Mouloud Mammeri University of Tizi–Ouzou
Faculty of Sciences
Mathematics Department

PROBABILITY AND STATISTICS

Lecture notes for Engineers (ST Domains)

HESSAS Fatima

September 2025

PROBABILITY AND STATISTICS

Lecture notes for Engineers (ST Domain)

This course has been developed as an educational resource for second-year Bachelor's degree students specialising in Technological Sciences (ST) at the Faculty of Construction Engineering at Mouloud Mammeri University in Tizi Ouzou. Drawing on over ten years of teaching experience at this institution, it provides comprehensive guidance on the fundamental principles of statistics and probability, in line with the official LMD programme for the Civil Engineering discipline, which is taught during the third semester of the university course. The book has been carefully designed to emphasise clarity, breaking down complex concepts into their most basic elements while avoiding mathematical proofs. It has been supplemented with illustrations, elucidating remarks and examples of application to enhance comprehension and accessibility.

The first part of the book is dedicated to descriptive statistics, explaining how to represent univariate and bivariate statistical series and how to describe and analyse their characteristics numerically.

The second part of the book covers the theory of probability, introducing all the necessary terminology to understand random events. The course provides concrete, illustrated examples of calculating probability distributions for discrete and continuous random variables, as well as their characteristics, such as mathematical expectation and variance. Some of the most commonly used probability distributions for discrete and continuous random variables are detailed, and their corresponding statistical tables are listed in the appendix.

Contents

Part I

Descriptive Statistics

CHAPTER 1

BASICS OF STATISTICS

1 Introduction

Statistics constitutes a branch of mathematics. This essential tool is designed to facilitate the comprehension and management of complex and random phenomena. It aims to collect and process data from the observation of these phenomena in order to summarise the representation of the contained information (e.g. coding, graphs, tables and characteristics) for analysis, interpretation, decision-making, problem-solving and process design in many areas, including scientific research, public health, business and economy, engineering and agriculture.

The two broad categories of statistics are ***descriptive statistics*** and ***inferential statistics***. The former summarise and represent a set of data with the purpose of displaying and describing its characteristics, while the latter is used to test hypotheses and make decisions and predictions on a large population on the basis of limited data (samples) from this population.

In this first part of the course, the focus will be on descriptive statistics, which can have one dimension (univariate) if the data relate to only one variable, or two dimensions (bivariate) where two variables are concerned simultaneously.

2 Basic definitions and concepts

2.1 Population, sample and individual

It is imperative for any statistical study to be conducted within the confines of the statistical space in which the study is being carried out. This statistical space can be categorised into two distinct types: the ***population*** and the ***sample***.

Definition 1

A *population* is defined as a finite or infinite set of elements (people or objects) that is subject to statistical study. This set is denoted by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. The number N of the elements ω_i is the defined as the population size.

The observation of a whole population is frequently challenging, if not unfeasible. In general, the population under study is limited to a sample.

Definition 2

A *sample* is defined as a restricted collection of elements selected in an arbitrary manner from the population under consideration for the statistical investigation.

Definition 3

An element of the population or sample, which is the subject of the statistical study, is referred to as an *individual* or *statistical unit*. The number of individuals in a sample is denoted by $n < N$ and is referred to as the sample size.

Example 1

A civil engineer is tasked with analysing the compressive strength of concrete utilised in a construction project. The project involves the utilisation of 200 concrete blocks. However, due to practical constraints, the engineer decides to test a sample of 20 blocks, which are selected at random.

- The population under consideration is constituted by all the 200 concrete blocks that were used in the project.
- The sample under consideration is constituted by the 50 concrete blocks that were randomly selected for the purposes of testing. The sample size is $n = 50$.
- The individual is the concrete blocks from which the construction is assembled.

2.2 Statistical variables and modalities

In the context of the population or sample under consideration, individuals can be studied according to specific properties or characteristics.

Definition 4

A *statistical variable*, denoted X (or Y), is defined as any measurable or identifiable characteristic of a given individual, within a population or a sample.
The statistical variable X can be defined as an application:

$$\begin{array}{rll} X: & \Omega \rightarrow & \mathcal{C} \\ & \omega_i \rightarrow & X(\omega_i) \end{array}$$

$\mathcal{C}$is the set in which X takes its values.

It is important to note that statistical variables can be categorised into two distinct types: ***Quantitative*** and ***Qualitative*** variables.

- Quantitative variables are defined as measurable characteristics ($\mathcal{C} \subset \mathbb{R}$). These can be further categorised into two types:
    - *Discrete quantitative variables* where the *modalities* (different possible values) x_i of $\mathcal{C}$ are isolated:
      $$\mathcal{C} = \{x_1, x_2, \ldots, x_i, \ldots, x_k\}\,.$$
      The x_i values are assumed to be ordered $x_1 < x_2 < \ldots < x_k$.
    - *Continuous quantitative variables* where the modalities are continuous (specific intervals of $\mathbb{R}$). The modalities of $\mathcal{C}$ are different *classes*:
      $$\mathcal{C} = \{[c_1, c_2[\,, [c_2, c_3[\,, \ldots, [c_i, c_{i+1}[\,, \ldots, [c_k, c_{k+1}]\}\,.$$
      Each class $[c_i, c_{i+1}[$, with $i = \overline{1, k}$, is characterised by an interval, with its left side closed and its right side open.
      The width of each class $[c_i, c_{i+1}[$ is defined by the value $w_i = c_{i+1} - c_i$.
      The centre of each class $[c_i, c_{i+1}[$ is defined by the value $C_i = \dfrac{c_i + c_{i+1}}{2}$.
- Qualitative variables are defined as non-measurable characteristics ($\mathcal{C}$ not a subset of $\mathbb{R}$), which are expressed through an individual's membership of a specific category.

Example 2

- For a sample of 30 students from a population of 150 students in a particular section, the character X, defined as "the mark obtained by each student at an examination", is a discrete quantitative statistical variable. For $i = \overline{1, 4}$, its modalities are:
  $$\mathcal{C} = \{5, 8, 12, 15\}\,.$$
- For the same previous sample of 30 students, the character X, defined as "the size in cm of each student", is a continuous quantitative statistical variable. For $i = \overline{1, 4}$, its modalities are the different classes:
  $$\mathcal{C} = \{[159, 164[\,, [164, 169[\,, [169, 174[\,, [174, 179[\}\,.$$
  The amplitude, or width, of each class is $w_i = 5\,\text{cm}$, and the centre of the first class is $\dfrac{159 + 164}{2} = 161.5\,\text{cm}$.
- For the same previous sample of 30 students, the character X, defined as "the eye's colour for each student", is a qualitative statistical variable. Its modalities are :
  $$\mathcal{C} = \{\text{green}, \text{black}, \text{brown}, \text{blue}\}\,.$$

CHAPTER 2

UNIVARIATE STATISTICAL SERIES

1 Introduction

Univariate statistics is the branch of descriptive statistics that focuses on the analysis of a single variable. The purpose of this section is to provide a comprehensive description of the characteristics of the variable in question. This approach facilitates the discernment of patterns within a singular dataset, independently of external comparison with other datasets. In this chapter, the methodology for the representation of the aforementioned dataset in tabular and graphical form will be examined, with a view to determining characteristics and parameters such as the arithmetic mean, the median and the variance.

2 Frequency and cumulative frequency

In order to take into account the occurrence of each of the different modalities of a statistical variable, it is necessary to define a number of other concepts including the ***frequency*** (or ***absolute frequency***), the ***relative frequency***, the ***cumulative frequency*** and the ***relative cumulative frequency***.

Definition 1

The *frequency* (or *absolute frequency*) of each modality x_i(or the class $[c_i, c_{i+1}[$), denoted by n_i, is defined as the number of occurrences of the modality x_i (or a value in the class $[c_i, c_{i+1}[$) that have been observed for $i = \overline{1, k}$.
We write:

$$\sum_{i=1}^{i=k} n_i = n, \tag{2.1}$$

where n denotes the sample size.

Definition 2

The *relative frequency* of each modality x_i(or class $[c_i, c_{i+1}[$), denoted by f_i, is defined as the quotient of the frequency of the modality x_i (or the class $[c_i, c_{i+1}[$) to the sample size.
We write:

$$f_i = \frac{n_i}{n} \quad ; i = \overline{1, k}. \tag{2.2}$$

We note that $\sum_{i=1}^{i=k} f_i = 1$.

Definition 3

The *cumulative frequency* in each modality x_i(or class $[c_i, c_{i+1}[$), denoted by N_i, is defined as follows:

$$N_i = \sum_{j=1}^{j=i} n_j \quad ; i = \overline{1, k}. \tag{2.3}$$

Definition 4

The *cumulative relative frequency* of each modality x_i(or class $[c_i, c_{i+1}[$), denoted by F_i, is defined as follows:

$$F_i = \sum_{j=1}^{j=i} f_j = \frac{n_j}{n} \quad ; i = \overline{1, k}. \tag{2.4}$$

We note that $F_k = \sum_{j=1}^{j=k} f_j = 1$.

3 From discrete modalities to classes

In the context of continuous quantitative variables, it is frequently observed that multiple values or modalities x_i, in a statistical series of size n, may exhibit a high degree of proximity to one another. Rather than studying them separately, they are grouped into k classes, with extremities $c_1, c_2, \ldots, c_k$. For each class, the corresponding frequency is noted in the interval $[c_i, c_{i+1}]$. The following steps must be followed in order to determine the relevant classes:

- The range of the statistical series is defined by : $R = x_{max} - x_{min}$.
- The number of classes is the integer k, that is the closest to $\sqrt{n}$.
- The amplitude a of each class is such that $a \geq \dfrac{R}{k}$.

Example 1

Let's determine the classes corresponding to a series of 40 discrete values, with $x_{min} = 5.3$ and $x_{max} = 32.6$.

- The range of this statistical series is: $R = 32.6 - 5.3 = 27.3$.
- The number of classes is the integer k closest to $\sqrt{40} = 6.32$, i.e. $k = 6$.
- The amplitude of each class is : $a \geq \dfrac{27.3}{6} = 4.55 \implies a = 5$.

Thus, the 6 classes are $[5.3, 10.3[$, $[10.3, 15.3[$, $[15.3, 20.3[$, $[20.3, 25.3[$, $[25.3, 30.3[$, $[30.3, 35.3[$.

Remark 1

In certain cases, the initial class (and its amplitude) is predetermined. In this particular instance, the number of classes and the classes themselves can be readily deduced.

In the preceding example, if the first class is designated as $[5, 10.5[$ (with an amplitude $a = 5.5$), the remaining classes are as follows: $[10.5, 16[$, $[16, 21.5[$, $[21.5, 27[$, $[27, 32.5[$, $[32.5, 38[$.

4 Representation of univariate statistical series

Once the data has been collected through consistent observation, it is subsequently arranged, summarised, and classified in order to represent it in tabular form through statistical tables and in visual form through graphs.

4.1 Statistical tables

The frequency (or relative frequency) table is a synthetic representation of data, presented in tabular form. This representation is employed for either a discrete or continuous statistical variable.

Example 2

For the same previous sample of 30 students from a population of 150 students in a particular section (see *Chapter 1, Example 2*), the detailed values related to the character X, defined as "the mark obtained by each student at an examination" are given in the following series:

$$12; 8; 12; 5; 12; 12; 8; 5; 12; 8; 15; 12; 8; 5; 12; 8; 5; 8; 12; 8; 15; 8; 5; 8; 15; 8; 5; 8; 5; 8$$

The following statistical table provides a summary of the series, presenting the frequencies of the four modalities $(i = \overline{1, 4})$:

x_i	5	8	12	15	Total
n_i	7	12	8	3	30

where the sum of the frequencies is checked to ascertain whether it is equal to the sample size:

$$\sum_{i=1}^{i=4} n_i = 7 + 12 + 8 + 3 = 30.$$

From equations (2.2) and (2.4), we can complete the above table by the corresponding relative frequencies and cumulative relative frequencies:

x_i	5	8	12	15	Total
n_i	7	12	8	3	30
f_i	$\frac{7}{30} = 0.233$	$\frac{12}{30} = 0.400$	$\frac{8}{30} = 0.267$	$\frac{3}{30} = 0.100$	1
F_i	$f_1 = 0.233$	$f_1 + f_2 = 0.633$	$f_1 + f_2 + f_3 = 0.900$	$f_1 + f_2 + f_3 + f_4 = 1$	/

Example 3

For the same previous sample of 30 students from a population of 150 students in a particular section (see *Chapter 1, Example 2*), the detailed values related to the character X, defined as "the size in cm of each student" are a continuous statistical series given in the following statistical table:

Size in cm	[159, 164[	[164, 169[	[169, 174[	[174, 179[	Total
n_i	9	6	12	3	30

From equations (2.2) and (2.4), we can complete the above table by the corresponding relative frequencies and cumulative relative frequencies:

x_i	[159, 164[	[164, 169[	[169, 174[	[174, 179[	Total
n_i	9	6	12	3	30
f_i	0.3	0.2	0.4	0.1	1
F_i	0.3	0.5	0.9	1	/

4.2 Graphic representation of data

The graphical representation of the obtained (qualitative or quantitative) data statistics is achieved through the use of various symbols, including lines on a line graph, bars on a bar chart, or slices of a pie chart. This representation is more efficient in facilitating comprehension and comparison of data than its tabular counterpart. The utilisation of graphical representations facilitates the organisation, classification and presentation of data in a manner that is readily comprehensible to a broader audience. This visual representation has been found to facilitate clarity, comparison and understanding of numerical data.

4.2.1 Graphic representation of qualitative statistical series

In the context of qualitative statistical series, the most prevalent graphic representations of the frequency (or relative frequency) tables are the bar graph (illustrated in ***bar charts***) and the circular graph (illustrated in ***pie charts***).

- The graph representation employed in the bar chart method assigns a rectangular shape with a fixed base and a height proportional to the frequency (or relative frequency) of the modality of the qualitative character under consideration.
- The graph representation in the pie chart is such that each modality of the character is associated with an angular sector of a disk. The central angles of these sectors are proportional to the frequency (or relative frequency) of each modality. In this representation, a circle with a radius of unity is constructed. The frequencies are then represented by the areas of the sectors corresponding to the modalities of the angular variable:

$$\alpha_i = 2\pi \cdot \frac{n_i}{n},$$

with:

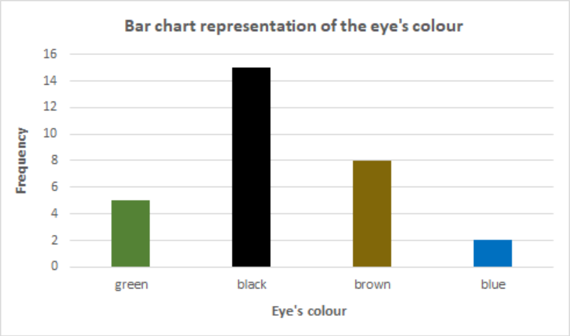
$$\sum_{i=1}^{i=k} \alpha_i = 2\pi.$$

Example 4

For the same previous sample of 30 students from a population of 150 students in a particular section (see *Chapter 1, Example 2*), the statistical series related to the character X, defined as“the eye’s colour for each student” are summarised in the following table:

eye’s colour	green	black	brown	blue	Total
n_i	5	15	8	2	30

As illustrated by the bar charts above, the frequency (we can also do it for the relative frequency) is indicated on the vertical axis, with the modality represented on the horizontal axis.



It is imperative to consider the following principles:

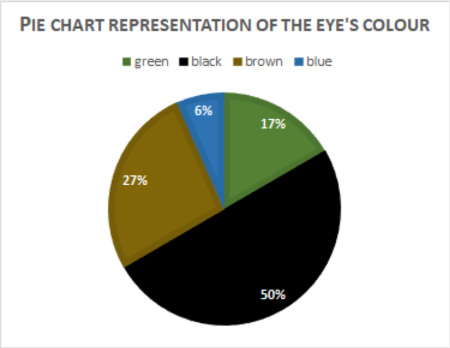
- Equal spacing must be be maintained on each axis.
- The bars should be of uniform width, and there is no contact between them.
- Each axis must be accompanied by a label, in addition to the designation of the graph by a title.

Example 5

For the same previous sample of 30 students from a population of 150 students in a particular section (see *Chapter 1, Example 2*), the statistical series related to the character X, defined as“the eye’s colour for each student” are summarised in the following table:

eye’s colour	green	black	brown	blue	Total
n_i	5	15	8	2	30
α_i	$2\pi \cdot \frac{5}{30} = \frac{\pi}{3}$	$2\pi \cdot \frac{15}{30} = \pi$	$2\pi \cdot \frac{8}{30} = \frac{8\pi}{15}$	$2\pi \cdot \frac{2}{30} = \frac{2\pi}{15}$	2π

As illustrated by the pie charts above, the frequency (we can also do it for the relative frequency) is indicated on sectors for which the the areas correspond to the angular values associated with the different modalities.



It is imperative to consider the following principles:

- Each sector should be designed by the value (or corresponding percentage) of each modality.
- The legend for each modality, in addition to the title of the graph, should be clearly visible.

Remark 2

Pie charts have been shown to be a useful graphical tool for the purpose of comparing the size of modalities. Bar charts offer a parallel representation of the data. The specific type of device used is of little consequence. The determination of the most appropriate approach is ultimately a subjective matter, contingent on individual preferences and the specific information to be addressed. However, pie charts are most efficacious when the number of categories is minimal and the data can be expressed as a percentage.

4.2.2 Graphic representation of quantitative statistical series

In the context of quantitative statistical series, the most common graphic representations of the frequency (or relative frequency) tables, depending on the character nature, are the stick diagram (illustrated in ***stick charts***) for the discrete character, and the ***histogram*** for the continuous character.

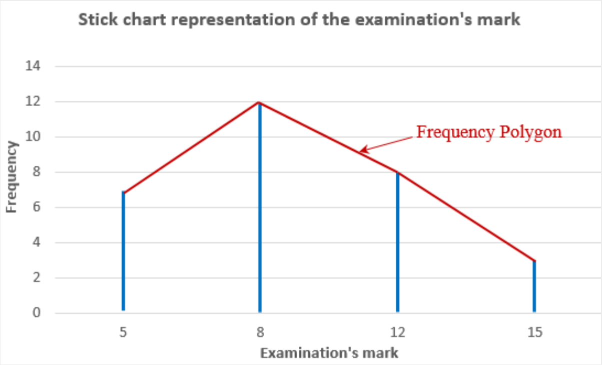
- The stick diagram representation assigns a segment (stick) with a height that is proportional to the frequency (or relative frequency) of the modality of the discrete quantitative character under consideration. It is important to note that the modalities are ordered in ascending order.
- The histogram assigns adjacent rectangles, with each one rectangle's area being proportional to the frequency (or relative frequency) of the class of the quantitative character under consideration. It is important to note that, in instance where the classes possess equal amplitudes, the height of each class is directly proportional to the frequency (or relative frequency) of that class.

Example 6

The statistical table from *Example 2,* concerning "the mark obtained by each student at an examination" on a sample of 30 students out of a total population of 150 students in a specific section, shall be considered.

x_i	5	8	12	15	Total
n_i	7	12	8	3	30

As illustrated by the stick charts above, the frequency (we can also do it for the relative frequency) is indicated on the vertical axis, with the modality represented on the horizontal axis.



The frequency (or relative frequency) polygon is a type of line graph used to represent the frequency (or relative frequency) distribution. It can be constructed by the process of joining the extremities of the sticks with linear segments.

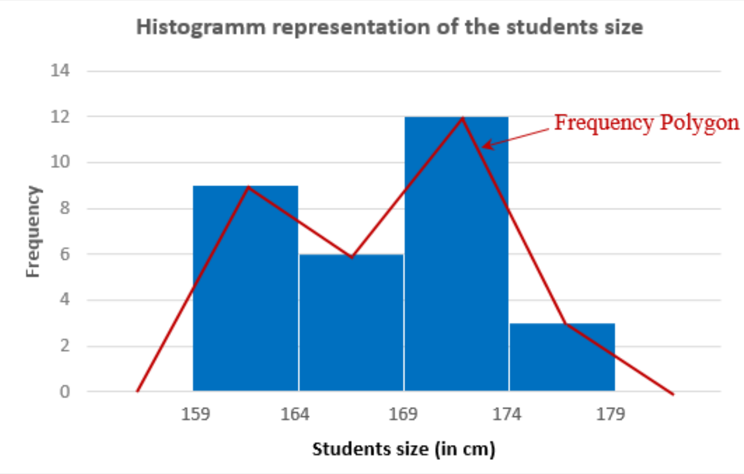
As for the bar chart representation, each axis must be accompanied by a label, in addition to the designation of the graph by a title.

Example 7

The statistical table from *Example 3,* concerning "the size in cm of each student", taken from a sample of 30 students out of a total population of 150 students in a particular section, can be considered:

Size in cm	$[159, 164[$	$[164, 169[$	$[169, 174[$	$[174, 179[$	Total
n_i	9	6	12	3	30

As illustrated by the histogram (with the frequency polygon), the frequency (we can also do it for the relative frequency) is indicated by vertical rectangles of equal basis, with the extremities of the classes represented on the horizontal axis.



As for the previous representations, each axis must be accompanied by a label, in addition to the designation of the graph by a title.

Remark 3

In the event of classes with unequal amplitudes $r_1, r_2, \ldots, r_i, \ldots, r_k$, the heights of the vertical rectangles are to be regarded as the corrected frequencies (or relatives frequencies) per amplitude, defined as follows:

$$n_{i_c}\,(\text{or} f_{i_c}) = n_i \frac{\min\left(r_1, r_2, \ldots, r_k\right)}{r_i}. \tag{2.5}$$

4.3 Cumulative distribution function - Cumulative curves

Cumulative relative frequency curves are a type of statistical curve that is employed to illustrate the cumulative distribution function, defined as follows:

Definition 5

For any statistical variable X, the cumulative distribution function, denoted F, is defined as follows:

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

- In the event of a discrete quantitative statistical variable with its modalities $x_1, x_2, \ldots x_i, \ldots, x_k$:

$$F(x) = \begin{cases} 0 & ; x < x_1 \\ \sum_{j=1}^{j=i-1} f_j & ; x \in [x_{i-1}, x_i[\text{ with } i \geq 2 \\ 1 & ; x \geq x_k. \end{cases} \tag{2.6}$$

- In the event of a continuous quantitative statistical variable with its classes $[c_1, c_2[\,, [c_2, c_3[\,, \ldots, [c_i, c_{i+1}[\,, \ldots, [c_k, c_{k+1}]$:

$$F(x) = \begin{cases} 0 & ; x \leq c_1 \\ \sum_{j=1}^{j=i-1} f_j & ; x = c_i \text{ with } i \geq 2 \\ 1 & ; x \geq x_{k+1}. \end{cases} \tag{2.7}$$

The cumulative distribution function is characterised by the following properties:

- $0 \leq F(x) \leq 1$.
- F is increasing.
- F is right-continuous.
- $\lim_{x \to +\infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$.

Example 8

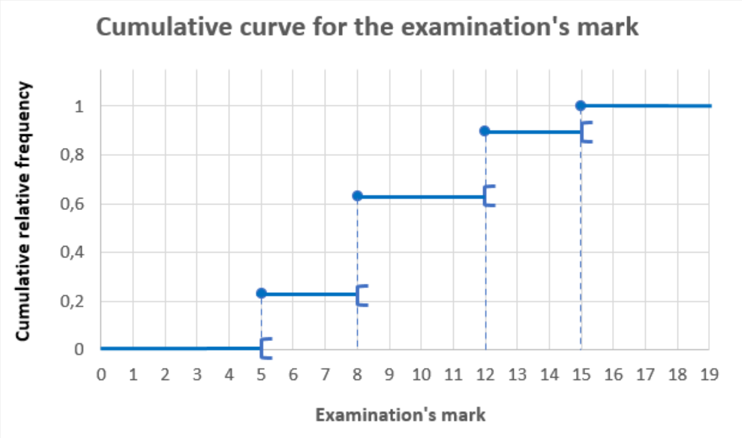
1- In the discrete case, we turn our attention to *Example 2,* which deals with "the mark obtained by each student at an examination" on a sample of 30 students from a total population of 150 students in a specific section. The statistical table, which presents the relative and cumulative relative frequencies, is presented below.

x_i	5	8	12	15	Total
n_i	7	12	8	3	30
f_i	$\frac{7}{30}=0.233$	$\frac{12}{30}=0.400$	$\frac{8}{30}=0.267$	$\frac{3}{30}=0.100$	1
F_i	$f_1=0.233$	$f_1+f_2=0.633$	$f_1+f_2+f_3=0.900$	$f_1+f_2+f_3+f_4=1$	/

The cumulative distribution function can be written as follows:

$$F\left(x\right)=\begin{cases}0 & ;x<5\\ 0.233 & ;x\in[5,8[\\ 0.633 & ;x\in[8,12[\\ 0.900 & ;x\in[12,15[\\ 1 & ;x\geq 15.\end{cases}$$

The corresponding cumulative curve, illustrated below, is invariably a staircase curve.



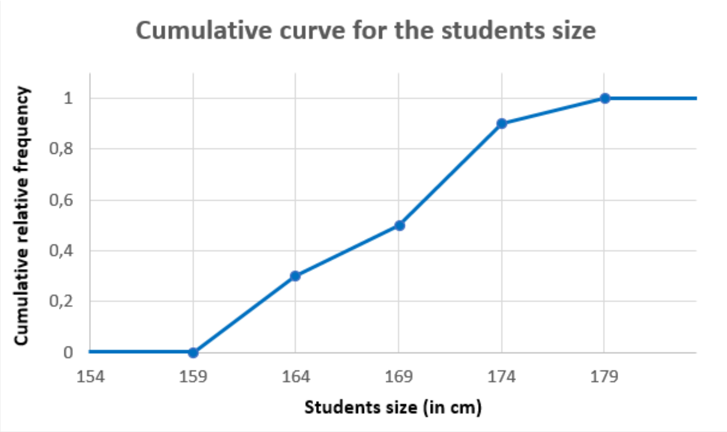
In the continuous case, let us consider the *Example 3,* which concerns "the size in cm of each student" from a sample of 30 students from a population of 150 students in a particular section. The statistical table with the relative and cumulative relative frequencies, is presented below.

x_i	$[159,164[$	$[164,169[$	$[169,174[$	$[174,179[$	Total
n_i	9	6	12	3	30
f_i	0.3	0.2	0.4	0.1	1
F_i	0.3	0.5	0.9	1	/

The cumulative distribution function can be written as follows:

$$F\left(x\right)=\begin{cases}0 & ;x\leq 159\\ 0.3 & ;x=164\\ 0.5 & ;x=169\\ 0.9 & ;x=174\\ 1 & ;x\geq 179.\end{cases}$$

The corresponding cumulative curve, illustrated below, is a continuous curve.



5 Numerical description : Position and dispersion parameters

The process of analysing a table or its graphical representation can often be a time-consuming one, and may not provide a clear representation of the observed statistical distribution. The numerical description of the statistical series consists of measurable quantities and characteristic parameters, such as the position and dispersion parameters, which are employed to provide an objective characterisation of the entire dataset under study.

5.1 Position parameters

The Position parameters are measures that estimate the central tendency of a population.These values are indicative of the typical value around which the observations are distributed.

5.1.1 Arithmetic mean

The ***arithmetic mean*** is the most common measure of central tendency, indicating an average of a given data collection. The following definition is provided for this position parameter.

