, 6, 8, 10, 11, 12

Exercise 2

The observed sequence of 50 DNA bases is: G, G, A, T, A, G, C, T, A, G, G, A, T, G, C, C, T, G, C, T, A, G, T, A, G, A, T, C, G, A, G, C, T, G, C, T, A, C, C, T, C, C, G, A, T, C, G, C, T, C.

- **Population:** Nitrogen bases of DNA.
- **Sample size:** $n = 50$.
- **Variable:** Type of base (A, G, C, T) — qualitative.

Base	n_i	$f_i = n_i/n$
A	10	0.20
G	14	0.28
C	14	0.28
T	12	0.24
Total	50	1

Exercise 3

- **Population:** Students of the SEI Faculty.
- **Sample size:** $n = 510$.
- **Variable:** Means of transportation — qualitative.

Transportation	Frequency n_i	f_i	Angle
Bus	300	0.59	212.4°
$Taxi$	165	0.32	115.2°
$Walking$	45	0.09	32.4°

Exercise 4

Dataset of 30 observations: 23, 23, 23, 34, 39, 39, 39, 39, 39, 39, 39, 39, 45, 45, 45, 45, 48, 48, 52, 52, 52, 52, 52, 55, 62, 62, 62, 62, 62, 62.

x_i	n_i	f_i	$N_i^{\%}$	$N_i^{\&}$
23	3	3/30	3	30
34	1	1/30	4	27
39	8	8/30	12	18
45	4	4/30	16	18
48	2	2/30	18	14
52	5	5/30	23	12
55	1	1/30	24	7
62	6	6/30	30	6

Results:

- Mode: $\text{Mo} = 39$
- Median: $\text{Me} = 45$
- Quartiles: $Q_1 = 39,\ Q_2 = 45,\ Q_3 = 52$
- $D_3 = 39,\ C_{90} = 62$

Exercise 6

Blood glucose level sample ($n = 32$), using 7 classes of width 0.06. Modal class: $[1.03, 1.09[$, with $n_m = 9$ and $c_m = 1.06$.

$$\overline{x} = \frac{32.54}{32} = 1.016\,\text{g/l}, \quad \overline{x}_G = 1.011\,\text{g/l}, \quad \overline{x}_H = 1.006\,\text{g/l}.$$

Median (interpolated): Me ≈ 1.036 g/l. Mode (interpolated): Mo ≈ 1.06 g/l.

Chapter 3

Probability calculus

Combinatorial analysis is a branch of mathematics that studies how to count objects. It provides counting methods that are particularly useful in probability theory.

3.1 Introduction

The aim of combinatorial analysis is to count the dispositions (layouts) that can be formed from the elements of a finite set of objects. An object is characterized by:

- the place it occupies in the disposition;
- the number of times it can appear.

3.1.1 Notion of repetition

If an element appears more than once in a layout, the layout is said to have *repetition*; otherwise, the layout is said to have *no repetition*.

3.1.2 Notion of order

A layout is said to be *ordered* when each time an element changes place (or position), the layout changes.

Example 3.1. Consider a set E with three elements $E = \{a, b, c\}$. Choosing two elements from this set can be done in several different ways.

The following table shows all possible cases:

Disposition with repetition	Without repetition	
With order (ordered)	$aa, ab, ac, ba, bb, bc, ca, cb, cc$	ab, ac, ba, bc, ca, cb
Out of order (unordered)	aa, ab, ac, bb, bc, cc	ab, ac, bc

3.1.3 Factorial of an integer n

Let n be a natural integer, the factorial of n, noted $n!$, is:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1.$$

Conventionally, $0! = 1$ and $1! = 1$.

Example 3.2.

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120, \qquad 10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800.$$

3.2 Arrangements

Definition 3.1. Given a set E of n objects, an *arrangement* of p of these objects is an ordered sequence of p objects taken from these n objects.

There are two types of arrangements: with and without repetition.

3.2.1 Arrangement without repetition

An arrangement without repetition of p objects chosen from n objects is an ordered layout of p objects taken from the n objects without repetitions.

The number of arrangements without repetition, denoted A_n^p, is:

$$A_n^p = n \times (n-1) \times (n-2) \times \cdots \times (n-p+1) = \frac{n!}{(n-p)!}, \qquad 1 \leq p \leq n.$$

In an arrangement without repetition, the p objects in the list are all distinct. This corresponds to a draw without replacement with order.

Example 3.3 (Example 4.3)**.** How many three-letter words containing no more than one repeated letter can be formed using the letters of the alphabet?

$$A_{26}^{3} = \frac{26!}{(26-3)!} = 26 \times 25 \times 24 = 15600 \text{ words.}$$

3.2.2 Arrangement with repetition

An arrangement with repetition of p objects chosen from n objects is an ordered layout of p objects taken from the n objects with repetitions.

The number of arrangements with repetition, denoted n^p, is:

$$n^p = \underbrace{n \times n \times \cdots \times n}_{p \text{ times}}, \qquad 1 \leq p.$$

This corresponds to a draw with replacement with order.

Example 3.4. How many two-letter words can be made with the letters of the alphabet?

$$26^2 = 26 \times 26 = 676 \text{ words.}$$

3.3 Permutations

Definition 3.2. Let E be a set of n objects. A *permutation* of n distinct objects is any ordered sequence of those n objects (an arrangement of n from n).

3.3.1 Permutation without repetition

This is the special case $p = n$ of arrangements without repetition.

The number of permutations of n distinct objects is:

$$P_n = n!.$$

Remark 3.1.

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

Example 3.5. The number of ways to seat eight diners around a table is:

$$P_8 = 8! = 40320.$$

3.3.2 Permutation with repetition

If among n objects there are repeated objects (for instance k identical objects), then the number of distinct permutations is

$$\frac{n!}{k!}.$$

More generally, if counts of identical objects are $n_1, n_2, \dots, n_r$ with $\sum_i n_i = n$, then

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Example 3.6 (Example 4.6)**.** The number of possible words (with or without meaning) that can be formed by permuting the 8 letters of the word "Quantity" is

$$\frac{8!}{2!} = 20160,$$

because there are two t's.

For the word "Swimming" (8 letters) with repeated letters i twice and m twice:

$$\frac{8!}{2!\,2!} = 10080.$$

3.4 Combinations

3.4.1 Combination without repetitions

Definition 3.3. Given a set E of n objects, a combination of p objects is a subset of p objects taken from n without order and without replacement.

The number of combinations of p objects among n, denoted $\binom{n}{p}$, is

$$\binom{n}{p} = \frac{n!}{(n-p)!\,p!}, \qquad 1 \le p \le n.$$

Remark 3.2.

$$\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{(n-p)!\,p!}.$$

Example 3.7. The random drawing of 5 cards from a deck of 32 cards (poker hand) is a combination with $p = 5$ and $n = 32$:

$$\binom{32}{5} = \frac{32!}{27!\,5!} = 409696.$$

Example 3.8. Forming a delegation of 2 students from a group of 20:

$$\binom{20}{2} = \frac{20!}{18!\,2!} = 190.$$

3.4.2 Combination with repetitions

The number of combinations of p objects among n with repetitions allowed (multiset combinations) is

$$\binom{n+p-1}{p} = \frac{(n+p-1)!}{p!\,(n-1)!}.$$

Example 3.9. Make 3-letter words from a 5-letter alphabet with repetition. The number of words is

$$\binom{5+3-1}{3} = \binom{7}{3} = 35.$$

One can verify by counting cases: three distinct letters, two same one different, or all three same.

Example 3.10. Consider domino halves labeled $0, 1, \ldots, 6$ (7 values). Each domino is an unordered pair of values (repetition allowed). Total dominoes:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 = 28.$$

Equivalently, number of multisets of size 2 from 7 values:

$$\binom{7+2-1}{2} = \binom{8}{2} = 28.$$

3.4.3 Properties of combinations and Newton's binomial

$$\binom{0}{n} = \binom{n}{0} = 1, \qquad \binom{1}{n} = n, \qquad \binom{2}{n} = \frac{n(n-1)}{2}.$$

Symmetry: $\binom{n}{p} = \binom{n}{n-p}$.

