PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
IBN KHALDOUN UNIVERSITY OF TIARET
FACULTY OF APPLIED SCIENCES
DEPARTMENT OF SCIENCE AND TECHNOLOGY
Field: Science and Technology
Branch: Science and Technology
Specialization: Science and Technology

Course Material

COURSE OF PROBABILITY AND STATISTICS

For Second-Year Science and Technology Common Core Students

Prepared by:

Dr. MARIH Salah ***Associate Professor Class B***

Reviewed by:

Name	Title	Affiliation
Pr. SAHLI Belgacem	Professor	Ibn Khaldoun University of Tiaret
Dr. ELKHIRI Laid	Associate Professor Class A	Ibn Khaldoun University of Tiaret

Academic Year: 2025 / 2026

Semestre: 3
Unité d'enseignement: UEM2.1
Matière 1: Probabilités et statistiques
VHS: 45h00 (Cours: 1h30, TD: 1h30)
Crédits: 4
Coefficient: 2

Objectifs de la matière

Ce module permet aux étudiants de voir les notions essentielles da la probabilité et de la statistique, à savoir : les séries statistiques à une et à deux variables, la probabilité sur un univers fini et les variables aléatoires.

Connaissances préalables recommandées

Mathématiques 1 et Mathématiques 2

Contenu de la matière:

Partie A : Statistiques

Chapitre 1: Définitions de base **(1 semaine)**

A.1.1 Notions de population, d'échantillon, variables, modalités
A.1.2 Différents types de variables statistiques : qualitatives, quantitatives, discrètes, continues.

Chapitre 2: Séries statistiques à une variable **(3 semaines)**

A.2.1 Effectif, Fréquence, Pourcentage.
A.2.2 Effectif cumulé, Fréquence cumulée.
A.2.3 Représentations graphiques : diagramme à bande, diagramme circulaire, diagramme en bâton. Polygone des effectifs (et des fréquences). Histogramme. Courbes cumulatives.
A.2.4 Caractéristiques de position
A.2.5 Caractéristiques de dispersion : étendue, variance et écart-type, coefficient de variation.
A.2.6 Caractéristiques de forme.

Chapitre 3: Séries statistiques à deux variables **(3 semaines)**

A.3.1 Tableaux de données (tableau de contingence). Nuage de points.
A.3.2 Distributions marginales et conditionnelles. Covariance.
A.3.3 Coefficient de corrélation linéaire. Droite de régression et droite de Mayer.
A.3.4 Courbes de régression, couloir de régression et rapport de corrélation.
A.3.5 Ajustement fonctionnel.

Partie B : Probabilités

Chapitre 1 : Analyse combinatoire **(1 Semaine)**

B.1.1 Arrangements
B.1.2 Combinaisons
B.1.3 Permutations.

Chapitre 2 : Introduction aux probabilités **(2 semaines)**

B.2.1 Algèbre des évènements
B.2.2 Définitions
B.2.3 Espaces probabilisés
B.2.4 Théorèmes généraux de probabilités

Chapitre 3 : Conditionnement et indépendance **(1 semaine)**

B.3.1 Conditionnement,
B.3.2 Indépendance,

B.3.3 Formule de Bayes.

Chapitre 4 : Variables aléatoires **(1 Semaine)**

B.4.1 Définitions et propriétés,
B.4.2 Fonction de répartition,
B.4.3 Espérance mathématique,
B.4.4 Covariance et moments.

Chapitre 5 : Lois de probabilité discrètes et continues usuelles **(3 Semaines)**

Bernoulli, binomiale, Poisson, ... ; Uniforme, normale, exponentielle, ...

Mode d'évaluation :

Contrôle continu : 40 % ; Examen final : 60 %.

Références bibliographiques:

1. D. Dacunha-Castelle and M. Duflo. Probabilités et statistiques : Problèmes à temps fixe. Masson, 1982.
2. J.-F. Delmas. Introduction au calcul des probabilités et à la statistique. Polycopié ENSTA, 2008.
3. W. Feller. an Introduction to Probability Theory and its Applications, Volume 1. Wiley & Sons, Inc., 3rd edition, 1968.
4. G. Grimmett, D. Stirzaker, Probability and Random Processes, Oxford University Press, 2nd edition, 1992.
5. J. Jacod and P. Protter, Probability Essentials, Springer, 2000.
6. A. Montfort. Cours de statistique mathématique. Economica, 1988.
7. A. Montfort. Introduction à la statistique. Ecole Polytechnique, 1991

Probability and Statistics Syllabus Translation

Semester: 3
Teaching Unit: UEM2.1
Subject 1: Probability and Statistics
VHS: 45h00 (Lectures: 1h30, Tutorials: 1h30)
Credits: 4
Coefficient: 2

Objectives of the Course

This module allows students to grasp the essential concepts of probability and statistics, namely: one-variable and two-variable statistical series, probability on a finite universe, and random variables.

Recommended Prerequisite Knowledge

Mathematics 1 and Mathematics 2

Course Content

Part A: Statistics

Chapter 1: Basic Definitions *(1 week)*

A.1.1 Concepts of population, sample, variables, modalities
A.1.2 Different types of statistical variables: qualitative, quantitative, discrete, continuous

Chapter 2: One-Variable Statistical Series *(3 weeks)*

A.2.1 Count, Frequency, Percentage
A.2.2 Cumulative Count, Cumulative Frequency
A.2.3 Graphical representations: bar chart, circular chart, histogram, polyline of counts (and frequencies). Histograms. Cumulative curves.
A.2.4 Position indicators
A.2.5 Dispersion indicators: range, variance and standard deviation, coefficient of variation
A.2.6 Shape indicators

Chapter 3: Two-Variable Statistical Series *(3 weeks)*

A.3.1 Data tables (contingency tables). Scatter plots
A.3.2 Marginal and conditional distributions. Covariance
A.3.3 Linear correlation coefficient. Regression line and Mayer's line
A.3.4 Regression curves, regression color chart, correlation coefficient
A.3.5 Functional adjustment

Part B: Probability

Chapter 1: Combinatorial Analysis *(1 week)*

B.1.1 Arrangements
B.1.2 Combinations
B.1.3 Permutations

Chapter 2: Introduction to Probability *(2 weeks)*

B.2.1 Algebra of events
B.2.2 Definitions
B.2.3 Probabilized spaces
B.2.4 General theorems of probability

Chapter 3: Conditioning and Independence *(1 week)*

B.3.1 Conditioning
B.3.2 Independence
B.3.3 Bayes' Formula

Chapter 4: Random Variables *(1 week)*

B.4.1 Definitions and properties
B.4.2 Distribution function
B.4.3 Mathematical expectation
B.4.4 Covariance and moments

Chapter 5: Usual Discrete and Continuous Probability Laws *(3 weeks)*

Bernoulli, binomial, Poisson, … ; Uniform, normal, exponential, …

Assessment Method

Continuous assessment: **40%** ; Final exam: **60%**

Bibliographic References

1. D. Dacunha-Castelle and M. Duflo. *Probability and Statistics: Problems with Fixed Time.* Masson, 1982.
2. J.-F. Delmas. *Introduction to the Calculation of Probability and Statistics.*, ENSTA, 2008.
3. W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 1.* Wiley & Sons, Inc., 3rd edition, 1968.
4. G. Grimmett, D. Stirzaker. *Probability and Random Processes*, Oxford University Press, 2nd edition, 1992.
5. J. Jacod and P. Protter. *Probability Essentials*, Springer, 2000.
6. A. Montfort. *Course on Mathematical Statistics.* Economica, 1988.
7. A. Montfort. *Introduction to Statistics.* École Polytechnique, 1991.

Semester: 3

UEM 2.1
Subject 1: Probability & Statistics (VHS: 45h00, Lecture: 1h30, Tutorial: 1h30)

Table of Contents

Evaluation Method

Continuous assessment: 40%
Final exam: 60%

Objectives of the Course

This module allows students to study the essential concepts of probability and statistics, namely: statistical series with one and two variables, probability on a finite universe, and random variables.

Recommended prior knowledge:
Foundations of programming acquired in Math 1 and Math 2.

PART A: STATISTICS

Introduction

the science of learning from data. In an increasingly data-driven world, the ability to collect, analyze, interpret, and present data is an indispensable skill for engineers, scientists, and researchers. This discipline provides the fundamental tools for transforming raw, often overwhelming, information into meaningful insights that drive decision-making, innovation, and discovery.

This course begins with **Descriptive Statistics**, where we will learn the art and science of summarizing and presenting data. We start with **Basic Definitions**, establishing the vocabulary and concepts that form the language of statistics—populations, samples, variables, and data types. Understanding the nature of your data is the critical first step in any analysis.

We then delve into **Univariate Statistical Series**, focusing on methods to describe and understand a single variable. You will learn to use graphical representations like histograms and box plots to visualize data, and numerical summaries like mean, median, variance, and standard deviation to quantify its central tendency and variability. These tools allow us to tell the story of a dataset with clarity and precision.

The course progresses to **Bivariate Statistical Series**, where we explore the relationships between two variables. Through scatter plots, correlation coefficients, and regression lines, we will investigate how variables influence one another. This forms the basis for predictive modeling and understanding complex, interconnected systems, which is at the heart of engineering and scientific analysis.

Statistics is not merely a set of calculations; it is a framework for critical thinking. It equips you to handle the variability inherent in real-world measurements, to design effective experiments, and to make informed conclusions in the face of uncertainty. The skills you develop here will serve as a foundation for more advanced topics in data analysis, machine learning, and quality control, making you a more effective and analytical problem-solver. Let's embark on this journey to uncover the stories hidden within data.

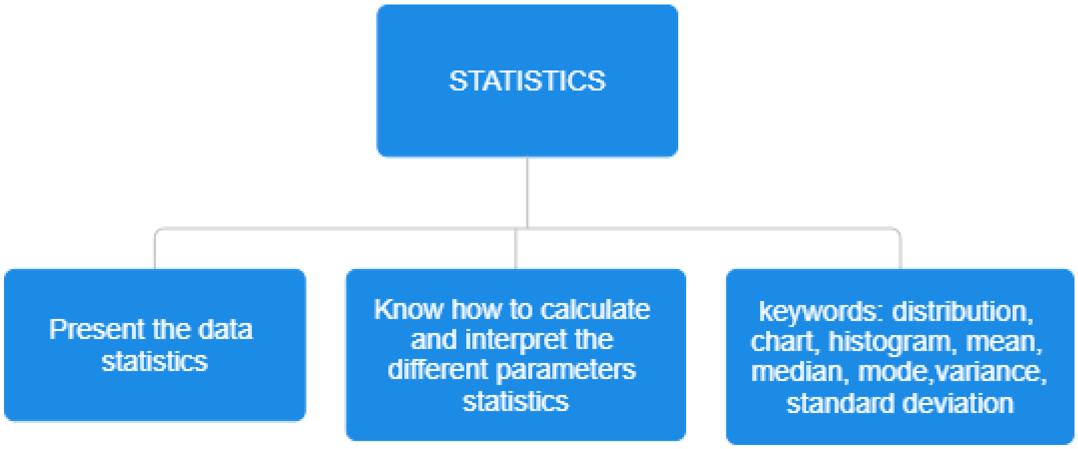

Figure 1: statistical processing

The place of this course in the students' future profession:

- Data analysis (scientific tools to summarize a set of data to highlight information).
- Simulations (stochastic process - time variable).
- Prediction and decisions (risk or occurrence probabilities).

CHAPTER I: BASIC DEFINITIONS

A.1.1 Generalities on Statistics

Definition: Statistics is the study of data collection, their analysis, their processing, the **interpretation of results, and their presentation to make the data understandable by everyone. It is simultaneously a science, a method, and a set of techniques.**

Data analysis:

Data analysis is used to describe the phenomena studied, make predictions, and make decisions about them. In this, statistics is an essential tool for understanding and managing complex phenomena.

The data studied can be of any nature, which makes **statistics useful in all disciplinary fields and explains why it is taught in all university programs,** from economics to biology, through psychology and, of course, engineering sciences. Statistics consists of:

- Collecting data.
- Presenting and summarizing this data.
- Drawing conclusions about the studied population and aiding decision-making.
- In the presence of time-dependent data, we try to make forecasts.

Vocabulary:

Statistics consist of various data classification methods such as tables, histograms, and graphs, allowing the organization of a large amount of data. Statistics developed in the second half of the 19th century in the field of human sciences (sociology, economics, anthropology, ...). They have developed a specific vocabulary.

a) Population:

In statistics, we work on populations. This term comes from the fact that demography, the study of human populations, held a central place in the beginnings of statistics, particularly through population censuses. But, in statistics, the term population applies to any studied statistical object, whether it is students (of a university or a country), households, or any other set on which statistical observations are made. We define the notion of population.

A **population** is the set on which our statistical study is based. This set is denoted by **P**.

Example

– Consider the set of students in section A. We are interested in the number of siblings of each student. In this case

P = set of students.

– If we are now interested in automobile traffic in a city, the population is then constituted by the set of vehicles likely to circulate in that city on a given date. In this case

– The set of cars in Algeria.

P = set of vehicles.

P = set of cars

b) Individual (statistical unit):

A population is composed of individuals. The individuals that make up a statistical population are called statistical units.

An **individual** is any element of the population **P**.

Remark: The set can be a set of people, things, or animals... The statistical unit is an object for which we are interested in collecting information.