Definition 6

For a statistical variable X, the *arithmetic mean* of a dataset of n values, denoted $\overline{X}$, is defined as follows:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{i=n} x_i. \tag{2.8}$$

- In the case of a discrete quantitative statistical variable X, with its k modalities $x_1, x_2, \dots x_i, \cdots, x_k$, the arithmetic mean is determined by the following expression:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{i=k} n_i x_i = \sum_{i=1}^{i=k} f_i x_i. \tag{2.9}$$

- In the case of a continuous quantitative statistical variable X, with its k classes $[c_1, c_2[\,, [c_2, c_3[\,, \dots, [c_i, c_{i+1}[\,, \dots, [c_k, c_{k+1}]$, the arithmetic mean is determined by the following expression:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{i=k} n_i C_i = \sum_{i=1}^{i=k} f_i C_i, \tag{2.10}$$

  where C_i is the centre of the i^{th}class.

5.1.2 Mode

Another measure of central tendency is the ***mode***, which identifies the most common value or class in a distribution.

Definition 7

- In the case of a discrete quantitative statistical variable X, with its k modalities $x_1, x_2, \dots x_i, \cdots, x_k$, the *mode*, denoted Mo, is defined as the modality with the highest frequency (or the highest relative frequency).

- For a continuous quantitative statistical variable X, with its k classes $[c_1, c_2[, [c_2, c_3[, \dots, [c_i, c_{i+1}[, \dots, [c_k, c_{k+1}]$, we first determine the modal class, denoted C_{Mo} as the class with the highest frequency (or the highest relative frequency). We then calculate the mode Mo by linear interpolation.

$$Mo = B_{low}(\mathrm{C}_{Mo}) + a(\mathrm{C}_{Mo}) \frac{\Delta_p}{\varDelta_p + \varDelta_n}, \tag{2.11}$$

where:
$B_{low}(\mathrm{C}_{Mo})$ is the lower boundary of the modal class.
$a(\mathrm{C}_{Mo})$ is the amplitude of the modal class,
Δ_p is the difference between the frequency (or relative frequency) of the modal class and that of the preceding class.
Δ_n is the difference between the frequency (or relative frequency) of the modal class and that of the next class.

Remarks 4

- It is evident that in the event of all possible values x_i (or classes in the continuous case) appearing with equal frequencies in the data set, there is no mode.

- There are several modes when the highest frequency has been observed for several different modalities.

- In instances where a mode is absent or multiple modes are present, the use of a mode to ascertain the centre of the distribution is not possible.

- In the case of a continuous statistic variable, where the amplitudes of the classes are different, the modal class is determined according to the corrected frequencies (see Equation 2.5). The mode is calculated based on the difference between the corrected frequency of the modal class and the corrected frequencies of the preceding and succeeding classes.

5.1.3 Median

The ***median*** is a more robust indicator of central tendency than the arithmetic mean, as it is less sensitive to extreme values. This is particularly beneficial in cases where the presence of extreme values has the potential to compromise the integrity of the arithmetic mean.

Definition 8

The median, denoted Me, is defined as the value of the statistical variable that divides the observations into two equal parts.
This value represents the modality for which the cumulative distribution function $F(Me) = 0.5$.

- In the case of a discrete quantitative variable X, The n values in a statistical series are arranged in either ascending or descending order, and the median is defined as the value that is positioned precisely at the midpoint of the series:

$$Me = \begin{cases} x_{k+1} & ;\text{if } n = 2k+1 \\ \frac{x_k + x_{k+1}}{2} & ;\text{if } n = 2k. \end{cases} \tag{2.12}$$

- For a continuous quantitative statistical variable X, with its k classes $[c_1, c_2[\,, [c_2, c_3[\,, \ldots, [c_i, c_{i+1}[\,, \ldots, [c_k, c_{k+1}]$, we first determine the median class $[c_i, c_{i+1}[$, such that:

$$F(c_i) \leq F(Me) = 0.5 \leq F(c_{i+1}). \tag{2.13}$$

  Then $Me \in [c_i, c_{i+1}[$ and is calculated by linear interpolation:

$$Me = c_i + (c_{i+1} - c_i)\frac{0.5 - F(c_i)}{F(c_{i+1}) - F(c_i)}. \tag{2.14}$$

5.1.4 Quantiles

Further indicators of central tendency include ***quantiles***, which may be defined as a generalisation of the median.

Definition 9

The *quartile*, denoted q_α, with $\alpha \in \,]0,1[$, is defined as the value of the statistical variable that divides the observations into $\frac{1}{\alpha}$ equal parts.
This value represents the modality for which the cumulative distribution function $F(q_\alpha) = \alpha$.

- In the case of a discrete quantitative statistical variable X, The n values in a statistical series are arranged in either ascending or descending order, and the quantile is defined as the value :

$$q_\alpha = \begin{cases} \frac{x_{n\alpha} + x_{n\alpha+1}}{2} & ;\text{if } n\alpha \in \mathbb{N} \\ x_{\lfloor n\alpha \rfloor + 1} & ;\text{if } n\alpha \notin \mathbb{N}, \end{cases} \tag{2.15}$$

  with $\lfloor n\alpha \rfloor$ denote the integer part of $n\alpha$.

- For a continuous quantitative statistical variable X, with its k classes $[c_1, c_2[\,, [c_2, c_3[\,, \ldots, [c_i, c_{i+1}[\,, \ldots, [c_k, c_{k+1}]$, we first determine the class $[c_i, c_{i+1}[$, such that:

$$F(c_i) \leq F(q_\alpha) \leq F(c_{i+1}). \tag{2.16}$$

  Then $q_\alpha \in [c_i, c_{i+1}[$ and is calculated by linear interpolation:

$$q_\alpha = c_i + (c_{i+1} - c_i)\frac{\alpha - F(c_i)}{F(c_{i+1}) - F(c_i)}. \tag{2.17}$$

Remarks 5

- For $\alpha = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$, the quantiles are designated as *quartiles*, denoted $Q_1 = q_{\frac{1}{4}}$ (*First quartile*), $Q_2 = q_{\frac{2}{4}}$ (*Second quartile*) and $Q_3 = q_{\frac{3}{4}}$ (*Third quartile*), respectively.
- For $\alpha = \frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10}$, the quantiles are designated as *deciles*, denoted $D_1 = q_{\frac{1}{10}}$ (*First decile*), $D_2 = q_{\frac{2}{10}}$ (*Second decile*), …,$D_9 = q_{\frac{9}{10}}$ (*Ninth decile*), respectively.
- For $\alpha = \frac{1}{100}, \frac{2}{100}, \ldots, \frac{99}{100}$, the quantiles are designated as *percentiles*, denoted $P_1 = q_{\frac{1}{100}}$ (*First percentile*), $P_2 = q_{\frac{2}{100}}$ (*Second percentile*), …,$P_{99} = q_{\frac{99}{100}}$ (*Ninety-ninth percentile*), respectively.
- It is evident that the second quartile, the fifth decile and the fiftieth percentile are equivalent to the median:

$$Q_2 = D_5 = P_{50} = Me.$$

Example 9

The number of employees who visited the health centre was recorded by the company over a period of 50 days. The statistical table associated with this study, given below, has been completed with $n_i \cdot x_i$ values to facilitate its subsequent exploitation.

x_i	0	1	2	3	4	5	Total
n_i	3	3	9	10	21	4	50
f_i	0.06	0.06	0.18	0.2	0.42	0.08	1
N_i	3	6	15	25	46	50	/
$n_i \cdot x_i$	0	3	18	30	84	20	155

- The arithmetic mean of X is determined as follows:

$$\overline{X} = \frac{1}{50} \sum_{i=1}^{i=6} n_i x_i = \frac{155}{50} = 3.1.$$

- The mode is determined by the modality which exhibits the highest frequency:

$$Mo = 4.$$

- As n is even ($n = 50 = 2 \cdot 25$), the median is:

$$Me = \frac{x_{25} + x_{26}}{2} = \frac{3+4}{2} = 3.5.$$

- For the first quartile $Q_1 = q_{\frac{1}{4}}$, we have $n\alpha = 50 \cdot \frac{1}{4} = 12.5$. As $n\alpha \notin \mathbb{N}$:

$$Q_1 = x_{\lfloor 12.5 \rfloor + 1} = x_{13} = 2.$$

- For the third quartile $Q_3 = q_{\frac{3}{4}}$, we have $n\alpha = 50 \cdot \frac{3}{4} = 37.5$. As $n\alpha \notin \mathbb{N}$:

$$Q_3 = x_{\lfloor 37.5 \rfloor + 1} = x_{38} = 4.$$

Example 10

The following statistical table, completed with $n_i \cdot c_i$ values to facilitate its subsequent exploitation, was yielded by a study of the distribution of 30 workers in a company according to their weekly salary:

weekly salary ($\times 1000\,\text{DA}$)	$[10, 12[$	$[12, 14[$	$[14, 16[$	$[16, 18[$	Total
n_i	4	6	15	5	30
f_i	0.13	0.20	0.50	0.17	1
F_i	0.13	0.33	0.83	1	/
c_i	11	13	15	17	/
$n_i \cdot c_i$	44	87	225	85	432

- The arithmetic mean of X is determined as follows:

$$\overline{X} = \frac{1}{30} \sum_{i=1}^{i=6} n_i c_i = \frac{432}{30} = 14.4 \times 1000\,\text{DA}.$$

- The modal class, which is the class with the highest frequency, is $\text{C}_{Mo} = [14, 16[$. The differences between the frequency of C_{Mo} and that of the preceding and next classes are $\Delta_p = 15 - 6 = 9$, and $\Delta_n = 15 - 5 = 10$. Thus we calculate the mode Mo by linear interpolation:

$$Mo = 14 + (16 - 14)\,\frac{9}{9 + 10} = 14.95 \times 1000\,\text{DA}.$$

- the median class is $[14, 16[$ because $F(14) = 0.33 \leq F(Me) = 0.5 \leq F(16) = 0.83$. Then the median $Me \in [14, 16[$ and is calculated by linear interpolation:

$$Me = 14 + (16 - 14)\frac{0.5 - 0.33}{0.83 - 0.33} = 14.95 \times 1000\,\text{DA}.$$

- The first quartile $Q_1 \in [12, 14[$because $F(12) = 0.13 \leq F(Q_1) = 0.25 \leq F(14) = 0.33$. Then the first quartile is calculated by linear interpolation:

$$Q_1 = 12 + (14 - 12)\frac{0.25 - 0.13}{0.33 - 0.13} = 13.20 \times 1000\,\text{DA}.$$

  The third quartile $Q_3 \in [14, 16[$because $F(14) = 0.33 \leq F(Q_1) = 0.75 \leq F(14) = 0.83$. Then the third quartile is calculated by linear interpolation:

$$Q_3 = 14 + (16 - 14)\frac{0.75 - 0.33}{0.83 - 0.33} = 15.68 \times 1000\,\text{DA}.$$

5.2 Dispersion parameters

The dispersion parameters are measures that quantify the extend and distribution of variation within a dataset. These parameters provide information regarding the dispersion of individual values from the central tendency. In summary, the position parameters provide information regarding the centre of the data, while the dispersion parameters indicate the data spread around this centre.

5.2.1 Range

Previously used in section 3, the ***range*** represents the most elementary dispersion parameter for quantifying the distribution of data.

Definition 10

The *range*, denoted R, indicates the distance between the highest value x_{max} and the the lowest value x_{min} in a sample:

$$R = x_{max} - x_{min}. \tag{2.18}$$

5.2.2 Variance and standard deviation

The concepts of ***variance*** and the ***standard deviation*** are both used to quantify the dispersion observed in a given dataset. However, these two measures are expressed in different units. The prevailing preference for standard deviation is attributable to the fact that it is expressed in the same units as the original data, thereby facilitating interpretation.

Definition 11

- For a statistical variable X, the *variance* of a dataset of n values, denoted σ_X^2 or σ^2, is defined as the arithmetic mean of the squares of the differences between each value and in the dataset and the mean:

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{i=n}\left(x_i - \overline{X}\right)^2. \tag{2.19}$$

- In the event of a discrete quantitative statistical variable X, with its k modalities $x_1, x_2, \ldots x_i, \cdots, x_k$, the variance is expressed as follows:

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{i=k} n_i\left(x_i - \overline{X}\right)^2 = \sum_{i=1}^{i=k} f_i\left(x_i - \overline{X}\right)^2. \tag{2.20}$$

- In the case of a continuous quantitative statistical variable X, with its k classes $[c_1, c_2[\,, [c_2, c_3[\,, \ldots, [c_i, c_{i+1}[\,, \ldots, [c_k, c_{k+1}]$, the variance is given by:

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{i=k} n_i\left(C_i - \overline{X}\right)^2 = \sum_{i=1}^{i=k} f_i\left(C_i - \overline{X}\right)^2, \tag{2.21}$$

  where C_i is the centre of the i^{th}class.

- The *standard deviation*, denoted σ_X or σ, is defined as the square root of the variance:

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{i=n}\left(x_i - \overline{X}\right)^2.} \tag{2.22}$$

Property

We can demonstrate an alternative formulation of the variance as follows:

$$\begin{aligned}
\sigma_X^2 &= \frac{1}{n}\sum_{i=1}^{i=k} n_i\left(x_i-\overline{X}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{i=k}\left(n_i x_i^2+n_i\overline{X}^2-2n_i x_i\overline{X}\right) \\
&= \frac{1}{n}\sum_{i=1}^{i=k} n_i x_i^2+\overline{X}^2\left(\frac{1}{n}\sum_{i=1}^{i=k} n_i\right)-2\overline{X}\left(\frac{1}{n}\sum_{i=1}^{i=k} n_i x_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{i=k} n_i x_i^2+\overline{X}^2-2\overline{X}^2,
\end{aligned}$$

and finally:

$$\sigma_X^2=\frac{1}{n}\sum_{i=1}^{i=k} n_i x_i^2-\overline{X}^2. \tag{2.23}$$

Interquartile range and interval

The ***interquartile range*** and the ***interquartile interval*** are two related parameters that are frequently employed to measure the dispersion of data around the median. These tools facilitate comprehension of the data distribution and enable the identification of outliers.

Definition 12

- The *interquartile range,* denoted IQR, is defined as the difference between the last quartile Q_3 and the first quartile Q_1:
$$IQR=Q_3-Q_1. \tag{2.24}$$
- The *interquartile interval,* denoted IQI, refers to the interval that is delimited by the two quartiles Q_1 and Q_3:
$$IQI=\left[Q_1,Q_3\right]. \tag{2.25}$$

Remark 6

Interdecile and interpercentile ranges and intervals, IDR, IPR, IDI, and IPI respectively, can be defined in a similar manner:

$$\begin{aligned}
IDR &= D_9-D_1 \\
IDI &= [D_1,D_9] \\
IPR &= P_{99}-P_1 \\
IPI &= [P_1,P_{99}]
\end{aligned}$$

5.2.3 Coefficient of variation

The ***coefficient of variation*** is a quantitative measure employed to quantify the dispersion of data points around the arithmetic mean. This dispersion is typically expressed as a percentage, allowing for quantitative analysis and comparison between data series. It is evident that as the value of the coefficient of variation diminishes, the homogeneity of the series increases. It is also evident that as the value increases, the dispersion around the mean value also increases in proportion.

Definition 13

The *coefficient of variation*, denoted CV_X, is defined as the ratio of the standard deviation to the arithmetic mean:

$$CV_X = \frac{\sigma_X}{\overline{X}}. \tag{2.26}$$

Example 11

Let's take a look at the statistical table from *Example 9*, where we've added the$n_i \cdot x_i^2$ values.

x_i	0	1	2	3	4	5	Total
n_i	3	3	9	10	21	4	50
f_i	0.06	0.06	0.18	0.2	0.42	0.08	1
N_i	3	6	15	25	46	50	/
$n_i \cdot x_i$	0	3	18	30	84	20	155
$n_i \cdot x_i^2$	0	3	36	90	336	100	565

- The range of X is determined as follows:

$$R = x_{max} - x_{min} = 5 - 0 = 5.$$

- We calculate the variance of X as follows:

$$\sigma_X^2 = \frac{1}{50}\sum_{i=1}^{i=6} n_i x_i^2 - \overline{X}^2 = \frac{565}{50} - (3.1)^2 = 1.69.$$

  Thus, we deduce its standard deviation :

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{1.69} = 1.3.$$

- The interquartile range is:

$$IQR = Q_3 - Q_1 = 4 - 2 = 2,$$

  and the interquartile interval is:

$$IQI = [Q_1, Q_3] = [2, 4].$$

- The value of the coefficient of variation is then:

$$CV_X = \frac{\sigma_X}{\overline{X}} = \frac{1.3}{3.1} = 0.42.$$

Example 12

Now, let's take a look at the statistical table of *Example 10*, which is filled out with $n_i \cdot c_i^2$ values.

Weekly salary (×1000 DA)	$[10, 12[$	$[12, 14[$	$[14, 16[$	$[16, 18[$	Total
n_i	4	6	15	5	30
f_i	0.13	0.20	0.50	0.17	1
F_i	0.13	0.33	0.83	1	/
c_i	11	13	15	17	/
$n_i \cdot c_i$	44	78	225	85	432
$n_i \cdot c_i^2$	484	1044	3375	1445	66318

- The range of X is determined as follows:

$$R = x_{max} - x_{min} = 18 - 10 = 8 \times 1000\,\text{DA}.$$

- The variance and the standard deviation of X are calculated as follows

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{i=4} n_i c_i^2 - \overline{X}^2 = \frac{66318}{30} - (14.4)^2 = 3.24 \times (1000\,\text{DA})^2,$$

and

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{3.24} = 1.8 \times 1000\,\text{DA}.$$

- The interquartile range and interquartile interval are:

$$IQR = Q_3 - Q_1 = 15.68 - 13.20 = 2.48 \times 1000\,\text{DA},$$

and

$$IQI = [Q_1, Q_3] = [13\,200, 156\,800]\,.$$

- We can also determine the coefficient of variation as:

$$CV_X = \frac{\sigma_X}{\overline{X}} = \frac{1.8}{14.4} = 0.125.$$

CHAPTER 3

BIVARIATE STATISTICAL SERIES

1 Introduction

A discussion of the methods for summarising and representing univariate statistical series was presented in the preceding chapter. However, there are cases where the analysis of an individual can be carried out on more than a single statistical variable. For instance, a statistical study might undertake the monitoring of a child's growth by observing both its height and weight. In this particular instance, it is imperative to examine the interrelationship between these two variables.

The present chapter will be dedicated to the analysis of quantitative bivariate statistical series. The utilisation of these descriptive statistical methods is a common practice in research and quality of life studies, with the objective of ascertaining any potential correlation, association or dependency between two variables.

In this context, the two statistical variables X and Y can be defined as an application:

$$\begin{aligned} (X,Y): \quad \Omega &\to \quad \mathbb{R}^2 \\ \omega_i &\to \quad (X(\omega_i), Y(\omega_i)), \end{aligned}$$

For each individual ω_i, we associate the pair of values (x_i, y_i) such that: $x_i = X(\omega_i)$, and $y_i = Y(\omega_i)$.

(X, Y) is the ***statistical variable pair***, with X and Y having the capacity to be either discrete or continuous. For instance, in a statistical study concerning the number of annual absences in a sample of students, based on their obtained annual average marks, the number of absences is discrete while the annual final mark (any value within a range) is continuous.

2 Representation of bivariate statistical series

The bivariate discrete statistical series of values $x_1, x_2, \dots x_i, \dots, x_n$, and $y_1, y_2, \dots, y_i, \dots, y_n$ can be represented in a classical table as follows:

ω	ω_1	ω_2	$\dots$	ω_i	$\dots$	ω_n
X	x_1	x_2	$\dots$	x_i	$\dots$	x_n
Y	y_1	y_2	$\dots$	y_i	$\dots$	y_n

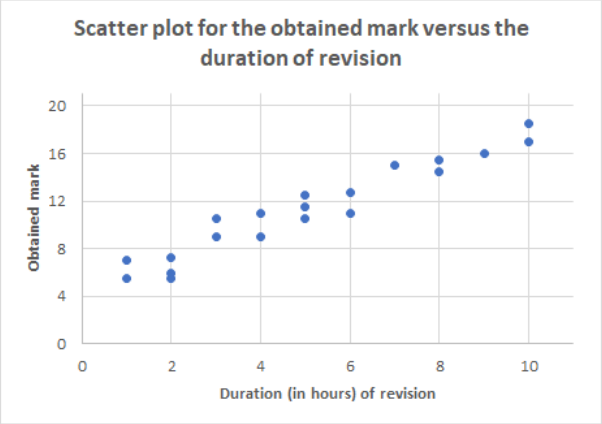
The graphic representation of such a statistical table is known as the ***scatter plot***, which is achieved by plotting each pair (x_i, y_i) as a dot on a Cartesian coordinate system. The position of each dot on the horizontal and vertical axis is indicative of the indicates x_i and y_i values, respectively.

Example 1

The following table presents the findings of a statistical investigation into the relationship between the duration of revision (in hours) x_i and the marks y_i obtained mark in an examination for a sample of 20 students:

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X	1	2	5	3	2	4	5	3	6	1	7	8	4	9	6	10	2	5	8	10
Y	5.5	7.25	12.5	10.5	6	9	11.5	9	12.75	7	15	15.5	11	16	11	18.5	5.5	10.5	14.5	17

The graphic representation of this statistical table is given by the scatter plot below.



2.1 Joint distribution - Contingency table

As for univariate statistical series, the representation of the statistical variable pair (X, Y) can be achieved by taking into account the occurrence of each of the different modalities $x_1, x_2, \dots x_i, \dots, x_k$, and $y_1, y_2, \dots, y_j, \dots, y_p$ of X and Y respectively. It is therefore imperative to establish clear definitions of key concepts including the ***joint frequency***, and the ***relative joint frequency***.

Definition 1

The *joint frequency* of each modalities pair (x_i, y_j), denoted by n_{ij}, is defined as the number of times that the modalities x_i and y_j occur together in a dataset, for $i = \overline{1,k}$, and $j = \overline{1,p}$.
We write:

$$\sum_{i=1}^{i=k}\sum_{j=1}^{j=p} n_{ij} = n, \tag{3.1}$$

where n denotes the sample size.

Definition 2

The *relative joint frequency* of each modalities pair (x_i, y_j), denoted by f_{ij}, is defined as the quotient of the frequency of (x_i, y_j) to the sample size.
We write:

$$f_{ij} = \frac{n_{ij}}{n} \quad ; i = \overline{1,k} \text{ and } j = \overline{1,p}. \tag{3.2}$$

We note that $\sum_{i=1}^{i=k}\sum_{j=1}^{j=p} f_{ij} = 1$.

Definition 3

The *joint distribution* of the statistical variable pair (X, Y) is defined by providing all the modalities pairs (x_i, y_j) and the corresponding joint frequencies (or relative frequencies).

Th joint distribution associated to the statistical variable pair (X, Y) (where X and Y are either discrete or continuous) can thus be represented within a ***contingency table*** (or ***double-entry table***), which is a table of k rows and p columns. The table is presented in matrix format, thereby offering a visual representation of the multivariate frequency distribution of the statistical variables.

$X \backslash Y$	y_1	...	y_j	...	y_p	Totals
x_1	n_{11}(or f_{11})	...	n_{1j} (or f_{1j})	...	n_{1p} (or f_{1p})	$\sum_{j=1}^{j=p} n_{1j}$ (or $\sum_{j=1}^{j=p} f_{1j}$)
x_2	n_{21} (or f_{21})	...	n_{2j} (or f_{2j})	...	n_{2p} (or f_{2p})	$\sum_{j=1}^{j=p} n_{2j}$ (or $\sum_{j=1}^{j=p} f_{2j}$)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1} (or f_{i1})	...	n_{ij} (or f_{ij})	...	n_{ip} (or f_{ip})	$\sum_{j=1}^{j=p} n_{ij}$ (or $\sum_{j=1}^{j=p} f_{ij}$)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1} (or f_{k1})	...	n_{kj} (or f_{kj})	...	n_{kp} (or f_{kp})	$\sum_{j=1}^{j=p} n_{kj}$ (or $\sum_{j=1}^{j=p} f_{kj}$)
Totals	$\sum_{i=1}^{i=k} n_{i1}$ (or $\sum_{i=1}^{i=k} f_{i1}$)	...	$\sum_{i=1}^{i=k} n_{ij}$ (or $\sum_{i=1}^{i=k} f_{ij}$)	...	$\sum_{i=1}^{i=k} n_{ip}$ (or $\sum_{i=1}^{i=k} f_{ip}$)	n (or 1)

2.2 Marginal distributions

Referring to the contingency table above, the final column denotes the sum of the frequencies (or relative frequencies) associated with each modality x_i of the statistical variable X, while the final row indicates the sum of the frequencies (or relative frequencies) associated with each modality y_i of the statistical variable Y. These sums are designated as ***marginal frequencies***.

Definition 4

- The *marginal frequency* and *marginal relative frequency* associated with each modality x_i of the statistical variable X, denoted by $n_{i.}$and $f_{i.}$ respectively, are defined as follows:

$$n_{i.} = \sum_{j=1}^{j=p} n_{ij} \tag{3.3}$$

$$f_{i.} = \frac{n_{i.}}{n} = \sum_{j=1}^{j=p} f_{ij}. \tag{3.4}$$

- The *marginal frequency* and *marginal relative frequency* associated with each modality y_j of the statistical variable Y, denoted by $n_{.j}$ and $f_{.j}$ respectively, are defined as follows:

$$n_{.j} = \sum_{i=1}^{i=k} n_{ij} \tag{3.5}$$

$$f_{.j} = \frac{n_{.j}}{n} = \sum_{i=1}^{i=k} f_{ij}. \tag{3.6}$$

We note that $\sum_{i=1}^{i=k} n_{i.} = \sum_{j=1}^{j=p} n_{.j} = \sum_{i=1}^{i=k}\sum_{j=1}^{j=p} n_{ij} = n$, and $\sum_{i=1}^{i=k} f_{i.} = \sum_{j=1}^{j=p} f_{.j} = \sum_{i=1}^{i=k}\sum_{j=1}^{j=p} f_{ij} = 1$.

Definition 5

The *marginal distribution* of the statistical variable X (or Y) is defined by providing all the modalities x_i, and the corresponding marginal frequencies $n_{i.}$(or $n_{.j}$) and marginal relative frequencies $f_{i.}$(or $f_{.j}$).

The marginal distribution of the statistical variable X can be represented by the table below.

X	$n_{i.}$	$f_{i.} = \frac{n_{i.}}{n}$
x_1	$n_{1.}$	$f_{1.}$
x_2	$n_{2.}$	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$
x_i	$n_{i.}$	$f_{i.}$
$\vdots$	$\vdots$	$\vdots$
x_k	$n_{k.}$	$f_{k.}$
Total	n	1

The marginal distribution of the statistical variable Y can be represented by the table below.

Y	$n_{.j}$	$f_{.j} = \dfrac{n_{.j}}{n}$
y_1	$n_{.1}$	$f_{.1}$
y_2	$n_{.2}$	$f_{.2}$
$\vdots$	$\vdots$	$\vdots$
y_j	$n_{.j}$	$f_{.j}$
$\vdots$	$\vdots$	$\vdots$
y_p	$n_{.p}$	$f_{.p}$
Total	n	1

Example 2

The following contingency table illustrates the distribution of the number of annual absences X for 25 students as a function of the final mark obtained Y.

$X\backslash Y$	$[0, 5[$	$[5, 10[$	$[10, 15[$	$[15, 20[$	$n_{i.}$
0	2	0	3	0	5
1	0	1	2	3	6
2	0	0	1	1	2
3	4	3	0	0	7
4	1	0	4	0	5
$n_{.j}$	7	4	10	4	25

The marginal distribution of the statistical variable X can be represented by the following table:

X	$n_{i.}$	$f_{i.} = \dfrac{n_{i.}}{n}$
0	5	0.20
1	6	0.24
2	2	0.08
3	7	0.28
4	5	0.20
Total	25	1

The following tabular representation provides a depiction of the marginal distribution of the statistical variable Y:

Y	$n_{.j}$	$f_{.j} = \dfrac{n_{.j}}{n}$
$[0, 5[$	7	0.28
$[5, 10[$	4	0.16
$[10, 15[$	10	0.40
$[15, 20[$	4	0.16
Total	25	1

2.3 Conditional distributions

In the contingency table, the joint frequencies corresponding to the modalities pairs (x_i, y_j) can be used to define ***conditional distributions***, as they are divided by the marginal frequencies associated with the x_i or y_j modalities.