Pascal's identity (compound combination):

$$\binom{n-1}{p-1} + \binom{n-1}{p} = \binom{n}{p}.$$

Theorem 3.1 (Binomial Theorem)**.** For integer $n \geq 0$,

$$(a+b)^n = \sum_{p=0}^{n} \binom{n}{p} a^{n-p} b^p.$$

Example 3.11. For $n = 4$:

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.$$

For $n = 5$:

$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5.$$

Remark 3.3. Setting $a = b = 1$ gives $2^n = \sum_{p=0}^{n} \binom{n}{p}$.

3.5 Corrected exercises

3.5.1 Exercises

Exercise 3.1. Prove that:

$$C_n^p + C_n^{p+1} = C_{n+1}^{p+1}$$

Exercise 3.2. Solve in $\mathbb{N}$:

$$C_n^2 = 3n$$

Exercise 3.3. Compute:

$$\sum_{k=0}^{10} C_{10}^k$$

Exercise 3.4. Simplify:

$$\frac{C_{n+1}^{p+1}}{C_n^p}$$

Exercise 3.5. In a class of 12 students, we choose a group of 5 students that must include two fixed students A and B. How many groups are possible?

Exercise 3.6. Prove Pascal's identity:

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

Exercise 3.7. A keypad allows you to enter a building code using a letter followed by a 3-digit number, digits may repeat.

(1) How many different codes can be formed?

(2) How many codes are there without the digit 1?

(3) How many codes are there containing at least one digit 1?

(4) How many codes are there containing distinct digits?

Exercise 3.8. How many distinct words, with or without meaning, can be formed with all the letters of each of the words:

(1) `maths`;

(2) `statistics`.

Exercise 3.9. We take 3 bulbs out of 15 simultaneously, 5 of which are defective.

(1) In how many different ways can this draw be made?

(2) In how many different ways can at least one defective bulb be obtained?

Exercise 3.10. A coat rack has 5 coat hangers in a row. How many distinct layouts (dispositions) without stacking coats can be hung on it:

(1) three coats?

(2) five coats?

(3) six coats?

Exercise 3.11. Four mathematicians and two physicists sit on a six-seater bench.

(1) How many layouts are possible?

(2) Same question if mathematicians are on one side and physicists on the other?

(3) Same question if each physicist sits between two mathematicians?

(4) Same question if the physicists want to stay next to each other?

Exercise 3.12. An urn contains one white ball, three black balls and four red balls. Three balls are drawn at random.

3.5.2 Corrections

Exercise 3.1 Prove that $C_n^p + C_n^{p+1} = C_{n+1}^{p+1}$.

We use the definition:

$$C_n^p = \frac{n!}{p!(n-p)!}, \quad C_n^{p+1} = \frac{n!}{(p+1)!(n-p-1)!}$$

Bring to a common denominator:

$$C_n^p + C_n^{p+1} = \frac{n!(p+1)}{(p+1)!(n-p)!} + \frac{n!(n-p)}{(p+1)!(n-p)!}$$

Factor:

$$= \frac{n![(p+1)+(n-p)]}{(p+1)!(n-p)!}$$

$$= \frac{n!(n+1)}{(p+1)!(n-p)!}$$

$$= \frac{(n+1)!}{(p+1)!((n+1)-(p+1))!}$$

So:

$$C_n^p + C_n^{p+1} = C_{n+1}^{p+1}$$

Exercise 3.2 Solve $C_n^2 = 3n$.

$$C_n^2 = \frac{n(n-1)}{2}$$

So:

$$\frac{n(n-1)}{2} = 3n \Rightarrow n(n-1) = 6n \Rightarrow n^2 - n - 6n = 0$$

$$n^2 - 7n = 0 \Rightarrow n(n-7) = 0$$

Thus:

$$n = 0 \text{ or } n = 7$$

Since $n \geq 2$, we keep:

$$n = 7$$

Exercise 3.3 Compute $\sum_{k=0}^{10} C_{10}^k$.

Using the binomial identity:

$$\sum_{k=0}^{n} C_n^k = 2^n$$

So:

$$\sum_{k=0}^{10} C_{10}^k = 2^{10} = 1024$$

Exercise 3.4 Simplify $\dfrac{C_{n+1}^{p+1}}{C_n^p}$.

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

$$\frac{C_{n+1}^{p+1}}{C_n^p} = \frac{(n+1)!}{(p+1)!} \cdot \frac{p!}{n!}$$

$$= \frac{n+1}{p+1}$$

Exercise 3.5 Fixed elements problem.

A and B are fixed, so we choose 3 students from the remaining 10:

$$C_{10}^3 = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$$

Exercise 3.6 Prove Pascal's identity:

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

We consider a fixed element:

- If it is not chosen: C_{n-1}^p
- If it is chosen: C_{n-1}^{p-1}

Thus total combinations:

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

Exercise 3.7 The code is Letter $d_1 d_2 d_3$. There are 26 letters and 10 digits $0, \ldots, 9$.

(1) Total codes (order important, repetitions allowed):

$$N_1 = 26 \times 10^3 = 26{,}000.$$

(2) Codes without the digit 1: digits come from $\{0, 2, 3, 4, 5, 6, 7, 8, 9\}$ (9 digits)

$$N_2 = 26 \times 9^3 = 26 \times 729 = 18{,}954.$$

If digits are required to be distinct (and no 1), then

$$N_2' = 26 \times A_9^3 = 26 \times 9 \times 8 \times 7 = 13{,}104.$$

(3) Codes containing at least one digit 1:

$$N_3 = N_1 - N_2 = 26{,}000 - 18{,}954 = 7{,}046.$$

(4) Codes with distinct digits (any digit allowed):

$$N_4 = 26 \times A_{10}^3 = 26 \times 10 \times 9 \times 8 = 18{,}720.$$

Exercise 3.8 (1) `maths`: 5 distinct letters, permutations:

$$5! = 120.$$

(2) `statistics`: letters counts: $n = 10$, with repeats: i appears 2 times, s 3 times, t 3 times. So

$$\frac{10!}{2!\,3!\,3!} = 50{,}400.$$

Exercise 3.9 (1) Total simultaneous draws of 3 from 15 (order not important):

$$\binom{15}{3} = \frac{15!}{12!\,3!} = 455.$$

(2) At least one defective (there are 5 defective and 10 good): Count cases with exactly 1,2,3 defective:

$$N_2 = \binom{5}{1}\binom{10}{2}+\binom{5}{2}\binom{10}{1}+\binom{5}{3}\binom{10}{0} = (5\times45)+(10\times10)+(10\times1) = 335.$$

Exercise 3.10 (1) Three coats on 5 hooks (order matters, without stacking): arrangements $A_5^3 = 5 \times 4 \times 3 = 60$.

(2) Five coats: permutations: $5! = 120$.

(3) Six coats: impossible to hang 6 coats on 5 hangers without stacking.

Exercise 3.11 There are 4 mathematicians (M) and 2 physicists (P). Total $n = 6$.

(1) Total permutations: $6! = 720$.

(2) Mathematicians one side and physicists the other: two block orders (MMMMPP or PPMMMM). Internally the mathematicians and physicists can permute: $2 \times 4! \times 2! = 96$.

(3) Each physicist sits between two mathematicians: counted cases give $3 \times 4! \times 2! = 144$.

(4) Physicists next to each other (treated as a block) – there are 5 possible block positions and internal permutations:

$$5 \times 4! \times 2! = 240.$$

Exercise 3.12 Urn: 1 white, 3 black, 4 red (8 balls total).

(1) Successive draws of three balls without replacement (order matters):

$$A_8^3 = 8 \times 7 \times 6 = 336.$$

(2) Simultaneous draws without replacement (order doesn't matter):

$$\binom{8}{3} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56.$$

(3) Simultaneous draws of three balls of different colors (1 white, 1 black, 1 red):

$$\binom{1}{1}\binom{3}{1}\binom{4}{1} = 1 \times 3 \times 4 = 12.$$

(4) Successive draws with replacement: two red and one white in any order. There are 3 permutations (WRR, RWR, RRW). Each draw probability counting combinations (counting ways): Number of sequences: $3 \times 1 \times 4^2 = 48$.