Example:

1. In the example above, an individual is any student in the section.
2. If we study the annual production of a metal drink can factory. The population is the set of cans produced during the year, and a can constitutes an individual.
3. The set of countries in the world. An individual is a country.

c) Sample:

This is a subset of a population. These samples are, in principle, chosen randomly from the population. Observations will be made on the sample to extrapolate the results to the entire population.

d) Character (Statistical Variable):

"Descriptive" statistics, as the name implies, seeks to describe a given population. We are interested in the characteristics of the units that can take different values.

Height, temperature, nationality, sex, eye color, socio-professional category, weight, number of children, their age.

e) Modalities:

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

- Variable is "marital status "Modalities are "single, married, divorced"
- Variable is "sex" Modalities are "male, female"

• Variable is "switch status" Modalities are "0 and 1".

• Variable is "socio-professional categories "Modalities are "Employees, workers, retirees, ..."

• Variable is "color" Modalities are "black, green, blue, red, ..."

Remark:

Modalities are the different situations in which individuals can be with respect to the considered characteristic.

A.1.2 Types of Characteristics (Statistical Variable):

We distinguish two categories of characteristics:

Qualitative characteristics and quantitative characteristics.

.

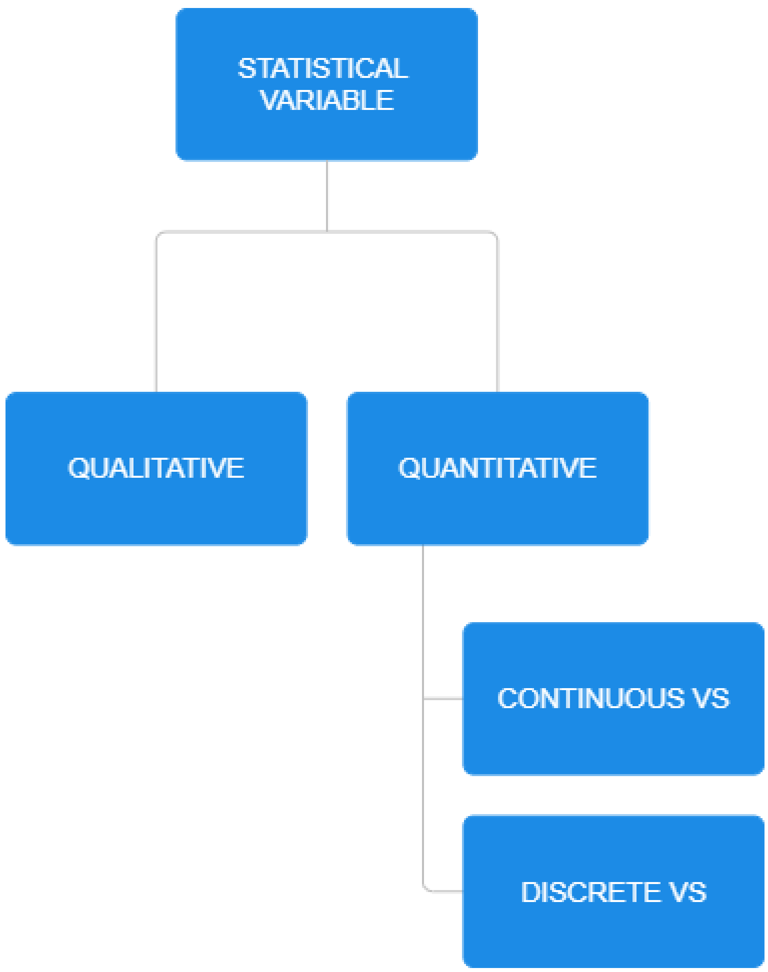

Figure 2: types of statistical variables

a) Qualitative Statistical Variable:

Qualitative characteristics are those whose modalities cannot be ordered, meaning that if we consider two randomly taken characteristics, we cannot say that one is less than or equal to the other. More precisely, we have the following definition.

Definition
A qualitative variable is a variable that does not take a numerical value, i.e., it is not measurable.

i) Nominal Qualitative Statistical Variable

Definition
The variable is called nominal qualitative when the modalities cannot be ordered.

ii) Ordinal Qualitative Statistical Variable

Definition
The variable is called ordinal qualitative when the modalities can be ordered.

- The condition of a house: we can consider the following modalities: Old, dilapidated, new.
- The condition of a person: Sex, color, nationality, activity sector, illness……

b) Quantitative Statistical Variable:

Quantitative characteristics are characteristics whose modalities can be ordered. Thus, age, life expectancy, or an individual's salary are quantitative characteristics. Therefore, we have the following definition.

Definition:

When the modalities of a statistical variable are measurable quantities or the set of values is represented by numbers.

Example:

Height, temperature, age, number of children, annual salary……

i) Continuous Quantitative Statistical Variable:

The modalities of the statistical variable can take all values within a given interval, notably an infinite number of values.

Example: Height

ii) Discrete Quantitative Statistical Variable:

The possible values of the variable are isolated.

Example: Number of children

Remark:

In general, a discrete quantitative variable is a variable taking only integer values (more rarely decimal). The number of distinct values of such a variable is usually quite low. For example, the number of houses per neighborhood in a city. A quantitative variable is said to be continuous when the observations associated with it are not precise values, but intervals. This is the case when we have a large number of distinct observations.

Descriptive statistics aims to synthesize the information contained in datasets through tables, figures, or numerical summaries. Statistical variables are analyzed differently according to their nature (quantitative, qualitative).

CHAPTER II: UNIVARIATE STATISTICAL SERIES

Presentation of Statistical Variable Series

A.2.1 Statistical Series:

A statistical series is the sequence of observations of one (or more) variable(s), recorded on the individuals of a population.

Modalities	X_1	X_2	X_3	………	X_R
Frequency	n_1	n_2	n_3	………	N_r

Table 1: Statistical Series

a) Partial Frequency - Cumulative Frequency:

- The number of individuals in the population is called the **total frequency**, denoted by **N**.
- The **absolute frequency** is the number of times this value of the characteristic was observed, denoted by **"n"**.

We have

$$n_1 + n_2 + \cdots + n_r = N$$

b) Partial Frequency (Relative Frequency):

For each value X_i, we define by definition:
f_i: is called the **partial frequency** of X_i.

$$f_i = n_i/N$$

The frequency of a value is the ratio of the frequency of that value to the total frequency.

$$f_1 + f_2 + f_3 + \cdots + f_r = \frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} + \cdots + \frac{n_r}{N}$$

c) Increasing and Decreasing Cumulative Frequencies:

The **increasing cumulative frequencies** (resp. **increasing cumulative relative frequencies**) of a value are obtained by adding to each frequency (resp. relative frequency) the frequencies (resp. relative frequencies) of the values that precede it.

A.2.2 Graphical Representation of Statistical Series:

The methods of representing a statistical variable are distinguished based on the nature of this variable (qualitative or quantitative). The recommended and most frequent representations are tables and diagrams (graphs).

a) Representations of Statistical Series of Qualitative Variables:

Based on the observation of a qualitative variable, two diagrams allow representing this variable: the bar chart (called pipe organ) and the sector diagram (called pie chart).

i) Pipe Organ Diagram:

We place the modalities on the x-axis, arbitrarily. We place rectangles on the y-axis whose length is proportional to the frequencies, or relative frequencies, of each modality.

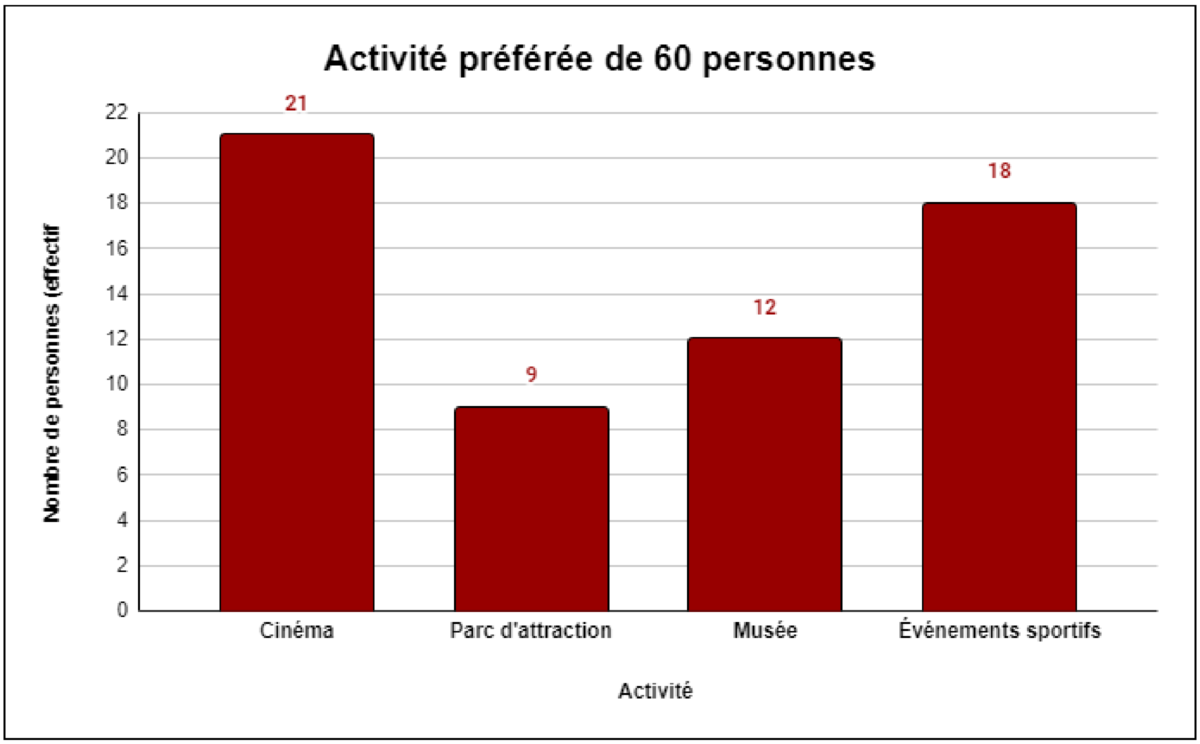

Figure 3: *Pipe Organ Diagram*

ii) Sector Diagram (Pie Chart):

Pie charts, or semi-circular charts, consist of dividing a disk or a semi-disk into slices, or sectors, corresponding to the observed modalities and whose area is proportional to the frequency, or relative frequency, of the modality.

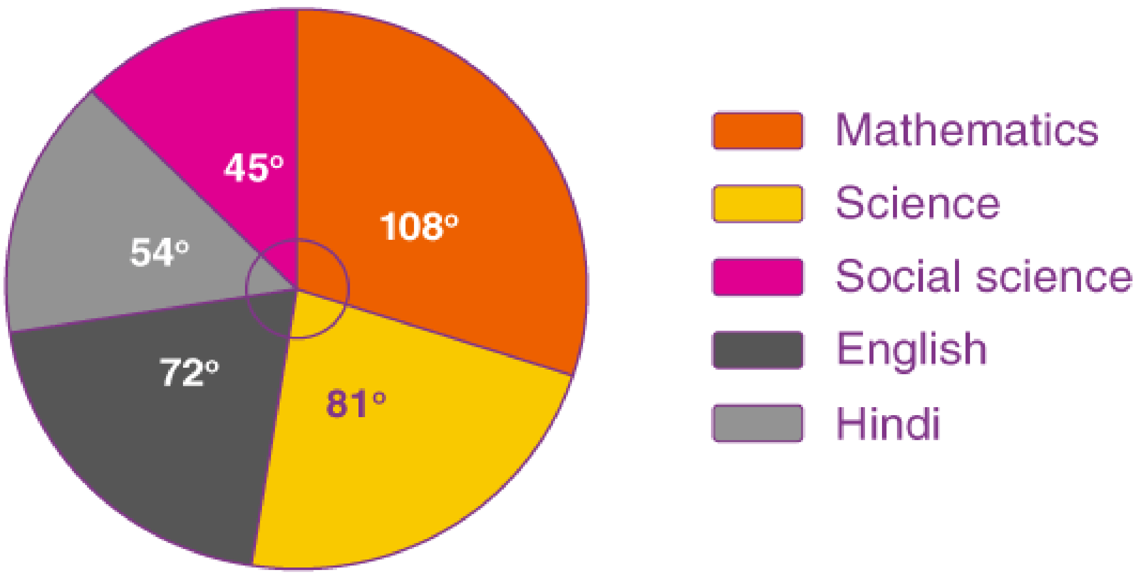

Figure 4: Sector Diagram

The degree of a sector is determined using the rule of three as follows:

N -- 360°

n_i -- d_i (degree of modality)

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

b) Representations of Statistical Series of Discrete Quantitative Variables:

Based on the observation of a discrete quantitative variable, the diagram that allows representing this variable: the bar chart.

i) Bar Chart:

We want to represent this distribution in the form of a bar chart. Each mark corresponds to a bar. The heights of the bars are proportional to the represented frequencies.