Definition 6

- The *conditional relative frequency* associated with each modality x_i of the statistical variable X, given that $Y = y_j$, is defined as follows:

$$f_{i|j} = \frac{n_{ij}}{n_{.j}} \; ; \, i = \overline{1, k}. \tag{3.7}$$

- The *conditional relative frequency* associated with each modality y_j of the statistical variable Y, given that $X = x_i$, is defined as follows:

$$f_{j|i} = \frac{n_{ij}}{n_{i.}} \; ; \, j = \overline{1, p}. \tag{3.8}$$

We note that $\sum_{i=1}^{i=k} f_{i|j} = \sum_{j=1}^{j=p} f_{j|i} = 1$.

Definition 7

The *conditional distribution* of the statistical variable X (or Y) given that $Y = y_j$ (or $X = x_i$), denoted by $X \mid Y = y_j$ (or $Y \mid X = x_i$), is defined by providing all the modalities x_i (or y_j), the corresponding joint frequencies n_{ij}, and the conditional relative frequencies $f_{i|j}$ (or $f_{j|i}$).

Remark 1

- In contrast to the marginal distribution, which provides a description of the distribution of a single variable across the entire population, the conditional distribution examines the distribution of one variable within a specific subset of the population, as determined by another variable.
- A conditional distribution provides a more precise and nuanced comprehension of the relationship between two variables by considering the distribution of one variable within the context of the other.

The conditional distribution of $X \mid Y = y_j$ can be represented by the table below.

X	n_{ij}	$f_{i\|j} = \frac{n_{ij}}{n_{.j}}$
x_1	n_{1j}	$f_{1\|\mathbf{j}}$
x_2	n_{2j}	$f_{2\|\mathbf{j}}$
$\vdots$	$\vdots$	$\vdots$
x_i	n_{ij}	$f_{i\|\mathbf{j}}$
$\vdots$	$\vdots$	$\vdots$
x_k	n_{kj}	$f_{k\|\mathbf{j}}$
Total	$n_{.j}$	1

The following tabular representation provides a depiction of the conditional distribution of $Y \mid X = x_i$.

Y	n_{ij}	$f_{j\|i} = \dfrac{n_{ij}}{n_{i.}}$
y_1	n_{i1}	$f_{j\|1}$
y_2	n_{i2}	$f_{j\|2}$
$\vdots$	$\vdots$	$\vdots$
y_j	n_{ij}	$f_{j\|i}$
$\vdots$	$\vdots$	$\vdots$
y_p	n_{ip}	$f_{j\|p}$
Total	$n_{i.}$	1

Example 3

The contingency statistical table from *Example 2*, which illustrates the distribution of the number of annual absences X for 25 students as a function of the final mark obtained Y, is to be considered.

$X \backslash Y$	$[0, 5[$	$[5, 10[$	$[10, 15[$	$[15, 20[$	$n_{i.}$
0	2	0	3	0	5
1	0	1	2	3	6
2	0	0	1	1	2
3	4	3	0	0	7
4	1	0	4	0	5
$n_{.j}$	7	4	10	4	25

The conditional distribution of $X \mid Y \in [10, 15[$ can be represented by the table below.

X	n_{ij}	$f_{i\|j} = \dfrac{n_{ij}}{n_{.j}}$
0	3	$\dfrac{3}{10}$
1	2	$\dfrac{2}{10}$
2	1	$\dfrac{1}{10}$
3	0	0
4	4	$\dfrac{4}{10}$
Total	10	1

The conditional distribution of $Y \mid X = 2$ can be represented by the table below.

Y	n_{ij}	$f_{j\|i} = \frac{n_{ij}}{n_{i.}}$
$[0, 5[$	0	0
$[5, 10[$	0	0
$[10, 15[$	1	$\dfrac{1}{2}$
$[15, 20[$	1	$\dfrac{1}{2}$
Total	2	1

3 Numerical description

In a manner consistent with the approach employed for univariate statistics, the statistical representation of a statistical variable pair (X, Y) is characterised by the provision of a numerical representation, incorporating arithmetic means and variances in both marginal and conditional distributions. In the latter case, the independence condition is also exposed.

3.1 Arithmetic mean and variance

The most common position and dispersion parameters are the arithmetic mean and variance, respectively. For a statistical variable pair (X, Y), the marginal arithmetic means and variables are defined as follows.

Definition 8

For a statistical variable pair (X, Y), the *marginal arithmetic means*, denoted $\overline{X}$ and $\overline{Y}$ respectively, are defined as follows:

$$\overline{X} = \tfrac{1}{n}\textstyle\sum_{i=1}^{i=k} n_{i.}x_i = \displaystyle\sum_{i=1}^{i=k} f_{i.}x_i \tag{3.9}$$

$$\overline{Y} = \tfrac{1}{n}\textstyle\sum_{j=1}^{j=p} n_{.j}y_j = \displaystyle\sum_{j=1}^{j=p} f_{.j}y_j. \tag{3.10}$$

Definition 9

For a statistical variable pair (X, Y), the *marginal variance*s, denoted σ_X^2 and σ_Y^2 respectively, are defined as follows:

$$\sigma_X^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^{i=k} n_{i.}x_i^2 - \overline{X}^2 = \displaystyle\sum_{i=1}^{i=k} f_{i.}x_i^2 - \overline{X}^2. \tag{3.11}$$

$$\sigma_Y^2 = \tfrac{1}{n}\textstyle\sum_{j=1}^{j=p} n_{.j}y_j^2 - \overline{Y}^2 = \displaystyle\sum_{j=1}^{j=p} f_{.j}y_j^2 - \overline{Y}^2. \tag{3.12}$$

Remark 2

In the case of a classical table representation (see *Example 1*), the means and variances of each statistical variable are calculated independently, as for two univariate statistical series:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{i=n} x_i. \tag{3.13}$$

$$\overline{Y} = \frac{1}{n}\sum_{j=1}^{j=n} y_j. \tag{3.14}$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{i=n} x_i^2 - \overline{X}^2. \tag{3.15}$$

$$\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^{j=n} y_j^2 - \overline{Y}^2. \tag{3.16}$$

For a statistical variable pair (X, Y), the conditional arithmetic means and variables are defined as follows.

Definition 10

- For the conditional distribution $X \mid Y = y_j$, the *conditional arithmetic mean*, denoted $\overline{X_j}$, is defined as follows:

$$\overline{X_j} = \frac{1}{n_{.j}} \textstyle\sum_{i=1}^{i=k} n_{ij}x_i \displaystyle = \sum_{i=1}^{i=k} f_{i|j}x_i \quad ; \forall j = \overline{1,p}. \tag{3.17}$$

- For the conditional distribution $Y \mid X = x_i$, the *conditional arithmetic mean*, denoted $\overline{Y_i}$, is defined as follows:

$$\overline{Y_i} = \frac{1}{n_{i.}} \textstyle\sum_{j=1}^{j=p} n_{ij}y_j \displaystyle = \sum_{j=1}^{j=p} f_{j|i}y_j \quad ; \forall i = \overline{1,k}. \tag{3.18}$$

Definition 11

The *conditional variances* for the conditional distributions $X \mid Y = y_j$ and $Y \mid X = x_i$, denoted $\sigma_{X_j}^2$ and $\sigma_{Y_i}^2$ respectively, are defined as follows:

$$\sigma_{X_j}^2 = \frac{1}{n_{.j}} \textstyle\sum_{i=1}^{i=k} n_{ij}x_i^2 - \overline{X_j}^2 \displaystyle = \sum_{i=1}^{i=k} f_{i|j}x_i^2 - \overline{X_j}^2 \quad ; \forall j = \overline{1,p} \tag{3.19}$$

$$\sigma_{Y_i}^2 = \frac{1}{n_{i.}} \textstyle\sum_{j=1}^{j=p} n_{ij}y_j^2 - \overline{Y_i}^2 \displaystyle = \sum_{j=1}^{j=p} f_{j|i}y_j^2 - \overline{Y_i}^2 \quad ; \forall i = \overline{1,k}. \tag{3.20}$$

Remark 3

In the case of a continuous quantitative statistical variable X (or Y), the expressions of the marginal arithmetic means and variances, and conditional arithmetic means and variances, are obtained by replacing x_i (or y_j) with the centre C_i (or C_j) of the class $[c_i, c_{i+1}[$ (or $[c_j, c_{j+1}[$) .

Example 4

In considering the *Example 2* once more with regard to the annual number of absences X, the arithmetic mean and variance of X, as well as the standard deviation, will be calculated.
The marginal distribution of X is determined by adding the the numbers $n_{i.}x_i$ and $n_{i.}x_i^2$.

X	$n_{i.}$	$f_{i.} = \frac{n_{i.}}{n}$	$n_{i.}x_i$	$n_{i.}x_i^2$
0	5	0.20	0	0
1	6	0.24	6	6
2	2	0.08	4	8
3	7	0.28	21	63
4	5	0.20	20	80
Total	25	1	51	157

- The marginal arithmetic mean of X is determined as follows:

$$\overline{X} = \frac{1}{25}\sum_{i=1}^{i=5} n_{i.}x_i = \frac{51}{25} = 2.04.$$

- We calculate the marginal variance of X as follows:

$$\sigma_X^2 = \frac{1}{25}\sum_{i=1}^{i=5} n_{i.}x_i^2 - \overline{X}^2 = \frac{157}{25} - (2.04)^2 = 2.1184.$$

  Thus, we deduce its marginal standard deviation :

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2.1184} = 1.45.$$

In a similar manner, the marginal distribution for the continuous statistical variable Y associated with the final mark obtained, is determined by taking into account the centres C_j of the classes.

Y	$n_{.j}$	$f_{.j} = \frac{n_{.j}}{n}$	C_j	$n_{.j}C_j$	$n_{.j}C_j^2$
$[0, 5[$	7	0.28	2.5	17.5	43.75
$[5, 10[$	4	0.16	7.5	30	225
$[10, 15[$	10	0.40	12.5	125	1562.5
$[15, 20[$	4	0.16	17.5	70	1225
Total	25	1	-	242.5	3056.25

- The marginal arithmetic mean of Y is determined as follows:

$$\overline{Y} = \frac{1}{25}\sum_{j=1}^{j=4} n_{.j}C_j = \frac{242.5}{25} = 9.7.$$

- The marginal variance and the standard deviation of Y are calculated as follows

$$\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^{j=4} n_{.j}C_j^2 - \overline{Y}^2 = \frac{3056.25}{25} - (9.7)^2 = 28.16,$$

  and

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{28.16} = 5.30.$$

The conditional distribution of $X \mid Y \in [10, 15[$can be represented by the table below, by adding the numbers $n_{i3}x_i$, $f_{i|3}x_i$, and $f_{i|3}x_i^2$.

X	n_{i3}	$f_{i\|3} = \frac{n_{i3}}{n_{.3}}$	$n_{i3}x_i$	$f_{i\|3}x_i$	$f_{i\|3}x_i^2$
0	3	$\frac{3}{10} = 0.3$	0	0	0
1	2	$\frac{2}{10} = 0.2$	2	0.2	0.2
2	1	$\frac{1}{10} = 0.1$	2	0.2	0.4
3	0	0	0	0	0
4	4	$\frac{4}{10} = 0.4$	16	1.6	6.4
Total	10	1	20	2	7

- The value of the conditional arithmetic mean $\overline{X_3}$ is:

$$\overline{X_3} = \frac{1}{n_{.3}} \sum_{i=1}^{i=5} n_{i3} x_i = \sum_{i=1}^{i=5} f_{i|3} x_i = \frac{20}{10} = 2.$$

- The value of the conditional variance $\sigma^2_{X_3}$is:

$$\sigma^2_{X_3} = \sum_{i=1}^{i=5} f_{i|3} x_i^2 - \overline{X_3}^2 = 7 - (2)^2 = 3.$$

In a similar manner, the conditional distribution of $Y \mid X = 2$ can be represented by the table below.

Y	n_{2j}	$f_{j\|2} = \frac{n_{2j}}{n_{2.}}$	C_j	$n_{2j}C_j$	$f_{j\|2}C_j$	$f_{j\|2}C_j^2$
$[0, 5[$	0	0	2.5	0	0	0
$[5, 10[$	0	0	7.5	0	0	0
$[10, 15[$	1	$\frac{1}{2}$	12.5	12.5	6.25	78.125
$[15, 20[$	1	$\frac{1}{2}$	17.5	17.5	8.75	153.125
Total	2	1	-	30	15	231.25

- The value of the conditional arithmetic mean $\overline{Y_2}$ is:

$$\overline{Y_2} = \frac{1}{n_{2.}} \sum_{j=1}^{j=4} n_{2j} C_j = \sum_{j=1}^{j=4} f_{j|2} C_j = \frac{30}{2} = 15.$$

- The value of the conditional variance $\sigma^2_{Y_2}$is:

$$\sigma^2_{Y_2} = \sum_{j=1}^{j=4} f_{j|2} C_j^2 - \overline{Y_2}^2 = 231.25 - (15)^2 = 6.25.$$

3.2 Independence condition

The independence of two statistical variables is defined as the absence of a statistical link between them. It is important to note that awareness of the value of one variable provides no information about the value of the other variable. Statistical independence can thus be defined as the absence of a link between variables, and the knowledge of the frequencies associated with one does not influence the frequencies of the other. The independence condition can thus be defined as follows.

Definition 12

The statistical variables X and Y are said to be *independent* if:

$$f_{ij} = f_{i.} \times f_{.j} \quad ; \forall i = \overline{1, k} \text{ and } \forall j = \overline{1, p}. \tag{3.21}$$

If X and Y are independent then:

- $f_{i|j} = f_{i.} \quad ; \forall i = \overline{1, k}$ and $\forall j = \overline{1, p}$.
- $f_{j|i} = f_{.j} \quad ; \forall i = \overline{1, k}$ and $\forall j = \overline{1, p}$.

Example 5

From the contingency table of *Example 2*, we can deduce, for the modalities pair $(0, [0.5[)$, that $f_{11} = \frac{n_{11}}{n} = \frac{2}{25} = 0.08$. From the marginal distributions of X and Y, we can also deduce the values $f_{1.} = 0.20$ and $f_{.1} = 0.28$.

It can thus be concluded that the condition of independence has not been verified, since:

$$\exists i = 1, j = 1 \text{ such that } f_{11} \neq f_{1.} \times f_{.1}.$$

Consequently, the statistical variables X and Y are dependent.

The statistical study revealed a correlation between the number of annual absences and the final annual result. It is evident that further quantification of this dependence is required.

3.3 Notion of Covariance and Pearson correlation coefficient

In the context of bivariate statistical analysis, the ***covariance***, and the ***correlation coefficient*** are defined as metrics which quantifies the direction and strength of the relationship between two statistical variables.

Definition 13

The *covariance* between two statistical variables X and Y, denoted $\operatorname{Cov}(X, Y)$, is defined as follows:

$$\begin{aligned} \operatorname{Cov}(X, Y) &= \frac{1}{n} \sum_{i=1}^{i=k} \sum_{j=1}^{j=p} n_{ij} \left(x_i - \overline{X}\right) \left(y_j - \overline{Y}\right) \\ &= \frac{1}{n} \sum_{i=1}^{i=k} \sum_{j=1}^{j=p} n_{ij} x_i y_j - \overline{X}\overline{Y}. \end{aligned} \tag{3.22}$$

Remarks 4

- The covariance of two variables X and Y can be positive, negative, or equal to zero, depending on the nature of relationship between these two variables.
  - If $\operatorname{Cov}(X, Y) > 0$, the variables X and Y exhibit a tendency to move in the same direction.
  - If $\operatorname{Cov}(X, Y) < 0$, the variables X and Y exhibit a tendency to move in opposite directions.
  - If $\operatorname{Cov}(X, Y) = 0$, the variables X and Y are not likely to exhibit a predictable tendency to move together.
- The measurement of the covariance of two variables X and Y is expressed in units, which are calculated through the multiplication of the units of the two variables in the event that they exist.

Remark 5

In the case of a classical table representation (see *Example 1*), the covariance between the two statistical variables is defined as follows:

$$\operatorname{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{i=n} \left(x_i - \overline{X}\right) \left(y_i - \overline{Y}\right) = \frac{1}{n} \sum_{i=1}^{i=n} x_i y_i - \overline{X}\overline{Y}. \tag{3.23}$$

Definition 14

The *Pearson correlation coefficient* between two statistical variables X and Y, referred to as Pearson's r, is defined as follows:

$$\text{Pearson's } r = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{3.24}$$

Remarks 6

- The Pearson's r coefficient is a statistical measure used to quantify the strength of linear relationships between two statistical variables.
- The Pearson's r coefficient is a value that varies between -1 and $+1$.
  - Pearson's $r = +1$ indicates a perfect positive linear relationship between the two variables, i.e. an increase or decrease in one variable is accompanied by an increase or decrease in the other.
  - Pearson's $r = 0$ indicates no linear relationship between the two variables.
  - Pearson's $r = -1$ indicates a perfect negative linear relationship between the two variables, i.e. an increase or decrease in one variable is accompanied by a decrease or increase in the other.
  - When Pearson's r approaches ± 1, the linear relationship between the two variables is considered strong, and when Pearson's r approaches 0, the linear relationship between the two variables is considered weak.

Example 6

From the contingency table of *Example 2*, and the values of marginal arithmetic means $\overline{X}$ and $\overline{Y}$ calculated in *Example 4*, we can evaluate the covariance of the statistical variable pair (X,Y) associated to the annual number of absence (discrete variable X) and the final mark (Continuous variable Y), using the equation 3.22.

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \frac{1}{25}\sum_{i=1}^{i=5}\sum_{j=1}^{j=4} n_{ij} x_i C_j - \overline{X}\overline{Y} \\
&= \frac{1}{25}(1\times 1\times 7.5 + 2\times 1\times 12.5 + 3\times 1\times 17.5 + 1\times 2\times 12.5 + 1\times 2\times 17.5 \\
&= +4\times 3\times 2.5 + 3\times 3\times 7.5 + 1\times 4\times 2.5 + 4\times 4\times 12.5) - (2.04\times 9.7) \\
&= -1.688.
\end{aligned}$$

The calculated value of the covariance, which is negative, indicates that the two variables tend to move in opposite directions.

From the calculated values of σ_X and σ_Y in *Example 4*, we deduce the value of the Pearson's r, using the equation 3.24:

$$\text{Pearson's } r = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{-1.688}{1.45\times 5.30} = -0.21.$$

It is evident that, given the close proximity of Pearson's r to 0 on the scale, it can be deduced that a linear relationship between the two variables is not to be expected.

4 Linear Regression

The ***regression analysis*** is predicated on statistical methodologies employed to estimate the correlation between a dependent variable and one or more independent variables, with the purpose of evaluating its strength and predicting missing values.

In instances where an approximately linear relationship is observed between two statistical variables (X and Y), a ***simple linear regression*** model is utilised to predict the value of the dependent variable (Y, for example) based on a single independent variable (X, for example). This relationship is expressed as ***Regressing*** Y ***onto*** X and is mathematically defined as:

$$Y = aX + b, \tag{3.25}$$

where the coefficients a and b represent the *slope* and the *intercept*, respectively.

Subsequent to the estimation of these parameters, the calculation of $\hat{y}_i = a\hat{x}_i + b$, representing an estimate of Y for a particular value $\hat{x}_i$ of X, can be performed.

In the field of linear regression modelling, the ***Ordinary Least Squares*** (OLS) method has emerged as the prevailing approach for estimating parameters a and b. This method involves the selection of coefficients a and b that optimise the minimisation of the sum of squares of the vertical deviations from the P_i data points of the scatter plot to the regression line of best fit. This sum is referred to as ***Residual Sum of Squares*** (RSS) and is defined as follows:

$$RSS = \sum_{i=1}^{i=n} \left(y_i - (ax_i + b)\right)^2 . \tag{3.26}$$

The subsequent minimisation process yields the following expression a and b:

$$a = \frac{\operatorname{Cov}(X,Y)}{\sigma_X^2}. \tag{3.27}$$

$$b = \overline{Y} - a\overline{X}. \tag{3.28}$$

Remarks 7

- In the same manner, in the event that we want to predict the value of the dependent variable X, based on a single independent variable Y, the ***Regressing*** X ***onto*** Y is defined as:

  $$X = a'Y + b', \tag{3.29}$$

  where the slope and intercept are:

  $$a' = \frac{\operatorname{Cov}(X,Y)}{\sigma_Y^2}. \tag{3.30}$$

  $$b' = \overline{X} - a'\overline{Y}. \tag{3.31}$$

- The point $G(\overline{X}, \overline{Y})$ is known as the ***average point***, or ***centre of gravity***, of the scatter plot.

Example 7

We reconsider the classical representation of *Example 1*.

X	1	2	5	3	2	4	5	3	6	1
Y	5.5	7.25	12.5	10.5	6	9	11.5	9	12.75	7
X	7	8	4	9	6	10	2	5	8	10
Y	15	15.5	11	16	11	18.5	5.5	10.5	14.5	17

In order to quantify the direction and strength of the the relationship between the two statistical variables X and Y, we need to determine the covariance and the correlation coefficient:

- The arithmetic means $\overline{X}$, $\overline{Y}$ using equations 3.13 and 3.14:

$$\begin{aligned}\overline{X} &= \frac{1}{20}\sum_{i=1}^{i=20} x_i = \frac{101}{20} = 5.05\\ \overline{Y} &= \frac{1}{20}\sum_{j=1}^{j=20} y_j = \frac{225.5}{20} = 11.275\end{aligned}$$

- The variances σ_X^2 and σ_Y^2 are calculated, using equations 3.15 and 3.16, respectively:

$$\sigma_X^2 = \frac{1}{20}\sum_{i=1}^{i=20} x_i^2 - \overline{X}^2 = \frac{669}{20} - (5.05)^2 = 7.9475$$

$$\sigma_Y^2 = \frac{1}{20}\sum_{j=1}^{j=20} y_j^2 - \overline{Y}^2 = \frac{2836.375}{20} - (11.275)^2 = 14.6931$$

- Thus, we deduce the covariance between the two statistical variables. Its value is calculated using the equation 3.23:

$$\text{Cov}\,(X,Y) = \frac{1}{20}\sum_{i=1}^{i=20} x_i y_i - \overline{XY} = \frac{1347.5}{20} - (5.05 \times 11.275) = 10.43625$$

- The *Pearson correlation coefficient* between the two statistical variables is finally calculated using the equations 3.24 as follows:

$$\text{Pearson's } r = \frac{\text{Cov}\,(X,Y)}{\sigma_X \sigma_Y} = \frac{10.43625}{\sqrt{7.9475} \times \sqrt{14.6931}} = 0.97 \approx 1.$$

Given the close proximity of Pearson's r to 1 on the scale, it can be deduced that a linear relationship between the two variables is to be expected.
The equation of the regression line Y onto X given by :

$$D_{(Y \text{onto} X)} : Y = aX + b,$$

with

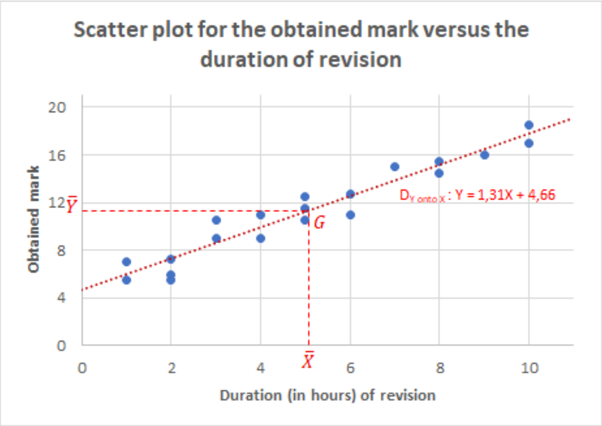
$$a = \frac{\text{Cov}\,(X,Y)}{\sigma_X^2} = \frac{10.43625}{7.9475} \approx 1.3131,$$

and

$$b = \overline{Y} - a\bar{X} = 11.275 - (1.31 \times 5.05) \approx 4.66.$$

Therefore

$$D_{(Y \text{onto} X)} : Y = 1.31X + 4.66.$$



Remarks 8

Since we have:

$$\text{Pearson's } r = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y},$$

then

$$(\text{Pearson's } r)^2 = \frac{\operatorname{Cov}(X,Y)^2}{\sigma_X^2 \sigma_Y^2}.$$

Therefore

$$|\text{Pearson's } r| = \sqrt{a\,a'}. \tag{3.32}$$

Example 8

A scatter plot has been fitted for a pair of statistical variables (X, Y). The obtained linear regressions are as follows:

$$\begin{aligned} D_{(Y \text{onto} X)}: \quad Y &= \quad X + 30 \\ D_{(X \text{onto} Y)}: \quad X &= \quad \frac{Y}{4} + 60. \end{aligned}$$

- The Pearson correlation coefficient is then:

$$|\text{Pearson's } r| = \sqrt{a\,a'} = \sqrt{1 \times \frac{1}{4}} = \frac{1}{2}.$$

- The arithmetic mean of the variables X and Y are determined as follows:

$$\begin{cases} \overline{Y} = \overline{X} + 30 \\ \overline{X} = \frac{\overline{Y}}{4} + 60. \end{cases}$$

  Solving this system of equations yields the following:

$$\begin{aligned} \overline{X} &= 90 \\ \overline{Y} &= 120. \end{aligned}$$

- Knowing that the variance of Y is $\sigma_Y^2 = 40$, we can determine the covariance between the two statistical variables:

$$a' = \frac{\operatorname{Cov}(X,Y)}{\sigma_Y^2} \implies \operatorname{Cov}(X,Y) = a' \times \sigma_Y^2 = \frac{1}{4} \times 40 = 10.$$

Part II

Probability theory

CHAPTER 4

COMBINATORIAL ANALYSIS

1 Introduction

Combinatorial analysis is a branch of mathematics that studies the enumeration and structure of discrete objects or elements. The focus of this study is the various methodologies employed for arranging and selecting these elements, with a particular emphasis on finite sets. The following chapter will focus on arrangements, permutations and combinations. The field of combinatorial analysis has applications in a variety of disciplines, including mathematics such as probability theory, number theory and graph theory, and computer science such as algorithmics and cryptography. In this second part of the course, the emphasis will be on probability calculations in situations that require the enumeration of possible events. This necessitates a foundation in combinatorial analysis, which will be established in the initial part of the course.

2 Preliminaries

In this section, we will briefly review some fundamental concepts of set theory, including the definition of sets and their subsets, the concept of cardinality, and the basic operations on sets. These concepts are essential for understanding probability theory, as they provide the mathematical language and framework for describing and calculating probabilities.

Definitions 1

- A *set*, named and represented by the use of capital letter, is defined as collection of distinct objects, which are referred to as *elements* or *members*. These objects may take on any form, including numbers, letters, symbols, or even other sets.
- A set with a finite or countable number of elements is referred to as a *finite set*, while a set with an infinite number of elements is designated as an *infinite set*.
- A *cardinal* of a finite set A, denoted $\#A$, is defined as the total number of elements that comprise that particular set.
- An *empty set*, denoted by the symbol $\emptyset$, is defined as a set that contains no elements.
- The symbol $\in$ is employed to denote the inclusion of an element within a set. If an element is not a member of a set, then it is denoted using the symbol $\notin$.
- In the context of two sets, A and B, the condition that every element in A is also present in B is indicative of *set A being a subset of set B*, written as $A \subset B$.