3.6 Supplementary exercises

Exercises on Binomial Coefficients and Applications

Exercise 3.13. Compute:

$$\sum_{k=0}^{8} C_8^k$$

Exercise 3.14. Simplify:

$$\sum_{k=0}^{n} C_n^k 3^k$$

Exercise 3.15. Compute:

$$\sum_{k=0}^{n} k C_n^k$$

Exercise 3.16. Prove:

$$\sum_{k=0}^{n} (-1)^k C_n^k = 0 \quad (n \geq 1)$$

Exercise 3.17. Show that:

$$\sum_{k=0}^{n} C_n^k = 2^n$$

using the binomial theorem.

Exercise 3.18. Simplify:

$$C_n^p + C_n^{p+1} + C_n^{p+2}$$

Exercise 3.19. In a class of 15 students, how many ways are there to choose a committee of 4 students?

Exercise 3.20. In a group of 12 students, we must choose a team of 5 students containing exactly 2 girls from a group of 6 girls and 6 boys. How many teams are possible?

Exercise 3.21. A password consists of 3 letters followed by 2 digits. How many passwords are possible if repetition is allowed?

Exercise 3.22. A company has 10 employees. How many ways can we choose a manager, assistant manager, and secretary?

Exercise 3.23. From a set of 9 elements, how many subsets contain exactly 3 elements?

Exercise 3.24. From a group of 8 people, how many ways can we form a committee of at least 1 person?

Exercise 3.25. In how many ways can we distribute 5 identical balls into 3 distinct boxes?

Exercise 3.26. A test has 10 questions. A student chooses to answer exactly 6 questions. How many choices are possible?

Exercise 3.27. A code is formed using 2 letters and 3 digits. How many codes are possible?

Exercise 3.28. From 7 men and 5 women, how many committees of 4 people contain at least 1 woman?

3.7 Random Experiment and Event

3.7.1 Random Experiment

A random experiment (r.e.) is any experiment whose outcome is governed by chance.

Example 3.12. Tossing a coin and observing the upper face is a random experiment that results in two possible outcomes: Heads (H) or Tails (T).

Definition 3.4. The set of all possible outcomes of an r.e. is called the sample space, denoted as Ω.

Example 3.13. When rolling a die (with six numbered faces), if we are interested in the number on the upper face, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

3.7.2 Event

An event in Ω is a subset of Ω. An event can be elementary (a single element) or compound (several elements).

Example 3.14. When rolling a six-sided die, $\Omega = \{1, 2, 3, 4, 5, 6\}$, the event A: "getting a 2" is an elementary event $A = \{2\} \subset \Omega$. The event B: "getting an even number" is a compound event $B = \{2, 4, 6\} \subset \Omega$.

3.8 Relations and Operations Between Events

3.8.1 Inclusion

Let A and B be two events associated with a random experiment. We say that A is included in B (or A implies B) if the occurrence of A necessarily implies the occurrence of B. This is denoted as $A \subset B$ or $A \Rightarrow B$.

Example 3.15. In the previous example (Example 3.3), $A = \{2\} \subset B = \{2, 4, 6\}$, meaning that if A occurs, then B also occurs.

3.8.2 Complementary Event

The complementary event of an event A is the complement of A in Ω, denoted $\bar{A}$, the event that occurs when A does not occur, and vice versa.

Remark 3.4. Let A be an event in Ω and $\bar{A}$ its complementary event.

If A occurs, then $\bar{A}$ does not occur.

If A does not occur, then $\bar{A}$ occurs.

The sample space Ω is always the certain event, and its complementary event is the impossible event, denoted $\emptyset$.

Example 3.16.

Example 3.17. If we take the event B in the previous example, "getting an even number" $B = \{2, 4, 6\}$, then its complementary event $\bar{B}$ is "getting an odd number" $\bar{B} = \{1, 3, 5\}$.

3.8.3 Union (Disjunction)

We say that the event "A or B", denoted $(A \cup B)$, occurs if at least one of the two events occurs (i.e., A occurs or B occurs).

3.8.4 Intersection (Conjunction)

The event "A and B", denoted $(A \cap B)$, occurs when both A and B occur.

3.8.5 Incompatible (Disjoint) Events

Events A and B are said to be incompatible if the occurrence of one excludes the occurrence of the other. In other words, if one occurs, the other cannot occur. Two events are incompatible if:

$$A \text{ and } B \text{ are incompatible} \Leftrightarrow A \cap B = \emptyset.$$

Example 3.18. B and $\bar{B}$ are incompatible.

3.8.6 Complete System of Events

Let Ω be the sample space associated with a random experiment, and let $A_1, A_2, \ldots, A_n$ be events in Ω. The events $A_1, A_2, \ldots, A_n$ form a complete system of events if the following conditions are met:

1. The A_i are realizable ($A_i \neq \emptyset$) for all $i = 1, 2, \ldots, n$.

2. The A_i are pairwise incompatible: $A_i \cap A_j = \emptyset$ for all $(i, j) \in \{1, 2, \ldots, n\}$ with $i \neq j$.

3. $\bigcup_{i=1}^{n} A_i = \Omega$.

Example 3.19. In the roll of a die (once), we have $\Omega = \{1, 2, 3, 4, 5, 6\}$. The events $A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{4\}, A_5 = \{5\}, A_6 = \{6\}$ form a complete system of events.

3.9 Axiomatic Definition of Probability

Let A and B be two events in Ω.

1. The probability of an event A is a number between 0 and 1:

$$0 \leq p(A) \leq 1$$

2. The probability of the certain event Ω is 1, and the probability of the impossible event $\emptyset$ is 0:

$$p(\Omega) = 1 \quad \text{and} \quad p(\emptyset) = 0.$$

3. If A and B are incompatible events ($A \cap B = \emptyset$), then:

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

Remark 3.5. We have

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

1. If $A \subseteq B$, then $p(A) \leq p(B)$.

2. Generalization to n events: Let $A_1, A_2, \ldots, A_n$ be pairwise incompatible events. Then:

$$p(A_1 \cup A_2 \cup \cdots \cup A_n) = p(A_1) + p(A_2) + \cdots + p(A_n), \text{ i.e., } p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i).$$

Consequence: Let A be an event and $\bar{A}$ its complement, then:

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

Indeed,

$$A \cup \bar{A} = \Omega, \text{ so } p(A \cup \bar{A}) = p(\Omega) \Leftrightarrow p(A) + p(\bar{A}) = p(\Omega) = 1 \Leftrightarrow p(\bar{A}) = 1 - p(A).$$

Remark 3.6. Any calculation that leads to negative or values greater than 1 for probabilities is incorrect.

Example 3.20. Reconsider the roll of a balanced six-sided die, where:

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

with events:

$$A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{4\}, A_5 = \{5\}, A_6 = \{6\}.$$

The events $A_i, i = 1, 2, \ldots, 6$ are all incompatible, and $p(A_i) = \frac{1}{6}$ for all $i = \overline{1, 6}$. $A_1 \cup A_2 \cup \cdots \cup A_n = \{1, 2, 3, 4, 5, 6\} = \Omega$. Thus:

$$p(A_1 \cup A_2 \cup \cdots \cup A_n) = p(A_1) + p(A_2) + \cdots + p(A_n) = 6 \times \frac{1}{6} = 1.$$

Classical Definition of Probabilities

For each event A of a random experiment (r.e.), a number is associated, denoted $p(A)$, which measures the probability of the occurrence of A. If an r.e. has N possible outcomes, and among these N outcomes, there are n outcomes favorable to the occurrence of event A, the probability of event A is defined as:

$$p(A) = \frac{\text{number of favorable outcomes}}{\text{number of possible outcomes}} = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega}.$$

Alternatively,

$$p(A) = \frac{n}{N}.$$

Example 3.21. In the roll of a balanced six-sided die, let A be the event "getting an even number." The number of favorable outcomes for this event is 3 ($A = \{2, 4, 6\}$), and the total number of outcomes is 6, so:

$$p(A) = \frac{3}{6} = \frac{1}{2}.$$

Conditional Probabilities

Consider the roll of two fair dice, and let A be the event: "the sum of the points obtained is at least 10".
The outcomes that give a sum of at least 10 are:

$$A = \{(4,6), (5,5), (5,6), (6,4), (6,5)\}, \text{ and } p(A) = \frac{6}{36} = \frac{1}{6}.$$

(a) Suppose that the first die shows a 3 (event B: "getting the number 3 on the top face of the first die"). Then, event A becomes impossible (A and B are mutually exclusive). We say that the probability of A given that B occurs is zero, and we write $p(A/B) = 0$.