► With the example of grades:

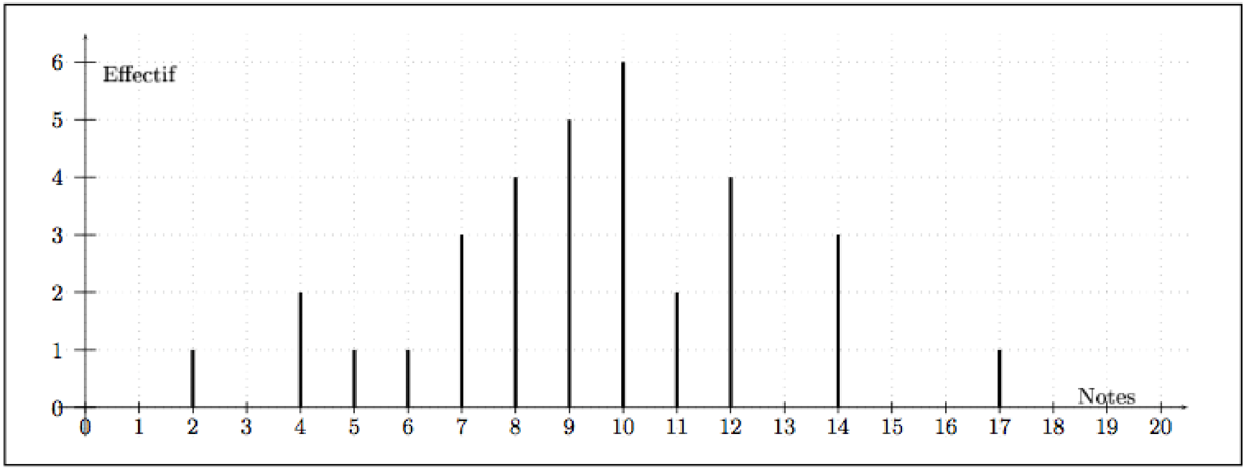

Figure 5: Bar Chart

c) Representations of Statistical Series of Continuous Quantitative Variables

Based on the observation of a continuous quantitative variable, the diagram that allows representing this variable: Frequency Histogram.

i) Frequency (or Frequency) Histogram:

We can represent the statistical table by a histogram. We report the classes on the x-axis and, above each of them, we draw a rectangle whose area is proportional to the frequency fi (or the frequency ni) associated. This graph is called the frequency histogram.

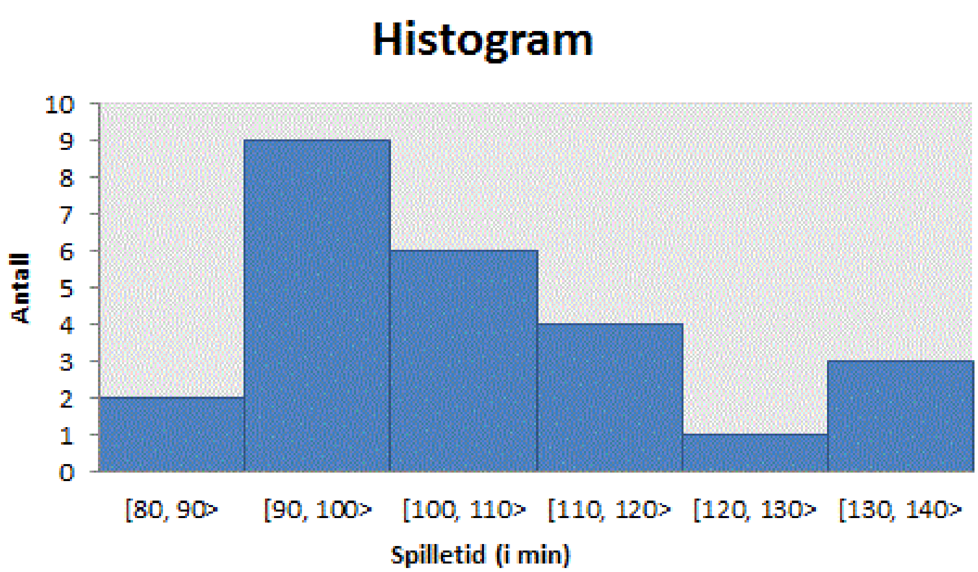

Figure 6: *Frequency Histogram*

A.2.3 Location Parameters (Central Tendency):

a) The Mode (Discrete Case):

The mode of a discrete statistical variable is the value that has the greatest partial frequency (or the greatest partial relative frequency) and is denoted by M_0

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

***Table 2**: Mode Discrete Case*

The mode M_0=2

b) The Mode (Continuous Case):

The following definition allows understanding the approach to follow to calculate the mode accurately, which is found in one of the classes called the "modal class."

Definition:

We define the **modal class** as being the class of values of X that has the greatest partial frequency (or the greatest partial relative frequency).

M_0= (center of the modal class): The modal class being the class that contains the greatest frequency.

Example:

Classes	[20;22]	[22;25]	[25;31]	[31;36]	[36;41]
Frequency (n_i)	8	38	3	1	3

***Table 3**: Mode Continuous Case*

M_0 = 23.5: The modal class being the class [22;25] that contains the greatest frequency 38. (The center of the class is (22+25)/2 = 23.5)

c) The Median:

The **median** is the value that divides the series into two series with the same total frequency.

Depending on whether N is an even or odd total number.

1. Case: N = 2p+1 (odd) therefore the median is the value of order p+1.

2. Case: N = 2p (even) therefore the median is the value of order p and p+1 divided by 2 (the average of the values at positions p and p+1).

d) The Mean:

The mean is calculable for numerical variables, whether discrete or continuous. It is obtained simply by adding all the values and dividing this sum by the number of values. This calculation can be done from raw data or from a frequency table. Here are some calculation examples.

Discrete Variable:

$$\bar{X} = \frac{\sum_{i=1}^{k} n_i x_i}{N}$$

Continuous Variable:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{N}$$

with f_i frequency of the characteristic xi (or frequency of the i-th class), m_i midpoint of the i-th class and n the number of individuals in the sample (N is the total frequency).

A.2.4 Dispersion Parameters (Variability):

The usual statistical indicators of dispersion are the range, variance, standard deviation, and the dispersion parameters associated with the median.

a) The Range

The difference between the largest value and the smallest value of the characteristic, given by the quantity

$$E = X_{\max} - X_{\min}$$

Is called the **range** of the statistical variable. The calculation of the range is very simple. It gives a first idea of the dispersion of the observations. It is a very rudimentary indicator.

b) The Variance (S^2 or Var(X)):

We call **variance** of this statistical series X, the number

Discrete Variable:

$$S^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{N}$$

Continuous Variable:

$$S^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \bar{x})^2}{N - 1}$$

(Note: The formula for the continuous variable uses class midpoints m_i. The mention $N-1$ in the denominator is unusual for descriptive variance; it's typically for sample variance).

Properties of Variance:

1. $Var(X + a) = Var(X)$
2. $Var(a \times X) = a^2 \times Var(X)$

c) The Standard Deviation

The **standard deviation** is useful when comparing the dispersion of two datasets of similar size that have approximately the same mean. The spread of values around the mean is less important in the case of a dataset with a smaller standard deviation.

The quantity: $\sigma(X) = \sqrt{Var(X)} = \sqrt{S^2}$

d) The Dispersion Parameters Associated with the Median.

Definition:

The general idea is to divide the population into four parts of equal frequency.

Given a statistical series with median M whose list of values is arranged in ascending order (it is the same list used to determine the median).

By cutting the list into two sub-series of equal frequency (Attention: when the total frequency is odd, the median should not be included in the sub-series):

- We call **first quartile** the real number denoted **Q1** equal to the median of the lower sub-series.
- We call **third quartile** the real number denoted **Q3** equal to the median of the upper sub-series.
- The **interquartile range** is equal to **Q3 - Q1**.
- **]Q1;Q3[** is called the **interquartile interval**.

The **box plot** (box-and-whisker plot) of a statistical series is then constructed as follows: (the values of the characteristic are on the abscissa - min and max represent the minimum and maximum values of the characteristic) quantile of order α is:

$$q_\alpha = L_i + \frac{(\alpha N - F_{i-1})}{f_i} \times C$$

with L_i lower boundary of the class containing the quantile of order α (if α=0.5 then q_α is the median M_e)

α=0.25 is the quantile 1 (Q1),

α=0.75 is the quantile 3 (Q3),

α=0.1 is the decile 1 (D1), *etc.*…

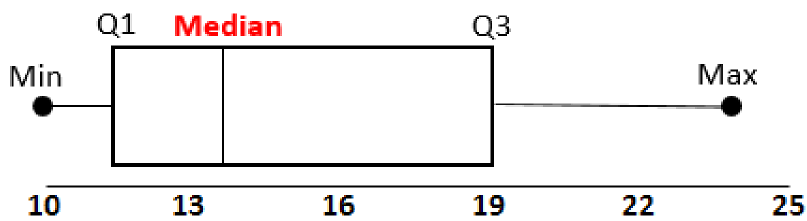

Figure 7: box plot

Interpretation:

- 25% of the population has a value of the characteristic between min and Q1
- 25% of the population has a value of the characteristic between Q1 and M (the median)
- 25% of the population has a value of the characteristic between M and Q3
- 25% of the population has a value of the characteristic between Q3 and max

A.2.5 Stem-and-Leaf Displays

Consider a numerical data set *x*1, *x*2, . . . , *xn* for which each *xi* consists of at least two digits. A quick way to obtain an informative visual representation of the data set is to construct a *stem-and-leaf display.*

Steps for Constructing a Stem-and-Leaf Display

a. Select one or more leading digits for the stem values. The trailing digits become the leaves.

b. List possible stem values in a vertical column.

c. Record the leaf for every observation beside the corresponding stem value.

d. Indicate the units for stems and leaves someplace in the display.

If the data set consists of exam scores, each between 0 and 100, the score of 83 would have a stem of 8 and a leaf of 3. For a data set of automobile fuel efficiencies (mpg), all between 8.1 and 47.8, we could use the tens digit as the stem, so 32.6 would then have a leaf of 2.6. In general, a display based on between 5 and 20 stems is recommended.

The use of alcohol by college students is of great concern not only to those in the academic

community but also, because of potential health and safety consequences, to society at large. The article "Health and Behavioral Consequences of Binge Drinking in College" (*J. of the Amer. Med. Assoc.,* 1994: 1672–1677) reported on a comprehensive study of heavy drinking on campuses across the United States. A binge episode was defined as five or more drinks in a row for males and four or more for females. Table 4 shows a stem-and-leaf display of 140 values of x _ the percentage of undergraduate students who are binge drinkers. (These values were not given in the cited article, but our display agrees with a picture of the data that did appear.)

0 |	4	
1 |	1345678889	
2 |	12234566667778889999	Stem: tens digit
3 |	011223334455566667777888899999	Leaf: ones digit
4 |	111222223344445566666677788888999	
5 |	00111222223345566666777888899	
6 |	01111244455666778	

Table 4: Stem-and-leaf display

Table 4 Stem-and-leaf display for percentage binge drinkers at each of 140 colleges

The first leaf on the stem 2 row is 1, which tells us that 21% of the students at one of the colleges in the sample were binge drinkers. Without the identification of stem digits and leaf digits on the display, we wouldn't know whether the stem 2, leaf 1 observation should be read as 21%, 2.1%, or .21%.

When creating a display by hand, ordering the leaves from smallest to largest on each line can be time-consuming. This ordering usually contributes little if any extra information. Suppose the observations had been listed in alphabetical order by school name, as 16%, 33%, 64%, 37%, 31% . . .

Then placing these values on the display in this order would result in the stem 1 row having 6 as its first leaf, and the beginning of the stem 3 row would be 3 | 371 . . .

The display suggests that a typical or representative value is in the stem 4 row, perhaps in the mid-40% range. The observations are not highly concentrated about this typical value, as would be the case if all values were between 20% and 49%.

The display rises to a single peak as we move downward, and then declines; there are no gaps in the display. The shape of the display is not perfectly symmetric, but instead appears to stretch out a bit more in the direction of low leaves than in the direction of high leaves. Lastly, there are no observations that are unusually far from the bulk of the data (no *outliers*), as would be the case if one of the 26% values had instead been 86%. The most surprising feature of this data is that, at

most colleges in the sample, at least one-quarter of the students are binge drinkers.

The problem of heavy drinking on campuses is much more pervasive than many had suspected.

A stem-and-leaf display conveys information about the following aspects of

the data:

• identification of a typical or representative value

• extent of spread about the typical value

• presence of any gaps in the data

• extent of symmetry in the distribution of values

• number and location of peaks

• presence of any outlying values

Example 2 : Stem-and-Leaf Plot Statistical data

Statistical data, generated in large masses, can be very useful for studying the behavior of the distribution if presented in a combined tabular and graphic display called a stem-and-leaf plot. To illustrate the construction of a stem-and-leaf plot, consider the data of Table 1.4, which specifies the "life" of 40 similar car batteries recorded to the nearest tenth of a year. The batteries are guaranteed to last 3 years. First, split each observation into two parts consisting of a stem and a leaf such that the stem represents the digit preceding the decimal and the leaf corresponds to the decimal part of the number. In other words, for the number 3.7, the digit 3 is designated the stem and the digit 7 is the leaf. The four stems 1, 2, 3, and 4 for our data are listed vertically on the left side in Table 1.5; the leaves are recorded on the right side opposite the appropriate stem value. Thus, the leaf 6 of the number 1.6 is recorded opposite the stem 1; the leaf 5 of the number 2.5 is recorded opposite the stem 2; and so forth. The number of leaves recorded opposite each stem is summarized under the frequency column.

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Table 5: Car battery life

Stem	Leaf	Frequency
1	69	2
2	25669	5
3	0011112223334445567778899	25
4	11234577	8

Table 6: Stem-and-Leaf Plot of battery life

Class Interval	Class Midpoint	Frequency, *f*	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Table 7: relative frequency distribution of battery life

CHAPTER III: BIVARIATE STATISTICAL SERIES

The objective is to analyze the distribution of the values of variables x and y and the potential relationship between them.