Remark 1

The representation of a set is of different forms.The most common form is known as the *roster notation*, in which the elements are enclosed in curly brackets and separated by commas. The *set builder notation* is characterised by a set of rules or statements that collectively delineate the common feature of all elements within a given set. This notation employs a vertical bar to depict its representation, accompanied by a text that outlines the characteristics of the elements constituting the set.

Example 1

- The set $\mathbb{N}$ of natural numbers is an infinite set while the set $A = \{n \mid 2 \leq n \leq 10 \text{ and } n \in \mathbb{N}\}$ is finite.
- The set $A = \{n \mid 2 \leq n \leq 10 \text{ and } n \in \mathbb{N}\}$ is represented in the builder notation. It can be written in the roster notation as $A = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Its cardinal is $\#A = 9$.
- By considering the set $B = \{n \mid n \equiv 0\,[3] \text{ and } 0 < n \leq 10\} = \{3, 6, 9\}$, we can deduce that $B \subset A$, and $10 \in A$, and $10 \notin B$.

2.1 Operations on sets

Set operations are considered to be fundamental procedures in the field of set theory. These operations are employed for the purpose of combining or manipulating elements from one or more sets. The fundamental operations comprise union, intersection, and difference, in conjunction with the complement of a given set.

- In the event of two sets containing identical elements, they are designated as *equal sets*.
- The *union of sets* A and B, denoted by $A \cup B$, is defined as the set of elements contained in both sets A and B:

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}\,. \tag{4.1}$$

- The *intersection of sets* A and B, denoted $A \cap B$, is defined as the set of elements that are common to both sets A and B:

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}\,. \tag{4.2}$$

- The *set difference* between two sets A and B, denoted by $A - B$ (or $B - A$), constitutes the set of elements belonging to set A (or B) that are not present in set B (or A):

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}\,. \tag{4.3}$$

$$B - A = \{x \mid x \in B \text{ and } x \notin A\}\,. \tag{4.4}$$

- Two sets A and B, which constitute subsets of a set E, are considered to be *complementary* if:

$$A \cap B = \emptyset, \quad \text{and } A \cup B = E. \tag{4.5}$$

  We write:

$$B = \complement_E\left(A\right) = E - A = \{x \mid x \in E \text{ and } x \notin A\}\,. \tag{4.6}$$

$$A = \complement_E\left(B\right) = E - B = \{x \mid x \in E \text{ and } x \notin B\}\,. \tag{4.7}$$

- Two sets A and B, are said to be *disjoint* if their intersection is an empty set:

$$A \cap B = \emptyset. \tag{4.8}$$

- The *power set* of a given set A is defined as the set containing the empty set, the set A itself, and all subsets of A:

$$\mathbb{P}\left(A\right) = \{\emptyset, A, S \mid S \subset A\}\,. \tag{4.9}$$

Example 2

Consider the sets $A = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $B = \{3, 6, 9\}$ of *Example 1*.

- The union of the two sets is $A \cup B = \{2, 3, 4, 5, 6, 7, 8, 9, 10\} = A$ because $B \subset A$.
- The intersection of the two sets is $A \cap B = \{3, 6, 9\} = B$ because $B \subset A$.
- The differences between the two sets are $A - B = \{2, 4, 5, 7, 8, 10\}$, and $B - A = \emptyset$.
- The complement of the set A, with respect to the natural numbers set, is $\complement_{\mathbb{N}}\left(A\right) = \{n \mid 0 \leq n < 2 \text{ and } n > 10 \text{ and } n \in \mathbb{N}\}$.
- The power set of the set B is $\mathbb{P}\left(B\right) = \{\emptyset, \{3.6.9\}, \{3\}, \{6\}, \{9\}, \{3, 6\}, \{3, 9\}, \{6, 9\}\}$.

Properties

For 3 sets A, B and C, we can demonstrate the following assertions:

- $A \cap B = B \cap A$, and $A \cup B = B \cup A$. Intersection an reunion operators are commutative.
- $(A \cap B) \cap C = A \cap (B \cap C)$, and $(A \cup B) \cup C = A \cup (B \cup C)$. Intersection an reunion operators are associative.
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. Intersection an reunion operators are distributive.
- $A - (B \cap C) = (A - B) \cup (A - C)$, and $A - (B \cup C) = (A - B) \cap (A - C)$.

2.2 Patterns formed using elements from a set

Definition 2

A *pattern* is made up of elements that have been selected from the n elements of a given set E. Therefore, an element of a pattern is characterised by the following:

- The number of times it appears (repetition);
- Its position in the pattern (order).

Let's consider a family of patterns.

- If a given element in each pattern of the family can only appear 0 or 1 times, then the patterns are said to be *without repetition*.
- If an element can appear more than once in certain patterns of the family, we say that these patterns *have repetition*.
- If the elements that appear in a given pattern play the same role, the pattern is said to be *unordered*.
- If the elements in a given pattern do not play the same role, the pattern is said to be *ordered*.

Example 3

Let us consider the families of patterns made up of 2 elements from the set $E = \{a, b, c, d\}$.

- The family of ordered patterns with repetition is as follows:

  $$aa, ab, ac, ad, ba, bb, bc, bd, ca, cb, cc, cd, da, db, dc, dd.$$

  The total number of patterns is 16.
- The family of ordered patterns without repetition is as follows:

  $$ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc.$$

  The total number of patterns is reduced to 12.
- The family of unordered patterns without repetition is as follows:

  $$ab, ac, ad, bc, bd, cd.$$

  The total number of patterns is reduced to 6.

2.3 Fundamental principle of combinatorial analysis

The ***fundamental principle of combinatorial analysis***, frequently referred to as the ***rule of product*** or ***multiplication principle***, constitutes a foundational stone of combinatorics, offering a systematic approach to quantifying possibilities in scenarios involving multiple choices. The assertion is that :

In the event of n ways to complete one experiment and m ways to complete another, it can be deduced that there are $n \times m$ ways to complete both experiments in combination.
In the general case, if n_i ways exist to complete the i^{th} experiment out of a total of p experiments, then the total number of possible outcomes for the entire experiment is given by:

$$\prod_{i=1}^{i=p} n_i = n_1 \times n_2 \times \ldots \times n_p$$

The principle, in its fundamental aspect, facilitates the enumeration of the possibilities for the occurrence of a sequence of two or more events. This fundamental principle assists in elucidating the following:

- To ascertain the total number of possible outcomes when multiple choices are made in successive sequences.
- To calculate the number of outcomes when one event is followed by another.
- To understand the underlying logic of the processes involved in the enumeration of arrangements, combinations and permutations within a range of circumstances.

Example 4

Imagine an hypothetical menu, presented by a culinary establishment, which features a selection of 4 appetisers, 3 main courses and 5 desserts. The objective is to ascertain the total number of possible meals.
It is evident that each selection of an appetiser, main course and dessert is considered to be independent of the others. Therefore, the fundamental principle of combinatorial analysis facilitates the systematic enumeration of the total number of distinct meal combinations without the necessity of listing them exhaustively:

$$\text{The total possible meal combinations} = 4\,(\text{appetisers}) \times 3\,(\text{main courses}) \times 5\,(\text{desserts}) = 60.$$

3 Counting formulas of combinatorial analysis

The following section is devoted to the topic of combinatorial analysis, and in particular, the most common counting formulas that are employed in this field of study.

3.1 Arrangements

Let consider a finite set E of n elements and an integer p such that $0 < p \leq n$.

Definition 3

- *Arrangements with repetition* of p elements from n can be defined as a family of ordered patterns of p elements with repetition.
  The number of these arrangements, denoted a_n^p, is given by:
  $$a_n^p = n^p. \tag{4.10}$$
- *Arrangements without repetition* of p elements from n can be defined as a family of ordered patterns of p elements without repetition.
  The number of these arrangements, denoted A_n^p, is given by:
  $$A_n^p = \frac{n!}{(n-p)!}, \tag{4.11}$$
  where $n!$ and $(n-p)!$ are the factorials of positive integer numbers n and $n-p$.

Remarks 1

- We recall the the factorial of a positive integer number n (or $n-p$) is defined as the product of all positive integers less than or equal to n (or $n-p$):
  $$\begin{aligned} n! &= n \times (n-1) \times (n-2) \times \ldots \times (n-p) \times \ldots \times 1 \\ &= n \times (n-1) \times \ldots \times (n-p+1) \times (n-p)! \end{aligned}$$
  Thus, the total number of arrangements without repetition can be written as follows:
  $$A_n^p = \frac{n!}{(n-p)!} = n \times (n-1) \times \ldots \times (n-p+1)\,.$$
- We recall that, by convention, the factorial of 0 is $0! = 1$.
- The set of arrangements without repetition, of cardinal A_n^p, forms a subset of the set of arrangements with repetitions, of cardinal a_n^p:
  $$A_n^p \leq a_n^p.$$

Calculation of A_n^p and a_n^p

The construction of arrangements without repetition of p elements from the n elements of the finite set E is an experiment O which can be decomposed into p ordered elementary experiments $O_1, O_2, \ldots, O_p$:

O_1: Selection of the first element, which presents a total of n possible options.

O_2: Selection of the second element, which presents a total of $(n-1)$ possible options.

$\vdots$

O_p: Selection of the p^{th} element, which also present a total of $(n-p)$ possible options.

In accordance with the fundamental principle of combinatorial analysis, the total numbers of possibilities is given as follows:

$$A_n^p = \prod_{i=1}^{i=p} n_i = n \times (n-1) \times \ldots \times (n-p)\,. \tag{4.12}$$

In a similar manner, the construction of arrangements with repetition of p elements from the n elements of the finite set E can also be considered as experiment O which can be decomposed into p ordered elementary experiments $O_1, O_2, \ldots, O_p$ with the same number n of possible options for each experiment O_i. Thus, the total numbers of possibilities is given, using the fundamental principle of combinatorial analysis, as follows :

$$a_n^p = \prod_{i=1}^{i=p} n_i = \underbrace{n \times n \times \ldots \times n}_{p \text{times}} = n^p. \tag{4.13}$$

Example 5

The objective is to devise a 4-digit code that will allow access to a block of flats. The set of digits is:

$$E = \{0, 1, 2, \ldots, 9\},$$

such that its cardinal is $\#E = n = 10$.

Its imperative to acknowledge the significance of the sequence of the digits within the code.

- In the event of the consideration of the possibility of repeated digits in the code, the total number of the 4-digit codes is equivalent to the number of arrangements of 4 digits from 10 with repetitions:

$$a_{10}^4 = 10^4 = 10\,000.$$

- In the event of the consideration of different digits in the code, the total number of the 4-digit codes is equivalent to the number of arrangements of 4 digits from 10 without repetitions:

$$A_{10}^4 = \frac{10!}{(10-4)!} = 10 \times 9 \times 8 \times 7 = 5\,040.$$

3.2 Permutations

Let consider a finite set E of n elements.

Definition 4

- *Permutations without repetition* of the n elements of E are defined as a family of ordered patterns of the n elements without repetition.
  The number of these permutations, denoted P_n, is given by:

$$P_n = n!. \tag{4.14}$$

- In the event of the set E being such that there are α_1 identical elements of type 1, α_2 identical elements of type 2, …, and α_k identical elements type k, the family of ordered patterns of the n elements is referred to as *Permutations with repetitions.*
  The number of these permutations, denoted $P_{n,k}$, is given by:

$$P_{n,k} = \frac{n!}{\alpha_1! \times \alpha_2! \times \ldots \times \alpha_k!}, \tag{4.15}$$

with $\sum_{i=1}^{i=k} \alpha_i = n$.

Remark 2

- Permutations without repetition of the n elements of the set E are equivalent to the arrangements without repetition of $p = n$ elements from the n elements of E:
$$P_n = A_n^n = \frac{n!}{0!} = n!. \tag{4.16}$$
- In the event that all the elements of the set E are repeated only once ($\alpha_1 = \alpha_2 = \ldots = \alpha_n = 1$), the counting formula for permutations with repetitions leads to the counting formula for permutations without repetitions:
$$P_{n,n} = \frac{n!}{\alpha_1! \times \alpha_2! \times \ldots \times \alpha_n!} = \frac{n!}{1! \times 1! \times \ldots \times 1!} = n!. \tag{4.17}$$

Example 6

- Given the word “STUDY”, the number of distinct strings that we can form is the number of permutations without repetitions of the 5 elements of the set $E = \{S, T, U, D, Y\}$.
This number is given by:
$$P_5 = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120.$$
- Given the word “MATHEMATICS”, the number of distinct strings that we can form is the number of permutations with repetitions of the 11 elements of the set $E = \{(M, M), (A, A), (T, T), (H), (E), (I), (C), (S)\}$.
This number is given by:
$$P_{11,8} = \frac{11!}{2! \times 2! \times 2! \times 1! \times 1! \times 1! \times 1! \times 1!} = 4\,989\,600,$$

3.3 Combinations

Let consider a finite set E of n elements and an integer p such that $0 < p \leq n$.

Definition 5

- *Combination* of p elements from n can be defined as a family of unordered patterns of p elements without repetition.
The number of these combinations, denoted C_n^p, is given by:
$$C_n^p = \frac{n!}{p!\,(n-p)!}. \tag{4.18}$$
- *Combinations with repetition* of p elements from n can be defined as a family of unordered patterns of p elements with repetition.
The number of these combinations, denoted K_n^p, is given by:
$$K_n^p = \frac{(n+p-1)!}{p!\,(n-1)!}. \tag{4.19}$$

Remarks 3

- In combinations, the order of elements is of no consequence. For each combination, there exist $p!$ different ways to arrange those elements, so the total number of arrangements corresponds to $p!$ combinations:

$$p! \times C_n^p = \frac{n!}{(n-p)!} = A_n^p. \tag{4.20}$$

- The number of combinations with repetitions is equivalent to the number of combinations of p elements from $(n+p-1)$ elements:

$$K_n^p = C_{n+p-1}^p = \frac{(n+p-1)!}{p!\,(n-1)!}. \tag{4.21}$$

Example 7

- Given a card deck of 32 cards, the poker game is based on 5-card hands without regard to order. The total number of different 5-card hands is the number of combinations of the 5 cards from 32 cards, given by:

$$C_{32}^5 = \frac{32!}{5! \times 27!} = 201\,376.$$

Within this number, if we want to determine how many 5-card hands contain exactly a pair of 2 aces and 2 queens, it is necessary to decompose the experiment into 3 unordered elementary experiments O_1, O_2, O_3:

O_1: Selection of the 2 aces from the 4 in the deck:

$$C_4^2 = \frac{4!}{2! \times 2!} = 6.$$

O_2: Selection of the 2 queens from the 4 in the deck:

$$C_4^2 = \frac{4!}{2! \times 2!} = 6.$$

O_3: Selection of the 5^{th} card from the 24 remaining cards in the deck:

$$C_{24}^1 = \frac{24!}{1! \times 23!} = 24.$$

In accordance with the fundamental principle of combinatorial analysis, the total numbers of 5-card hands containing exactly 2 aces and 2 queens is given by:

$$C_4^2 \times C_4^2 \times C_{24}^1 = 864.$$

- In the event of being required to select 3 ice cream scoops from a total of 4 available flavours, the sequence in which the flavours are chosen in inconsequential. It permissible for a single flavour to be selected on multiple times. Thus, the total number of distinct combinations of cups with 3 scoops is a combination with repetition of 3 flavours from 4, given by:

$$K_4^3 = C_6^3 = \frac{6!}{3! \times 3!} = 20.$$

Properties

We can demonstrate the following assertions:

- The selection of n elements from a set of n elements is equivalent to the selection of 0 elements from n:
$$C_n^n = C_n^0 = 1. \tag{4.22}$$
- The *symmetry in combinations* can be attributed to the fact that the selection of p elements from a set of n elements is equivalent to the choosing of $n-p$ elements from that same set:
$$C_n^p = C_n^{n-p}. \tag{4.23}$$
In the case where $p=1$, the following holds true:
$$C_n^1 = C_n^{n-1} = n. \tag{4.24}$$
- *Pascal's formula*, also known as *Pascal's Identity* postulates that the number of combinations of p elements from a total of n elements is equivalent to the sum of the number of combinations of p elements from $n-1$ total elements and the number of combinations of $p-1$ elements from $n-1$ total elements:
$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1}. \tag{4.25}$$
- The *Newton's binomial formula* is used to expand binomials when they are raised to a power: $(a+b)^n = \sum_{p=0}^{p=n} C_n^p \times a^p \times b^{n-p} \quad ; (a,b) \in \mathbb{R}^2$and $n \in \mathbb{N}^*$ (Newton's binomial formula).

CHAPTER 5

INTRODUCTION TO PROBABILITY THEORY

1 Introduction

Probability theory is a branch of mathematics that has been developed for the purpose of analysing random events and quantifying the associated uncertainty. The objective of this chapter is to introduce the conceptualisation of assigning numerical values to the likelihood of different outcomes in uncertain situations. This approach facilitates a logical and systematic analysis of chance, thereby offering a valuable framework for decision-making in uncertain scenarios. In essence, this approach constitutes a method of comprehending and anticipating the probability of an event materialising.

The subsequent discussion will be focused on the different terminologies of the probability theory, followed by an examination of conditional probabilities and independent events.

2 Probability terminology

The field of probability is predicated on a number of key concepts, including outcomes, sample space, events, sigma-algebra, and probability space. These concepts are instrumental in the characterisation of random experiments.

2.1 Random experiment

In the domain of probability theory, the concept of ***random experiment*** is associated with a process that yields multiple potential results, which cannot be determined with certainty prior to execution of the experiment.

Definition 1

In probability, a *random event*, or *random experiment*, denoted $\mathcal{E}$, is defined as any experiment whose outcome cannot be predicted in advance.

A random experiment is characterised by the following attributes:

- Uncertainty: The outcome of such an event is not determined in advance and is instead dependent upon chance.
- Well-definition: The subsequent process is described in a clear and methodical manner, with a defined set of possible outcomes.
- Repetition: The experiment can be conducted multiple times under identical conditions.
- Probability: It is axiomatic that each possible outcome has a given probability of occurrence.

Example 1

The process of rolling a 6-sided die (numbered 1 to 6) is known as a ***die-roll***. The random nature of this experiment is evidenced by the inability to predict the outcome of the top face of the die.

The potential outcomes of this experiment are clearly established, with each outcome corresponding to one of the six sides of the die.

It is important to note that this experiment can be replicated, and it is noteworthy that, for a fair die, each of the six faces has the same probability of occurring.

2.2 Sample space

In the context of a random experiment, it is possible to define the set of all possible results.

Definition 2

In probability, the *sample space*, denoted Ω, is defined as the set of all potential outcomes ω_i of a random experiment.

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_n\}.$$

Example 2

- In the die-roll, the sample space corresponding to the outcome of the top face of the die is well-defined. It is given by:

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

- A double die-roll is defined as the process of rolling two 6-sided dies. The sample space of the random experiment corresponding to the two top faces of the dies is also well-defined. It is given by:

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

2.3 Event

Within the field of probability, the concept of ***event*** is closely associated with a particular outcome or a series of outcomes derived from a random experiment.

Definition 2

In probability, an *event* is defined as a subset of the sample space, and it *occurs* when the outcome of the random experiment is an element of that subset.

Properties

For a random experiment:

- The union of 2 events A and B, denoted $A \cup B$ and written "A or B", comprises all possible outcomes that are in either event A, or event B, or in both events.
- The intersection of 2 events A and B, denoted $A \cap B$ and written "A and B", comprises only the possible outcomes that are common to both events A and B.
- An event is to be considered *elementary*, or *simple*, if it refers to a single outcome in the sample space.
- The *contrary*, or *opposite*, event of a given event A is defined as the event such that A does not occur, and refers to its complement, denoted $\overline{A}$:

  $$\overline{A} = \complement_\Omega (A)\,. \tag{5.1}$$

  The two events A and $\overline{A}$ are said to be *mutually exclusive*, or *disjoint*:

  $$A \cap \overline{A} = \emptyset. \tag{5.2}$$

  The occurrence of A is contingent upon the non-occurrence of $\overline{A}$.
  More generally, k events $A_1, A_2, \dots A_i, \dots, A_k$ are said to be *pairwise disjoints* if:

  $$A_i \cap A_j = \emptyset \quad ; \forall i \neq j. \tag{5.3}$$

- Two useful properties are given by *de Morgan's laws*, which states that:
  - For two events A and B, "not $(A$ or $B)$" is equivalent to "(not $A)$ and (not $B)$" :

    $$\overline{A \cup B} = \overline{A} \cap \overline{B}. \tag{5.4}$$

  - For two events A and B, " not $(A$ and $B)$" is equivalent to "(not $A)$ or (not $B)$" :

    $$\overline{A \cap B} = \overline{A} \cup \overline{B}. \tag{5.5}$$

  - More generally, for k events $A_1, A_2, \dots A_i, \dots, A_k$:

    $$\overline{\bigcup_{i=1}^{i=k} A_i} = \bigcap_{i=1}^{i=k} \overline{A_i}. \tag{5.6}$$

    $$\overline{\bigcap_{i=1}^{i=k} A_i} = \bigcup_{i=1}^{i=k} \overline{A_i}. \tag{5.7}$$

- An *impossible event* is defined as an event which is incapable of occurring under any given circumstances. It is represented by the empty set $\emptyset$.
- An event is said to be *sure* if it is guaranteed to occur. It is represented by the sample set Ω.

Example 3

In the die-roll random experiment, we have $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Consider the events:

$$A : \text{"Getting an even number on the top face"}$$

$$B : \text{"Getting an odd number on the top face"}$$

$$C : \text{"Getting the number 3 on the top face"}$$

- $A = \{2, 4, 6\}$, $B = \{1, 3, 5\}$ and $C = \{3\}$ are subsets of Ω.
- The event C is an elementary event.
- The events A and B are mutually exclusive events, since $A = \overline{B}$ or $B = \overline{A}$, and $A \cap B = \emptyset$.
- The event "A or B or C" is $A \cup B \cup C = \Omega$. it is a sure event.
- The event "A and C" is $A \cap C = \emptyset$. It is an impossible event.
- The event "not $(A$ or $C)$" is equivalent to the event "(not A) and (not C)":

$$\begin{aligned} \overline{A \cup C} &= \complement_\Omega\left(\{2.3, 4, 6\}\right) = \{1, 5\}\,. \\ \overline{A} \cap \overline{C} &= \{1, 3, 5\} \cap \{1, 2, 4, 5, 6\} = \{1, 5\}\,. \end{aligned}$$

Complete system of events

Definition 3

Consider the sample space Ω of a random experiment. The events $A_1, A_2, \ldots, A_i, \ldots, A_k$ are said to constitute a *complete system of events* if the following conditions are satisfied:

$$\begin{cases} \bigcup_{i=1}^{i=k} A_i = \Omega \\ A_i \cap A_j = \emptyset \quad ; \forall i \neq j. \end{cases} \tag{5.8}$$

A complete system of events is defined by a set of mutually exclusive and exhaustive events. This implies that when an experiment is conducted, one of these events (A_i) is certain to occur, and it is impossible for two events $(A_i$ and A_j, $i \neq j)$ to occur simultaneously.

The set of the mutually exclusive and exhaustive events forms a ***partition*** of the sample state.

2.4 σ-algebra

A ***σ-algebra***, also referred to as a ***tribe***, is a fundamental notion in probability theory that defines the events to which probabilities can be assigned.

Definition 4

In probability, a *σ-algebra* on a sample space Ω, is defined by a set $\mathcal{F}$ of k subsets A_i of Ω, that satisfy the following conditions:

- $\emptyset \in \mathcal{F}$,
- $\Omega \in \mathcal{F}$,
- $\forall A_i \in \mathcal{F}$, $\complement_\Omega\left(A_i\right) \in \mathcal{F}$,
- $\forall A_i \in \mathcal{F}; \displaystyle\bigcup_{i=1}^{i=k} A_i \in \mathcal{F}$.

Remark 1

From the properties of a σ-algebra:

- If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $\overline{A}\text{and}\overline{B} \in \mathcal{F}$.
- If $\overline{A}\text{and}\overline{B} \in \mathcal{F}$, then $\overline{A} \cup \overline{B} \in \mathcal{F}$, then $\complement_\Omega \left(\overline{A} \cup \overline{B}\right) \in \mathcal{F}$.

As a consequence of de Morgan's law, we can deduce that:

$$\complement_\Omega \left(\overline{A} \cup \overline{B}\right) = \complement_\Omega \left(\overline{A \cap B}\right) = A \cap B \in \mathcal{F}.$$

More generally: $\forall A_i \in \mathcal{F}; \bigcap\limits_{i=1}^{i=k} A_i \in \mathcal{F}$.

We can say that σ-algebra of a non-empty sample set is closed under complementation, countable unions, and countable intersections. This means that if we know the measure to a set in the σ-algebra, we also know the measure of its complement set. Additionally, if we know the measures of a a collection of sets in the the σ-algebra, we can then assign a measure to their union.

Example 4

For a sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_n\}$:

- The set $\mathcal{F} = \{\emptyset, \Omega\}$ is a σ-algebra since :

$$\emptyset \in \mathcal{F}, \text{ and } \Omega \in \mathcal{F},$$

$$\complement_\Omega (\emptyset) = \Omega \in \mathcal{F}, \text{ and } \complement_\Omega (\Omega) = \emptyset \in \mathcal{F},$$
$$\emptyset \cup \Omega = \Omega \in \mathcal{F}.$$

- The set $\mathcal{F} = \left\{\emptyset, \Omega, A, \overline{A}\right\}$ is a σ-algebra since :

$$\emptyset \in \mathcal{F}, \text{ and } \Omega \in \mathcal{F},$$

$$\complement_\Omega (A_i) \in \mathcal{F},$$
$$\emptyset \cup \Omega \cup A \cup \overline{A} = \Omega \in \mathcal{F}.$$

- The power set of Ω is a σ-algebra:

$$\mathcal{F} = \mathbb{P}(\Omega) = \{\emptyset, \Omega, A \mid A \subset \Omega\}.$$

2.5 Probability space

The σ-algebra is of great importance in theory of probability. In fact, if the measure is the ***probability*** in a random experiment and the sets in the σ-algebra are the events whose probabilities we wish to calculate, then the definition of the σ-algebra enables us to assign their probabilities unambiguously. In other words, the σ-algebra excludes events for which the probability is not clearly defined.

We can use the next definition to assign probabilities to the events in the σ-algebra.

Definition 5

Let $\mathcal{F}$ the *σ-algebra* on a sample space Ω. A *probability*, denoted P, is defined as follows:

$$\begin{array}{rrl} P: & \mathcal{F} \to & [0,1] \\ & A \to & P(A), \end{array}$$

such that:

- $P(\Omega) = 1$.
- if $A_1, A_2, \ldots, A_i, \ldots, A_k$ are pairwise disjoint sets in $\mathcal{F}$, then:

$$P\left(\bigcup_{i=1}^{i=k} A_i\right) = \sum_{i=1}^{i=k} P(A_i). \tag{5.9}$$

A probability is an application that assigns a value between 0 and 1, both inclusive, to each element (event) in the σ-algebra.

Properties

We can demonstrate the following assertions:

- $P(\emptyset) = 0$.
- For 2 events A, $B \in \mathcal{F}$:
  - $P(\overline{A}) = 1 - P(A)$.
  - If $A \subset B$, then $P(A) \leq P(B)$.
  - $P(A - B) = P(A) - P(A \cap B)$.
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

The framework for mathematically expressing and manipulating probabilities and random events is provided by the ***probability space***, which is defined using the definitions of the σ-algebra and probability.

Definition 6

In probability theory, a *probability space* is defined as a triple comprising a sample space Ω, its *σ-algebra* $\mathcal{F}$, and the associated probability P. It is denoted $(\Omega, \mathcal{F}, P)$.