(b) Now suppose that the first die shows a 6 (event C). To reach or exceed a sum of 10, the second die must show 4, 5, or 6. There are 3 favorable outcomes out of 6, so $p(A/B) = \frac{3}{6} = \frac{1}{2}$.

Definition 3.2. Let A and B be two events such that $p(B) \neq 0$. The conditional probability of A given that B occurs is given by the formula:

$$p(A/B) = \frac{p(A \cap B)}{p(B)}$$

Example 3.22. An urn contains 2 red balls and 3 white balls. We draw one ball, keep it, and then draw another.

(i) What is the probability of drawing a red ball on the second draw, given that a red ball was drawn on the first draw?

(ii) What is the probability of drawing two red balls during the two draws (one red ball in each draw)?

Solution

Let us define the following events:

A_1: "drawing a red ball on the first draw".

A_2: "drawing a red ball on the second draw".

$A_1 \cap A_2$: "drawing a red ball in both draws (two red balls)".

(i) We have

$$p(A_1) = \frac{2}{5}$$

and the probability of drawing a red ball on the second draw, given that a red ball was drawn on the first draw, is:

$$p(A_2/A_1) = \frac{1}{4}$$

(ii) By definition:

$$p\left(A_2/A_1\right) = \frac{p\left(A_1 \cap A_2\right)}{p\left(A_1\right)}$$

So,

$$p\left(A_1 \cap A_2\right) = p\left(A_1\right) \times p\left(A_2/A_1\right) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}.$$

Formula of Compound Probabilities

For any events A and B such that $p(A) \neq 0$ and $p(B) \neq 0$, we have:

$$p(A/B) = \frac{p(A \cap B)}{p(B)}, \text{ then } p(A \cap B) = p(A/B)p(B)$$
$$p(B/A) = \frac{p(A \cap B)}{p(A)}, \text{ then } p(A \cap B) = p(B/A)p(A)$$

From these two formulas, we deduce:

$$p(A \cap B) = p(A/B)p(B) = p(B/A)p(A)$$

Example 3.23. An urn contains three white balls and two black balls. We draw two balls successively without replacement.
What is the probability that the first ball is black and the second is white?

Solution

Let us define the following events:
A: "drawing a black ball on the first draw".
B: "drawing a white ball on the second draw".
We have

$$p(A) = \frac{2}{5} \text{ and } p(B/A) = \frac{3}{4},$$

so

$$p(A \cap B) = p(B/A)p(A) = \frac{3}{4} \times \frac{2}{5} = \frac{3}{10}.$$

Independence of Events

Definition of Independence

Two events A and B are said to be independent if the occurrence of one does not affect the probability of the occurrence of the other. In mathematical terms, events A and B are independent if:

$$p(A \cap B) = p(A) \times p(B)$$

If this condition holds, then A and B are independent; otherwise, they are dependent.

Example 3.24. Consider the case of rolling two fair dice. Let event A be "the sum of the dice is even" and event B be "the first die shows a 3". These two events are independent because the outcome of the first die does not affect whether the sum is even or odd.

We can verify this by calculating:

$$p(A \cap B) = p(\text{sum is even and first die is 3}) = \frac{3}{36} = \frac{1}{12}$$

The probability of A is the probability of the sum being even, which is $\frac{18}{36} = \frac{1}{2}$, and the probability of B is the probability that the first die shows a 3, which is $\frac{6}{36} = \frac{1}{6}$.

Now check if $p(A \cap B) = p(A) \times p(B)$:

$$\frac{1}{12} = \frac{1}{2} \times \frac{1}{6}$$

Since this equality holds, we can conclude that events A and B are independent.

Independence of EventsBayes' Theorem

Bayes' Theorem provides a way to update the probability of an event, given new information. If events $A_1, A_2, \ldots, A_n$ form a partition of the sample space Ω, then for any event B:

$$p(A_i \mid B) = \frac{p(B \mid A_i)p(A_i)}{p(B)} \quad \text{for } i = 1, 2, \ldots, n$$

where:

- $p(A_i \mid B)$ is the conditional probability of A_i given B.
- $p(B \mid A_i)$ is the likelihood of B given A_i.
- $p(A_i)$ is the prior probability of A_i.
- $p(B)$ is the total probability of B, given by:

$$p(B) = \sum_{i=1}^{n} p(B \mid A_i)p(A_i)$$

3.9.1 Bayes' Theorem

It allows us to revise the probability of an event based on new evidence.

1. Example of Bayes' Theorem

   Suppose there are two types of light bulbs in a factory: 80% of the bulbs are good, and 20% are defective. A testing machine is 90% accurate in identifying good bulbs and 95% accurate in identifying defective bulbs. If a bulb is tested and the result is positive (i.e., the machine says the bulb is good), what is the probability that the bulb is actually good?

   Let:

   - A_1 be the event "the bulb is good".
   - A_2 be the event "the bulb is defective".

- B be the event "the test result is positive".

We are asked to find $p(A_1 \mid B)$, the probability that the bulb is good given the test result is positive.

From Bayes' Theorem:

$$p(A_1 \mid B) = \frac{p(B \mid A_1)p(A_1)}{p(B)}$$

We know:

- $p(A_1) = 0.80$, $p(A_2) = 0.20$.
- $p(B \mid A_1) = 0.90$ (the probability of a correct positive test for a good bulb).
- $p(B \mid A_2) = 0.05$ (the probability of a false positive test for a defective bulb).

The total probability of a positive test is:

$$p(B) = p(B \mid A_1)p(A_1) + p(B \mid A_2)p(A_2) = (0.90 \times 0.80) + (0.05 \times 0.20) = 0.72 + 0.01 = 0.73$$

Thus:

$$p(A_1 \mid B) = \frac{(0.90 \times 0.80)}{0.73} = \frac{0.72}{0.73} \approx 0.986$$

So, given that the test result is positive, the probability that the bulb is actually good is approximately 98.6%.

3.10 Exercices

Exercice 1: Describe the sample space and all 16 events for a trial in which two coins are thrown and each shows either a head or a tail.

Exercice 2: A fair coin is tossed, and a fair die is thrown. Write down sample spaces for

(a) the toss of the coin;

(b) the throw of the die;

(c) the combination of these experiments.

Let A be the event that a head is tossed, and B be the event that an odd number is thrown. Directly

from the sample space, calculate $P(A \cap B)$ and $P(A \cup B)$.

Exercice 3: Are the following assertions true or false?

1. Two incompatible events are independent.
2. Two independent events are incompatible.
3. If $P(A) + P(B) = 1$, then $A = \bar{B}$ (the complement of B).
4. If A and B are two independent events, then $P(A \cup B) = P(A) + P(B)$.

Execice 4: A neighbor has two children, and we are unaware of their genders. Define the following events:

- A: The two children are of different genders.
- B: The older child is a girl.
- C: The younger child is a boy.

Assume that the probability of a child being a girl (or a boy) at birth is equal to $1\backslash 2$. Show that the events A, B, and C are pairwise independent but not mutually independent.

Exercice 5: A bag contains 3 red and 7 black balls. Two balls are drawn at randon without replacement. If the second ball is red, what is the probability that the first ball is also red?

Exercice 6: If a family has two children, what is the conditional probability that both are girls if there is atleast one girl?

Exercice 7: A dice and a coin are tossed simultaneously. Find the probability of obtaining a 6, given that a head came up.

Exercice 8: I have in my pocket ten coins. Nine of them are ordinary coins with equal chances of coming up head

and tail when tossed and the tenth has two heads.

(a) If I take one of the coins at random from my pocket, what is the probability that it is the coin with two heads ?

(b) If I toss the coin and it comes up heads, what is the probability that it is the coin with two heads ?

(c) If I toss the coin one further time and it comes up tails, what is the probability that it is one of

the nine ordinary coins ?