A.3.1 Scatter Plot:

A graph that represents the two statistical series using a 2-dimensional diagram. Let x and y be two numerical statistical variables observed on k individuals. In an orthogonal coordinate system $\left(o, \vec{i}, \vec{j}\right)$, the set of k points with coordinates $(x_i; y_i)$ forms the scatter plot associated with this statistical series.

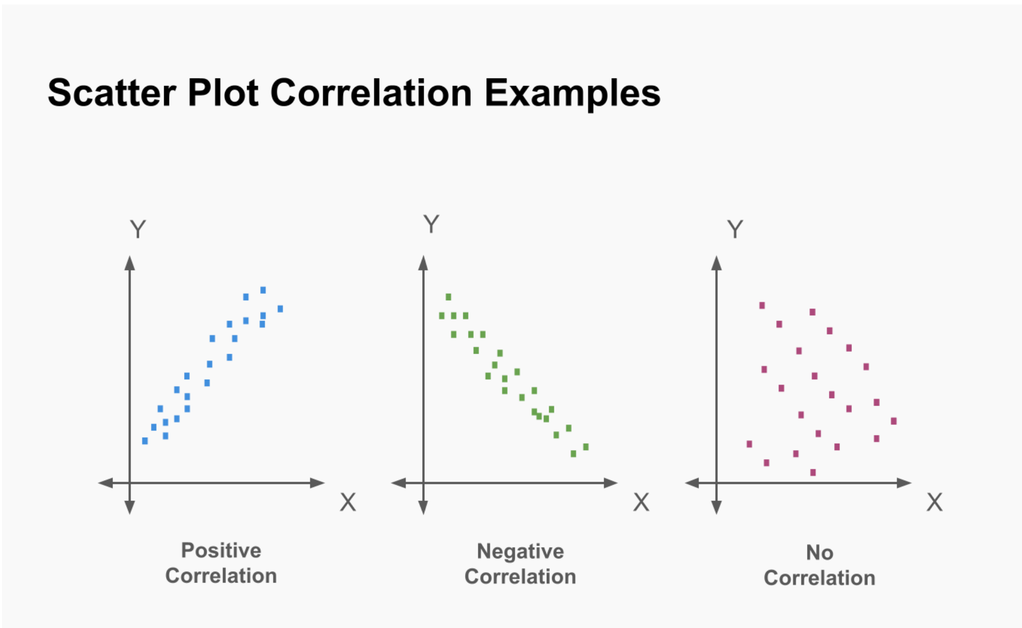

Figure 8: Scatter Plot

A.3.2 Data Tables (Contingency Table)

- **Contingency table of frequencies:**
- **Discrete Variables:** Pairs of values = (x_1, y_1); (x_2, y_2); (x_3, y_3); ... ; (x_n, y_n)

Each local frequency n_{ij} corresponds to the number of individuals having the abscissa x_i and the ordinate y_j.

Y \ X	y_1	y_2	...	y_j	...	y_q
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}
.	.	.	...	.	...	.
.	.	.	...	.	...	.
.	.	.	...	.	...	.
x_i	n_{i1}	n_{i2}	...	n_ij	...	n_iq
.	.	.	...	.	...	.
.	.	.	...	.	...	.
x_p	n_{p1}	n_{p2}	...	n_pj	...	n_pq

Table 8: Contingency Table for "Discrete Variables"

- **Continuous Variables:** These are the central modalities $(x_1, x_2, x_3, \ldots, x_p)$ and $(y_1, y_2, y_3, \ldots, y_p)$ of the classes that replace the discrete modalities.

Each local frequency nij corresponds to the number of individuals whose x values belong to the class [a_{i-1}, a_i] and whose y values belong to the class [b_{j-1}, b_j].

Y \ X	$[b_0,b_1]$ Y_1	$[b_1,b_2]$ Y_2	…	[b_{j-1},b_j] Y_j	…	[b_{q-1},b_q] Y_q
$[a_0,a_1]$	x_1	n_{11}	n_{12}	…	n_{1j}	…
$[a_1,a_2]$	x_2	n_{21}	n_{22}	…	n_{2j}	…
. . .	. . .	. . .	. . .	…	. . .	…
[a_{i-1},a_i]	x_i	n_{i1}	n_{i2}	…	n_ij	…
. . .	. . .	. . .	. . .	…	. . .	…
[a_{p-1},a_p]	x_p	n_{p1}	n_{p2}	…	n_pj	…

Table 9: Contingency Table for "Continuous Variables"

- **Contingency table of relative frequencies:** We keep the previous tables and divide all the frequencies (local, marginal) by the total frequency n.

A.3.3 Marginal and Conditional Distributions, Covariance:

a) Marginal Distributions:

We add the row and column totals to the contingency table.

Y \ X	y_1	y_2	…	y_j	…	y_q	**Marginal Distribution of Y** $n_{i\bullet}$
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2q}	$n_{2\bullet}$
.	.	.	…	.	…	.	.
.	.	.	…	.	…	.	.
.	.	.	…	.	…	.	.
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iq}	$n_{i\bullet}$
.	.	.	…	.	…	.	.
.	.	.	…	.	…	.	.
.	.	.	…	.	…	.	.
x_p	n_{p1}	n_{p2}	…	n_{pj}	…	n_{pq}	$n_{p\bullet}$
---	---	---	---	---	---	---	---
Marginal Distribution of X $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet q}$	$n_{\bullet\bullet} = n$

Sum of nij in the column Total Frequency

Table 10: Marginal Distributions for "Discrete Variables"

Y \ X	$[b_0,b_1]$ y_1	$[b_1,b_2]$ y_2	…	$[b_{j-1},b_j]$ y_j	…	$[b_{q-1},b_q]$ y_q	**Marginal Distribution of Y** $n_{i\bullet}$
$[a_0,a_1]$	x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1q}
$[a_1,a_2]$	x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2q}
---	---	---	---	---	---	---	---
.	.	.	.	…	.	…	.
.	.	.	.	…	.	…	.
.	.	.	.	…	.	…	.
$[a_{i-1},a_i]$	x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iq}
---	---	---	---	---	---	---	---
.	.	.	.	…	.	…	.
.	.	.	.	…	.	…	.
.	.	.	.	…	.	…	.
$[a_{p-1},a_p]$	x_p	n_{p1}	n_{p2}	…	n_{pj}	…	n_{pq}
---	---	---	---	---	---	---	---
Marginal Distribution of X $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet q}$	$n_{\bullet\bullet} = n$

Sum of nij in the column Total Frequency

Table 11: Marginal Distributions for "Continuous Variables"

- **In the right margin (row totals):**
  the distribution of X: for each index i, the frequency n_{i•} is the total number of observations of the modality x_i of X, regardless of the modality of Y. That is $n_{i*} = \sum_{j=1}^{q} n_{ij}$ = Total of row i.

The p pairs (x_i, n_{i•}) define the marginal distribution of the variable X.

- **In the bottom margin (column totals):**
  the distribution of Y: for each index j, the frequency n_{•j} is the total number of observations of the modality y_j of Y, regardless of the modality of X. That is $n_{*j} = \sum_{i=1}^{p} n_{ij}$ = Total of column j.

The q pairs (y_j, n_{•j}) define the marginal distribution of the variable Y.

Remark:

$$\sum_{i=1}^{p} n_{i*} = \sum_{j=1}^{q} n_{*j} = n$$

b) Conditional Distributions:

- The distribution of observations according to the modalities of the variable Y, given that the variable X takes the modality x_i, is called the conditional distribution of Y for X = x_i. In row i of the contingency table, we read the distribution of the variable Y given that X = x_i, denoted Y|X=x_i.

- The distribution of observations according to the modalities of the variable X, given that the variable Y takes the modality y_j, is called the conditional distribution of X for Y = y_j. In column j of the contingency table, we read the distribution of the variable X given that Y = y_j, denoted X|Y=y_j.

c) Covariance:

The covariance of the double statistical series of variables x and y is the real number defined by:

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})\,(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - \overline{xy}$$

Remark:

$Cov(x, x) = V(x)$ (Variance of x)

A.3.4 Linear Correlation Coefficient:

The linear correlation coefficient is a number used to determine the strength of a linear relationship between two quantitative variables. The linear correlation coefficient of a statistical series of variables x and y is the number r defined by: $\rho = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

Remarks:

- The correlation coefficient is a unitless value that is always between -1 and +1.
- A positive linear correlation coefficient indicates a positive linear relationship, whereas if ρ is negative, the linear relationship between the two variables is negative.
- The closer the value of ρ is to -1 or +1, the stronger the linear relationship between the two variables.

A.3.5 Regression Line and Mayer's Line

a) Regression Line

A regression line is the line that best fits a scatter plot showing a linear correlation. The regression line is used for making predictions. We speak of a linear correlation when the points in a scatter plot tend to align. The stronger the trend, the stronger the linear correlation. The line D with equation y = ax + b is called the regression line of y on x for the statistical series if the following quantity is minimized: $S = \sum_{i=1}^{n}[\,y_i - (ax_i + b)]^2$ To define the coefficients a and b; we expand S and consider it successively as a trinomial in b; then, with b determined, as a trinomial in a. We find:

$$a = \frac{cov(x,y)}{\sigma_x^2} \quad b = \overline{y} - a\overline{x}$$

b) Mayer's Line

This fitting method consists of determining the line passing through two mean points of the scatter plot.

A.3.6 Regression Curves, Regression Strip, and Correlation Ratio

a) Regression Curves

A regression curve allows for the analysis of the relationship between two variables (explanatory variable and explained variable) and highlights the nature of this relationship without making any prior assumption about its form. It matches the values of the first variable with the conditional means of the second. Therefore, from two variables X and Y, we can construct two regression curves:

- The curve of X on Y: it relates the values of Y (y_i) and the conditional means of X.
- The curve of Y on X: it relates the values of X (x_i) and the conditional means of Y.

Apart from giving an image, a representation of the form of the link existing between two variables, regression curves have important properties:

They best summarize the shape of the scatter plot since they are on average the closest to the points of this cloud (the sum and the means of the squares of the distances between the points of the cloud and the regression curves are minimal).

They intersect at a point which represents the center of gravity of the scatter plot (i.e., a point with approximate coordinates equal to the marginal means of the two variables).

b) Correlation Ratio

To study the relationship between a qualitative variable and a quantitative variable, the total variation is decomposed into inter-group (or inter-class) variation and intra-group (or intra-class) variation. To measure the intensity of the relationship, one can calculate a parameter called the correlation ratio.

A.3.7 Functional Fitting

When we want to model the data, the next question is to find the equation of the regression curves. Regression curves generally correspond to complicated functions, but we can try to approximate them with simpler functions. The general principle is to start with a known function form and seek the parameters that best fit the obtained curves to the regression curves. For example, if we start from the idea that the regression curve, aside from measurement errors, would be a straight line, then it would be characterized by an equation of the type $y = ax + b$

We would then need to identify the parameters a and b to know the equation of the line. If we imagine that the regression curve corresponds to a parabola, the sought equation would be of the form: $y = ax^2 + bx + c$ And in this case, we would need to find the values of the three parameters, a, b, and c.

a) Power Fitting

Power fitting is based on the curve represented by the equation of the type $y = ax^b$ We note that $\ln y = b\ln x + \ln a$ Let $V = \ln y$ and $U = \ln x$, we determine the equation of the regression line of v on u using Mayer's method or the least squares method. The obtained equation is of the form $v = Au + B$, from which we deduce the equation of the power function curve: $y = ax^b$ since $A = b$ and $B = \ln a$.

b) Exponential Fitting

Exponential fitting has a curve of the exponential function with the equation: $y = ab^x$ We note that $\ln y = x\ln b + \ln a$ Let $z = \ln y$, we determine the equation of the regression line of z on x using Mayer's method or the least squares method. The obtained equation is of the form $z = Ax + B$. From this, we deduce the equation of the exponential function curve: $y = ab^x$ Since $A = \ln b$ and $B = \ln a$

PART B: PROBABILITY THEORY

Introduction

This field forms the essential backbone of data science, machine learning, scientific research, and informed decision-making in a world filled with uncertainty. The ability to model, quantify, and analyze randomness is a critical skill across countless disciplines, from finance and engineering to medicine and artificial intelligence.

This course is structured to provide a deep and intuitive understanding of both the theoretical foundations and practical applications of probability theory. We begin by building the mathematical machinery needed to count and arrange objects systematically in **Combinatorial Analysis**. This foundation is crucial for calculating probabilities in complex scenarios where simple enumeration is impossible.

With these tools in hand, we delve into the core of **Probability Foundations**, where we formalize the concepts of experiments, outcomes, and events. We will explore the axioms that govern probability and the powerful theorems that derive from them, establishing a rigorous framework for reasoning about uncertainty.

The journey continues with **Conditional Probability and Independence**, concepts that are fundamental to understanding relationships between events. Here, we introduce Bayes' Theorem, a profound result that allows us to update our beliefs in light of new evidence, with far-reaching applications in diagnostics, machine learning, and legal reasoning.

We then bridge the gap between abstract theory and quantifiable measures by introducing **Random Variables**. This powerful concept allows us to translate probabilistic outcomes into numerical values, enabling us to compute averages, variances, and other descriptive characteristics. We will thoroughly investigate both **Discrete and Continuous Probability Distributions**, each modeling different types of real-world phenomena, from the number of system failures to the timing of events and measurement errors.

By the end of this course, you will have developed a strong conceptual and practical grasp of probability theory. You will be equipped to model uncertain processes, make probabilistic predictions, and lay the groundwork for the subsequent study of statistical inference, where we learn about populations from data samples. Let's begin this journey into the mathematics of chance.