All possible outcomes (sample space), along with the collection of their subsets (the σ-algebra, or event space) and the probability measure (the application that assigns probabilities to events), are all included in the probability space.

Within a probability space $(\Omega, \mathcal{F}, P)$, the probability law is defined by the probabilities of the elementary events. In particular, The calculation of the probability of an event $\{\omega_1, \omega_2, \ldots, \omega_k\}$ is achieved through the summation of the probabilities of the individual outcomes that constitute the event:

$$P(\{\omega_1, \omega_2, \ldots, \omega_k\}) = P(\omega_1) + P(\omega_2) + \ldots + P(\omega_k).$$

If all elementary events possess an equal likelihood of occurring, they are said to be ***equiprobable events***, and their probabilities are considered equal. Therefore, the probability of any event $A \in \mathcal{F}$ is given by:

$$P(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega} = \frac{\#A}{\#\Omega}.$$

Example 5

1- In a double die-roll random experiment, two fair dies are thrown and the probability that the sum of the results obtained is greater than or equal to 4 is calculated.
The probability space is defined by giving:

- The sample space $\Omega = \{(x, y) \mid x, y \in \{1, 2, 3, 4, , 5, 6\}\}$ with $\#\Omega = 6^2 = 36$,
- the σ-algebra $\mathcal{F} = \mathbb{P}(\Omega) = \{\emptyset, \Omega, A \mid A \subset \Omega\}$,
- the probability for an elementary event $P((x, y)) = \dfrac{1}{\#\Omega} = \dfrac{1}{36} \quad ; \forall (x, y) \in \Omega$.

The calculation will be undertaken to determine the probability of the event A: "Getting 2 top faces with a sum is greater or equal to 4", given by the subset: $A = \{(x, y) \mid x, y \in \{1, 2, 3, 4, , 5, 6\} \text{ and } x + y \geq 4\}$ within $\mathcal{F}$.
Its more easier to determine the probability of:

$$\begin{aligned} \overline{A} &= \{(x, y) \mid x, y \in \{1, 2, 3, 4, , 5, 6\} \text{ and } x + y < 4\} \\ &= \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}, \end{aligned}$$

such that:

$$\#\overline{A} = 6.$$

We deduce that:

$$\begin{aligned} P(A) &= 1 - P\left(\overline{A}\right) \\ &= 1 - \frac{\#\overline{A}}{\#\Omega} \\ &= 1 - \frac{6}{36} \\ &= \frac{5}{6}. \end{aligned}$$

2- In probability, an ***urn experiment*** refers to a specific type of random experiment in which objects (e.g., balls of different colours) are selected from a designated container (the urn) to investigate the probability of obtaining specific outcomes.
In an urn containing 5 red balls and 3 black balls, two balls are removed from the urn without being returned. The objective is to calculate the probability that the two balls are of different colours.

In the event of the balls being selected simultaneously, the probability space is defined by giving:

- The sample space $\Omega = \{(b_1, b_2) \mid b_1, b_2 \in \text{ urn}\}$with $\#\Omega$ given by the number of combinations of 2 balls from the 8 within the urn:
$$\#\Omega = C_8^2 = \frac{8!}{2! \times 6!} = 28,$$
- the σ-algebra $\mathcal{F} = \mathbb{P}(\Omega) = \{\emptyset, \Omega, A \mid A \subset \Omega\}$,
- the probability for an elementary event $P((b_1, b_2)) = \dfrac{1}{\#\Omega} = \dfrac{1}{28} \quad ; \forall (b_1, b_2) \in \Omega$.

The calculation will be undertaken to determine the probability of the event A: "Getting 2 balls of different colours", given by the subset $A = \{(b_1, b_2) \mid b_1, b_2 \in \text{ urn and } b_1 \neq b_2, \}$ within $\mathcal{F}$. Its cardinal is equivalent to the multiplication of the number of combinations of 1 ball from the 5 red balls and the number of combinations of 1 ball from the 3 black balls:

$$\#A = C_5^1 \times C_3^1 = 5 \times 3 = 15.$$

Thus:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{15}{28}.$$

In the event of the balls being selected successively, the probability space is defined by giving:

- The sample space $\Omega = \{(b_1, b_2) \mid b_1, b_2 \in \text{ urn}\}$with $\#\Omega$ given by the number of arrangements without repetition of 2 balls from the 8 within the urn:
$$\#\Omega = A_8^2 = \frac{8!}{6!} = 56,$$

- the σ-algebra $\mathcal{F} = \mathbb{P}(\Omega) = \{\emptyset, \Omega, A \mid A \subset \Omega\}$,
- the probability for an elementary event $P((b_1, b_2)) = \frac{1}{\#\Omega} = \frac{1}{56} \quad ; \forall (b_1, b_2) \in \Omega$.

The cardinal of the event $A = \{(b_1, b_2) \mid b_1, b_2 \in \text{ urn and } b_1 \neq b_2,\}$ is equivalent to the multiplication of the number of arrangements without repetition of 1 ball from the 5 red balls and the number of arrangements without repetition of 1 ball from the 3 black balls, taking into account the order of selection:

$$\#A = \left(A_5^1 \times A_3^1\right) + \left(A_3^1 \times A_5^1\right) = 2 \times 5 \times 3 = 30.$$

Thus:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{30}{56} = \frac{15}{28}.$$

It is important to note that, in terms of probability, drawing the balls successively or simultaneously, without returning the balls, is precisely equivalent.

3 Conditional probability

Conditional probability is the branch of probability theory that deals with the likelihood of an event happening given that another event has already taken place.

Definition 7

Consider a probability space $(\Omega, \mathcal{F}, P)$ and an event B such that $P(B) \neq 0$.
The conditional probability of any event A, given the event B, denoted $P(A \mid B)$ or $P_B(A)$, is defined as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{5.10}$$

The conditional probability is a concept that allows the recalculation of the probability of an event based on the introduction of new relevant information. This calculus provides an estimation of the likelihood of occurrence of event A, given the prior occurrence of event B.

The definition of the conditional probability allows to express:

$$P(A \cap B) = P(A \mid B) \times P(B), \tag{5.11}$$

or, by permutations involving the roles of A and B:

$$P(A \cap B) = P(B \mid A) \times P(A). \tag{5.12}$$

Example 6

1- In the context of a urn random experiment, the contains of an urn are such that there are 6 red balls and 4 white balls. Two balls are removed from the urn successively without being returned. In the given context, the objective is to calculate the probability that the the second removed ball is red given that the fist removed ball is white.

In the probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = \{(b_1, b_2) \mid b_1, b_2 \in \text{ urn}\}$ with $\#\Omega = A_{10}^2 = \frac{10!}{8!} = 90$, $\mathcal{F} = \mathbb{P}(\Omega)$, and $P((b_1, b_2)) = \frac{1}{90}$, we define the events:

A : "The second removed ball is red"

B : "The second removed ball is white" such that:

$$A \cap B \quad = \quad \{(b_1, b_2) \mid b_1 \text{is red and} b_2 \text{ is white}\}.$$

Its cardinal is the multiplication of the the number of arrangements without repetition of 1 ball from the 6 red balls and the number of arrangements without repetition of 1 ball from the 4 white balls:

$$\#(A \cap B) = A_6^1 \times A_4^1 = 6 \times 4 = 24.$$

Thus:

$$P(A \cap B) = \frac{\#(A \cap B)}{\#\Omega} = \frac{24}{90} = \frac{4}{15}.$$

The cardinal of the event B is the multiplication of the number of arrangements without repetition of 1 ball from 9 balls and the number of arrangements without repetitions of 1 ball from the the 4 white balls:

$$\#B = A_9^1 \times A_4^1 = 9 \times 4 = 36.$$

Thus:

$$P(B) = \frac{\#B}{\#\Omega} = \frac{36}{90} = \frac{6}{15}.$$

Finally, the probability of the event A given the event B is calculated as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{4}{6} = \frac{2}{3}.$$

3.1 Total probability theorem and Bayes's formula

The total probability theorem states that:

Let a complete system of events $A_1, A_2, \ldots, A_i, \ldots, A_k$ constituting a partition of a sample state, and assume that $P(A_i) \neq 0$, $\forall i = \overline{1, k}$. Then, the probability of any event B is given by:

$$P(B) = \sum_{i=1}^{i=k} P(B \mid A_i) \times P(A_i) \tag{5.13}$$

We can demonstrate this theorem as follows:

$A_1, A_2, \ldots, A_k$ is a complete system for Ω, then $\Omega = \bigcup_{i=1}^{i=k} A_i$ and $A_i \cap A_j = \emptyset\,;\, \forall i \neq j$.

By considering an event $B \in \mathcal{F}$, and then $B \subset \Omega$, we can write:

$$\begin{aligned} B &= B \cap \Omega \\ &= B \cap \left(\bigcup_{i=1}^{i=k} A_i\right) \\ &= B \cap (A_1 \cup A_2 \cup \ldots \cup A_k) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_i) \cup \ldots \cup (B \cap A_k). \end{aligned}$$

Then:

$$\begin{aligned} P(B) &= P[(B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_i) \cup \ldots \cup (B \cap A_k)] \\ &= P\left[\bigcup_{i=1}^{i=k} (B \cap A_i)\right] \end{aligned}$$

Since $A_i \cap A_j = \emptyset\,;\,\forall i \neq j$, $P\left(B \cap A_i \cap A_j\right) = 0$, and then:

$$\begin{aligned} P\left(B\right) &= P\left(B \cap A_1\right) + P\left(B \cap A_2\right) + \ldots + P\left(B \cap A_i\right) + \ldots + P\left(B \cap A_k\right) \\ &= \sum_{i=1}^{i=k} P\left(B \cap A_i\right). \end{aligned}$$

The reformulation of the conditional probability, as given by equation 5.12, allows us to achieve the desired result:

$$P\left(B\right) = \sum_{i=1}^{i=k} P\left(B \mid A_i\right) \times P\left(A_i\right).$$

Bayes's formula

Bayes's formula constitutes a foundational principle within the domain of probability theory, elucidating the process of updating beliefs in light of new evidence. It enables the calculation of the conditional probability of an event given evidence, thereby inverting the logic of the traditional probability formula. It states that:

Under the same conditions of the total probability theorem, with $P\left(B\right) \neq 0$, we can write:

$$P\left(A_j \mid B\right) = \tfrac{P(A_j \cap B)}{P(B)} = \frac{P\left(B \mid A_j\right) \times P\left(A_j\right)}{\displaystyle\sum_{i=1}^{i=k} P\left(B \mid A_i\right) \times P\left(A_i\right)} \quad ; \forall j. \qquad (5.14)$$

In other words, the Bayes's formula provides a method of calculating the probability of event A_j given an event B, contingent upon the knowledge of the probability of B given A_j, and the prior probabilities of A and B.

Example 7

In the context of manufacturing setting, it is to be considered that 3 distinct machines, M_1, M_2 and M_3, are involved in the production of articles. The respective proportions of these machines are determined as 20 %, 30 % and 50 % respectively. The probability of defective articles being produced by the 3 machines is as follows: 2 % for M_1, 4 % for M_2 and 3 %for M_3.
The objective is to calculate the probability that a random article is produced by the machine M_2, given that it is defective.
Firstly, the following events must be considered:

$$M_i : \text{"The choosen article is made by } M_i\text{"}$$

$$D : \text{"The choosen article is defective"}$$

so that:

$$\begin{aligned} P\left(M_1\right) &= 0.2 \\ P\left(M_2\right) &= 0.3 \\ P\left(M_3\right) &= 0.5, \end{aligned}$$

and:

$$\begin{aligned} P(D \mid M_1) &= 0.02 \\ P(D \mid M_2) &= 0.04 \\ P(D \mid M_3) &= 0.03. \end{aligned}$$

Therefore, the probability of choosing a random article produced by M_2, given that it is defective is $P(M_2 \mid D)$, which is determined by applying Bayes's formula:

$$\begin{aligned} P(M_2 \mid D) &= \frac{P(D \mid M_2) \times P(M_2)}{\sum_{i=1}^{i=3} P(D \mid M_i) \times P(M_i)} \\ &= \frac{0.04 \times 0.3}{[(0.02 \times 0.2) + (0.04 \times 0.3) + (0.03 \times 0.5)]} \\ &\approx 0.39. \end{aligned}$$

The probability that a defective article is produced by machine M_2 is approximately 39 %.

4 Independence of events

Independence of two events in probability is considered when knowledge of one event provides no information regarding the occurrence of the other. Mathematically, this can be defined as follows:

Definition 8

In a probability space $(\Omega, \mathcal{F}, P)$, two events A and B are defined as *independent events* if:

$$P(A \cap B) = P(A) \times P(B). \tag{5.15}$$

Moreover, in the event that $P(B) \neq 0$, A and B are defined as *independent events* if:

$$P(A \mid B) = P(A). \tag{5.16}$$

More generally, the events $A_1, A_2, \ldots, A_i, \ldots, A_k$ are said to be *mutually independent* if:

$$P\left(\bigcap_{i=1}^{i=k} A_i\right) = \prod_{i \in \overline{1,k}} P(A_i). \tag{5.17}$$

In other words, if two events A and B are independent, then the occurrence of B will not affect the probability of A occurring.

Remark 2

- If A and B are two independent events then:
  - A and $\overline{B}$ are also independent.
  - $\overline{A}$ and B are also independent.
  - $\overline{A}$ and $\overline{B}$ are also independent.
- If the events $A_1, A_2, \ldots, A_i, \ldots, A_k$ are pairwise independent, this does not guarantee that they are mutually independent.

Example 8

For a double die-roll random experiment, we define the probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = \{(x,y) \mid x, y \in \{1,2,3,4,5,6\}\}$ with $\#\Omega = a_6^2 = 6^2 = 36$, $\mathcal{F} = \mathbb{P}(\Omega)$, and $P((x,y)) = \dfrac{1}{36}$.

In this context, consider the events:

$$\begin{gathered} A_1 : \text{”The top face of the first die is even”}, \\ A_2 : \text{”The top face of the second die is odd”}, \\ A_1 : \text{”The top face of the two dies are of same parity”}, \end{gathered}$$

defined by the subsets:

$$\begin{aligned} A_1 &= \{(x,y) \mid x \in \{2,4,6\} \text{ and } y \in \{1,2,3,4,5,6\}\} \\ A_2 &= \{(x,y) \mid x \in \{1,2,3,4,5,6\} \text{ and } y \in \{1,3,5\}\} \\ A_3 &= \{(x,y) \mid x, y \in \{2,4,6\} \text{ or } x, y \in \{1,3,5\}\}, \end{aligned}$$

such that :

$$\begin{aligned} P(A_1) &= \frac{\#A_1}{\#\Omega} = \frac{18}{36} = \frac{1}{2} \\ P(A_2) &= \frac{\#A_2}{\#\Omega} = \frac{18}{36} = \frac{1}{2} \\ P(A_3) &= \frac{\#A_3}{\#\Omega} = \frac{18}{36} = \frac{1}{2}. \end{aligned}$$

Consider the following subsets:

$$\begin{aligned} A_1 \cap A_2 &= \{(x,y) \mid x \in \{2,4,6\} \text{ and } y \in \{1,3,5\}\} \\ A_1 \cap A_3 &= \{(x,y) \mid x, y \in \{2,4,6\}\} \\ A_2 \cap A_3 &= \{(x,y) \mid x, y \in \{1,3,5\}\} \\ A_1 \cap A_2 \cap A_3 &= \emptyset, \end{aligned}$$

such that:

$$\begin{aligned} P(A_1 \cap A_2) &= \frac{\#(A_1 \cap A_2)}{\#\Omega} = \frac{9}{36} = \frac{1}{4} \\ P(A_1 \cap A_3) &= \frac{\#(A_1 \cap A_3)}{\#\Omega} = \frac{9}{36} = \frac{1}{4} \\ P(A_2 \cap A_3) &= \frac{\#(A_2 \cap A_3)}{\#\Omega} = \frac{9}{36} = \frac{1}{4} \\ P(A_1 \cap A_2 \cap A_3) &= \frac{\#(A_1 \cap A_2 \cap A_3)}{\#\Omega} = \frac{0}{36} = 0. \end{aligned}$$

Thus, we can conclude that :

$$A_1 \text{ and } A_2 \text{ are independent since } P(A_1 \cap A_2) = P(A_1) \times P(A_2),$$

$$\begin{gathered} A_1 \text{ and } A_3 \text{ are independent since } P(A_1 \cap A_3) = P(A_1) \times P(A_3), \\ A_2 \text{ and } A_3 \text{ are independent since } P(A_2 \cap A_3) = P(A_2) \times P(A_3), \\ A_1, A_2, A_3 \text{ are not mutually independent since } P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1) \times P(A_2) \times P(A_3). \end{gathered}$$

CHAPTER 6

RANDOM VARIABLES

1 Introduction

The previous chapter introduced the language and tools of probability for describing uncertainty. It introduced the concept of the sample space, which describes the possible outcomes of a random process. However, the elements of the sample space are often not of direct interest, particularly if the sample space is large or infinite. It is usually more convenient to work with subsets of these elements.

For instance, the sample space for observing the rolls of two dies contains 36 elements. However, we may be more interested in the sum of the two obtained top faces than in the specific elements of the sample space that produced that sum. In other words, we may be more interested in rolling a sum of 5 than in how that sum was obtained. We can collect various elements of the sample space and treat them collectively as what we are interested in.

More generally, grouping elements of the sample space together is useful, and we can assign a real number to that group. This leads to the concept of a random variable.

2 Basic concepts

A ***random variable*** is a variable that associates a numerical value with each possible outcome of a random experiment. It can therefore be defined as follows:

Definition 1

Let a given probability space $(\Omega, \mathcal{F}, P)$.

- A real *random variable*, denoted r. v. X, is defined as follows:

$$\begin{aligned} X: \quad \Omega &\rightarrow \quad \mathbb{R} \\ \omega_i &\rightarrow \quad X\left(\omega_i\right) = x_i, \end{aligned}$$

such that $\forall x_i \in \mathbb{R} : X^{-1}\left(]-\infty, x_i]\right) = \{\omega_i \in \Omega \mid X(\omega_i) \leq x_i\} \in \mathcal{F}$.

- The set $X(\Omega) = \{X(\omega_i) \mid \omega_i \in \Omega\}$ is referred to as the *defining set* (or *image*) of the r. v. X.

A real random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ is :

- discrete if its defining set $X(\Omega)$ is either finite or an infinite countable set.
- continuous if its defining set $X(\Omega)$ is either an interval of $\mathbb{R}$ or the union of several intervals $\mathbb{R}$.

Properties

- If X is a r. v., then $aX + b$ is also a r. v. $\forall a, b \in \mathbb{R}$.
- If X and Y are 2 r. v., then $X + Y$, $X - Y$, $\frac{X}{Y}(Y \neq 0)$ are also r. v.
- If X is a r. v. then a function $\phi(X)$ defines another r. v.

Example 1

In a balanced coin toss experiment, the possible outcomes are heads (H) or tails (T). If the coin is tossed 3 times, the sample space associated to this experiment is:

$$\Omega = \{(x, y, z) \mid x, y, z \in \{H, T\}\} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

with $\#\Omega = a_2^3 = 2^3 = 8$, $\mathcal{F} = \mathbb{P}(\Omega)$, and $P((x, y, z)) = \frac{1}{\#\Omega} = \frac{1}{8} \quad ; \forall (x, y, z) \in \Omega$.

Let X be the r. v. representing the number of heads obtained. Its defining set is :

$$X(\Omega) = \{0, 1, 2, 3\}.$$

As the defining set is finite, the r. v. X is discrete.

3 Probability distribution and cumulative distribution function for discrete random variable

In the context of discrete random variables, we will introduce two important concepts: the ***probability distribution*** and the ***cumulative distribution function*** of the random variable. The former is useful when we are interested in the likelihood of each individual outcome, while the latter is useful when we are interested in the likelihood of a range of outcomes up to a specific point.

3.1 Probability distribution

Definition 2

Let X a discrete r. v. on the probability space $(\Omega, \mathcal{F}, P)$.
The *probability distribution* (also known as *probability mass function*) of X, denoted $P_X(\cdot)$, is defined as follows:

$$\begin{aligned} P_X: \quad X(\Omega) \rightarrow \quad & [0,1] \\ x_i \rightarrow \quad & P_X(x_i) = P(X = x_i). \end{aligned}$$

The simplest way to represent a discrete probability distribution is in the form of a table, with a row (or a column) listing the possible values x_i and another row (or column) listing their corresponding probabilities $P_X(x_i)$. Alternatively, it can be represented graphically, plotting the probabilities $P_X(x_i)$ on the y-axis versus the possible values x_i on the x-axis. In some cases, a formula can be employed to calculate the probability $P_X(x_i)$ of any given value x_i from the defining set $X(\Omega)$.

Property

- For a defining set $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$, the events $X = x_i$, $i = \overline{1, n}$ constitute a complete system of events. Thus the sum of the probabilities $P(X = x_i)$ is the probability of a certain event:

$$\sum_{i=1}^{i=n} P(X = x_i) = 1. \tag{6.1}$$

Example 2

In the random experiment of *Example 1*, the sample space associated to this experiment is:

$$\Omega = \{(x,y,z) \mid x,y,z \in \{H,T\}\} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

and the defining set of the r. v. X representing the number of heads is :

$$X(\Omega) = \{0,1,2,3\}.$$

For each of the events $X = x_i$, $x_i \in X(\Omega)$, we calculate the probabilities $P(X = x_i)$ as follows:

$$\begin{aligned} P(X=0) &= P(\text{Getting } 0\ H) = P(TTT) = \frac{1}{8} \\ P(X=1) &= P(\text{Getting } 1\ H) = P(HTT, THT, TTH) = \frac{3}{8} \\ P(X=2) &= P(\text{Getting } 2\ H) = P(HHT, HTH, THH) = \frac{3}{8} \\ P(X=3) &= P(\text{Getting } 3\ H) = P(HHH) = \frac{1}{8}. \end{aligned}$$

We verify that $\sum_{i=1}^{i=4} P(X = x_i) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$
The discrete probability distribution can be represented by the following table:

x_i	0	1	2	3	$\sum$
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

3.2 Cumulative distribution function

Definition 3

Let X a discrete r. v. on the probability space $(\Omega, \mathcal{F}, P)$.
The *cumulative distribution function* of X, denoted $F_X(\cdot)$, is defined as follows:

$$\begin{array}{rrl} F_X: & \mathbb{R} \to & [0,1] \\ & x \to & F_X(x) = P(X \le x). \end{array}$$

In other words, the cumulative distribution function of a discrete random variable provides the probability of the variable taking on a value that is less than or equal to a given value. It accumulates the probabilities of the variable taking on all values up to and including the given value.

In practice, for a defining set $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ with $x_1 < x_2 < \dots < x_i < \dots < x_n$, the cumulative distribution function is defined for all the values x of the set $\mathbb{R} =]-\infty, x_1[\cup]x_1, x_2[\cup \dots \cup]x_n, +\infty[$ as follows:

$$F_X(x) = \begin{cases} 0 & ; x < x_1 \\ P(X = x_1) & ; x_1 \le x < x_2 \\ P(X = x_1) + P(X = x_2) & ; x_2 \le x < x_3 \\ \vdots & \vdots \\ \sum_{i=1}^{i=n} P(X = x_i) = 1 & ; x \ge x_n. \end{cases}$$

The cumulative distribution function can be represented graphically by plotting $F_X(x)$ on the y-axis and the possible values x_i on the x-axis. As with univariate statistical series, the resulting graph is a cumulative curve, invariably in the form of a staircase (see Chapter 2, subsection 4.3).

Properties

The cumulative distribution function of a discrete r. v. X displays the following properties:

- $\lim_{x \longrightarrow -\infty} F_X(x) = 0$.
- $\lim_{x \longrightarrow +\infty} F_X(x) = 1$.
- F_X is increasing : $\forall x_1, x_2 \in \mathbb{R}$, $x_2 > x_1 \Longrightarrow F_X(x_2) > F_X(x_1)$.
- F_X is right-continuous.
- $\forall a, b \in \mathbb{R}$ such that $a < b$, $P(a < X \le b) = F_X(b) - F_X(a)$.

Example 3

In the random experiment of *Example 1*, the discrete probability distribution can be represented by the following table:

x_i	0	1	2	3	$\sum$
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

The cumulative distribution function is as follows:

$$F_X(x) = \begin{cases} 0 & ; x < 0 \\ P(X=0) = \frac{1}{8} & ; 0 \leq x < 1 \\ P(X=0) + P(X=1) = \frac{4}{8} & ; 1 \leq x < 2 \\ P(X=0) + P(X=1) + P(X=2) = \frac{7}{8} & ; 2 \leq x < 3 \\ \sum_{i=1}^{i=4} P(X = x_i) = 1 & ; x \geq 3. \end{cases}$$

The definition and properties of the cumulative distribution function allow us to calculate particular probabilities of the r. v. X, as follows:

For $a, b \in \mathbb{R}$ such that $a < b$:

- $P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$.
- $P(X < a) = P(X \leq a - 1) = F_X(a - 1)$.
- $P(a < X < b) = P(a < X \leq b) - P(X = b) = F_X(b) - F_X(a) - P(X = b)$.
- $P(a \leq X \leq b) = P(a < X \leq b) + P(X = a) = F_X(b) - F_X(a) + P(X = a)$.

4 Probability density function and cumulative distribution function for continuous random variable

The probability that a random variable will take on a particular value within a given interval is described by a function known as the ***probability density function***. This can be defined as follows:

Definition 4

Let X a continuous r. v. on the probability space $(\Omega, \mathcal{F}, P)$.
The *probability density function* of X, denoted $f_X(\cdot)$, is defined as follows:

$$\begin{aligned} f_X : \ \mathbb{R} \to & \ \mathbb{R}^+ \\ x \to & \ f_X(x) = f(x). \end{aligned}$$

such that f is integrable and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

In contrast to the probability mass function for the discrete random variable, the probability density function does not directly provide the probability of a particular value (which is actually zero for continuous variables). Instead, it shows how likely it is that the variable will lie within a given interval.

Subsequently, the cumulative distribution function for a continuous r. v. can be defined as follows:

Definition 5

Let X a continuous r. v. on the probability space $(\Omega, \mathcal{F}, P)$.

The *cumulative distribution function* of X, denoted $F_X(\cdot)$, is defined as follows:

$$\begin{aligned} F_X : \quad \mathbb{R} &\to \quad [0,1] \\ x &\to \quad F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t. \end{aligned}$$

Properties

- All the properties stated for the cumulative distribution function of a discrete r. v. are satisfied by the cumulative distribution function of a continuous r. v., which is continuous over $\mathbb{R}$.

- In addition, we have:
  - F_X is continuous.
  - $P(X = x) = 0\ \forall, x \in \mathbb{R}$.
  - $P(X \leq a) = P(X < a) = F_X(a)$.
  - $P(X \geq a) = P(X > a) = 1 - F_X(a)$.
  - $\forall\, a, b \in \mathbb{R}$ such that $a < b$, $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x = F_X(b) - F_X(a)$.

Example 4

The probability density function of a continuous r. v. is given as follows:

$$f(x) = \begin{cases} k\left(4x - 2x^2\right) & ;\ 0 < x < 2 \text{ and } k \in \mathbb{R} \\ 0 & ;\ \text{otherwise.} \end{cases}$$

- We can determine the the value of k for which $f(x)$ is a probability density function as follows:

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1 \iff \int_0^2 k\left(4x - 2x^2\right)\mathrm{d}x = 1 \implies \boxed{k = \frac{3}{8}}.$$

- The cumulative distribution function is determined as follows:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \forall, x \in \mathbb{R}$$

  - If $x < 0$ then $F_X(x) = \int_{-\infty}^{x} 0\mathrm{d}t = 0$.
  - If $0 \leq x < 2$ then $F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{0} f(t)\,\mathrm{d}t + \int_0^x f(t)\,\mathrm{d}t = \int_0^x \frac{3}{8}\left(4t - 2t^2\right)\mathrm{d}t = \frac{3}{4}x^2 - \frac{1}{4}x^3$.
  - If $x \geq 2$ then $F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{0} f(t)\,\mathrm{d}t + \int_0^2 f(t)\,\mathrm{d}t + \int_2^x f(t)\,\mathrm{d}t = \int_0^2 \frac{3}{8}\left(4t - 2t^2\right)dt = 1$.