Exercice 9: An urn contains r red balls and b blue balls, $r \geq 1$ and $b \geq 3$. Three balls are selected, without replacement, from the urn. Using the notion of conditional probability to simplify the problem, find the probability of the sequence Blue, Red, Blue.

Exercice 10: Let n be an integer greater than or equal to 3. We have two urns, U and V.

Urn **U** contains 2 white balls and n black balls; urn **V** contains n white balls and 2 black balls.

One of the two urns is chosen at random, then two balls are drawn from that urn, successively and without replacement.

We denote by **U** the event: "urn U is chosen", by **V** the event: "urn V is chosen", and by **B** the event: "both drawn balls are white".

- 1. Show that:

$$P(B \cap U) = \frac{1}{(n+2)(n+1)}$$

  2. Deduce

$$P(B) = \frac{n(n-1)+2}{2(n+2)(n+1)}$$

  3.

$$P(U \mid B) = \frac{2}{n(n-1)+2}$$

4. Prove that to ensure that

$$P(U \mid B) \leq 0.1,$$

it is sufficient to take $n \geq 5$.

Exercise 11: A class of 18 students in the 1^{st} university year of SMI faculty is distributed as follows:

- 6 students in mathematics: 4 males and 2 females
- 5 students in computer sciences: 3 males and 2 females
- 4 students in physic: all males
- 3 students in chemistry: all females

I. Committee Formation: We want to randomly form a committee of 4 students from the class.

1. Calculate the probability of each of the following events:
    - Event **A**: "The committee is composed of students from the same specialty"
    - Event **B**: "The committee is composed of students of the same gender"
    - Event **C**: "The committee includes students from all specialties"
    - Event **D**: "The committee includes all chemistry students"
2. Show that $P(A \cap B) = \frac{1}{1530}$ and deduce that the probability that the committee includes students either from the same specialty or of the same gender.

II. Random Variable on Pair Selection

Now, we randomly choose two students from the class: one will be designated as president, and the other as vice-president.

Let X be the random variable that gives the number of mathematics students in the selected committee.

(1) Justify that the possible values of X are $\{0, 1, 2\}$. Then, determine the probability distribution (law) of X.

(2-a) Compute the expected value $\mathbb{E}(X)$ of the random variable X.

(2-b) Find the real number α such that:

$$\mathbb{E}(2025X + \alpha) = 2024,$$

3.11 Corrected exercices

Exercice 1: The sample space is $S = \{hh, ht, th, tt\}$. As this has 4 elements there are $2^4 = 16$ subsets, namely

∅, hh, ht, th, tt, {hh, ht}, {hh, th}, {hh, tt}, {ht, th}, {ht, tt}, {th, tt}, {hh, ht, th}, {hh, ht, tt},

{hh, th, tt}, {ht, th, tt} and finally {hh, ht, th, tt}.

Exercice 2:

(a) $\{Head, Tail\}$

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

(c) $\{(1\backslash Head), (1\backslash Tail), ..., (6\backslash Head), (6\backslash Tail)\}$

Clearly $P(A) = 1/2 = P(B)$. We can assume that the two events are independent, so $P(A\backslash B) = P(A)P(B) = 1/4$.

Alternatively, we can examine the sample space above and deduce that three of the twelve equally likely events comprise A and B.

Also, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/4$, where this probability can also be determined by noticing from the sample space that nine of twelve equally likely events comprise $A \cup B$.

Exercice 3:

1. False! It's enough to take:

A with $0 < P(A) < 1$and $B = \bar{A}$. Then A and B are incompatible, but they are not independent!

2. Not necessarily! If you roll a perfectly balanced die, and consider the two events:

$A =$ "Getting an even number" and

$B =$ "Getting 5 or 6".

Then A and B are independent. Indeed, $P(A) = 1/2, P(B) = 1/3$, and we have $P(A \cap B) = 1/6$, but they are not incompatible.

In fact, independent events are almost never incompatible. At the very least, the probability of one of them would have to be zero!

3. False! It's enough that $P(A \cap B) = 0$ with $A \cap B \neq \emptyset$ for it to not be true.

4. False (take the example above with the die...).

5. That's true!

Exercice 4

Sample Space

The possible gender combinations for the two children, considering the order (older, younger), are:

Outcome	Probability
(G, G)	$\frac{1}{4}$
(G, B)	$\frac{1}{4}$
(B, G)	$\frac{1}{4}$
(B, B)	$\frac{1}{4}$

Probabilities of Individual Events

- $P(A) = P(G, B) + P(B, G) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
- $P(B) = P(G, G) + P(G, B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
- $P(C) = P(G, B) + P(B, B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

Pairwise Independence

Two events X and Y are independent if $P(X \cap Y) = P(X) \cdot P(Y)$.

- $P(A \cap B) = P(G, B) = \frac{1}{4}$; $P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

- $P(A \cap C) = P(G, B) = \frac{1}{4}$; $P(A) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
- $P(B \cap C) = P(G, B) = \frac{1}{4}$; $P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Since in each case $P(X \cap Y) = P(X) \cdot P(Y)$, the events are pairwise independent.

Mutual Independence

For mutual independence, we require:

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

Calculate:

- $P(A \cap B \cap C) = P(G, B) = \frac{1}{4}$
- $P(A) \cdot P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

Since $\frac{1}{4} \neq \frac{1}{8}$, the events are not mutually independent.

Exercice 5: Let A: event of selecting a red ball in first draw

B: event of selecting a red ball in second draw

$P(A \cap B) = P(\text{selecting both red balls}) = 3/10 \times 2/9 = 1/15$.

$P(B) = P(\text{selecting a red ball in second draw}) = P(\text{red ball and rad ball or black ball and red ball})$

$= P(\text{red ball and red ball}) + P(\text{black ball and red ball}) = 3/10 \times 2/9 + 7/10 \times 3/9 = 3/10$. Consequently

$$P(A|B) = P(A \cap B)/P(B) = 1/15 \div 3/10 = 2/9.$$

Exercice 6: Let A: both being girls

B: Atleast one girl

$n(A) = 1$, $n(B) = 3$, $n(A \cap B) = 1$, then

$$P(A|B) = n(A \cap B)/n(B) = 1/3.$$

Exercice 7: Let A: six coming with a heads

B: coin shows a head

$A = \{(6, H))$

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

$n(A \cap B) = 1$ and $n(B) = 6$

Probability of getting a six when there is a head is given by

$$P(A|B) = n(A \cap B)/n(B) = 1/6.$$

Exercise 8:

Class Composition

A class of 18 students is distributed as follows:

- Mathematics: 6 students (4 males, 2 females)
- Computer Science: 5 students (3 males, 2 females)
- Physics: 4 students (4 males)
- Chemistry: 3 students (3 females)

I. Committee Formation: We choose 4 students at random.

1. Probability Calculations

Total number of committees:

$$C_{18}^4 = 3060$$

Event A: Same specialty

$$\text{Mathematics: } C_6^4 = 15$$
$$\text{Computer Science: } C_5^4 = 5$$
$$\text{Physics: } C_4^4 = 1$$
$$\text{Chemistry: } C_3^4 = 0$$
$$\Rightarrow P(A) = \frac{15 + 5 + 1}{3060} = \frac{21}{3060}$$

Event B: Same gender

$$\text{Males (11): } C_{11}^4 = 330$$
$$\text{Females (7): } C_7^4 = 35$$
$$\Rightarrow P(B) = \frac{330+35}{3060} = \frac{365}{3060}$$

Event C: All specialties represented

$$6 \times 5 \times 4 \times 3 = 360 \Rightarrow P(C) = \frac{360}{3060}$$

Event D: All 3 chemistry students in committee

$$C_{15}^1 = 15 \Rightarrow P(D) = \frac{15}{3060}$$

2. Probability of A or B

$$\text{Event A} \cap \text{B}: \text{ Math males: } C_4^4 = 1, \text{ Physics males: } C_4^4 = 1$$
$$\Rightarrow P(A \cap B) = \frac{1+1}{3060} = \frac{2}{3060}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{21+365-2}{3060} = \frac{384}{3060}$$

II. Random Variable: Pair Selection

We choose a president and vice-president from 18 students.