CHAPTER I: COMBINATORIAL ANALYSIS

B.1.1 Introduction to Combinatorial Analysis

Definition:

Combinatorial analysis is the branch of mathematics concerned with counting, arranging, and selecting objects according to specified rules.

Historical Context:

Originally developed to solve gambling problems in the 17th century, it now forms the foundation of probability theory, computer science, and operations research.

Key Applications:

- Calculating probabilities in games of chance
- Designing computer algorithms
- Optimizing network routes
- Cryptography and security
- Genetic sequencing

Detailed Explanation:

Combinatorics provides systematic methods for counting without exhaustive enumeration. When faced with problems like "how many possible outcomes" or "in how many ways," combinatorial techniques offer efficient solutions rather than listing all possibilities.

Extended Example:

- **Simple Counting:** A pizza restaurant offers 3 crust types, 2 sauce types, and 5 toppings. How many different pizzas can be made? Using basic multiplication: $3 \times 2 \times 5 = 30$ different pizzas.

- **Complex Counting:** How many different protein sequences of length 10 can be made from 20 amino acids? Using arrangements with repetition: $20^{10} = 1.024 \times 10^{13}$ possibilities.

B.1.2 Fundamental Principle of Multiplication

Theorem Statement:

If a process consists of k successive steps, and step 1 can be performed in n_1 ways, step 2 in n_2 ways (regardless of how step 1 was performed), step 3 in n_3 ways (regardless of previous steps), and so on, then the total number of ways to complete the entire process is:

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

Detailed Explanation:

This principle applies when choices are independent and sequential. Each choice point multiplies the possibilities. It's foundational for understanding more complex counting problems.

Detailed Example: Password Creation

- A password must be 4 characters long
- First character: uppercase letter (26 possibilities)
- Second character: digit (10 possibilities)
- Third character: lowercase letter (26 possibilities)
- Fourth character: special symbol (8 possibilities)

$$\text{Total passwords} = 26 \times 10 \times 26 \times 8 = 54{,}080$$

Detailed Example: Travel Routes To travel from City A to City D:

- A → B: 3 roads
- B → C: 2 roads
- C → D: 4 roads

$$\text{Total routes} = 3 \times 2 \times 4 = 24$$

Additional Example: Computer Configuration

- Processor: 4 options
- RAM: 3 options
- Storage: 5 options
- Graphics card: 2 options
- Total configurations: 4 × 3 × 5 × 2 = 120 different computers

B.1.3 Factorial Notation

Definition:

Factorials grow extremely fast (super-exponential growth). They represent the number of ways to arrange n distinct objects in sequence.

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

Mathematical Properties:

- Recursive definition: $n! = n \times (n-1)!$ with $0! = 1$
- Gamma function extension: $n! = \Gamma(n+1)$
- Stirling's approximation: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

Special Values:

$$\begin{aligned} 0! &= 1 \\ 1! &= 1 \\ 5! &= 120 \\ 10! &= 3{,}628{,}800 \\ 52! &\approx 8.0658 \times 10^{67} \end{aligned}$$
(Number of ways to shuffle a deck of cards)

Practical Example:

- 10 books on a shelf: 10! = 3,628,800 arrangements
- 20 students in a line: $20! \approx 2.43 \times 10^{18}$ arrangements
- 52-card deck shuffles: $52! \approx 8.07 \times 10^{67}$ arrangements (more than atoms on Earth)

B.1.4 Arrangements (Permutations of Subsets)

a) Arrangements Without Repetition

Definition: Ordered selection of r distinct objects from n objects.

When order matters and items cannot be reused/repeated. Common in race rankings, password generation (without character reuse), and committee roles.

Formula Derivation:

- First position: n choices
- Second position: (n-1) choices
- ...
- r-th position: (n-r+1) choices

$$P(n,r) = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

Example:

- In a race with 8 runners, how many ways to award gold, silver, and bronze medals?

$$P(8,3) = 8 \times 7 \times 6 = 336$$

- From 15 candidates, select President, Vice-President, and Secretary: P(15,3) = 15 × 14 × 13 = 2,730 ways
- 7 songs played from 100-song playlist without repeats: P(100,7) ≈ 8.09×10^{13} ways

b) Arrangements With Repetition

Definition: Ordered selection of r objects from n objects, allowing repetition.

When order matters and items can be reused. Common in PIN codes, license plates, and DNA sequences.

Formula Derivation:

Each of the r positions has n independent choices:

$$P_{\text{rep}}(n,r) = n^r$$

Example:

- How many 4-digit PIN codes are possible?

$$10^4 = 10{,}000$$

- Binary strings of length 8: 2^8 = 256 possible strings
- 3-character codes using A-Z: 26^3 = 17,576 codes
- Dice rolls: 3 dice rolled: 6^3 = 216 possible outcomes

B.1.5 Permutations

a) Permutations Without Repetition

Definition: Arrangements of all n distinct objects.

Special case of arrangements where r = n (all objects used).

Formula:

$$P_n = n!$$

Example:

- How many ways to arrange 5 books on a shelf?

$$5! = 120$$

- 8 people around circular table: (8-1)! = 7! = 5,040 arrangements (circular permutations)
- Anagram of "LOGIC": 5! = 120 arrangements

b) Permutations With Repetition (Multiset Permutations)

Definition: Arrangements of n objects where some objects are identical.

Accounts for identical items. Reduces total arrangements since swapping identical items doesn't create new arrangements.

Formula:

If there are k types with counts $n_1, n_2, ..., n_k$ where $n_1 + n_2 + \cdots + n_k = n$:

$$P = \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}$$

Example:

- How many distinct arrangements of the word "BANANA"?

  B: 1, A: 3, N: 2

$$\frac{6!}{1! \times 3! \times 2!} = \frac{720}{1 \times 6 \times 2} = 60$$

- "MISSISSIPPI": 11 letters, M=1, I=4, S=4, P=2 → 11!/(1!4!4!2!) = 34,650 arrangements
- 10 flags: 4 red, 3 blue, 2 green, 1 yellow → 10!/(4!3!2!1!) = 12,600 arrangements

B.1.6 Combinations

a) Combinations Without Repetition

Definition: Unordered selection of r objects from n objects.

When order doesn't matter. Common in committee selection, lottery numbers, and sampling.

Formula Derivation:

Start with arrangements: $P(n,r) = \frac{n!}{(n-r)!}$ Since order doesn't matter, divide by r! to eliminate duplicate arrangements:

$$C(n,r) = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Properties of Binomial Coefficients:

- Symmetry: $\binom{n}{r} = \binom{n}{n-r}$
- Pascal's Rule: $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$
- Sum: $\sum_{r=0}^{n}\binom{n}{r} = 2^n$

Example:

- How many ways to choose 3 committee members from 10 people?

$$\binom{10}{3} = \frac{10!}{3!\,7!} = 120$$

- Poker hand from 52-card deck: C(52,5) = 2,598,960 hands
- Choose 3 toppings from 10 options: C(10,3) = 120 combinations
- Select 4 patients from 20 for clinical trial: C(20,4) = 4,845 ways

b) Combinations With Repetition

Definition:

Unordered selection of r objects from n types, allowing multiple selections of the same type.

When order doesn't matter and items can be chosen multiple times. "Stars and bars" method visualizes this.

Formula Derivation (Stars and Bars Method):

Represent selection as r stars and (n-1) bars:

$$C_{\text{rep}}(n,r) = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!}$$

Example:

- How many ways to buy 5 donuts from 3 types?

$$\binom{3+5-1}{5} = \binom{7}{5} = 21$$

- Buy 10 donuts from 5 types: C(5+10-1,10) = C(14,10) = 1,001 ways
- Distribute 20 identical candies to 4 children: C(4+20-1,20) = C(23,20) = 1,771 ways
- Non-negative integer solutions to x+y+z=8: C(3+8-1,8) = C(10,8) = 45 solutions

CHAPTER II: PROBABILITY FOUNDATIONS - DETAILED THEORY

B.2.1 Algebra of Events

A trial is a random experiment "whose outcome is uncertain. The potential results of a trial generally involve chance. The set of potential results (the possible outcomes, the eventualities) is called the sample space.

Events are subsets of the sample space. Event algebra provides rules for combining events using set operations (union, intersection, complement) with corresponding probabilistic interpretations.

B.2.2 Definitions

Probability theory provides a mathematical framework for describing chance and variability, as well as for reasoning in uncertain environments.

The purpose of probability theory is to provide a precise mathematical formalism, suitable for describing situations involving "chance", i.e., situations in which, given a certain number of conditions (the causes), several consequences are possible (the effects) without being able to know a priori which one will occur.

Probability = "mathematization of chance".

Probability theory provides mathematical models for the study of experiments whose outcome cannot be predicted with total certainty. Here are some examples:

Experiment	Observable Result
Rolling a die	An integer $k \in 1,2,3,4,5,6$
100-question binary questionnaire	A sequence ω of 100 answers $\omega \in yes, no^{100}$
Tossing a coin until the first head appears	An integer: the waiting time for the first success

Table 1: Some examples of experiments

Extended Examples:

- **Complement:** Event A = "rain tomorrow", A^c = "no rain tomorrow"
- **Union:** A = "temperature > 30°C", B = "humidity > 80%", A∪B = "either hot OR humid"
- **Intersection:** A∩B = "both hot AND humid"

B.2.3 Basic Probability Concepts

a) Random Experiment

Formal Definition: A random experiment is a process that:

- Can be repeated under identical conditions
- Has well-defined possible outcomes
- Has uncertain individual outcomes
- Exhibits statistical regularity over many repetitions

Examples:

- Tossing a fair coin
- Rolling a die
- Measuring the height of a randomly selected person
- Observing daily stock price changes
- **Manufacturing:** Checking if a produced item is defective
- **Survey:** Asking a randomly selected person their opinion
- **Biology:** Observing if a seed germinates under controlled conditions

b) Sample Space (Ω)

Definition: The set of all possible outcomes of a random experiment.

Types of Sample Spaces:

- **Finite:** Countable number of elements
    - Coin toss: $\Omega = \{H, T\}$
    - Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Countably Infinite:** Infinite but countable
    - Number of customers arriving at a store: $\Omega = \{0, 1, 2, ...\}$
- **Uncountably Infinite:** Continuous outcomes
    - Time until a radioactive atom decays: $\Omega = [0, \infty)$
- **Continuous:** Measure a person's height: $\Omega = [50\text{cm}, 250\text{cm}]$ (practical range)
- **Multidimensional:** Record temperature and pressure: $\Omega = \mathbb{R}^2$

c) Events

Definition: Any subset of the sample space.

Types of Events:

- **Elementary Event:** Single outcome
- **Compound Event:** Multiple outcomes
- **Certain Event:** The entire sample space Ω
- **Impossible Event:** Empty set ∅
- **Complementary Event:** $A^c = \Omega - A$

Visual Example with Dice:

- $\Omega = \{1,2,3,4,5,6\}$
- A = "even number" = {2,4,6}
- B = "number > 3" = {4,5,6}
- $A \cup B = \{2,4,5,6\}$ (even OR >3)
- $A \cap B = \{4,6\}$ (even AND >3)
- $A^c = \{1,3,5\}$ (odd numbers)

Event Operations:

- **Union ($A \cup B$):** Outcomes in A or B or both
- **Intersection ($A \cap B$):** Outcomes in both A and B
- **Difference (A - B):** Outcomes in A but not in B

d) Relations and Operations on Events:

The following table shows the correspondence between two languages: set theory language and probability language

Notation	Set Theory Vocabulary	Probability Vocabulary
∅	Empty set	Impossible event
Ω	Entire set / Universe	Certain event
$\omega \in \Omega$	Element of Ω	Elementary event
$A \subset \Omega$	Subset of Ω	Event
$\omega \in A$	ω belongs to A	The event A occurs
$A = B$	Sets A and B are equal	Events A and B are identical

Notation	Set Theory Vocabulary	Probability Vocabulary
A ⊂ B	A is a subset of B (A implies B)	A cannot occur without B occurring
A ∪ B	Union of A and B (A or B)	At least one of the two events occurs
A ∩ B	Intersection of A and B (A and B)	Both events occur
A ∩ B = Ø	A and B are disjoint	Events A and B are incompatible (mutually exclusive)
$\overline{A} = \Omega \setminus A$ or A^c	Complement of A in Ω	The event A does not occur

Table 2: Relations and Operations on Events

B.2.4 Probability Definitions

a) Classical (Laplacian) Probability

Definition: If all outcomes are equally likely:

$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}} = \frac{|A|}{|\Omega|}$$

Requirements:

- Finite sample space
- All outcomes equally likely
- Known and countable outcomes

Example:

- Probability of rolling an even number on a fair die:

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

- Fair coin: P(H) = 1/2
- Fair die: P(prime number) = P({2,3,5}) = 3/6 = 1/2
- Card draw: P(heart) = 13/52 = 1/4
- Lottery: P(win jackpot) = 1/(total combinations)

b) Axiomatic Definition (Kolmogorov)

Three Axioms:

- **Non-negativity:** Probabilities can't be negative (makes no sense)
  $P(A) \geq 0$ for any event A
- **Normalization:** Something must happen (total probability = 1)
  $P(\Omega) = 1$
- **Additivity:** For mutually exclusive events, probabilities add
  If A_1, A_2, ... are mutually exclusive events, then:

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

Example Application: For mutually exclusive events A and B:

- $P(A \cup B) = P(A) + P(B)$
- If $P(A) = 0.3$, $P(B) = 0.4$, then $P(A \cup B) = 0.7$

B.2.5 Probability Theorems and Properties

a) Introduction

Once the set of events of interest is defined, we try to translate their likelihood of occurrence into a number.

b) Probability Measure

1. Definition

Let $P(\Omega)$ be the set of all subsets of Ω (the power set).