  In summary, we have:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{3}{4}x^2 - \frac{1}{4}x^3 & \text{if } 0 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}.$$

- Thus, we can for example determine $P(1 < X < 2)$ and $P(X > 1)$ as follows:

$$\begin{aligned} P(1 < X < 2) &= F_X(2) - F_X(1) = 1 - \left(\frac{3}{4} - \frac{1}{4}\right) = \frac{1}{2}. \\ P(X > 1) &= 1 - P(X \leq 1) = 1 - F_X(1) = 1 - \left(\frac{3}{4} - \frac{1}{4}\right) = \frac{1}{2}. \end{aligned}$$

5 Mathematical expectation and variance of a random variable

The ***mathematical expectation*** (or ***expected value*** or ***mean***) of a r. v. is the average value (which is equivalent to the mean in statistical series), obtained by taking into account the probabilities of each of the possible outcomes. As for the statistical series description, the variance is a parameter that measures the dispersion of the r. v, providing an indication of how much the individual values deviate from the mathematical expectation.

5.1 Mathematical expectation

Definition 6

The *mathematical expectation*, or *mean*, of a r. v. X, denoted $E\left(X\right)$, is a real value that may or may not exist, and is defined as follows:

- $E\left(X\right)=\sum\limits_{x_i\in X(\Omega)} x_i\cdot P_X\left(x_i\right)$ for a discrete r. v.
- $E\left(X\right)=\displaystyle\int_{-\infty}^{+\infty} x\cdot f\left(x\right)\mathrm{d}x$ for a continuous r. v.

The mathematical expectation is a measure of the central tendency of a r. v., indicating its average. For a discrete r. v., it is determined by adding together the product of each value x_i and its probability mass function $P\left(X=x_i\right)$. For a continuous r. v., it is determined by integrating the product of the variable x and its probability density function $f\left(x\right)$.

Properties

- $\forall\,, a\in\mathbb{R}$, $E\left(a\right)=a$.
- For a discrete r. v. $Y=\phi\left(X\right)$, the mathematical expectation is given by:

$$E\left(\phi\left(X\right)\right)=\sum_{x_i\in X(\Omega)}\phi\left(x_i\right)\cdot P_X\left(x_i\right). \tag{6.2}$$

- For a continuous r. v. $Y=\phi\left(X\right)$, the mathematical expectation is given by:

$$E\left(\phi\left(X\right)\right)=\int_{-\infty}^{+\infty}\phi\left(x\right)\cdot f\left(x\right)\mathrm{d}x \tag{6.3}$$

- For a r. v. $Y=a\,X+b$, $\forall\,, a,b\in\mathbb{R}$, The mathematical expectation is given by:

$$E\left(Y\right)=a\,E\left(X\right)+b. \tag{6.4}$$

- For n r. v. X_k, each of which has a finite mathematical expectation $E\left(X_k\right)$:

$$E\left(\sum_{k=1}^{k=n}X_k\right)=\sum_{k=1}^{k=n}E\left(X_k\right). \tag{6.5}$$

- For n independent r. v. X_k, each of which has a finite mathematical expectation $E\left(X_k\right)$:

$$E\left(\prod_{k=1}^{k=n}X_k\right)=\prod_{k=1}^{k=n}E\left(X_k\right). \tag{6.6}$$

5.2 Variance

Definition 7

The variance of a r. v. X, denoted $V(X)$ or σ_X^2, is a non-negative real value that may or may not exist, and is defined as follows:

$$V(X) = \sigma_X^2 = E\left[(X - E(X))^2\right], \tag{6.7}$$

or:

$$V(X) = \sigma_X^2 = E\left(X^2\right) - E(X)^2. \tag{6.8}$$

σ_X is the standard deviation of the r. v. X.

The variance (and standard deviation) quantifies the spread of a r. v. around its mathematical expectation. It represents the averaged squared difference between each value of the r. v. X and its mathematical expectation $E(X)$.

Properties

- $\forall, a \in \mathbb{R}$, $V(a) = 0$.
- For a r. v. $Y = a\,X + b$, $\forall, a, b \in \mathbb{R}$, the variance is given by:

$$V(Y) = a^2 V(X). \tag{6.9}$$

.

- For a discrete r. v. X , we can write:

$$V(X) = \sum_{x_i \in X(\Omega)} x_i^2 \cdot P_X(x_i) - E(X)^2. \tag{6.10}$$

- For a continuous r. v. X, we can write:

$$V(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,\mathrm{d}x - E(X)^2. \tag{6.11}$$

- For n independent r. v. X_k, each of which has a finite variance $V(X_k)$:

$$V\left(\sum_{k=1}^{k=n} X_k\right) = \sum_{k=1}^{k=n} V(X_k). \tag{6.12}$$

Example 5

I- For the the discrete r. v. of *Example 1*, we recall the probability distribution represented as follows:

x_i	0	1	2	3	$\sum$
$P\left(X=x_i\right)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

- The mathematical expectation of X is calculated as follows:

$$\begin{aligned} E\left(X\right) &= \sum_{i=1}^{i=4} x_i \cdot P_X\left(x_i\right) \\ &= 0 \cdot P\left(X=0\right) + 1 \cdot P\left(X=1\right) + 2 \cdot P\left(X=2\right) + 3 \cdot P\left(X=3\right) \\ &= 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} \\ &= \frac{3}{2}. \end{aligned}$$

- The variance of X is calculated as follows:

$$\begin{aligned} V\left(X\right) &= E\left(X^2\right) - E\left(X\right)^2 \\ &= \left[0^2 \cdot P\left(X=0\right) + 1^2 \cdot P\left(X=1\right) + 2^2 \cdot P\left(X=2\right) + 3^2 \cdot P\left(X=3\right)\right] - \left[\frac{3}{2}\right]^2 \\ &= \left[0 + \frac{3}{8} + \frac{12}{8} + \frac{9}{8}\right] - \frac{9}{4} \\ &= \frac{3}{4}. \end{aligned}$$

II- For the the continuous r. v. of *Example 4*, we recall the probability density distribution represented as follows:

$$f\left(x\right) = \begin{cases} \frac{3}{2}x - \frac{3}{4}x^2 & ; 0 < x < 2 \\ 0 & ; \text{otherwise.} \end{cases}$$

- The mathematical expectation of X is calculated as follows:

$$\begin{aligned} E\left(X\right) &= \int_{-\infty}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= \int_{-\infty}^{0} x \cdot f\left(x\right) \mathrm{d}x + \int_{0}^{2} x \cdot f\left(x\right) \mathrm{d}x + \int_{2}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= 0 + \int_{0}^{2} x \cdot \left(\frac{3}{2}x - \frac{3}{4}x^2\right) \mathrm{d}x + 0 \\ &= 1. \end{aligned}$$

- The variance of X is calculated as follows:

$$\begin{aligned} V\left(X\right) &= E\left(X^2\right) - E\left(X\right)^2 \\ &= \int_{-\infty}^{+\infty} x^2 \cdot f\left(x\right) \mathrm{d}x - E\left(X\right)^2 \\ &= \int_{0}^{2} x^2 \cdot \left(\frac{3}{2}x - \frac{3}{4}x^2\right) \mathrm{d}x - 1 \\ &= \frac{1}{5}. \end{aligned}$$

6 Moments of random variables

In probability theory, the moments of a r. v. describe the shape of its distribution. They are classified into 2 types: ***non-central*** (or ***row***) ***moments*** and ***central moments***. The kth moment of a r. v., with k non-negative integer, is a statistical measure of the associated probability distribution.

6.1 Non-central moment

Definition 8

The *kth non-central moment* of a r. v. X, denoted m_k, is defined as follows:

$$m_k = E\left(X^k\right) = \begin{cases} \sum_{x_i \in X(\Omega)} x_i^k \cdot P_X(x_i) & ; \text{if } X \text{ is discrete.} \\ \\ \int_{-\infty}^{+\infty} x^k \cdot f(x)\,\mathrm{d}x & ; \text{if } X \text{is continuous.} \end{cases} \tag{6.13}$$

In other words, the kth non-central moment of a r. v. X is equivalent to the mathematical expectation of X raised to the power of k.

Remarks 1

- kth non-central moments provide an indication about the size and shape of the probability distribution around the origin, which is typically zero.
- The first non-central moment m_1 of a r. v. X is merely its mathematical expectation:

$$m_1 = E(X). \tag{6.14}$$

6.2 Central moment

Definition 9

The *kth central moment* of a r. v. X, denoted μ_k, is defined as follows:

$$\mu_k = E\left[(X - E(X))^k\right] = \begin{cases} \sum_{x_i \in X(\Omega)} (X - E(X))^k \cdot P_X(x_i) & ; \text{if } X \text{ is discrete.} \\ \\ \int_{-\infty}^{+\infty} (X - E(X))^k \cdot f(x)\,\mathrm{d}x & ; \text{if } X \text{is continuous.} \end{cases} \tag{6.15}$$

In other words, the kth central moment of a r. v. X is equivalent to the mathematical expectation of the kth power of the difference between the r. v. X and its expected value $E(X)$.

Remarks 2

- kth central moments provide an indication about the shape and dispersion of the probability distribution around the expected value.

- The first central moment $\mu_1$of a r. v. X is always zero:

$$\mu_1 = E\left[X - E\left(X\right)\right] = 0. \tag{6.16}$$

- The second central moment μ_2 of a r. v. X is its variance:

$$\mu_2 = E\left[\left(X - E\left(X\right)\right)^2\right] = V\left(X\right). \tag{6.17}$$

- The second central moment μ_2 of a r. v. X can be expressed in terms of the second and first non-central moments:

$$\mu_2 = E\left(X^2\right) - E\left(X\right)^2 = m_2 - m_1^2. \tag{6.18}$$

CHAPTER 7

USUAL PROBABILITY DISTRIBUTIONS

1 Introduction

Discrete (or continuous) random variables are associated with their own specific probability distributions (or density functions) and cumulative distribution functions. In this chapter, we will take a look at some of the most common ones, including the Bernoulli, binomial and Poisson distributions for discrete r. v., and the uniform, exponential and normal probability density functions for continuous r. v.

2 Discrete distributions

2.1 Bernoulli distribution

Consider a random experiment with two possible outcomes: success (labelled as 1) and failure (labelled as 0). If success is associated with event A with probability $P(A) = p$, then failure is associated with event $\overline{A}$ with the probability $P\left(\overline{A}\right) = 1 - p$.

We define the discrete r. v. X that can only take the two values associated with success or failure in this single trial:

$$X(\Omega) = \{0, 1\},$$

with:

$$\begin{aligned} P(X = 1) &= p \\ P(X = 0) &= 1 - p \end{aligned}$$

The r. v. X is said to follow a ***Bernoulli distribution*** with a parameter of p.

Definition 1

The *Bernoulli distribution* with a parameter of p is a discrete probability distribution in which the r. v. X takes the value 1, corresponding to a success, with probability $P(X=1)=p$, or the value 0, corresponding to a failure, with probability $P(X=0)=1-p$.

We write $X \rightsquigarrow \mathcal{B}(p)$.

- Its probability distribution is given by the following formula:

$$P(X=x_i)=p^{x_i}(1-p)^{1-x_i} \quad ; x_i \in \{0.1\}. \tag{7.1}$$

- Its cumulative distribution function is defined as follows:

$$F_X(x)=\begin{cases} 0 & ; x<0 \\ 1-p & ; 0 \leq x<1 \\ 1 & ; x \geq 1. \end{cases} \tag{7.2}$$

Mean and variance of Bernoulli distribution

From Definition 6 of Chapter 6, and equation 7.1, the mathematical expectation of the Bernoulli distribution can be written as follows:

$$\begin{aligned} E(X) &= \sum_{i=1}^{i=2} x_i \cdot p^{x_i}(1-p)^{1-x_i} \\ &= 0 \cdot p^0(1-p)+1 \cdot p(1-p)^0 \\ &= p. \end{aligned}$$

From Definition 7 of Chapter 6, the variance of the Bernoulli distribution can be written as follows:

$$\begin{aligned} V(X) &= E\left(X^2\right)-E(X)^2 \\ &= \left[0^2 \cdot p^0(1-p)+1^2 \cdot p(1-p)^0\right]-p^2 \\ &= p-p^2 \\ &= p(1-p). \end{aligned}$$

Hence the mean and variance of Bernoulli distribution are:

$$\boxed{\begin{aligned} E(X) &= \quad p \\ V(X) &= \quad p(1-p) \end{aligned}} \tag{7.3}$$

Example 1

In an urn containing 5 red balls and 3 green balls, one ball is removed from the urn. Let X the r. v. defined as follows:

$$X = \begin{cases} 1 & ; \text{ if the ball is green} \\ 0 & ; \text{ otherwise.} \end{cases}$$

We have $X \leadsto \mathcal{B}(p)$ with:

$$p = \frac{C_3^1}{C_8^1} = \frac{3}{8}.$$

Thus:

$$\begin{aligned} E(X) &= \frac{3}{8} \\ V(X) &= \frac{3}{8}\left(1 - \frac{3}{8}\right) = \frac{15}{64}. \end{aligned}$$

2.2 Binomial distribution

Consider n independent random experiments, each with two possible outcomes for each experiment: success, with probability p, and failure, with probability $(1-p)$.

We define the discrete r. v. X as the number of successes in these n independent trials:

$$X(\Omega) = \{0,\, 1,\, \ldots,\, n\},$$

with:

$$X = \sum_{k=1}^{k=n} X_k \quad ; X_k \leadsto \mathcal{B}(p). \tag{7.4}$$

The r. v. X is the sum of n independent r. v. X_k, each of which follows a Bernoulli distribution with a parameter p.

X is said to follow a ***Binomial distribution*** with parameters n and p.

Definition 2

The *Binomial distribution*, with parameters n and p, is a discrete probability distribution in which the r. v. X takes the number of successes, with probability p, obtained through n independent random experiments.

We write $X \leadsto \mathcal{B}(n,\, p)$.

- Its probability distribution is given by the following formula:

$$P(X = k) = C_n^k\, p^k \cdot (1-p)^{n-k} \quad ; k \in \{0,\, 1,\, \ldots,\, n\}. \tag{7.5}$$

- Its cumulative distribution function is defined as follows:

$$F_X(x) = \begin{cases} 0 & ; x < 0 \\ n(1-p)^n & ; 0 \leq x < 1 \\ n(1-p)^{n-1} & ; 1 \leq x < 2 \\ \vdots & \vdots \\ 1 & ; x \geq n. \end{cases} \tag{7.6}$$

In the binomial distribution formula:

- C_n^k is the binomial coefficient representing the number of ways of achieving k successes out of n independent trials.
- p^k is the probability of obtaining exactly k successes.
- $(1-p)^{n-k}$ is the probability of obtaining exactly $n-k$ failures.

Mean and variance of binomial distribution

From Equations 6.5 and 7.4, the mathematical expectation of the Binomial distribution can be written as follows:

$$\begin{aligned} E(X) &= E\left(\sum_{k=1}^{k=n} X_k\right) \\ &= \sum_{k=1}^{k=n} E(X_k) \\ &= \sum_{k=1}^{k=n} p. \\ &= np \end{aligned}$$

From Equations 6.12 and 7.4, the variance of the Binomial distribution can be written as follows:

$$\begin{aligned} V(X) &= V\left(\sum_{k=1}^{k=n} X_k\right) \\ &= \sum_{k=1}^{k=n} V(X_k) \\ &= \sum_{k=1}^{k=n} p(1-p) \\ &= np(1-p). \end{aligned}$$

Hence the mean and variance of Binomial distribution are:

$$\boxed{\begin{aligned} E(X) &= np \\ V(X) &= np(1-p) \end{aligned}} \tag{7.7}$$

Example 2

Consider the outcome of flipping a fair coin 10 times. Each flip is a trial with 2 possible outcomes: Head or Tail. Assuming a fair coin, the probability of getting Heads or Tails is the same for each flip ($p = 0.5$). The outcome of one coin flip does not influence the outcome of any other. These trials are therefore independent.

Let X the r. v. defined as "the number of Heads obtained after 10 trials".
We have $X \rightsquigarrow \mathcal{B}(10, 0.5)$ and its probability distribution is given by:

$$P(X = k) = C_{10}^{k}\ (0.5)^{k} \cdot (0.5)^{10-k} \quad ; k \in \{0, 1, \ldots, 10\}.$$

Thus:

- We can determine, for instance, the probability of getting 2 Heads:

$$\begin{aligned} P(X = 2) &= C_{10}^{2}\ (0.5)^{2} \cdot (0.5)^{8} \\ &= \frac{10!}{2! \cdot 8!}\,(0.5)^{10} \\ &\approx 0.0439. \end{aligned}$$

- We can also determine, for instance, the probability of getting at least 3 Heads:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[C_{10}^{0}\ (0.5)^{10} + C_{10}^{1}\ (0.5)^{10} + C_{10}^{2}\ (0.5)^{10}\right] \\ &= 1 - 56\,(0.5)^{10}. \\ &\approx 0.9453. \end{aligned}$$

- The mathematical expectation and variance of the r. v. are given by:

$$\begin{aligned} E(X) &= 10 \cdot 0.5 = 5 \\ V(X) &= 10 \cdot 0.5 \cdot (1 - 0.5) = 2.5. \end{aligned}$$

Remark 1

The need to calculate probabilities manually using the binomial distribution formula, particularly for samples of a large size, is eliminated by using a binomial distribution table (see Appendix 3.5.2), which contains precomputed probabilities for various numbers of successes within a fixed number of independent trials, given a specified probability of success for each trail.

2.3 Poisson's distribution

Poisson's distribution is a powerful modelling tool for random processes where the focus is on the number of events occurring within a given time interval, provided these events are both rare and independent of each other.

For instance, Poisson's distribution can be used to model the number of customers visiting a store per hour, or the number of phone calls received by a call centre per minute, or the number of suicides per year, ...

Thus, we can define a discrete r. v. X as the number of these independent events such that :

$$X(\Omega) = \{n \mid n \in \mathbb{N}\}.$$

Definition 3

The *Poisson's distribution*, with parameter $\lambda > 0$, is a discrete probability distribution in which the r. v. X takes a specific number of events occurring randomly and independently within a fixed interval T of time or space, given the average rate of occurrence θ.
We write $X \rightsquigarrow \mathcal{P}(\lambda)$.
Its probability distribution is given by the following formula:

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^{k}}{k!} \quad ; k \in \mathbb{N}, \tag{7.8}$$

with $\lambda = \theta \cdot T$.

Mean and variance of Poisson's distribution

From Definition 6 of Chapter 6, and equation 7.8, the mathematical expectation of the Poisson's distribution can be written as follows:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda} \cdot \lambda^{k}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} k \cdot \frac{\lambda^{k}}{k!} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k}}{k!} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j}}{j!} \quad ; j = k - 1. \end{aligned}$$

Since the sum in the second term of the final identity is the Taylor series of e^{λ}, we obtain:

$$E(X) = \lambda.$$

From Definition 7 of Chapter 6, the variance of the Poisson's distribution can be written as follows:

$$\begin{aligned} V(X) &= E\left(X^{2}\right) - E(X)^{2} \\ &= \sum_{k=0}^{\infty} k^{2} \cdot \frac{e^{-\lambda} \cdot \lambda^{k}}{k!} - \lambda^{2} \\ &= \sum_{k=1}^{\infty} k^{2} \cdot \frac{e^{-\lambda} \cdot \lambda^{k}}{k!} - \lambda^{2} \\ &= \lambda \left[\sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda} \cdot \lambda^{k-1}}{(k-1)!} \right] - \lambda^{2} \\ &= \lambda \left[\sum_{j=0}^{\infty} (j+1) \cdot \frac{e^{-\lambda} \cdot \lambda^{j}}{j!} \right] - \lambda^{2} \quad ; j = k - 1 \\ &= \lambda \left[\sum_{j=0}^{\infty} j \cdot \frac{e^{-\lambda} \cdot \lambda^{j}}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda} \cdot \lambda^{j}}{j!} \right] - \lambda^{2}. \end{aligned}$$

Since we have:

$$\sum_{j=0}^{\infty} j \cdot \frac{e^{-\lambda} \cdot \lambda^j}{j!} = E\left(X\right) = \lambda,$$

and

$$\sum_{j=0}^{\infty} \frac{e^{-\lambda} \cdot \lambda^j}{j!} = \sum_{j=0}^{\infty} P\left(X = j\right) = 1,$$

we can therefore write:

$$\begin{aligned} V\left(X\right) &= \lambda\left[\lambda + 1\right] - \lambda^2. \\ &= \lambda. \end{aligned}$$

Hence the mean and variance of Poisson's distribution are:

$$\boxed{E\left(X\right) = \quad V\left(X\right) = \lambda.} \tag{7.9}$$

Example 4

Given that the average number of customers visiting a store per hour is 4:

- Let us calculate the probability that there will be more than 4 customers visiting the store in 1 h.
  To do this, we will consider the r. v. X associated with the number of customers visiting the store per hour.
  We have $X \rightsquigarrow \mathcal{P}\left(\lambda\right)$ with $\lambda = 4$, and its probability distribution is as follows:

  $$P\left(X = k\right) = \frac{e^{-4} \cdot 4^k}{k!} \quad ; k \in \mathbb{N},$$

  such that the required probability is given by:

  $$\begin{aligned} P\left(X > 4\right) &= 1 - P\left(X \leq 4\right) \\ &= 1 - \left[P\left(X = 0\right) + P\left(X = 1\right) + P\left(X = 2\right) + P\left(X = 3\right) + P\left(X = 4\right)\right] \\ &= 1 - \left[\frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!} + \frac{e^{-4} \cdot 4^3}{3!} + \frac{e^{-4} \cdot 4^4}{4!}\right] \\ &= 1 - \frac{103}{3} e^{-4} \\ &= 0.371. \end{aligned}$$

- To calculate the probability of there being at least 3 customers visiting the store in 45 mn, we must consider the r. v. X associated with the number of customers visiting the store per 45 mn such that $X \rightsquigarrow \mathcal{P}\left(\lambda\right)$ with $\lambda = 4 \cdot \frac{3}{4} = 3$, and its probability distribution is as follows:

  $$P\left(X = k\right) = \frac{e^{-3} \cdot 3^k}{k!} \quad ; k \in \mathbb{N},$$

  The required probability is therefore given by:

$$\begin{aligned} P\left(X \geq 3\right) &= 1 - P\left(X < 3\right) \\ &= 1 - \left[P\left(X = 0\right) + P\left(X = 1\right) + P\left(X = 2\right)\right] \\ &= 1 - \left[\frac{e^{-3} \cdot 3^{0}}{0!} + \frac{e^{-3} \cdot 3^{1}}{1!} + \frac{e^{-3} \cdot 3^{2}}{2!}\right] \\ &= 1 - \frac{17}{2}e^{-3}. \\ &= 0.576. \end{aligned}$$

3 Continuous distributions

3.1 Continuous uniform distribution

A ***Continuous uniform distribution*** describes a situation in which all outcomes within a defined interval are equally likely to occur.

Definition 4

A continuous r. v. X is said to exhibit a *continuous uniform distribution* over the interval $[a, b]$, denoted $X \rightsquigarrow \mathcal{U}\left([a, b]\right)$, if :

- Its probability density function is defined as follows:

$$f\left(x\right) = \begin{cases} \frac{1}{b-a} & ; \text{if } x \in [a, b] \\ 0 & ; \text{otherwise.} \end{cases} \tag{7.10}$$

- Its cumulative distribution function is defined as follows:

$$F_X\left(x\right) = \begin{cases} 0 & ; \text{if } x < a \\ \frac{x-a}{b-a} & ; \text{if } a \leq x < b \\ 1 & ; \text{if } x \geq b. \end{cases} \tag{7.11}$$

This means that all numbers within the interval $[a, b]$ have the same probability density function $f\left(x\right)$. The total area under the $f\left(x\right)$ curve represents the total probability and is equal to 1.

Mean and variance of continuous uniform distribution

From Definition 6 of Chapter 6, and equation 7.10, the mathematical expectation of the continuous uniform distribution can be written as follows:

$$\begin{aligned} E\left(X\right) &= \int_{-\infty}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= \int_{-\infty}^{a} x \cdot f\left(x\right) \mathrm{d}x + \int_{a}^{b} x \cdot f\left(x\right) \mathrm{d}x + \int_{b}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= \int_{a}^{b} \frac{x}{b-a} \mathrm{d}x \\ &= \frac{1}{b-a}\left[\frac{b^2 - a^2}{2}\right] \\ &= \frac{b+a}{2}. \end{aligned}$$

Thus, the mean of the continuous uniform distribution over $[a, b]$ is the average of the parameters a and b,or the midpoint of the given interval.

From Definition 7 of Chapter 6, the variance of the continuous uniform distribution can be written as follows:

$$\begin{aligned} V(X) &= E\left(X^2\right) - E(X)^2 \\ &= \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,\mathrm{d}x - E(X)^2 \\ &= \int_a^b \frac{x^2}{b-a}\mathrm{d}x - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right] - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

Hence the mean and variance of continuous uniform distribution are:

$$\boxed{\begin{aligned} E(X) &= \frac{b+a}{2} \\ V(X) &= \frac{(b-a)^2}{12}. \end{aligned}} \tag{7.12}$$

Example 5

If a bus arrives every 2 hours, for instance between 8 amand 10 am, at a random time within this interval, the arrival time can be considered as a r. v. X following a continuous uniform distribution.

We write $X \rightsquigarrow \mathcal{U}([8,\ 10])$.
Its probability density function is as follows:

$$f(x) = \begin{cases} \frac{1}{2} & ;\text{ if } x \in [8,\ 10] \\ 0 & ;\text{ otherwise,} \end{cases}$$

and its cumulative distribution function is as follows:

$$F_X(x) = \begin{cases} 0 & ;\text{ if } x < 8 \\ \frac{x-8}{2} & ;\text{ if } 8 \leq x < 10 \\ 1 & ;\text{ if } x \geq 10. \end{cases}$$

Thus:

- We can determine the probability, for instance, that the arrival time will be before 9 am:

$$P(X < 9) = F_X(9) = \frac{9-8}{2} = 0.5.$$

- We can also determine, for instance, the probability that the arrival time is between 8 am and 9 am:

$$P(8 < X < 9) = F_X(9) - F_X(8) = 0.5.$$

- We can calculate the mean and the variance of the r. v. X given by:

$$\begin{aligned} E(X) &= \frac{8+10}{2} = 9 \\ V(X) &= \frac{(10-8)^2}{12} = \frac{1}{3}. \end{aligned}$$

3.2 Exponential distribution

The ***exponential distribution*** is commonly employed to describe the time it takes for an event to occur, given the average number of events per unit of time.

Definition 5

A continuous r. v. X is said to exhibit an *exponential distribution* with parameter $\lambda > 0$, denoted $X \rightsquigarrow Exp\left(\lambda\right)$, if :

- Its probability density function is defined as follows:

$$f\left(x\right) = \begin{cases} \lambda e^{-\lambda x} & ; \text{if } x \geq 0 \\ 0 & ; \text{otherwise.} \end{cases} \tag{7.13}$$

- Its cumulative distribution function is defined as follows:

$$F_X\left(x\right) = \begin{cases} 0 & ; \text{if } x < 0 \\ 1 - e^{-\lambda x} & ; \text{if } x \geq 0. \end{cases} \tag{7.14}$$

The exponential distribution formula shows how likely it is that a specific time period x will elapse before an event occurs.