1. Values and Probability Law of X : Let X be the number of mathematics students in the pair.

$$\text{Possible values of } X : \{0, 1, 2\}$$

Total unordered pairs:

$$A_{18}^2 = 306$$

Probabilities:

$$P(X=0):\ A_{12}^2 = 66$$

$$P(X=1):\ A_6^1 \cdot A_{12}^1 = 6 \cdot 12 = 72$$

$$P(X=2):\ A_6^2 = 15$$

$$\Rightarrow \text{Law of } X: \begin{cases} P(0) = \frac{132}{306} \\ P(1) = \frac{72}{306} \\ P(2) = \frac{30}{306} \end{cases}$$

2-a. Expected Value of X

$$\mathbb{E}(X) = 0 \cdot \frac{132}{306} + 1 \cdot \frac{72}{306} + 2 \cdot \frac{30}{306} = \frac{132}{306}$$

2-b. Find

α such that $\mathbb{E}(2025X + \alpha) = 2024$. We are given the equation:

$$E(2025X + \alpha) = 2024$$

and we need to find the real number α. The expected value operator has the following properties:

- $E(aX + b) = aE(X) + b$, where a and b are constants, and X is a random variable.

Applying this property to the given equation:

$$E(2025X + \alpha) = 2025E(X) + \alpha$$

We are given that $E(X) = \frac{2}{5}$. Substituting this value into the equation, we get:

$$2025 \times \frac{132}{306} + \alpha = 2024$$

$$\alpha = 2024 - 810 = 1150,47$$

Exercice 9:

1.

$$P(B \cap U) = P(U) \times P(B \mid U)$$

We have:

$$P(U) = \frac{1}{2}, \quad P(B \mid U) = \frac{2}{n+2} \times \frac{1}{n+1}$$

So:

$$P(B \cap U) = \frac{1}{2} \times \frac{2}{n+2} \times \frac{1}{n+1} = \frac{1}{(n+2)(n+1)}$$

2.Similarly:

$$P(B \cap V) = P(V) \times P(B \mid V)$$
$$P(V) = \frac{1}{2}, \quad P(B \mid V) = \frac{n}{n+2} \times \frac{n-1}{n+1}$$

So:

$$P(B \cap V) = \frac{1}{2} \times \frac{n(n-1)}{(n+2)(n+1)}$$

Now:

$$P(B) = P(B \cap U) + P(B \cap V) = \frac{1}{(n+2)(n+1)} + \frac{1}{2} \times \frac{n(n-1)}{(n+2)(n+1)}$$

$$P(B) = \frac{2 + n(n-1)}{2(n+2)(n+1)}$$

3.

$$P(U \mid B) = \frac{P(B \cap U)}{P(B)} = \frac{\frac{1}{(n+2)(n+1)}}{\frac{2+n(n-1)}{2(n+2)(n+1)}} = \frac{2}{n(n-1)+2}$$

4. We want

$$P(U \mid B) \leq 0.1 \Rightarrow \frac{2}{n(n-1)+2} \leq 0.1 \Rightarrow 2 \leq 0.1(n(n-1)+2)$$

$$\Rightarrow 20 \leq n(n-1) + 2 \Rightarrow 18 \leq n(n-1)$$

Try integer values:

$$\text{For } n = 4: \quad 4 \times 3 = 12 < 18 \text{ For } n = 5: \quad 5 \times 4 = 20 > 18$$

So, for $P(U \mid B) \leq 0.1$, it is sufficient that $n \geq 5$.

Chapter 4

Exams

Exam – University of Khemis Miliana 2022/2023

Exercise 1

A study is conducted on a population of 60 fish whose weights (in grams) are grouped into classes.

1. Identify:
   - The population and its size,
   - The statistical unit,
   - The statistical variable and its type.
2. The relative frequencies f_i are given in the table below. Complete the table by computing the cumulative frequencies F_i^+ and the corrected frequencies f_i'.

Classes	Class width a_i	f_i	F_i^+	$f_i' = \frac{f_i}{a_i}$
$[8, 8.5[$	0.5	0.05		
$[8.5, 9[$	0.5	0.20		
$[9, 9.4[$	0.4	0.40		
$[9.4, 9.6[$	0.2	0.25		
$[9.6, 10[$	0.4	0.10		

3.
   - Draw the histogram of the distribution.
   - Draw the frequency polygon.
4. Determine graphically the mode of the distribution and interpret it.
5. Determine the median using interpolation.
6. What percentage of fish have a weight less than the median?

Exercise 2

An urn contains 5 black balls and 5 white balls. We perform n successive draws **with replacement**.

1. Let:
   - E_1: event "all drawn balls are of the same color",
   - E_2: event "exactly one white ball is obtained".

   Compute $P(E_1)$ and $P(E_2)$.
2. Let:
   - A: event "at least one white ball is drawn",
   - B: event "at most one white ball is drawn".

   (a) Express $P(A \cap B)$, $P(A)$, and $P(B)$.

   (b) For which values of n are the events A and B independent?
3. Consider the sequence (u_n) defined by:

$$u_n = 2^{n-1} - (n+1)$$

   (a) Compute u_2, u_3, and u_4.

   (b) Study the monotonicity of the sequence (u_n).

   (c) Deduce for which value(s) of n the events A and B are independent.

Correction Exam – University of Khemis Miliana 2022/2023

Exercise 1

1.
   - Population: the fish, size $N = 60$.
   - Statistical unit: a fish.
   - Statistical variable: the weight.
   - Type of variable: quantitative continuous.

2. Frequencies and cumulative frequencies:

Classes	a_i	f_i	F_i^+	$f_i' = \frac{f_i}{a_i}$
$[8, 8.5[$	0.5	0.05	0.05	0.10
$[8.5, 9[$	0.5	0.20	0.25	0.40
$[9, 9.4[$	0.4	0.40	0.65	1.00
$[9.4, 9.6[$	0.2	0.25	0.90	1.25
$[9.6, 10[$	0.4	0.10	1.00	0.25
Total		1		

3.
   - The corrected frequencies are given by $f_i' = \frac{f_i}{a_i}$.
   - Draw the histogram using f_i'.
   - Draw the frequency polygon.

4. The mode is determined graphically:

$$\text{Mode} \approx 9.44$$

   This means that most fish have a weight around 9.44 grams.

5. Median:

   Since

$$0.25 \leq 0.50 \leq 0.65,$$

   the median class is $[9, 9.4[$.

Using interpolation:

$$M_e = e_2 + a_3 \left(\frac{0.5 - F_2^+}{f_3} \right)$$

$$M_e = 9 + 0.4 \left(\frac{0.5 - 0.25}{0.4} \right) = 9.25$$

Hence:

$$M_e = 9.25 \text{ grams}$$

6. The percentage of fish with weight less than 9.25 grams is:

$$50\%$$

since the median divides the population into two equal parts.

Exercise 2

(a) The experiment consists of drawing n times with replacement from an urn containing 5 black balls and 5 white balls.

The sample space has 10^n equiprobable outcomes.

- Event E_1: all drawn balls are of the same color.

  There are 5^n ways to obtain only black balls and 5^n ways to obtain only white balls. Hence:

$$P(E_1) = \frac{5^n + 5^n}{10^n} = \frac{2 \cdot 5^n}{10^n} = \frac{1}{2^{n-1}}$$

- Event E_2: exactly one white ball.