Probability is a function that maps each event (each element of $P(\Omega)$) to a real number. This function, denoted P, is called a "probability measure".

$P: P(\Omega) \to \mathbb{R}\ E_i \to P(E_i)$

The number $P(E_i)$ is called the probability of occurrence of event E_i.

2. Basic Theorems

Theorem 1: $P(\emptyset) = 0$

Theorem 2: $P(A^c) = 1 - P(A)$

Theorem 3: If $A \subseteq B$, then $P(A) \leq P(B)$

Theorem 4: For any event A, $0 \leq P(A) \leq 1$

3. Addition Rule

General Addition Rule:

Why Subtract Intersection? When adding P(A) + P(B), the intersection is counted twice. Must subtract once to correct.

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

For Mutually Exclusive Events:

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

Extension to Three Events:

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & +P(A \cap B \cap C) \end{aligned}$$

Extended Example: Survey of 100 people:

- Like coffee: 60 → P(C) = 0.6
- Like tea: 50 → P(T) = 0.5
- Like both: 30 → $P(C \cap T) = 0.3$
- Like coffee OR tea: $P(C \cup T) = 0.6 + 0.5 - 0.3 = 0.8$
- 80 people like at least one beverage

Three Events Addition Rule:

Example: Student club memberships:

- Math club: 40% → P(M) = 0.4
- Science club: 30% → P(S) = 0.3
- Debate club: 25% → P(D) = 0.25
- Math & Science: 15% → $P(M \cap S) = 0.15$
- Math & Debate: 10% → $P(M \cap D) = 0.10$
- Science & Debate: 8% → $P(S \cap D) = 0.08$
- All three: 5% → $P(M \cap S \cap D) = 0.05$

P(M∪S∪D) = 0.4 + 0.3 + 0.25 - 0.15 - 0.10 - 0.08 + 0.05 = 0.67 67% of students in at least one club

4. Inclusion-Exclusion Principle

For events A_1, A_2, ..., A_n:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P\left(A_i\right) - \sum_{1 \le i < j \le n} P\left(A_i \cap A_j\right) + \sum_{1 \le i < j < k \le n} P\left(A_i \cap A_j \cap A_k\right) - \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

CHAPTER III: CONDITIONAL PROBABILITY AND INDEPENDENCE - DEEP DIVE

B.3.1 Conditional Probability - Comprehensive Treatment

a) Conceptual Foundation

Definition: The probability of event A occurring given that event B has already occurred.

Conditional probability updates our beliefs based on new information. It restricts the sample space to only those outcomes where the conditioning event occurred.

Mathematical Definition:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0$$

Intuitive Understanding:

- We restrict our attention to the reduced sample space where B has occurred
- Only the portion of A that lies within B is relevant
- The conditional probability "re-normalizes" probabilities within the context of B

b) Detailed Examples

Example: Card Drawing From a standard 52-card deck:

- Event B: Card is a face card (J, Q, K) - 12 cards
- Event A: Card is a King - 4 cards
- A ∩ B: Card is a King and a face card - 4 cards

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{4/52}{12/52} = \frac{4}{12} = \frac{1}{3}$$

Example: Medical Testing

- Disease prevalence: P(D) = 0.01
- Test sensitivity: P(+|D) = 0.95
- Test specificity: P(-|D^c) = 0.90

Probability of having disease given positive test:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.10 \times 0.99} \approx 0.0876$$

Example: Dice Roll

- $\Omega = \{1,2,3,4,5,6\}$
- Event A = "number > 3" = {4,5,6}, P(A) = 3/6 = 1/2
- Event B = "even number" = {2,4,6}, P(B) = 3/6 = 1/2
- Given B occurred (we know it's even), what's P(A)?
- Reduced sample space: {2,4,6}
- $A \cap B = \{4,6\}$
- $P(A|B) = |A \cap B|/|B| = 2/3$

B.3.2 Multiplication Rule and Chain Rule

a) Basic Multiplication Rule

The multiplication rule calculates joint probabilities from conditional probabilities.

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

b) Extended Chain Rule

For multiple events:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \cdots \times P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Example:

Sequential Drawing Urn with 5 red, 3 blue balls. Draw 3 balls without replacement:

$$P(\text{Red}_1 \cap \text{Red}_2 \cap \text{Red}_3) = \frac{5}{8} \times \frac{4}{7} \times \frac{3}{6} = \frac{60}{336} \approx 0.1786$$

Quality Control Batch of 100 items: 5 defective, 95 good. Randomly select 3 items without replacement.

P(all 3 good) = P(first good) × P(second good|first good) × P(third good|first two good) = (95/100) × (94/99) × (93/98) ≈ 0.856

P(at least one defective) = 1 - 0.856 = 0.144

Chain Rule Application: For events A, B, C: $P(A \cap B \cap C) = P(A) \times P(B|A) \times P(C|A \cap B)$

Example: Password Security

- P(first character letter) = 0.9
- P(second digit|first letter) = 0.8
- P(third special|first two) = 0.7
- P(valid password) = 0.9 × 0.8 × 0.7 = 0.504

B.3.3 Law of Total Probability

a) Theorem Statement

Breaks complex probability into simpler conditional probabilities. Works by partitioning sample space into mutually exclusive, exhaustive events.

If $B_1, B_2, \ldots, B_n$ form a partition of Ω (mutually exclusive and exhaustive), then for any event A:

$$P(A) = \sum_{i=1}^{n} P\,(A|B_i) \times P(B_i)$$

b) Special Case: Two Events

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Application Example:

Factory Production

- Factory has 2 machines: M1 (60% of production), M2 (40% of production)
- Defect rates: M1: 2%, M2: 3%

Probability a randomly chosen item is defective:

$$P(D) = P(D|M_1)P(M_1) + P(D|M_2)P(M_2) = 0.02 \times 0.60 + 0.03 \times 0.40 = 0.024$$

Factory Production 2 Three machines produce widgets:

- M1: 50% production, defect rate 2%
- M2: 30% production, defect rate 3%
- M3: 20% production, defect rate 5%

P(defective) = P(D|M1)P(M1) + P(D|M2)P(M2) + P(D|M3)P(M3) = 0.02×0.5 + 0.03×0.3 + 0.05×0.2 = 0.01 + 0.009 + 0.01 = 0.029

B.3.4 Bayes' Theorem - Comprehensive Treatment

a) Theorem Statement

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P\,(A|B_j)P(B_j)} = \frac{P(A|B_i)P(B_i)}{P(A)}$$

b) Bayesian Interpretation

- **Prior Probability P(B_i):** Initial belief before evidence
- **Likelihood P(A|B_i):** How probable the evidence is under each hypothesis
- **Posterior Probability P(B_i|A):** Updated belief after considering evidence
- **Evidence:** Normalizing constant ensuring probabilities sum to 1

c) Advanced Applications

Application 1: Spam Filtering

- Prior: P(Spam) = 0.3, P(Not Spam) = 0.7
- Likelihood of word "free":
    - P("free"|Spam) = 0.6
    - P("free"|Not Spam) = 0.1

Posterior probability email is spam given it contains "free":

$$P(\text{Spam}|\text{"free"}) = \frac{0.6 \times 0.3}{0.6 \times 0.3 + 0.1 \times 0.7} = \frac{0.18}{0.25} = 0.72$$

Now email also contains "winner":

- P("winner"|Spam) = 0.4
- P("winner"|Ham) = 0.01

Update using previous posterior as new prior: New prior: P(Spam) = 0.72, P(Ham) = 0.28

P(Spam|"free"∩"winner") = (0.6×0.4×0.72)/(0.6×0.4×0.72 + 0.1×0.01×0.28) ≈ 0.998

Application 2: Courtroom Evidence

- Prior probability of guilt: P(G) = 0.001
- Evidence: DNA match
    - P(Match|Guilty) = 0.999
    - P(Match|Innocent) = 0.0001

Posterior probability of guilt given DNA match:

$$P(G|M) = \frac{0.999 \times 0.001}{0.999 \times 0.001 + 0.0001 \times 0.999} \approx 0.909$$

B.3.5 Independence

a) Definition of Independence

Two events A and B are independent if:

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

Equivalently: $P(A|B) = P(A)$ or $P(B|A) = P(B)$

Two events A and B are independent if any of these hold:

1. P(A∩B) = P(A)P(B)
2. P(A|B) = P(A)
3. P(B|A) = P(B)

Extended Examples:

- **Dependent:** Draw card from deck: A = "card is heart", B = "card is queen" P(A) = 1/4, P(B) = 1/13, P(A∩B) = 1/52 1/52 = (1/4)×(1/13) = 1/52 → INDEPENDENT actually!
- **Actually Dependent:** Draw without replacement: First draw heart, second draw heart P(second heart|first heart) = 12/51 ≠ 13/52 = P(heart)

b) Pairwise vs Mutual Independence

Pairwise Independence: Events A, B, C are pairwise independent if:

- P(A∩B) = P(A)P(B)
- P(A∩C) = P(A)P(C)
- P(B∩C) = P(B)P(C)

Mutual Independence: Events A, B, C are mutually independent if:

- They are pairwise independent AND
- P(A∩B∩C) = P(A)P(B)P(C)

Important: Mutual independence implies pairwise independence, but not vice versa.

Pairwise vs Mutual Independence Example: Three fair coins: A=coin1=H, B=coin2=H, C=coin3=H

Pairwise: $P(A \cap B)=1/4=P(A)P(B)$, etc. Mutual: $P(A \cap B \cap C)=1/8=P(A)P(B)P(C)=1/8$ → MUTUALLY INDEPENDENT

Counterexample: Two dice, sum events A=die1 even, B=die2 even, C=sum even $P(A)=1/2$, $P(B)=1/2$, $P(C)=1/2$ Pairwise: $P(A \cap B)=1/4$, $P(A \cap C)=1/4$, $P(B \cap C)=1/4$ But $P(A \cap B \cap C)=1/4 \neq 1/8$ → NOT mutually independent

c) Conditional Independence

A and B are conditionally independent given C if:

$$P(A \cap B|C) = P(A|C) \times P(B|C)$$

Example: Weather events A = "rain today", B = "rain tomorrow", C = "low pressure system" Without C, A and B dependent Given C (low pressure), A and B may become independent

CHAPTER IV: RANDOM VARIABLES - COMPREHENSIVE THEORY

B.4.1 Definitions and Properties

a) Introduction

A random variable (RV) is a function defined on the set of all possible outcomes of a random experiment, such that it is possible to determine the probability that it takes a given value or that it falls within a given interval. Originally, a variable was a payoff function, representing the gain obtained at the outcome of a game. For example, when rolling two dice, we might be interested in the sum of the numbers being equal to 7. The pairs (1,6), (6,1), (5,2), (2,5), (3,4), (4,3) result in a sum equal to 7. Since this sum depends on random values, it is a random variable.

b) Definition

Let (Ω, S, P) be a probability space. A random variable X on a sample space Ω is a function from Ω to $\mathbb{R}$

$X: \Omega \to \mathbb{R}\ \omega \to X(\omega)$

such that the inverse image of every interval in $\mathbb{R}$ under X is an event in S.

$I = [a, b] \subset \mathbb{R}\ X^{-1}([a, b]) = \omega \in \Omega / X(\omega) \in [a, b]$

X is called a real random variable.

Example:

Let's give a simple example of rolling two dice, which is equivalent to rolling one die twice. A first random variable X_1 gives the result of the first roll, a second X_2 gives the result of the second roll, meaning $X_1(\omega) \in 1,2,3,4,5,6$ and $X_2(\omega) \in 1,2,3,4,5,6$, which we note more simply as $X_1 \in 1,2,3,4,5,6$ and $X_2 \in 1,2,3,4,5,6$.

One might be interested in the sum of the two results, which can be denoted by a random variable $S \in 2,3,4,5,6,7,8,9,10,11,12$.

c) Types of Random Variables

We distinguish two groups of random variables: Quantitative RVs and Qualitative RVs.

- **Quantitative Variables:**
  - Here we find all numerical RVs. We distinguish two types of quantitative variables:

- **Discrete Random Variables (DRV):**
  The values are discrete; they are whole numbers (finite or countably infinite).
  - $X \in D_x = x_1, x_2, \ldots, x_n$ (Finite DRV)
  - $X \in D_x = x_1, x_2, \ldots, x_n, \ldots$ (Infinite DRV)
- **Continuous Random Variables (CRV):**
  All values within a defined interval are possible. $X \in [a, b] \subset \mathbb{R}$

- **Qualitative Variables:**
  - **Nominal or Lexical Variables:**
    The different categories of the variable cannot be ordered.
  - **Ordinal Variables:**
    The categories of the variable possess the property of order.

- **Mixed Random Variables:**
  - Has both discrete and continuous components

Detailed Type Classification with Examples:

Discrete Random Variables (countable values):

- **Finite:** Dice roll (1-6), coin toss count in 10 flips (0-10)
- **Countably Infinite:** Number of calls to call center in hour (0,1,2,...), number of trials until first success

Continuous Random Variables (uncountable values):

- Height, weight, temperature, time, distance
- Any measurement with precision limited only by instrument

Qualitative/Categorical Variables:

- **Nominal:** Colors, brands, blood types (no inherent order)
- **Ordinal:** Survey responses (poor, fair, good, excellent), education level

Mixed Variables: Insurance claims (0 with probability p, continuous positive amount with probability 1-p)

d) Probabilities of a Random Variable

Let X be a Random Variable defined by an application: $\Omega \to \mathbb{R}$ taking values in $D_x = x_1, x_2, \ldots, x_n$.