The rate parameter λ is positive constant indicating the average number of events occurring in a given time. The greater the value of λ, the greater the likelihood of an event occurring at an earlier time.

Mean and variance of Exponential distribution

From Definition 6 of Chapter 6, and equation 7.13, the mathematical expectation of the the exponential distribution, with rate parameter λ, can be written as follows:

$$\begin{aligned} E\left(X\right) &= \int_{-\infty}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= \int_{-\infty}^{0} x \cdot f\left(x\right) \mathrm{d}x + \int_{0}^{+\infty} x \cdot f\left(x\right) \mathrm{d}x \\ &= \int_{0}^{+\infty} x \cdot \lambda e^{-\lambda x} \mathrm{d}x. \end{aligned}$$

Applying the technique of integration by parts with $u = x$ and $\mathrm{d}v = \lambda e^{-\lambda x}\mathrm{d}x$, we obtain:

$$E\left(X\right) = \left[-xe^{-\lambda x}\right]_0^{\infty} + \int_0^{+\infty} e^{-\lambda x}\mathrm{d}x.$$

Since $\lim_{x\to+\infty} xe^{-\lambda x} = 0$, we get 0 for the fist term, and thus:

$$E\left(X\right) = \int_0^{+\infty} e^{-\lambda x}\mathrm{d}x = \left[-\frac{e^{-\lambda x}}{\lambda}\right]_0^{+\infty} = \frac{1}{\lambda}.$$

From Definition 7 of Chapter 6, the variance of the exponential distribution can be written as follows:

$$\begin{aligned}
V(X) &= E\left(X^2\right) - E(X)^2 \\
&= \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,\mathrm{d}x - E(X)^2 \\
&= \int_{0}^{+\infty} \lambda x^2 e^{-\lambda x}\mathrm{d}x - \frac{1}{\lambda^2}.
\end{aligned}$$

Applying twice the technique of integration by parts, we obtain:

$$\begin{aligned}
V(X) &= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x}\mathrm{d}x - \frac{1}{\lambda^2} \\
&= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + \left[-\frac{2x}{\lambda} e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} \frac{2}{\lambda} e^{-\lambda x}\mathrm{d}x - \frac{1}{\lambda^2} \\
&= \left[\left(-x^2 - \frac{2x}{\lambda} - \frac{2}{\lambda^2}\right) e^{-\lambda x}\right]_0^{+\infty} - \frac{1}{\lambda^2} \\
&= \left[0 + \frac{2}{\lambda^2}\right] - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}.
\end{aligned}$$

Hence the mean and variance of Exponential distribution are:

$$\boxed{\begin{aligned} E(X) &= \frac{1}{\lambda} \\ V(X) &= \frac{1}{\lambda^2}. \end{aligned}} \tag{7.15}$$

Example 6

In a call centre, the duration of phone calls is modelled using an exponential distribution to describe the time until the next event occurs.

Let X be the r. v. associated with this duration. We write $X \rightsquigarrow \mathit{Exp}(\lambda)$.
Given an average call duration of 4 mn, for instance, we can deduce the rate parameter as follows:

$$E(X) = \frac{1}{\lambda} \Longrightarrow \lambda = \frac{1}{E(X)} = \frac{1}{4}.$$

Then $X \rightsquigarrow \mathit{Exp}\left(\frac{1}{4}\right)$. Its probability density function is defined as follows:

$$f(x) = \begin{cases} \frac{1}{4} e^{-\frac{x}{4}} & ; \text{if } x \geq 0 \\ 0 & ; \text{otherwise,} \end{cases}$$

and its cumulative distribution function is defined as follows:

$$F_X(x) = \begin{cases} 0 & ; \text{if } x < 0 \\ 1 - e^{-\frac{x}{4}} & ; \text{if } x \geq 0. \end{cases}$$

Thus:

- We can calculate the probability of a call being longer than 5 mn, for example:

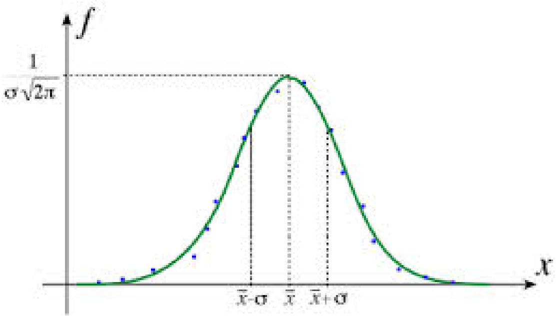
$$\begin{aligned} P\left(X>5\right) &= 1-P\left(X\leq 5\right) \\ &= 1-F_X\left(5\right) \\ &= 1-\left(1-e^{-\frac{5}{4}}\right) \\ &= e^{-\frac{5}{4}} \\ &\approx 0.287. \end{aligned}$$

- We can also calculate the probability of a call lasting between 2 mnand 5 mn:

$$\begin{aligned} P\left(2<X<5\right) &= F_X\left(5\right)-F_X\left(2\right) \\ &= \left(1-e^{-\frac{5}{4}}\right)-\left(1-e^{-\frac{2}{4}}\right) \\ &= e^{-\frac{1}{2}}-e^{-\frac{5}{4}} \\ &\approx 0.32. \end{aligned}$$

3.3 Normal distribution or Laplace-Gauss distribution

The ***normal distribution***, also referred to as the ***Gaussian distribution*** or the ***Laplace–Gauss distribution*** (in recognition of the contributions of Pierre-Simon Laplace and Carl Friedrich Gauss), is a continuous probability distribution characterised by a graph of a bell-shaped curve (see the opposite figure). The curve shows that most of the data points are concentrated around the mean and taper off towards the ends of the curve.



In the fields of probability theory and statistics, the normal distribution is the most widely used for modelling a variety of natural phenomena originating from many random events. Normal distributions correspond to the behaviour of a series of random experiments under certain conditions when the number of experiments is very high. This property means that the normal distribution can be applied to approximate other distributions, enabling many scientific studies to be modelled, such as error measurements or statistical tests.

Definition 6

A continuous r. v. X is said to exhibit a *normal, Gaussian* or *Laplace-Gauss distribution,* with real parameters μ and $\sigma>0$, denoted $X\leadsto\mathcal{N}\left(\mu,\,\sigma\right)$, if its probability density function is defined as follows:

$$f\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(x-\mu\right)^2}{2\,\sigma^2}\right). \tag{7.16}$$

μ is the mean of the distribution, while σ is the standard deviation.

$\mu=E\left(X\right)$ is located at the centre of the probability density function curve.

$\sigma=\sqrt{V\left(X\right)}$ indicates how spread out the data is around the mean.

Properties

If $X \rightsquigarrow \mathcal{N}(\mu_1,\, \sigma_1)$ and $Y \rightsquigarrow \mathcal{N}(\mu_2,\, \sigma_2)$ are two independent r. v., then:

- $(X+Y) \rightsquigarrow \mathcal{N}\left(\mu_1+\mu_2,\, \sqrt{\sigma_1^2+\sigma_2^2}\right)$.
- $(X-Y) \rightsquigarrow \mathcal{N}\left(\mu_1-\mu_2,\, \sqrt{\sigma_1^2+\sigma_2^2}\right)$.

Remarks 1

- The cumulative distribution function of the normal distribution, given by:

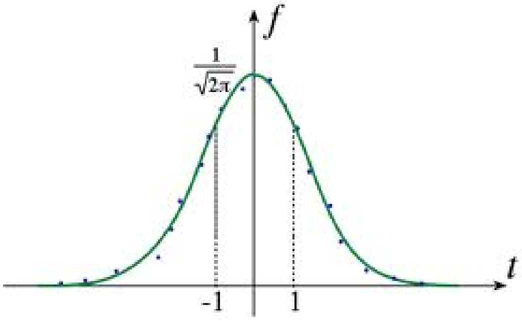
$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\,\sigma^2}\right) \mathrm{d}t, \tag{7.17}$$

  has no analytical expression that can be easily derived. Its values can be accessed via the usual statistical tables (see Appendix 3.5.2).

- The mean, median and mode are all equal in a normal distribution.
- Around 68 %of the data is within one standard deviation of the mean (between $\mu-\sigma$ and $\mu+\sigma$), 95 % is within two (between $\mu-2\sigma$ and $\mu+2\sigma$), and 99.7 % is within three (between $\mu-3\sigma$ and $\mu+3\sigma$). This is what we call the $68-95-99.7$ ***rule***.

3.4 Standard normal distribution

The ***standard normal distribution***, also known as the ***unit normal distribution***, is the simplest case of normal distribution when $\mu = 0$ and $\sigma = 1$ (see the opposite figure).



Definition 7

A continuous r. v. X is said to exhibit a *standard normal* or *unit distribution*, denoted $X \rightsquigarrow \mathcal{N}(0,\, 1)$, if its probability density function is defined as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{7.18}$$

The mean of the distribution is $\mu = 0$, and the standard deviation is $\sigma = 1$.

The probability density function of a standard normal distribution $\mathcal{N}(0,\, 1)$ is easily derived from a general normal distribution $\mathcal{N}(\mu,\, \sigma)$, by normalising the normal r. v. This involves the transformation of the original r. v. $X \rightsquigarrow \mathcal{N}(\mu,\, \sigma)$ by subtraction of the mean and division by the standard deviation. This results in a standard normal r. v. $Z = \frac{X-\mu}{\sigma} \rightsquigarrow \mathcal{N}(0,\, 1)$ with an expected value $E(Z) = \mu = 0$, and a standard deviation $\sigma = 1$.

From the cumulative distribution function of Z,given by:

$$\begin{aligned} F_Z(x) &= P(Z \leq x) \\ &= P\left(\frac{X-\mu}{\sigma} \leq x\right) \\ &= P(X \leq \sigma x + \mu) \\ &= F_X(\sigma x + \mu), \end{aligned}$$

we can deduce the probability density function as follows:

$$\begin{aligned} f_Z(x) &= \frac{\mathrm{d}F_Z(x)}{\mathrm{d}x} \\ &= \frac{\mathrm{d}F_X(\sigma x + \mu)}{\mathrm{d}x} \\ &= \sigma f_X(\sigma x + \mu) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\sigma x + \mu - \mu)^2}{2\,\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \end{aligned}$$

Properties

For the two r. v. $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$, and $Z = \frac{X-\mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$:

- $P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right)$.
- $P(Z < -x) = P(Z > x) = 1 - P(Z \leq x) = 1 - F_Z(x)$.
- $P(a < Z < b) = F_Z(b) - F_Z(a); \forall\, a, \in \mathbb{R}$.

Remark 2

The cumulative distribution function of any normal distribution $\mathcal{N}(\mu, \sigma)$ can be calculated in two steps.

- Firstly, the value x is converted to $z = \frac{x-\mu}{\sigma}$:

$$F_X(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z) = F_Z(z).$$

- Secondly, the standard normal cumulative distribution function $F_Z(z)$ from the statistical standard normal distribution table (see Appendix 3.5.2) is determined.

The cumulative distribution function $F_Z(z)$ associated with a given value z can be found in the standard normal distribution table, also known as z-table. Firstly, we locate the integer and decimal places of the z value in the column on the left, and the hundredths place in the row above. The intersection of the corresponding row and column indicates the value of the cumulative distribution function, representing the area under the probability density function curve from infinity to the z value.

Example 7

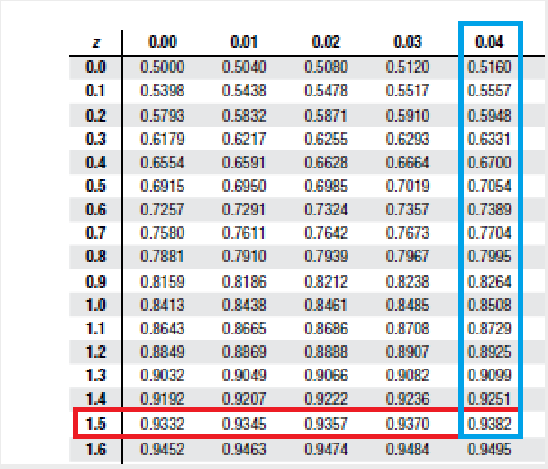
I. Let us consider a r. v. $Z \rightsquigarrow \mathcal{N}(0,\,1)$.
We will use the standard normal distribution table to determine $P(Z < 1.54)$, $P(Z < -1.54)$, and $P(0.7 < Z < 1.54)$:

- For the first probability, we have:

$$P(Z < 1.54) = F_Z(1.5)$$

The intersection of the row corresponding to the integer and decimal places and the column corresponding to the hundredths place of $z = 1.54$, in the standard normal distribution table, indicates that:

$$P(Z < 1.54) = F_Z(1.5) = 0.9382.$$



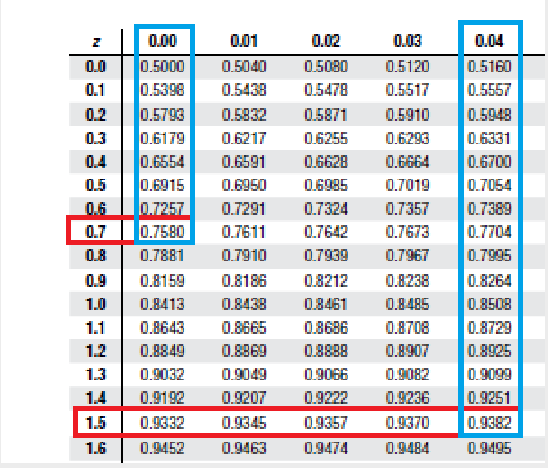
z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495

- For the second probability, we can write:

$$P(0.7 < Z < 1.54) = F_Z(1.54) - F_Z(0.7)\,.$$

In the same manner as for the value $z = 1.54$, the z-table indicates that $F_Z(0.7) = 0.7580$. Then:

$$P(0.7 < Z < 1.54) = 0.9382 - 0.7580 = 0.1802.$$



z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495

II. Let us consider a r. v. $X \rightsquigarrow \mathcal{N}(2,\,4)$.
We will use the standard normal distribution table to determine $P(X > 6)$.

- Firstly, we write:

$$P(X > 6) = 1 - P(X \leq 6) = 1 - P\left(Z \leq \frac{6-2}{4}\right) = 1 - P(Z \leq 1) = 1 - F_Z(1)\,,$$

with $Z \rightsquigarrow \mathcal{N}(0,\, 1)$.

- From the z-table, we locate $F_Z(1) = 0.8413$. Then:

$$P(X > 6) = 1 - 0.8413 = 0.1587.$$

III. Let us consider a r. v. $Z \rightsquigarrow \mathcal{N}(0,\, 1)$.
We will use the standard normal distribution table to determine the value z for which the probability is $P(Z < z) = 0.8670$.

From the z-table, we find that $F_Z(1.11) = 0.8665$, and $F_Z(1.12) = 0.8686$. We deduce that the approached value of z is $z = \dfrac{1.11 + 1.12}{2} = 1.115$.

3.5 Derived distributions from the standard normal distribution

The standard normal distribution can be used to derive several probability distributions. These include the chi-square distribution, the student's distribution, the F-distribution and the log-normal distribution. Here, we focus on the first two distributions.

3.5.1 Chi-squared (Pearson's) distribution

The ***chi-squared distribution*** is a continuous probability distribution that is particularly important in the context of statistical hypothesis testing. It is defined by an integer number of degrees of freedom, and used in tests such as the chi-squared test of independence.

Definition 8

If we consider n independent r. v. $X_1, X_2, \ldots, X_i, \ldots, X_n$, such that $X_i \rightsquigarrow \mathcal{N}(0,\, 1)\,;\, \forall\, i = \overline{1, n}$, then the r. v. $Y = \sum_{i=1}^{i=n} X_i^2$ is said to follow a *chi-squared distribution* with n degrees of freedom.

We write $Y \rightsquigarrow \chi^2(n)$.
Its probability density function is defined as follows:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)}\, x^{\frac{n}{2}-1} \cdot \exp\left(-\frac{x}{2}\right) & ;\text{ if } x \geq 0 \\ 0 & ;\text{ otherwise,} \end{cases} \tag{7.19}$$

with $\Gamma(a) = \int_0^{+\infty} x^{a-1} \cdot e^{-x} \mathrm{d}x;\ \forall\, a \in \mathbb{N}$.

The mathematical expectation of the the chi-squared distribution, with n degrees of freedom, can be written as follows:

$$\begin{aligned} E(Y) &= E\left(\sum_{i=1}^{i=n} X_i^2\right) \\ &= \sum_{i=1}^{i=n} E\left(X_i^2\right) \end{aligned}$$

Knowing that $X_i \rightsquigarrow \mathcal{N}(0,\, 1)$, we have $E(X_i) = 0$, and $V(X_i) = E\left(X_i^2\right) - E(X_i)^2 = 1$, then $E\left(X_i^2\right) = 1$, and:

$$E(Y) = \sum_{i=1}^{i=n} 1 = n.$$

The variance of the chi-squared distribution, with n degrees of freedom, can be written as follows:

$$\begin{aligned}V(Y) &= V\left(\sum_{i=1}^{i=n} X_i^2\right)\\ &= \sum_{i=1}^{i=n} V\left(X_i^2\right)\\ &= \sum_{i=1}^{i=n}\left[E\left(X_i^4\right) - E\left(X_i^2\right)^2\right]\\ &= \sum_{i=1}^{i=n}\left[E\left(X_i^4\right) - 1\right]\end{aligned}$$

with:

$$E\left(X_i^4\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^4 e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x = \frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty} x^4 e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x,$$

because the function $x^4 e^{\left(-\frac{x^2}{2}\right)}$ is even.

Applying the technique of integration by parts for this last identity, with $u = x^3$ and $\mathrm{d}v = xe^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x$, we obtain:

$$\begin{aligned}E\left(X_i^4\right) &= \frac{2}{\sqrt{2\pi}}\left(\left[-x^3 e^{\left(-\frac{x^2}{2}\right)}\right]_0^{+\infty} + 3\int_0^{+\infty} x^2 e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x\right)\\ &= \frac{6}{\sqrt{2\pi}}\int_0^{+\infty} x^2 e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x.\end{aligned}$$

Applying again the technique of integration by parts for this last identity, with $u = x$ and $\mathrm{d}v = xe^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x$, we obtain:

$$\begin{aligned}E\left(X_i^4\right) &= \frac{6}{\sqrt{2\pi}}\left(\left[-xe^{\left(-\frac{x^2}{2}\right)}\right]_0^{+\infty} + \int_0^{+\infty} e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x\right)\\ &= \frac{6}{\sqrt{2\pi}}\left(\frac{1}{2}\int_{-\infty}^{+\infty} e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x\right).\\ &= 3\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x\right)\end{aligned}$$

Since $\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x = 1$ (from the definition of a probability density function), we obtain $E\left(X_i^4\right) = 3$, and then:

$$V(Y) = \sum_{i=1}^{i=n}\left[E\left(X_i^4\right) - 1\right] = \sum_{i=1}^{i=n} 2 = 2n.$$

Thus, the mean and variance of chi-squared distribution, with n degrees of freedom, are:

$$\boxed{\begin{aligned}E(Y) &= \quad n\\ V(Y) &= \quad 2n.\end{aligned}} \tag{7.20}$$

Remark 3

As for the standard normal distribution, the cumulative distribution function of the chi-squared distribution, with n degrees of freedom can be calculated using a chi-squared distribution table (see Appendix 3.5.2). Firstly, we locate the number of degrees of freedom in the column on the left, and the significant level corresponding to the test in the row above. The intersection of the corresponding column and row indicates the critical chi-squared value for the test.

Unlike the z-table, the critical chi-squared values in this table correspond to the right area of the probability density function, representing $P(Y > y)$.

Example 8

I. Let us consider a r. v. $Y \rightsquigarrow \chi^2(30)$.
We will use the chi-squared distribution table to determine the values y_1 and y_2 such that $P(Y > y_1) = 0.01$, and $P(Y \leq y_2) = 0.5$.

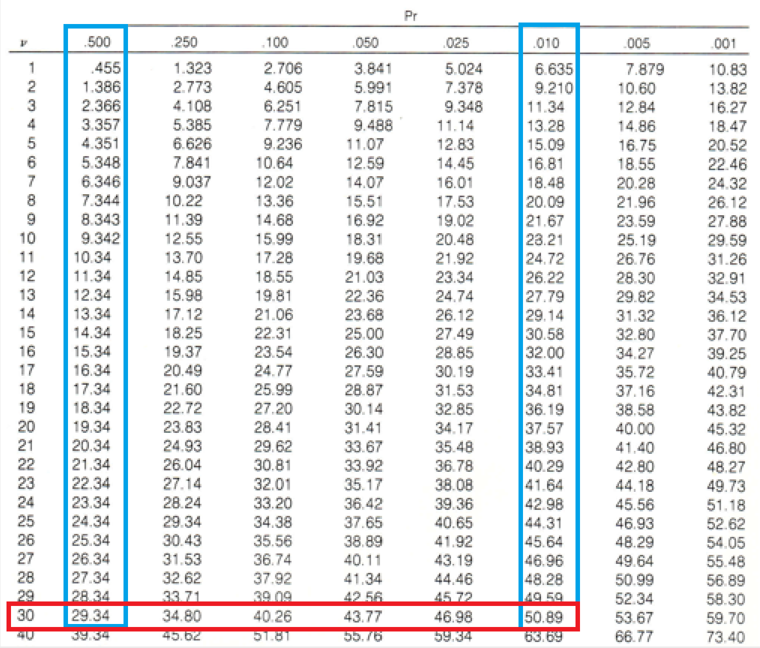
- For the first probability, y_1 is indicated by the intersection of the row corresponding to number of degrees of freedom and the column corresponding to the desired significant level for the test.
Thus, we find that $y_1 = 50.89$.
- For the second probability, we can write:

$$P(Y \leq y_2) = 1 - P(Y > y_2) = 0.5.$$

Then:

$$P(Y > y_2) = 1 - 0.5 = 0.5.$$

From the chi-squared table, we locate and find that $y_2 = 29.34$.



	Pr							
ν	.500	.250	.100	.050	.025	.010	.005	.001
1	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	7.344	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.342	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.79	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	20.34	24.93	29.62	33.67	35.48	38.93	41.40	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40

II. Let us consider a r. v. $Y \rightsquigarrow \chi^2(30)$.
We will use the chi-squared distribution table to determine the significant level α such that $P(Y > 45) = \alpha$.

From the table, we find that $P(Y > 43.77) = 0.05$, and $P(Y > 46.98) = 0.025$. We deduce that the approached value of α is $\alpha = \dfrac{0.05 + 0.025}{2} = 0.0375$.

3.5.2 Student's distribution

The ***student's distribution*** is a continuous probability distribution used in hypothesis testing and constructing confidence intervals, typically for populations with unknown standard deviations and small sample sizes. It is also defined by its degrees of freedom, which related to the sample size.

Definition 9

If we consider two independent r. v. $Y \rightsquigarrow \chi^2(n)$ and $Z \rightsquigarrow \mathcal{N}(0,\,1)$, then the r. v. $T = \dfrac{Z}{\sqrt{\frac{Y}{n}}}$ is said to follow a *student's distribution* with n degrees of freedom.

We write $T \rightsquigarrow T(n)$.

Its probability density function is defined as follows:

$$f_n(x) = \frac{1}{\sqrt{n \cdot \pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{7.21}$$

with $\Gamma(a) = \int_0^{+\infty} x^{a-1} \cdot e^{-x} \mathrm{d}x;\ \forall\, a \in \mathbb{N}$.

The n parameter represents the degrees of freedom and is usually equal to $n-1$ for a sample size of n. It affects the shape of the curve which is bell-shaped with heavier tails than the standard normal distribution. The greater the increase in degrees of freedom, the closer Student's distribution becomes to the standard normal distribution.

The expected value for the Student's distribution is defined for $n > 1$ and is always zero since it is symmetrical around zero. If $n = 1$, the distribution does not have a finite mean due to its heavy tails.

The variance of a Student's distribution is derived from its probability density function by integrating a specified expression. This results in the value $\dfrac{n}{n-2}$. This formula becomes undefined when $n \leq 2$, at which point the variance diverges.

Thus, the mean and variance of Student's distribution, with n degrees of freedom, are:

$$\boxed{\begin{array}{ll} E(T) = 0 & ;\, n > 1 \\ V(T) = \dfrac{n}{n-2} & ;\, n > 2. \end{array}} \tag{7.22}$$

Remark 4

The cumulative distribution function of the Student's distribution with n degrees of freedom can be found in a Student's distribution table (see Appendix 3.5.2). Firstly, we locate the degrees of freedom in the left-hand column and the calculated t-statistic within the corresponding row. Depending on whether the test is one-tailed or two-tailed, the intersection of the corresponding row and column indicates the p-value or $\frac{p}{2}$-value for a specific t-statistic in the two rows above.

As for the chi-squared table, the critical Student's values in this table correspond to the right area of the probability density function, representing $P(T > t)$ or the left area of the probability density function, representing $P(T < -t)$.

Example 9

I. Let us consider a r. v. $T \rightsquigarrow T\,(10)$.
We will use the Student distribution table to determine the probabilities $P\,(T > 1.372)$, and $P\,(T \leq 2.764)$.

- For the first probability, $t = 1.372$ is located in the first column of the row corresponding to number of degrees of freedom. Thus, we find that:

$$P\,(T > 1.372) = \tfrac{0.200}{2} = 0.100 \quad \text{if considering the are in two tails}$$
$$P\,(T > 1.372) = 0.100 \qquad \text{if considering the are in one tail.}$$

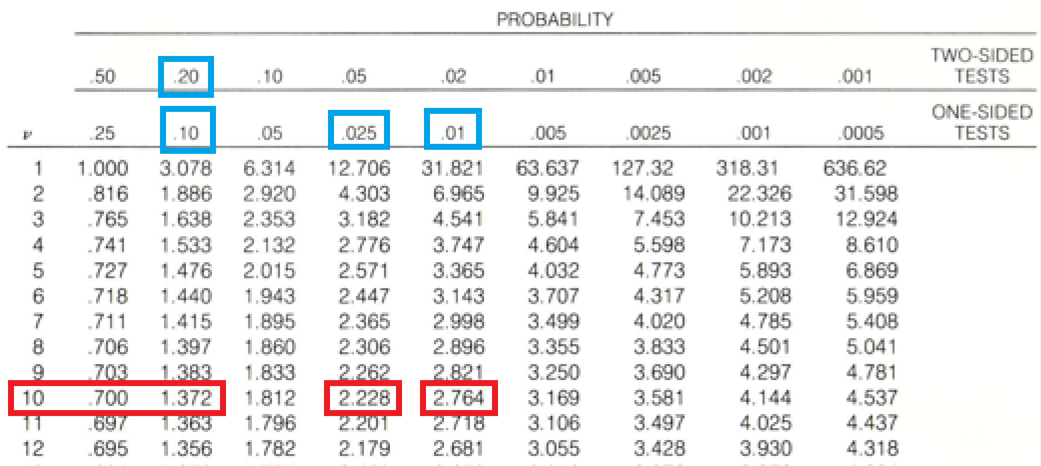
- For the second probability, we can write:

$$P\,(T \leq 2.764) = 1 - P\,(T > 2.764)\,.$$

From the Student table, we locate the value $t = 2.764$ in the row corresponding to the number of degrees of freedom. Thus, considering the area in one tail, we obtain its corresponding p-value:

$$P\,(T \leq 2.764) = 1 - 0.01 = 0.99.$$

.



	PROBABILITY									
	.50	.20	.10	.05	.02	.01	.005	.002	.001	TWO-SIDED TESTS
ν	.25	.10	.05	.025	.01	.005	.0025	.001	.0005	ONE-SIDED TESTS
1	1.000	3.078	6.314	12.706	31.821	63.637	127.32	318.31	636.62	
2	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598	
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924	
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408	
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537	
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	

II. Let us consider a r. v. $T \rightsquigarrow T\,(10)$.
We will use the Student distribution table to determine the t-statistic such that $P\,(T > t) = 0.025$.