  Choose the position of the white ball in n ways, and for each configuration:

$$P(E_2) = \frac{n \cdot 5^n}{10^n} = \frac{n}{2^n}$$

(b) i. • $A \cap B$ corresponds to "exactly one white ball", hence:

$$P(A \cap B) = P(E_2) = \frac{n}{2^n}$$

- A is the event "at least one white ball", i.e. the complement of E_1:

$$P(A) = 1 - P(E_1) = 1 - \frac{1}{2^{n-1}}$$

- B is the event "at most one white ball":

$$P(B) = P(\text{no white ball}) + P(E_2)$$

$$P(B) = \frac{1}{2^n} + \frac{n}{2^n} = \frac{n+1}{2^n}$$

ii. The events A and B are independent if and only if:

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

That is:

$$\frac{n}{2^n} = \left(1 - \frac{1}{2^{n-1}}\right)\frac{n+1}{2^n}$$

Simplifying:

$$n = (n+1)\left(\frac{2^{n-1} - 1}{2^{n-1}}\right) \iff 2^{n-1} = n+1$$

(c) Consider the sequence $u_n = 2^{n-1} - (n+1)$.

i.

$$u_2 = -1, \quad u_3 = 0, \quad u_4 = 4$$

ii.

$$u_{n+1} - u_n = (2^n - (n+2)) - \big(2^{n-1} - (n+1)\big) = 2^{n-1} - 1$$

Since $2^{n-1} - 1 > 0$ for all $n \geq 2$, the sequence (u_n) is strictly increasing.

iii. The independence condition is equivalent to:

$$2^{n-1} = n+1 \iff u_n = 0$$

Since (u_n) is strictly increasing and $u_3 = 0$, we conclude:

$$\boxed{n = 3}$$

Exam – University of Khemis Miliana 2023/2024

Exercise 1

We consider an urn (or experiment) involving two events B and C such that:

$$P(B) = \frac{2}{3}, \quad P(C \mid B) = \frac{1}{2}, \quad P(C \mid \overline{B}) = 0.$$

(a) Calculate $P(C \cap B)$ and $P(C \cap \overline{B})$.

(b) Deduce $P(C)$.

(c) Let F_n be the event "obtaining n consecutive successes (Heads)". Compute $P(F_n \cap B)$ and $P(F_n \cap \overline{B})$.

(d) Deduce $P(F_n)$.

(e) Compute $\lim_{n\to\infty} P(F_n)$ and interpret the result.

Exercise 2

The following tables give the distribution of speeds (in km/h) for two vehicles A and B.

Vehicle A

Speed interval	Frequency
$[40, 50)$	3
$[50, 60)$	6
$[60, 70)$	8
$[70, 80)$	2

Vehicle B

Speed interval	Frequency
$[40, 50)$	2
$[50, 60)$	5
$[60, 70)$	7
$[70, 80)$	6

(a) For each vehicle:

- Calculate the mean speed.
- Determine the median.
- Calculate the standard deviation.

(b) Represent the distributions graphically (histogram and/or frequency polygon).

(c) Compare the two vehicles in terms of:

- Average speed.
- Speed consistency.

Correction

Exercise 1

$$P(C \cap B) = P(C \mid B) \cdot P(B) = \frac{1}{2} \cdot \frac{2}{3}$$

$$P(C \cap \overline{B}) = P(C \mid \overline{B}) \cdot P(\overline{B}) = 0 \cdot \frac{1}{3} = 0$$

$$P(C) = P(C \cap B) + P(C \cap \overline{B}) = \frac{1}{3}$$

$$P(F_n \cap B) = P(F_n \mid B) \cdot P(B) = \left(\frac{1}{2}\right)^n \cdot \frac{2}{3}$$

$$P(F_n \cap \overline{B}) = P(F_n \mid \overline{B}) \cdot P(\overline{B}) = 1 \cdot \frac{1}{3}$$

$$P(F_n) = P(F_n \cap B) + P(F_n \cap \overline{B}) = \frac{2}{3}\left(\frac{1}{2}\right)^n + \frac{1}{3}$$

As $n \to +\infty$, $\left(\frac{1}{2}\right)^n \to 0$, so:

$$\lim_{n\to\infty} P(F_n) = \frac{1}{3}$$

Conclusion: As the number of tosses increases, the probability of obtaining all Heads approaches the probability of having selected the biased coin.

Exercise 2

1. Descriptive Statistics

Vehicle A:

Interval	Midpoint x_i	Frequency n_i	$f_i x_i$
$[40, 50)$	45	3	135
$[50, 60)$	55	6	330
$[60, 70)$	65	8	520
$[70, 80)$	75	2	150
Total		19	1135

$$\bar{x}_A = \frac{1135}{19} \approx 59.74 \text{ km/h}$$

Median class: $[60, 70)$

$$\text{Median}_A = 60 + \left(\frac{9.5 - 9}{8}\right) \cdot 10 = 60.625$$

$$\sum f_i x_i^2 = 3 \cdot 45^2 + 6 \cdot 55^2 + 8 \cdot 65^2 + 2 \cdot 75^2 = 69275$$

$$\sigma_A = \sqrt{\frac{69275}{19} - (59.74)^2} \approx \sqrt{76.74} \approx 8.76$$

Vehicle B:

Interval	Midpoint x_i	Frequency n_i	$f_i x_i$
$[40, 50)$	45	2	90
$[50, 60)$	55	5	275
$[60, 70)$	65	7	455
$[70, 80)$	75	6	450
Total		20	1270

$$\bar{x}_B = \frac{1270}{20} = 63.5 \text{ km/h}$$

Median class: $[60, 70)$

$$\text{Median}_B = 60 + \left(\frac{10-7}{7}\right) \cdot 10 = 64.29$$

$$\sum f_i x_i^2 = 2 \cdot 45^2 + 5 \cdot 55^2 + 7 \cdot 65^2 + 6 \cdot 75^2 = 82450$$

$$\sigma_B = \sqrt{\frac{82450}{20} - (63.5)^2} = \sqrt{90.25} = 9.5$$

2. Graphical Representations

Draw histograms and frequency polygons using the intervals and frequencies of both vehicles.

3. Comparison and Interpretation

a) Which vehicle drives faster on average?

Vehicle B has a higher mean speed (63.5 km/h compared to 59.74 km/h). This suggests that Vehicle B tends to travel at higher speeds on average.

b) Which vehicle has a more stable (consistent) speed?

Vehicle A has a smaller standard deviation (8.76 compared to 9.5), indicating more consistent speeds.

Exam-University of Khemis Miliana 2024/2025

Exercise 1

A study concerns companies in the automobile sector, where the variable under study is the annual turnover (in billions of DZD).

1. Specify:
   - The population,
   - The statistical variable,
   - The type of variable.
2. The distribution of turnover is given by:

Classes	Number of companies n_i
$[0, 0.25[$	80
$[0.25, 0.5[$	100
$[0.5, 1[$	180
$[1, 2.5[$	150
$[2.5, 5[$	50

   (a) Determine the class centers x_i.

   (b) Compute:
   - The arithmetic mean,
   - The variance,
   - The standard deviation.

3. (a) Construct the histogram of frequencies.

   (b) Determine graphically the mode.

   (c) Interpret the result.

4. (a) Determine the cumulative increasing frequencies N_i^+.

   (b) Determine the cumulative decreasing frequencies N_i^-.

(c) Plot both cumulative curves.

(d) Determine graphically the median.

5. Estimate the percentage of companies whose turnover exceeds 2 billion DZD.

Exercise 2

Solve the following combinatorial problems:

1. Compute the number of permutations of 10 elements.

2. Compute $4! \times 7!$.

3. Compute $10! - 4! \times 7!$.

4. Compute:

$$\binom{4}{3} + \binom{6}{3}$$

5. Compute:

$$\binom{10}{3} - \binom{4}{3} - \binom{6}{3}$$

Correction

Exercise 01

1. Population: companies in the automobile sector.

   Variable: turnover (gross revenue).

   Type: quantitative continuous.

2.