By definition, the probability that $X = x$ is the probability of the elements of Ω whose image is the value x under the application. Let $X = x_1, x_2, \dots, x_n; P = p_1, p_2, \dots, p_n$, with $p_i = P(X = x_i); i = \overline{1, n}$, we have $p_i = P(X = x_i) = P(\omega \in \Omega / X(\omega) = x_i)$.

B.4.2 Probability Distribution of a Random Variable

a) Probability Distribution of a Discrete Random Variable (DRV)

Let x be a DRV (finite). The probability distribution of x is given by:

- The set of possible values: $D_x = x_1, x_2, \dots, x_n$
- The probability of each value: $P = p_1, p_2, \dots, p_n / p_i = P(X = x_i)$.

We denote by f the probability distribution of x, which is written as: $\forall x_i, f(x_i) = P(X = x_i)$. The function f is called the distribution function of the DRV x.

Remarks

$\forall i: f(x_i) = P(X = x_i) \geq 0$

$$-\sum_{i=1}^{n(\infty)} f(x_i) = \sum_{i=1}^{n(\infty)} P(X = x_i) = 1$$

$$-P(a \leq x \leq b) = \sum_{x_i \in [a,b]} f(x_i)$$

Detailed PMF Properties:

1. $0 \leq p(x_i) \leq 1$ for all i
2. $\Sigma\, p(x_i) = 1$
3. $P(a \leq X \leq b) = \Sigma\_\{x_i \in [a,b]\}\ p(x_i)$

Extended Example: Dice Sum Distribution Two fair dice rolled, X = sum Possible values: 2,3,...,12

PMF:

- P(X=2) = 1/36
- P(X=3) = 2/36
- ...
- P(X=7) = 6/36
- ...
- P(X=12) = 1/36

$P(5 \leq X \leq 8) = P(5)+P(6)+P(7)+P(8) = (4+5+6+5)/36 = 20/36 = 5/9$

Example 2:

Let x be the Random Variable given by the following table:

x_i	$P(x_i)$
0	1/8
1	3/8
2	3/8
3	1/8

Table 3: Example on Probabilities

$$P(0 \leq x \leq 2) = \sum_{x \in [0,2]} f(x_i) = f(0) + f(1) + f(2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

b) Probability Distribution of a Continuous Random Variable (CRV)

Let x be a CRV. The probability distribution or probability density of x is the function f such that:

- $\forall x \in \mathbb{R}; f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

Remarks

Let X be a CRV with density f: $P(a \leq x \leq b) = \int_a^b f(x)dx$

This represents the area under the curve of f between a and b.

Detailed PDF Properties:

1. $f(x) \geq 0$ for all x
2. ∫_{-∞}^{∞} f(x)dx = 1
3. P(a ≤ X ≤ b) = ∫_{a}^{b} f(x)dx

Example Verification: f(x) = kx for 0≤x≤1, 0 otherwise

1. Find k: $\int_0^1 kx\,dx = k[x^2/2]_0^1 = k/2 = 1 \rightarrow k = 2$
2. P(0.5≤X≤1) = ∫*{0.5}¹ 2x dx = [x²]*{0.5}¹ = 1 - 0.25 = 0.75

Example 2:

Let f be a density function of a CRV x defined by: $f(x) = kx$, if $0 \leq x \leq 1$ and 0 otherwise.

- Determine the value of k so that f is a density.
- Calculate $P\left(\frac{1}{2} \leq x \leq 1\right)$.

B.4.3 Cumulative Distribution Function

a) Cumulative Distribution Function of a Discrete Random Variable (DRV)

Definition

Let x be a DRV (finite or infinite). The cumulative distribution function of x defined on (Ω, S, P) is the function F_x defined on ℝ by: $F_x\colon \mathbb{R} \to [0,1]\ x \to F_x(x) = P(X \leq x) = \sum_{x_i \leq x} P\,(X = x_i)$

b) Cumulative Distribution Function of a Continuous Random Variable (CRV)

Definition

Let x be a CRV with density f. The cumulative distribution function of x is the function F_x defined by: $F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f\,(t)dt$

Properties

- The graph of F_x is a continuous function (for CRVs).
- $P(a \leq x \leq b) = F_x(b) - F_x(a)$ (in the case where F_x is continuous)

Remark

If F_x is continuous on ℝ and if f is continuous, $\forall x \in \mathbb{R}$, then F_x is differentiable and we have: $F_x'(x) = f(x)$.

Detailed Properties:

1. $0 \leq F(x) \leq 1$
2. Non-decreasing: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
3. Right-continuous: lim_{h→0^+} F(x+h) = F(x)
4. lim_{x→-∞} F(x) = 0, lim_{x→∞} F(x) = 1
5. $P(a < X \leq b) = F(b) - F(a)$

Discrete Example: X = number of heads in 3 coin tosses Values: 0,1,2,3 with probabilities: 1/8, 3/8, 3/8, 1/8

CDF:

- $F(0) = P(X \leq 0) = 1/8$
- $F(1) = P(X \leq 1) = 1/8+3/8 = 4/8 = 1/2$
- $F(2) = P(X \leq 2) = 1/8+3/8+3/8 = 7/8$
- $F(3) = P(X \leq 3) = 1$

Continuous Example: Uniform [0,1] $F(x) = 0$ for $x<0$, $= x$ for $0 \leq x \leq 1$, $= 1$ for $x>1$ $P(0.3<X \leq 0.7) = F(0.7)-F(0.3) = 0.7-0.3 = 0.4$

B.4.4 Mathematical Expectation (Mean)

Let x be a Random Variable with distribution function f. The mathematical expectation of x, denoted by E(x), is defined by:

- For DRV:

$$E(x) = \sum_{i=1}^{n(\infty)} x_i\, f(x_i) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n).$$

- For CRV:

$$E(x) = \int_{-\infty}^{+\infty} x\, f(x) dx.$$

Properties of Expectation

1. **Linearity:** $E[aX + bY + c] = aE[X] + bE[Y] + c$
2. **Monotonicity:** If $X \leq Y$, then $E[X] \leq E[Y]$
3. **Expectation of Function:** $E[g(X)] = \sum g(x)p_x(x)$ or $\int g(x)f_x(x)dx$

Discrete Examples:

1. Fair die: $E[X] = (1+2+3+4+5+6)/6 = 3.5$
2. Bernoulli(p): $E[X] = 0\times(1-p) + 1\times p = p$
3. Game: Win 1 otherwise E[winnings] = $10\times0.1 + (-1)\times0.9 = 1 - 0.9 = 0.1$ (favorable game)

Continuous Examples:

1. Uniform[a,b]: $E[X] = (a+b)/2$
2. Exponential(λ): $E[X] = 1/\lambda$

Linearity Proof Example: $E[2X + 3Y - 5] = 2E[X] + 3E[Y] - 5$ Regardless of dependence between X and Y

B.4.5 Variance and Moments

a) Variance

$$Var(x) = E(x^2) - \left(E(x)\right)^2$$

- For DRV:
  $E(x^2) = \sum_{i=1}^{n} x_i^2\, f(x_i).$
- For CRV:
  $E(x^2) = \int_{-\infty}^{+\infty} x^2\, f(x)dx.$

b) Non-Central Moment of order k

$$m_k = E(x^k) = \int_{-\infty}^{+\infty} x^k\, f(x)dx \quad \text{(for CRV)}$$

c) Central Moment of order k

$$M_k = E[(x - m_1)]^k = \int_{-\infty}^{+\infty} (x - m_1)^k\, f(x)dx \quad \text{(for CRV)}$$

Examples:

1. Constant: Var(c) = 0
2. Bernoulli(p): Var(X) = p(1-p) Maximum at p=0.5: Var=0.25 Minimum at p=0 or 1: Var=0
3. Uniform[a,b]: Var(X) = $(b-a)^2/12$

Standard Deviation: σ = √Var (same units as X, easier interpretation)

Coefficient of Variation: CV = σ/μ (relative dispersion, unitless)

Moment Interpretation:

- First moment (mean): center
- Second central moment (variance): spread
- Third standardized moment (skewness): asymmetry
- Fourth standardized moment (kurtosis): tail heaviness

B.4.6 Variance and Covariance

a) Variance Properties

- Var(aX + b) = a^2Var(X)

- Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)
- If X, Y independent: Var(X + Y) = Var(X) + Var(Y)

b) Covariance and Correlation

Covariance:

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$$

Covariance Detailed Interpretation:

- Positive: X tends to be above mean when Y above mean
- Negative: X tends to be above mean when Y below mean
- Zero: No linear tendency (but could have nonlinear relationship)

Examples:

1. X=height, Y=weight: Cov > 0 (taller people tend heavier)
2. X=temperature, Y=heating cost: Cov < 0 (colder → higher cost)
3. X=coin toss 1, Y=coin toss 2: Cov = 0 (independent)

Correlation Coefficient:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1$$

Correlation Properties:

1. $-1 \le \rho \le 1$
2. $\rho = \pm 1$ iff perfect linear relationship: Y = aX + b
3. $\rho = 0$ for independent variables (but not conversely!)
4. Invariant to linear transformations

Example: Perfect Relationships

- Y = 2X + 3: $\rho = 1$
- Y = -0.5X + 1: $\rho = -1$
- $Y = X^2$ with X symmetric about 0: $\rho = 0$ (but clearly dependent!)

c) Moment Generating Functions

Definition:

$$M_X(t) = E[e^{tX}]$$

Properties:

- $E[X^k] = M_X^{(k)}(0)$
- If $M_x(t) = M_Y(t)$ for all t, then X and Y have the same distribution
- For independent X, Y: $M_{X+Y}(t) = M_X(t) \times M_Y(t)$

CHAPTER V: DISCRETE PROBABILITY DISTRIBUTIONS - COMPLETE ANALYSIS

B.5.1 Bernoulli Distribution

Definition:

A real random variable X follows a Bernoulli distribution with parameter p if it only takes the values 0 and 1 such that P(X = 1) = p; P(X = 0) = 1 - p = q. It is denoted by: X ~ B(p)

Characteristics:

$$E(X) = p;\ V(X) = pq;\ \sigma x = \sqrt{(pq)}$$

Proof:

We have $E(X) = \sum_i x_i P_i = 1 \times P(X = 1) + 0 \times P(X = 0) = p$

And: $E[(X - E(X))^k] = E[(X - p)^k] = (-p)^k\, q + (1 - p)^k\, p$

So: $E[(X - E(X))^k] = (-p)^k\, q + (1 - p)^k\, p$

In particular: $V(X) = E[(X - E(X))^2] = (-p)^2\, q + (1 - p)^2\, p = p^2\, q + (1-p)^2\, p = p\, q\, (p + q) = p\, q$.

PMF:

$$P(X = k) = \begin{cases} 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

Properties:

- E[X] = p
- Var(X) = p(1-p)
- MGF: M(t) = 1-p + pe^t

Applications: Single trial experiments, binary outcomes

Extended Applications:

- Single coin toss
- Yes/no survey response
- Success/failure of single trial
- Binary classification outcome

Example Quality Control: Item defective with probability 0.02 X ~ Bernoulli(0.02)

$E[X] = 0.02$, $Var(X) = 0.02 \times 0.98 = 0.0196$

B.5.2 Binomial Distribution

Definition:

A random variable X follows a Binomial distribution with parameters n and p if it has the following probability mass function:

For k = 0, 1, 2, ..., n; $P(X = k) = C_n^k\, p^k\, q^{n-k}$,

where $q = 1 - p$ It is denoted by: $X \sim B(n, p)$ A Bernoulli variable is a special case of a binomial variable: $X \sim B(1, p)$

Characteristics:

$E(X) = np$; $V(X) = n\, p(1-p) = n\, p\, q$; $\sigma x = \sqrt{(n\, p\, q)}$

PMF:

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

Using Newton's binomial formula, this gives: $\sum_{k=0}^{n} f(x) = (p + (1 - p))^n = 1$

Thus, if we take, for all integer k less than or equal to n; $P(X = k) = f(x)$ We define a discrete random variable and its probability distribution.

The definition of E(X) gives $E(X) = \sum_{k=0}^{n} k\, C_n^k\, p^k\, (1 - p)^{n-k} = \sum_{k=1}^{n} k\, C_n^k\, p^k\, (1 - p)^{n-k}$

Since we have $C_n^k = (n/k)\, C_{n-1}^{k-1}$, it follows

$$E(X) = np \sum_{k=1}^{n} C_{n-1}^{k-1}\, p^{k-1}(1-p)^{n-k}$$

We can therefore write, using the binomial formula again:

$$E(X) = np[p + (1-p)]^{n-1} = np$$

Similarly, we have:

$$E[X(X-1)] = \sum_{k=0}^{n} k\,(k-1) C_n^k p^k (1-p)^{n-k}$$

Since we have $C_n^k = \frac{n(n-1)}{k(k-1)} C_{n-2}^{k-2}$

It follows:

$$E[X(X-1)] = n(n-1)p^2 \sum_{k=0}^{n} C_{n-2}^{k-2}\, p^{k-2}(1-p)^{n-k} = n(n-1)p^2[p+(1-p)]^{n-2}$$
$$= n(n-1)p^2$$

Hence:

$$E(X^2) = E[X(X-1)] + E[X] = n(n-1)p^2 + np$$

Then:

$$V(X) = E(X^2) - [E(X)]^2 = np(1-p) = npq$$

Example: A fair coin is tossed 10 times in a row. What is the probability of getting a total of 8 heads? Let X be the random variable associating the number of heads to these 10 coin tosses.