From the table, the intersection of the row corresponding to the number of degrees of freedom and the column corresponding to p-value, considering the area in one tail, gives $t = 2.288$.

Appendix : Statistical tables

A- BINOMIAL DISTRIBUTION

Numerical entries represent $P(X = x)$.

						p				
n	**x**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
1	**0**	0.9000	0.8000	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1000
	1	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
2	**0**	0.8100	0.6400	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.1800	0.3200	0.4200	0.4800	0.5000	0.4800	0.4200	0.3200	0.1800
	2	0.0100	0.0400	0.0900	0.1600	0.2500	0.3600	0.4900	0.6400	0.8100
3	**0**	0.7290	0.5120	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.2430	0.3840	0.4410	0.4320	0.3750	0.2880	0.1890	0.0960	0.0270
	2	0.0270	0.0960	0.1890	0.2880	0.3750	0.4320	0.4410	0.3840	0.2430
	3	0.0010	0.0080	0.0270	0.0640	0.1250	0.2160	0.3430	0.5120	0.7290
4	**0**	0.6561	0.4096	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.2916	0.4096	0.4116	0.3456	0.2500	0.1536	0.0756	0.0256	0.0036
	2	0.0486	0.1536	0.2646	0.3456	0.3750	0.3456	0.2646	0.1536	0.0486
	3	0.0036	0.0256	0.0756	0.1536	0.2500	0.3456	0.4116	0.4096	0.2916
	4	0.0001	0.0016	0.0081	0.0256	0.0625	0.1296	0.2401	0.4096	0.6561
5	**0**	0.5905	0.3277	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.3281	0.4096	0.3602	0.2592	0.1563	0.0768	0.0284	0.0064	0.0005
	2	0.0729	0.2048	0.3087	0.3456	0.3125	0.2304	0.1323	0.0512	0.0081
	3	0.0081	0.0512	0.1323	0.2304	0.3125	0.3456	0.3087	0.2048	0.0729
	4	0.0005	0.0064	0.0284	0.0768	0.1563	0.2592	0.3602	0.4096	0.3281
	5	0.0000	0.0003	0.0024	0.0102	0.0313	0.0778	0.1681	0.3277	0.5905
6	**0**	0.5314	0.2621	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000
	1	0.3543	0.3932	0.3025	0.1866	0.0938	0.0369	0.0102	0.0015	0.0001
	2	0.0984	0.2458	0.3241	0.3110	0.2344	0.1382	0.0595	0.0154	0.0012
	3	0.0146	0.0819	0.1852	0.2765	0.3125	0.2765	0.1852	0.0819	0.0146
	4	0.0012	0.0154	0.0595	0.1382	0.2344	0.3110	0.3241	0.2458	0.0984
	5	0.0001	0.0015	0.0102	0.0369	0.0938	0.1866	0.3025	0.3932	0.3543
	6	0.0000	0.0001	0.0007	0.0041	0.0156	0.0467	0.1176	0.2621	0.5314
7	**0**	0.4783	0.2097	0.0824	0.0280	0.0078	0.0016	0.0002	0.0000	0.0000
	1	0.3720	0.3670	0.2471	0.1306	0.0547	0.0172	0.0036	0.0004	0.0000
	2	0.1240	0.2753	0.3177	0.2613	0.1641	0.0774	0.0250	0.0043	0.0002
	3	0.0230	0.1147	0.2269	0.2903	0.2734	0.1935	0.0972	0.0287	0.0026
	4	0.0026	0.0287	0.0972	0.1935	0.2734	0.2903	0.2269	0.1147	0.0230
	5	0.0002	0.0043	0.0250	0.0774	0.1641	0.2613	0.3177	0.2753	0.1240
	6	0.0000	0.0004	0.0036	0.0172	0.0547	0.1306	0.2471	0.3670	0.3720
	7	0.0000	0.0000	0.0002	0.0016	0.0078	0.0280	0.0824	0.2097	0.4783

						p				
n	*x*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
8	0	0.4305	0.1678	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	0.0000
	1	0.3826	0.3355	0.1977	0.0896	0.0313	0.0079	0.0012	0.0001	0.0000
	2	0.1488	0.2936	0.2965	0.2090	0.1094	0.0413	0.0100	0.0011	0.0000
	3	0.0331	0.1468	0.2541	0.2787	0.2188	0.1239	0.0467	0.0092	0.0004
	4	0.0046	0.0459	0.1361	0.2322	0.2734	0.2322	0.1361	0.0459	0.0046
	5	0.0004	0.0092	0.0467	0.1239	0.2188	0.2787	0.2541	0.1468	0.0331
	6	0.0000	0.0011	0.0100	0.0413	0.1094	0.2090	0.2965	0.2936	0.1488
	7	0.0000	0.0001	0.0012	0.0079	0.0313	0.0896	0.1977	0.3355	0.3826
	8	0.0000	0.0000	0.0001	0.0007	0.0039	0.0168	0.0576	0.1678	0.4305
9	0	0.3874	0.1342	0.0404	0.0101	0.0020	0.0003	0.0000	0.0000	0.0000
	1	0.3874	0.3020	0.1556	0.0605	0.0176	0.0035	0.0004	0.0000	0.0000
	2	0.1722	0.3020	0.2668	0.1612	0.0703	0.0212	0.0039	0.0003	0.0000
	3	0.0446	0.1762	0.2668	0.2508	0.1641	0.0743	0.0210	0.0028	0.0001
	4	0.0074	0.0661	0.1715	0.2508	0.2461	0.1672	0.0735	0.0165	0.0008
	5	0.0008	0.0165	0.0735	0.1672	0.2461	0.2508	0.1715	0.0661	0.0074
	6	0.0001	0.0028	0.0210	0.0743	0.1641	0.2508	0.2668	0.1762	0.0446
	7	0.0000	0.0003	0.0039	0.0212	0.0703	0.1612	0.2668	0.3020	0.1722
	8	0.0000	0.0000	0.0004	0.0035	0.0176	0.0605	0.1556	0.3020	0.3874
	9	0.0000	0.0000	0.0000	0.0003	0.0020	0.0101	0.0404	0.1342	0.3874
10	0	0.3487	0.1074	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.3874	0.2684	0.1211	0.0403	0.0098	0.0016	0.0001	0.0000	0.0000
	2	0.1937	0.3020	0.2335	0.1209	0.0439	0.0106	0.0014	0.0001	0.0000
	3	0.0574	0.2013	0.2668	0.2150	0.1172	0.0425	0.0090	0.0008	0.0000
	4	0.0112	0.0881	0.2001	0.2508	0.2051	0.1115	0.0368	0.0055	0.0001
	5	0.0015	0.0264	0.1029	0.2007	0.2461	0.2007	0.1029	0.0264	0.0015
	6	0.0001	0.0055	0.0368	0.1115	0.2051	0.2508	0.2001	0.0881	0.0112
	7	0.0000	0.0008	0.0090	0.0425	0.1172	0.2150	0.2668	0.2013	0.0574
	8	0.0000	0.0001	0.0014	0.0106	0.0439	0.1209	0.2335	0.3020	0.1937
	9	0.0000	0.0000	0.0001	0.0016	0.0098	0.0403	0.1211	0.2684	0.3874
	10	0.0000	0.0000	0.0000	0.0001	0.0010	0.0060	0.0282	0.1074	0.3487
11	0	0.3138	0.0859	0.0198	0.0036	0.0005	0.0000	0.0000	0.0000	0.0000
	1	0.3835	0.2362	0.0932	0.0266	0.0054	0.0007	0.0000	0.0000	0.0000
	2	0.2131	0.2953	0.1998	0.0887	0.0269	0.0052	0.0005	0.0000	0.0000
	3	0.0710	0.2215	0.2568	0.1774	0.0806	0.0234	0.0037	0.0002	0.0000
	4	0.0158	0.1107	0.2201	0.2365	0.1611	0.0701	0.0173	0.0017	0.0000
	5	0.0025	0.0388	0.1321	0.2207	0.2256	0.1471	0.0566	0.0097	0.0003
	6	0.0003	0.0097	0.0566	0.1471	0.2256	0.2207	0.1321	0.0388	0.0025
	7	0.0000	0.0017	0.0173	0.0701	0.1611	0.2365	0.2201	0.1107	0.0158
	8	0.0000	0.0002	0.0037	0.0234	0.0806	0.1774	0.2568	0.2215	0.0710
	9	0.0000	0.0000	0.0005	0.0052	0.0269	0.0887	0.1998	0.2953	0.2131
	10	0.0000	0.0000	0.0000	0.0007	0.0054	0.0266	0.0932	0.2362	0.3835
	11	0.0000	0.0000	0.0000	0.0000	0.0005	0.0036	0.0198	0.0859	0.3138

						p				
n	x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
12	0	0.2824	0.0687	0.0138	0.0022	0.0002	0.0000	0.0000	0.0000	0.0000
	1	0.3766	0.2062	0.0712	0.0174	0.0029	0.0003	0.0000	0.0000	0.0000
	2	0.2301	0.2835	0.1678	0.0639	0.0161	0.0025	0.0002	0.0000	0.0000
	3	0.0852	0.2362	0.2397	0.1419	0.0537	0.0125	0.0015	0.0001	0.0000
	4	0.0213	0.1329	0.2311	0.2128	0.1208	0.0420	0.0078	0.0005	0.0000
	5	0.0038	0.0532	0.1585	0.2270	0.1934	0.1009	0.0291	0.0033	0.0000
	6	0.0005	0.0155	0.0792	0.1766	0.2256	0.1766	0.0792	0.0155	0.0005
	7	0.0000	0.0033	0.0291	0.1009	0.1934	0.2270	0.1585	0.0532	0.0038
	8	0.0000	0.0005	0.0078	0.0420	0.1208	0.2128	0.2311	0.1329	0.0213
	9	0.0000	0.0001	0.0015	0.0125	0.0537	0.1419	0.2397	0.2362	0.0852
	10	0.0000	0.0000	0.0002	0.0025	0.0161	0.0639	0.1678	0.2835	0.2301
	11	0.0000	0.0000	0.0000	0.0003	0.0029	0.0174	0.0712	0.2062	0.3766
	12	0.0000	0.0000	0.0000	0.0000	0.0002	0.0022	0.0138	0.0687	0.2824
13	0	0.2542	0.0550	0.0097	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.3672	0.1787	0.0540	0.0113	0.0016	0.0001	0.0000	0.0000	0.0000
	2	0.2448	0.2680	0.1388	0.0453	0.0095	0.0012	0.0001	0.0000	0.0000
	3	0.0997	0.2457	0.2181	0.1107	0.0349	0.0065	0.0006	0.0000	0.0000
	4	0.0277	0.1535	0.2337	0.1845	0.0873	0.0243	0.0034	0.0001	0.0000
	5	0.0055	0.0691	0.1803	0.2214	0.1571	0.0656	0.0142	0.0011	0.0000
	6	0.0008	0.0230	0.1030	0.1968	0.2095	0.1312	0.0442	0.0058	0.0001
	7	0.0001	0.0058	0.0442	0.1312	0.2095	0.1968	0.1030	0.0230	0.0008
	8	0.0000	0.0011	0.0142	0.0656	0.1571	0.2214	0.1803	0.0691	0.0055
	9	0.0000	0.0001	0.0034	0.0243	0.0873	0.1845	0.2337	0.1535	0.0277
	10	0.0000	0.0000	0.0006	0.0065	0.0349	0.1107	0.2181	0.2457	0.0997
	11	0.0000	0.0000	0.0001	0.0012	0.0095	0.0453	0.1388	0.2680	0.2448
	12	0.0000	0.0000	0.0000	0.0001	0.0016	0.0113	0.0540	0.1787	0.3672
	13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0013	0.0097	0.0550	0.2542
14	0	0.2288	0.0440	0.0068	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.3559	0.1539	0.0407	0.0073	0.0009	0.0001	0.0000	0.0000	0.0000
	2	0.2570	0.2501	0.1134	0.0317	0.0056	0.0005	0.0000	0.0000	0.0000
	3	0.1142	0.2501	0.1943	0.0845	0.0222	0.0033	0.0002	0.0000	0.0000
	4	0.0349	0.1720	0.2290	0.1549	0.0611	0.0136	0.0014	0.0000	0.0000
	5	0.0078	0.0860	0.1963	0.2066	0.1222	0.0408	0.0066	0.0003	0.0000
	6	0.0013	0.0322	0.1262	0.2066	0.1833	0.0918	0.0232	0.0020	0.0000
	7	0.0002	0.0092	0.0618	0.1574	0.2095	0.1574	0.0618	0.0092	0.0002
	8	0.0000	0.0020	0.0232	0.0918	0.1833	0.2066	0.1262	0.0322	0.0013
	9	0.0000	0.0003	0.0066	0.0408	0.1222	0.2066	0.1963	0.0860	0.0078
	10	0.0000	0.0000	0.0014	0.0136	0.0611	0.1549	0.2290	0.1720	0.0349
	11	0.0000	0.0000	0.0002	0.0033	0.0222	0.0845	0.1943	0.2501	0.1142
	12	0.0000	0.0000	0.0000	0.0005	0.0056	0.0317	0.1134	0.2501	0.2570
	13	0.0000	0.0000	0.0000	0.0001	0.0009	0.0073	0.0407	0.1539	0.3559
	14	0.0000	0.0000	0.0000	0.0000	0.0001	0.0008	0.0068	0.0440	0.2288

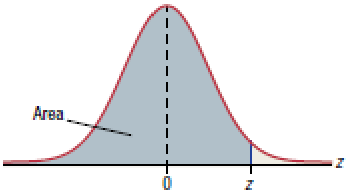
		p								
n	***x***	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
15	**0**	0.2059	0.0352	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3432	0.1319	0.0305	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.2669	0.2309	0.0916	0.0219	0.0032	0.0003	0.0000	0.0000	0.0000
	3	0.1285	0.2501	0.1700	0.0634	0.0139	0.0016	0.0001	0.0000	0.0000
	4	0.0428	0.1876	0.2186	0.1268	0.0417	0.0074	0.0006	0.0000	0.0000
	5	0.0105	0.1032	0.2061	0.1859	0.0916	0.0245	0.0030	0.0001	0.0000
	6	0.0019	0.0430	0.1472	0.2066	0.1527	0.0612	0.0116	0.0007	0.0000
	7	0.0003	0.0138	0.0811	0.1771	0.1964	0.1181	0.0348	0.0035	0.0000
	8	0.0000	0.0035	0.0348	0.1181	0.1964	0.1771	0.0811	0.0138	0.0003
	9	0.0000	0.0007	0.0116	0.0612	0.1527	0.2066	0.1472	0.0430	0.0019
	10	0.0000	0.0001	0.0030	0.0245	0.0916	0.1859	0.2061	0.1032	0.0105
	11	0.0000	0.0000	0.0006	0.0074	0.0417	0.1268	0.2186	0.1876	0.0428
	12	0.0000	0.0000	0.0001	0.0016	0.0139	0.0634	0.1700	0.2501	0.1285
	13	0.0000	0.0000	0.0000	0.0003	0.0032	0.0219	0.0916	0.2309	0.2669
	14	0.0000	0.0000	0.0000	0.0000	0.0005	0.0047	0.0305	0.1319	0.3432
	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0047	0.0352	0.2059
16	**0**	0.1853	0.0281	0.0033	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3294	0.1126	0.0228	0.0030	0.0002	0.0000	0.0000	0.0000	0.0000
	2	0.2745	0.2111	0.0732	0.0150	0.0018	0.0001	0.0000	0.0000	0.0000
	3	0.1423	0.2463	0.1465	0.0468	0.0085	0.0008	0.0000	0.0000	0.0000
	4	0.0514	0.2001	0.2040	0.1014	0.0278	0.0040	0.0002	0.0000	0.0000
	5	0.0137	0.1201	0.2099	0.1623	0.0667	0.0142	0.0013	0.0000	0.0000
	6	0.0028	0.0550	0.1649	0.1983	0.1222	0.0392	0.0056	0.0002	0.0000
	7	0.0004	0.0197	0.1010	0.1889	0.1746	0.0840	0.0185	0.0012	0.0000
	8	0.0001	0.0055	0.0487	0.1417	0.1964	0.1417	0.0487	0.0055	0.0001
	9	0.0000	0.0012	0.0185	0.0840	0.1746	0.1889	0.1010	0.0197	0.0004
	10	0.0000	0.0002	0.0056	0.0392	0.1222	0.1983	0.1649	0.0550	0.0028
	11	0.0000	0.0000	0.0013	0.0142	0.0667	0.1623	0.2099	0.1201	0.0137
	12	0.0000	0.0000	0.0002	0.0040	0.0278	0.1014	0.2040	0.2001	0.0514
	13	0.0000	0.0000	0.0000	0.0008	0.0085	0.0468	0.1465	0.2463	0.1423
	14	0.0000	0.0000	0.0000	0.0001	0.0018	0.0150	0.0732	0.2111	0.2745
	15	0.0000	0.0000	0.0000	0.0000	0.0002	0.0030	0.0228	0.1126	0.3294
	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0033	0.0281	0.1853

					p				
n	*x*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
17	**0**	0.1668	0.0225	0.0023	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3150	0.0957	0.0169	0.0019	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.2800	0.1914	0.0581	0.0102	0.0010	0.0001	0.0000	0.0000	0.0000
	3	0.1556	0.2393	0.1245	0.0341	0.0052	0.0004	0.0000	0.0000	0.0000
	4	0.0605	0.2093	0.1868	0.0796	0.0182	0.0021	0.0001	0.0000	0.0000
	5	0.0175	0.1361	0.2081	0.1379	0.0472	0.0081	0.0006	0.0000	0.0000
	6	0.0039	0.0680	0.1784	0.1839	0.0944	0.0242	0.0026	0.0001	0.0000
	7	0.0007	0.0267	0.1201	0.1927	0.1484	0.0571	0.0095	0.0004	0.0000
	8	0.0001	0.0084	0.0644	0.1606	0.1855	0.1070	0.0276	0.0021	0.0000
	9	0.0000	0.0021	0.0276	0.1070	0.1855	0.1606	0.0644	0.0084	0.0001
	10	0.0000	0.0004	0.0095	0.0571	0.1484	0.1927	0.1201	0.0267	0.0007
	11	0.0000	0.0001	0.0026	0.0242	0.0944	0.1839	0.1784	0.0680	0.0039
	12	0.0000	0.0000	0.0006	0.0081	0.0472	0.1379	0.2081	0.1361	0.0175
	13	0.0000	0.0000	0.0001	0.0021	0.0182	0.0796	0.1868	0.2093	0.0605
	14	0.0000	0.0000	0.0000	0.0004	0.0052	0.0341	0.1245	0.2393	0.1556
	15	0.0000	0.0000	0.0000	0.0001	0.0010	0.0102	0.0581	0.1914	0.2800
	16	0.0000	0.0000	0.0000	0.0000	0.0001	0.0019	0.0169	0.0957	0.3150
	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0023	0.0225	0.1668
18	**0**	0.1501	0.0180	0.0016	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3002	0.0811	0.0126	0.0012	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.2835	0.1723	0.0458	0.0069	0.0006	0.0000	0.0000	0.0000	0.0000
	3	0.1680	0.2297	0.1046	0.0246	0.0031	0.0002	0.0000	0.0000	0.0000
	4	0.0700	0.2153	0.1681	0.0614	0.0117	0.0011	0.0000	0.0000	0.0000
	5	0.0218	0.1507	0.2017	0.1146	0.0327	0.0045	0.0002	0.0000	0.0000
	6	0.0052	0.0816	0.1873	0.1655	0.0708	0.0145	0.0012	0.0000	0.0000
	7	0.0010	0.0350	0.1376	0.1892	0.1214	0.0374	0.0046	0.0001	0.0000
	8	0.0002	0.0120	0.0811	0.1734	0.1669	0.0771	0.0149	0.0008	0.0000
	9	0.0000	0.0033	0.0386	0.1284	0.1855	0.1284	0.0386	0.0033	0.0000
	10	0.0000	0.0008	0.0149	0.0771	0.1669	0.1734	0.0811	0.0120	0.0002
	11	0.0000	0.0001	0.0046	0.0374	0.1214	0.1892	0.1376	0.0350	0.0010
	12	0.0000	0.0000	0.0012	0.0145	0.0708	0.1655	0.1873	0.0816	0.0052
	13	0.0000	0.0000	0.0002	0.0045	0.0327	0.1146	0.2017	0.1507	0.0218
	14	0.0000	0.0000	0.0000	0.0011	0.0117	0.0614	0.1681	0.2153	0.0700
	15	0.0000	0.0000	0.0000	0.0002	0.0031	0.0246	0.1046	0.2297	0.1680
	16	0.0000	0.0000	0.0000	0.0000	0.0006	0.0069	0.0458	0.1723	0.2835
	17	0.0000	0.0000	0.0000	0.0000	0.0001	0.0012	0.0126	0.0811	0.3002
	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0016	0.0180	0.1501

		p								
n	*x*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
19	**0**	0.1351	0.0144	0.0011	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.2852	0.0685	0.0093	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.2852	0.1540	0.0358	0.0046	0.0003	0.0000	0.0000	0.0000	0.0000
	3	0.1796	0.2182	0.0869	0.0175	0.0018	0.0001	0.0000	0.0000	0.0000
	4	0.0798	0.2182	0.1491	0.0467	0.0074	0.0005	0.0000	0.0000	0.0000
	5	0.0266	0.1636	0.1916	0.0933	0.0222	0.0024	0.0001	0.0000	0.0000
	6	0.0069	0.0955	0.1916	0.1451	0.0518	0.0085	0.0005	0.0000	0.0000
	7	0.0014	0.0443	0.1525	0.1797	0.0961	0.0237	0.0022	0.0000	0.0000
	8	0.0002	0.0166	0.0981	0.1797	0.1442	0.0532	0.0077	0.0003	0.0000
	9	0.0000	0.0051	0.0514	0.1464	0.1762	0.0976	0.0220	0.0013	0.0000
	10	0.0000	0.0013	0.0220	0.0976	0.1762	0.1464	0.0514	0.0051	0.0000
	11	0.0000	0.0003	0.0077	0.0532	0.1442	0.1797	0.0981	0.0166	0.0002
	12	0.0000	0.0000	0.0022	0.0237	0.0961	0.1797	0.1525	0.0443	0.0014
	13	0.0000	0.0000	0.0005	0.0085	0.0518	0.1451	0.1916	0.0955	0.0069
	14	0.0000	0.0000	0.0001	0.0024	0.0222	0.0933	0.1916	0.1636	0.0266
	15	0.0000	0.0000	0.0000	0.0005	0.0074	0.0467	0.1491	0.2182	0.0798
	16	0.0000	0.0000	0.0000	0.0001	0.0018	0.0175	0.0869	0.2182	0.1796
	17	0.0000	0.0000	0.0000	0.0000	0.0003	0.0046	0.0358	0.1540	0.2852
	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0093	0.0685	0.2852
	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0011	0.0144	0.1351
20	**0**	0.1216	0.0115	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.2702	0.0576	0.0068	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.2852	0.1369	0.0278	0.0031	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.1901	0.2054	0.0716	0.0123	0.0011	0.0000	0.0000	0.0000	0.0000
	4	0.0898	0.2182	0.1304	0.0350	0.0046	0.0003	0.0000	0.0000	0.0000
	5	0.0319	0.1746	0.1789	0.0746	0.0148	0.0013	0.0000	0.0000	0.0000
	6	0.0089	0.1091	0.1916	0.1244	0.0370	0.0049	0.0002	0.0000	0.0000
	7	0.0020	0.0545	0.1643	0.1659	0.0739	0.0146	0.0010	0.0000	0.0000
	8	0.0004	0.0222	0.1144	0.1797	0.1201	0.0355	0.0039	0.0001	0.0000
	9	0.0001	0.0074	0.0654	0.1597	0.1602	0.0710	0.0120	0.0005	0.0000
	10	0.0000	0.0020	0.0308	0.1171	0.1762	0.1171	0.0308	0.0020	0.0000
	11	0.0000	0.0005	0.0120	0.0710	0.1602	0.1597	0.0654	0.0074	0.0001
	12	0.0000	0.0001	0.0039	0.0355	0.1201	0.1797	0.1144	0.0222	0.0004
	13	0.0000	0.0000	0.0010	0.0146	0.0739	0.1659	0.1643	0.0545	0.0020
	14	0.0000	0.0000	0.0002	0.0049	0.0370	0.1244	0.1916	0.1091	0.0089
	15	0.0000	0.0000	0.0000	0.0013	0.0148	0.0746	0.1789	0.1746	0.0319
	16	0.0000	0.0000	0.0000	0.0003	0.0046	0.0350	0.1304	0.2182	0.0898
	17	0.0000	0.0000	0.0000	0.0000	0.0011	0.0123	0.0716	0.2054	0.1901
	18	0.0000	0.0000	0.0000	0.0000	0.0002	0.0031	0.0278	0.1369	0.2852
	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0068	0.0576	0.2702
	20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0115	0.1216

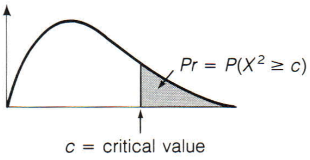
B- STANDARD NORMAL DISTRIBUTION

Numerical entries represent the probability that a standard normal random variable is between $-\infty$ and z where $z = \dfrac{x-\mu}{\sigma}$.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

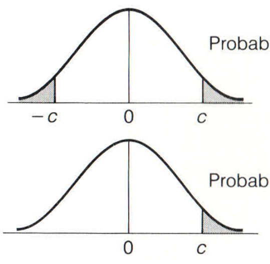
C- CRITICAL VALUES OF CHI-SQUARED DISTRIBUTION



	Pr							
ν	.500	.250	.100	.050	.025	.010	.005	.001
1	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	7.344	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.342	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.79	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	20.34	24.93	29.62	33.67	35.48	38.93	41.40	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	69.33	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	79.33	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	89.33	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	99.33	109.1	118.5	124.3	129.6	135.8	140.2	149.4

Source: Abridged from Table 8 of *Biometrika Tables for Statisticians*, Vol. 1, edited by E. S. Pearson and H. O. Hartley (London: Cambridge University Press, 1962).

D- CRITICAL VALUES OF STUDENT'S DISTRIBUTION



Probability $P(|t| \geq c)$ for two-sided tests

Probability $P(t \geq c)$ for one-sided tests

PROBABILITY

	.50	.20	.10	.05	.02	.01	.005	.002	.001	TWO-SIDED TESTS
ν	.25	.10	.05	.025	.01	.005	.0025	.001	.0005	ONE-SIDED TESTS
1	1.000	3.078	6.314	12.706	31.821	63.637	127.32	318.31	636.62	
2	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598	
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924	
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408	
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537	
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	
15	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	
17	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	
20	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.257	3.189	
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767	
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	
25	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707	
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690	
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674	
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659	
30	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	
60	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	

Bibliography

1. Bertsekas D. P., Tsitsiklis J. N., *Introduction to Probability: Lecture Notes.* MIT, Massachussets (2008).

2. Lejeune M, *La Statistique et ses Applications.* Springer-Verlag France (2010).

3. Devore J. *Probability and Statistics for Engineering and the Sciences (8th Edition).* Brooks/Cole CENGAGE Learning (2010).

4. Suhnov Y. M., Kelbert M., *Probability and Statistics by Example, Volume 1.* Cambridge University Press (2014).

5. Rohtagi V. K., Ehsanes Saleh A. K. Md., An Introduction to Probability and Statistics. John Wiley and Sons (2015).

6. Lecoutre J.-P., *Statistique et Probabilités. Cours et Exercices.* Dunod (2023).