Classes	$[0, 0.25[$	$[0.25, 0.5[$	$[0.5, 1[$	$[1, 2.5[$	$[2.5, 5[$
n_i	80	100	180	150	50
x_i	0.125	0.375	0.75	1.75	3.75
a_i	0.25	0.25	0.5	1.5	2.5
N_i^+	80	180	360	510	560
N_i^-	560	480	380	200	50
$n'_i = \frac{n_i}{a_i}$	320	400	360	100	20

We have:

$$\bar{x} = \frac{1}{560} \sum n_i x_i$$

$$\bar{x} = \frac{80 \times 0.125 + 100 \times 0.375 + 180 \times 0.75 + 150 \times 1.75 + 50 \times 3.75}{560} = 1.13$$

and

$$Var(X) = \left(\frac{1}{560} \sum n_i x_i^2 \right) - \bar{x}^2$$

$$= \frac{80 \times 0.125^2 + 100 \times 0.375^2 + 180 \times 0.75^2 + 150 \times 1.75^2 + 50 \times 3.75^2}{560} - \bar{x}^2$$

$$= 80 \times 0.016 + 100 \times 0.14 + 180 \times 0.562 + 150 \times 3.06 + 50 \times 14.06$$

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

3. (a) Histogram of frequencies:
    - Determination of n'_i
    - Graphical representation

   (b) Determination of mode graphically:

$$Mo = 0.4175$$

   This value indicates that the majority of companies have a turnover around:

$$0.4175$$

4. Graphs of N_i^+ and N_i^-:
   - Determination of N_i^+ and N_i^-
   - Graphical representation
   - Determination of median graphically:

$$Me = 0.8$$

5. Since $2 \in [1, 2.5[$, we add a new class $[2, 2.5[$.

   Length of this class:

$$0.5$$

   Using proportionality:

$$\begin{cases} 1.5 \to 150 \\ 0.5 \to x \end{cases}$$

   Thus:

$$n_i^* = x = \frac{0.5 \times 150}{1.5} = 50$$

   Therefore:

$$N^{-*} = 50 + 50 = 100$$

   Percentage of companies whose turnover exceeds 2 billion DZD:

$$\frac{100}{560} \times 100 = 17.86\%$$

Exercise 2

1. $$a = 10!$$

2. $$b = 4! \times 7!$$

3. $$c = 10! - 4! \times 7!$$

4. $$d = \binom{4}{3} + \binom{6}{3}$$

5. $$e = \binom{10}{3} - \binom{4}{3} - \binom{6}{3}$$

Exam – University of Khemis Miliana 2024/2025

Course Questions

1. Give the definition of a sigma-algebra (tribe).

2. Let $\Omega = \{1, 2, 3\}$ and

$$C = \{\emptyset, \Omega, \{1\}, \{2\}\}$$

Verify whether C is a sigma-algebra on Ω.

Exercise 1 (Statistics)

The daily wages (in hundreds of DA) of employees in factory A are distributed as follows:

Daily Wage	Frequency
$[5, 6[$	120
$[6, 7[$	190
$[7, 8.5[$	240
$[8.5, 9[$	100
$[9, 10[$	50

1. Identify:
   - Population
   - Variable
   - Nature of the variable

2. Draw histogram and corrected frequency polygon.

3. Determine the mode.

4. Determine the median and interpret it.

5. Compute number of employees earning between 600 DA and 800 DA.

Exercise 2 (Probability)

An urn contains:

- 2 white balls
- n green balls

One urn is selected randomly from two equiprobable urns, then 2 balls are drawn simultaneously.

Define:

- U: selecting the first urn
- V: selecting the second urn
- B: obtaining two white balls

1. Compute $P(B \cap U)$ and $P(B)$.
2. Find the smallest n such that:

$$P(B \mid U) \leq 0.1$$

Correction

Course Questions

1. A sigma-algebra $\mathcal{F}$ on Ω is a collection of subsets such that:
   - $\emptyset \in \mathcal{F}$
   - If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
   - If $(A_n) \subset \mathcal{F}$ then $\bigcup A_n \in \mathcal{F}$
2. C is not a sigma-algebra since:

$$\{1\} \cup \{2\} = \{1, 2\} \notin C$$

Exercise 1 (Statistics)

- Population: employees of factory A
- Variable: daily wage
- Nature: continuous quantitative variable

Corrected table:

Class	n_i	Width	Corrected Frequency	Cumulative	Midpoint
$[5, 6[$	120	1	120	120	5.5
$[6, 7[$	190	1	190	310	6.5
$[7, 8.5[$	240	1.5	160	550	7.75
$[8.5, 9[$	100	0.5	200	650	8.75
$[9, 10[$	50	1	50	700	9.5

Mode:

$$M_0 \approx 8.60 \Rightarrow 860 \text{ DA}$$

Median:

$$M_e = 7.25 \Rightarrow 725 \text{ DA}$$

$$n_{600-800} = 190 + 160 = 350$$

Exercise 2 (Probability)

1.

$$P(B \mid U) = \frac{\binom{2}{2}}{\binom{n+2}{2}} = \frac{2}{(n+2)(n+1)}$$

$$P(B \cap U) = \frac{1}{(n+2)(n+1)}$$

$$P(B) = \frac{2}{(n+2)(n+1)}$$

2.

$$\frac{2}{(n+2)(n+1)} \leq 0.1$$

$$20 \leq (n+2)(n+1)$$

$$n^2 + 3n - 18 \geq 0$$

$$n \geq 3$$

Proposed Exam 1

Course Questions

1. Define the notion of a sigma-algebra on a set Ω.

2. Let $\Omega = \{a, b, c, d\}$ and consider:

$$\mathcal{F} = \{\emptyset, \Omega, \{a, b\}, \{c, d\}, \{a, b, c\}\}$$

   Determine whether $\mathcal{F}$ is a sigma-algebra on Ω. Justify.

Exercise 1 (Statistics)

A company records the monthly production (in units) of 800 workers. The data are grouped as follows:

Production (units)	Frequency
$[10, 20[$	90
$[20, 30[$	160
$[30, 50[$	300
$[50, 70[$	180
$[70, 100[$	70

1. Identify the population, variable, and type of variable.

2. Compute class widths and relative frequencies.

3. Determine the modal class and explain your choice.

4. Estimate the median production.

5. Compute how many workers produce less than 50 units.

Exercise 2 (Probability)

An urn contains 3 red balls and n blue balls. A game is defined as follows:

- An urn is chosen at random between two urns U_1 and U_2 (equiprobable). - In U_1: 3 red and n blue balls - In U_2: 2 red and $(n + 1)$ blue balls - Two balls are drawn simultaneously.

Define:

- U_1: choosing urn 1
- U_2: choosing urn 2
- R: drawing two red balls

1. Compute $P(R \mid U_1)$ and $P(R \mid U_2)$.
2. Deduce $P(R)$.
3. Find the smallest integer n such that:

$$P(R \mid U_1) > P(R \mid U_2)$$

Proposed Exam 2

Course Questions

1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Define what a probability measure P is and state its main properties.

2. Let A and B be two events such that:

$$P(A) = 0.6, \quad P(B) = 0.5, \quad P(A \cap B) = 0.2$$

   Compute:

   - $P(A \cup B)$
   - $P(A^c \cap B)$

Exercise 1 (Statistics)

The heights (in cm) of 600 students are grouped as follows:

Height (cm)	Frequency
$[140, 150[$	50
$[150, 160[$	120
$[160, 170[$	200
$[170, 180[$	160
$[180, 190[$	70

1. Identify the population, variable, and its nature.

2. Compute cumulative frequencies.

3. Determine the modal class.

4. Estimate the mean height using class midpoints.

5. Determine the percentage of students shorter than 170 cm.

Exercise 2 (Probability)

A box contains:

- 4 defective items,
- n non-defective items.

Two items are selected randomly without replacement.
Define:

- D_1: first item is defective,
- D_2: second item is defective.

1. Compute $P(D_1)$ and $P(D_2 \mid D_1)$.
2. Compute $P(D_1 \cap D_2)$.
3. Find the smallest n such that:

$$P(D_1 \cap D_2) < 0.05$$

Bibliography

[1] Jay L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th Edition, Cengage Learning, 2011.

[2] Douglas C. Montgomery and George C. Runger, *Applied Statistics and Probability for Engineers*, 7th Edition, Wiley, 2020.

[3] Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 5th Edition, Academic Press, 2014.

[4] John E. Freund and Benjamin M. Perles, *Modern Elementary Statistics*, 12th Edition, Pearson, 2007.

[5] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, *Probability and Statistics for Engineers and Scientists*, 9th Edition, Pearson, 2017.

[6] George Casella and Roger L. Berger, *Statistical Inference*, 2nd Edition, Duxbury Press, 2002.

[7] Alexander M. Mood, Franklin A. Graybill, and Duane C. Boes, *Introduction to the Theory of Statistics*, 3rd Edition, McGraw-Hill, 1974.

[8] William Feller, *An Introduction to Probability Theory and Its Applications, Vol. I*, 3rd Edition, Wiley, 1968.