We have $X \sim B\left(10, \frac{1}{2}\right)$

$$P(X=8) = C_{10}^{8}\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{10-8} = 0.0439$$

Properties:

- E[X] = np
- Var(X) = np(1-p)
- MGF: M(t) = (1-p + pe^t)^n

Derivation: Sum of n independent Bernoulli trials

Applications:

- Quality control,
- survey sampling,
- clinical trials

Detailed Derivation: Sum of n independent Bernoulli trials

Extended Applications:

- Number of heads in n coin tosses
- Number of defective items in batch sampling

- Number of successful treatments in clinical trial
- Number of "yes" responses in survey

Quality Control Example: Sample 20 items, each defective with p=0.05 X ~ Binomial(20, 0.05) P(exactly 2 defectives) = C(20,2)×0.05^2×0.95^{18} ≈ 0.1887 P(at most 1 defective) = P(0)+P(1) ≈ 0.3585+0.3774 = 0.7359

Normal Approximation: For large n, Binomial ≈ Normal Rule of thumb: np ≥ 5 and n(1-p) ≥ 5

B.5.3 Poisson Distribution

Definition:

A random variable X follows a Poisson distribution with parameter $\lambda(\lambda > 0)$ if it has the following probability mass function:

PMF:

$$P(k,\lambda) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,2,\ldots$$

Proof: We have

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

So the relation $P(X = k) = f(x)$ defines a discrete random variable X and its probability distribution.

The definition of $E(X)$ gives

$$E(X) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

Since $\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{\lambda}$

it follows $E(X) = \lambda$

Similarly, we have

$$E[X(X-1)] = \sum_{k=0}^{\infty} k\,(k-1)\frac{e^{-\lambda}\lambda^k}{k!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}$$

Since $\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = e^{\lambda}$, we obtain $E[X(X-1)] = \lambda^2$

Consequently $E(X^2) = E[X(X-1)] + E[X] = \lambda^2 + \lambda$

It follows that $V(X) = E(X^2) - [E(X)]^2 = \lambda$

Properties:

- $E[X] = \lambda$
- $Var(X) = \lambda$
- $\sigma_X = \sqrt{\lambda}$
- MGF: M(t) = e^{λ(e^t - 1)}

Poisson Process Derivation:

- Events occur randomly in time/space
- Average rate λ events per unit
- Events independent across disjoint intervals

Applications:

- Call center arrivals
- Radioactive decay
- Traffic flow
- Natural disaster occurrences

Detailed Derivation from Binomial: Limit as $n\to\infty$, $p\to 0$, $np\to\lambda$

Poisson Process Characteristics:

1. Events occur independently
2. Average rate λ constant
3. No simultaneous events

Extended Applications:

- Call center: Calls per hour
- Website: Hits per minute
- Biology: Mutations per genome
- Finance: Trading orders per second
- Physics: Radioactive decays per second

Example: Call Center Average 10 calls per hour → $\lambda = 10$

P(15 calls in next hour) = $e^{-10} \times 10^{15}/15! \approx 0.0347$ P(≤5 calls) = $\Sigma_{k=0}^{5} e^{-10} \times 10^k/k! \approx 0.0671$

B.5.4 Geometric Distribution

PMF:

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1,2,3,\ldots$$

Properties:

- $E[X] = 1/p$
- $Var(X) = (1-p)/p^2$
- Memoryless property: $P(X > m+n \mid X > m) = P(X > n)$

Memoryless Property Detailed: $P(X > m+n \mid X > m) = P(X > n)$ "Future independent of past" - unique among discrete distributions

Applications:

- Number of coin tosses until first head
- Number of product tests until first failure
- Number of attempts until successful login

Example: Basketball Free Throws Player makes 70% of free throws X = attempts until first miss ~ Geometric(0.3) P(first miss on 4th attempt) = $0.7^3 \times 0.3 = 0.1029$ $E[X] = 1/0.3 \approx 3.33$ attempts

B.5.5 Negative Binomial Distribution

PMF:

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}, \quad k = r, r + 1, \ldots$$

Properties:

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

Relationship to Geometric: Sum of r independent Geometric(p)

Applications:

- Number of trials until r-th success

- Number of interviews until hiring 3 candidates
- Number of at-bats until 2 home runs

Example: Sales Calls Salesperson makes sale with probability 0.2 per call Want 3 sales X ~ Negative Binomial(r=3, p=0.2) P(need exactly 10 calls) = $C(9,2) \times 0.2^3 \times 0.8^7 \approx 0.0604$ E[X] = 3/0.2 = 15 calls expected

CHAPTER VI: CONTINUOUS PROBABILITY DISTRIBUTIONS

B.6.1 Uniform Distribution

Definition:

A continuous random variable X follows a uniform distribution on [a, b] if it has the following probability density function f defined by:

PDF:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Properties:

- E[X] = (a+b)/2
- Var(X) = (b-a)²/12
- $\sigma_X = \sqrt{\frac{(b-a)^2}{12}}$
- CDF: $F_x(x)$ = (x-a)/(b-a) for a ≤ x ≤ b

Proof: By definition of $E(X)$

- $E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$
- Similarly
- $E(X^2) = \int_{-\infty}^{+\infty} x^2\, f(x)dx = \int_a^b \frac{x^2}{b-a} dx = \frac{b^2+ab+a^2}{3}$
- Since $V(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$
- it follows $V(X) = \frac{(b-a)^2}{12}$

Detailed Properties:

- Maximum entropy distribution given only bounds
- All intervals of same length equally likely

Applications:

- Random number generation
- Rounding/quantization error

- Prior distribution in Bayesian analysis when only bounds known

Example: Bus Arrival Bus arrives every 10 minutes uniformly Wait time X ~ Uniform[0,10] P(wait ≤ 2 minutes) = 2/10 = 0.2 P(3 < wait < 8) = (8-3)/10 = 0.5 Expected wait: E[X] = 5 minutes

B.6.2 Exponential Distribution

Definition:

A continuous random variable X follows an exponential distribution with parameter λ if it has the following probability density function f defined by:

PDF:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Properties:

- $E[X] = 1/\lambda$
- $Var(X) = 1/\lambda^2$
- $\sigma_X = \frac{1}{\lambda}$
- Memoryless property: $P(X > s+t \mid X > s) = P(X > t)$

Proof: By definition of E(X)

- $E(X) = \int_{-\infty}^{+\infty} x\, f(x) dx = \int_0^{+\infty} \lambda\, x e^{-\lambda x} dx = \lim_{x \to +\infty} \int_0^x \lambda\, t e^{-\lambda t} dt$
- Integration by parts gives:
- $\int_0^x \lambda\, t e^{-\lambda t} dt = -x e^{-\lambda x} + \frac{1}{\lambda}\left(1 - e^{-\lambda x}\right)$
- Hence $E(X) = \frac{1}{\lambda}$

Memoryless Property Detailed: $P(X > s+t \mid X > s) = P(X > t)$ "If it hasn't happened yet, time to event distribution same as original"

Applications:

- Time between events in Poisson process
- Equipment lifetime (constant failure rate)
- Waiting times

- Radioactive decay

Example: Customer Service Time between customer arrivals ~ Exponential(λ=0.1 per minute) Mean time between arrivals = 1/0.1 = 10 minutes P(next arrival within 5 minutes) = 1 - e^{-0.1×5} ≈ 0.3935 P(wait > 15 minutes given already waited 10) = P(X > 5) = e^{-0.5} ≈ 0.6065

B.6.3 Normal Distribution

Definition:

A continuous random variable X follows a standard normal distribution, denoted by N(0, 1), if it has the following probability density function f defined on R by:

PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Since f is a density, the area between the x-axis and the curve y = f(x) is finite and equal to 1. The cumulative distribution function of X is: $F(x) = \int_{-\infty}^{x} f\,(t)dt$

a) Properties and Characteristics

- Bell-shaped curve
- Symmetric about mean μ
- Points of inflection at x = μ ± σ
- Total area under curve = 1

b) Standard Normal Distribution

- μ = 0, σ = 1
- PDF: φ(z) = \frac{1}{\sqrt{2π}} e^{-z²/2}
- CDF: Φ(z) = ∫_{-∞}^z φ(t) dt

c) Standardization

If X ~ N(μ, σ²), then:

$$Z = \frac{X - \mu}{\sigma}\ N(0,1)$$

d) Empirical Rule

- P(μ - σ < X < μ + σ) ≈ 0.6827
- P(μ - 2σ < X < μ + 2σ) ≈ 0.9545
- P(μ - 3σ < X < μ + 3σ) ≈ 0.9973

e) Central Limit Theorem

If $X_1, X_2, ..., X_n$ are i.i.d. with mean μ and variance σ^2, then:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

Properties:

- E[X] = 0
- Var(X) = 1
- $\sigma_X = 1$

Proof: According to the definition:

- $E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx$
- The function $x e^{-\frac{x^2}{2}}$ is odd, therefore E(X) = 0. Similarly:
- $V(X) = \int_{-\infty}^{+\infty} x^2\, f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx$
- The function $x^2 e^{-\frac{x^2}{2}}$ is even, therefore:
- $V(X) = 2\int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{\pi}} \int_{0}^{+\infty} x^2\, e^{-x^2} dx$
- Using a change of variables to polar coordinates, we easily obtain:
- $V(X) = \frac{2}{\sqrt{\pi}} \times \frac{\sqrt{\pi}}{2} = 1$

Applications:

- Measurement errors
- Biological measurements (height, weight)
- Test scores
- Financial returns (log returns)
- Quality control

Standardization Examples: X ~ N (70, 5^2) (height in inches, σ=5) P (X ≤ 65) = P (Z ≤ (65-70)/5) = P (Z ≤ -1) = 0.1587 P (68 ≤ X ≤ 72) = P (-0.4 ≤ Z ≤ 0.4) = 0.6554 - 0.3446 = 0.3108

Quality Control Example: Bottles filled with mean 500ml, σ=2ml Specification: 500 ± 5ml P(within spec) = P(495 ≤ X ≤ 505) = P(-2.5 ≤ Z ≤ 2.5) ≈ 0.9876 1.24% defective rate

B.6.4 Gamma Distribution

PDF:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

Properties:

- E[X] = α/β
- Var(X) = α/β^2
- $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,dt$ (Gamma function)

Relationship to Exponential: Sum of α independent Exponential(β)

Special Cases:

- Exponential: $\alpha = 1$
- Chi-square: $\alpha = \nu/2$, $\beta = ½$
- Erlang: α integer

Applications:

- Waiting time for α events in Poisson process
- Insurance claim amounts
- Rainfall modeling
- Bayesian conjugate prior for Poisson rate

Example: Call Center Time for 3 calls to arrive ~ Gamma (α=3, β=0.1) Mean time = 3/0.1 = 30 minutes P (time ≤ 20 minutes) requires gamma CDF calculation

B.6.5 Beta Distribution

PDF:

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad 0 \le x \le 1$$

Properties:

- E[X] = $\alpha/(\alpha+\beta)$
- Var(X) = $\alpha\beta/[(\alpha+\beta)^2 (\alpha+\beta+1)]$

Flexible Shapes: Can model many different distribution shapes on [0,1]

Bayesian Interpretation: Conjugate prior for binomial proportion

Applications:

- Modeling probabilities or proportions
- Bayesian A/B testing
- Task completion rates
- Project success probabilities

Example: Website Conversion Rate Prior: Beta(α=2, β=8) (belief: low conversion) Observe: 10 conversions in 100 visits Posterior: Beta(α=2+10=12, β=8+90=98)

Posterior mean = 12/(12+98) = 0.1091 95% credible interval requires beta quantiles

Bibliographic References

[1] Y. Croissant, *Descriptive Statistics and Probability*. Brussels, Belgium: De Boeck, 2010.

[2] J.-F. Delmas, *Introduction to Probability Theory and Statistics*. Berlin, Germany: Springer, 2014.

[3] M. Mountassir, *Probability and Statistics*, 2nd ed. Paris, France: Ellipses, 2011.

[4] M. Rouaud, *Probability, Statistics and Multicriteria Analysis*. Paris, France: Dunod, 2012.

[5] G. Saporta, *Probability, Data Analysis and Statistics*, 3rd ed. Paris, France: Technip, 2014.

[6] A.-J. Valleron, *Probability and Statistics*, 2nd ed. Paris, France: Flammarion, 2006.

[7] "*Cours de statistique descriptive*," Université Abderrahmane Mira de Bejaïa (PDF), available: http://fmi.univ-tiaret.dz/images/1lmd/2021-2022/S2/Cours_Statistique_discriptive.pdf (accessed Dec. 21, 2025). (Faculté MI)

[8] "*Statistics and Probability summaries*," Study smarter, available: https://www.studysmarter.fr/resumes/mathematiques/statistiques-et-probabilites/#:~:text=Les%20statistiques%20et%20probabilit%C3%A9s%20sont,qu'%C3%A0%20calculer%20des%20probabilit%C3%A9s (accessed Dec. 21, 2025).

[9] "*Analyzing categorical data – Probability and Statistics*," Khan Academy, available: https://fr.khanacademy.org/math/statistics-probability/analyzing-categorical-data (accessed Dec. 21, 2025). (Khan Academy)