الجمهــــورية الجزائـــــرية الديمقــــــراطية الشعـــــبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جـــامعة محمد بوضــياف – بالمســـيلة
Mohamed BOUDIAF University – M'Sila

جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila
كـلية التكنـــولـــوجيـــا
نيابة العمادة لما بعد التدرج والبحث العلمي والعلاقات الخارجية

جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila
Faculty of Technology
Vice Deanship of Post-Graduation, Scientific Research and External Relations

رقم: 487/ ن.ع.ب.ع /ك.ت /2025

المسيلة في: 16 ديسمبر 2025

شهادة ادارية

المصادقة على تقارير خبرة للموافقة على مطبوعة بيداغوجية

بعد الإطلاع على تقارير لجنة الخبراء للموافقة على المطبوعة البيداغوجية للأستاذ : **عابد أحسن– أستاذ محاضر قسم أ** ، بالقاعدة المشتركة بكلية التكنولوجيا بجامعة محمد بوضياف بالمسيلة والتي كانت كلها ايجابية ، تمّ تقرير التّالي:

1–المصادقة على تقارير لجنة الخبراء للموافقة المطبوعة البيداغوجية والمعنونة بـــ:

Statistics and Probability
Common Base ST– Cycle License

2– حيث تمّ تشكيل هذه اللجنة بناء على اجتماع المجلس العلمي للكلية المنعقد بتاريخ **2025/11/30** المكونة من السّادة الآتية أسماؤهم:

–**بوقرة عبد الرحمان**، أستاذ ، جامعة محمد بوضياف– المسيلة

– **حريزي عبد الغفور**، أستاذ محاضر "أ"، جامعة محمد بوضياف– المسيلة

– **بن دومية رضا**، أستاذ، جامعة البليدة–1

وتمت الموافقة بالاجماع على هذه المطبوعة.

رئيس المجلس العلمي للكلية

جامعة محمد بوضياف المسيلة - المجلس العلمي - كلية التكنولوجيا

د. علي جريو

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Mohamed BOUDIAF University - M'sila

جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila

Faculty of Technology

Statistics and Probability

Dr. ABED Ahcene

This course is designed for first-year students in electrotechnics, civil engineering, and mechanical engineering. It facilitates students' comprehension of the statistics and probability subject.

Material Polycopy: 1st year in Civil Engineering, Mechanical Engineering and Electrical Engineering 2025-2026

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed BOUDIAF University - M'sila

1985

جامعة محمد بوضياف - المسيلة

Université Mohamed Boudiaf - M'sila

Faculty of Technology

Statistics and Probability

Dr. ABED Ahcene

This course is designed for first-year students in electrotechnics, civil engineering, and mechanical engineering. It facilitates students' comprehension of the statistics and probability subject.

Material Polycopy: 1st year in Civil Engineering, Mechanical Engineering and Electrical Engineering 2025-2026

Contents

Preface

This course is meant to be a must-have study tool for first-year engineering students who are taking the Statistics and Probability course. It focuses on mechanical engineering, civil engineering, and electrotechnics. The content has been carefully reorganized into six full chapters that closely follow the main points in the specifications document. Each week, students will have one interesting lecture and one interactive tutorial session, which will help them understand the material better.

Students will learn important statistical and probabilistic methods in this course that are necessary for accurately modeling a wide range of technological applications. If you learn these ideas well, you will greatly improve your analytical skills and be better prepared for the challenges that come with being an engineer.

Author	Dr. ABED Ahcene
Email	abed.ahcene@univ-msila.dz
Faculty :	Technology
Department:	Common base
Institution:	Mohamed BOUDIAF University of M'sila – Algeria

Course Title:	Statistics and Probability
Semester:	S1
Weekly Hour Volume:	Course (1h30) and tutorials (1h30)
Evaluation Method:	Exam (60%) and Continuous Assessment (40%)

Chapter 1

Basic definitions

1.1 Statistics

Statistics involves scientific methods for collecting, organizing, summarizing, presenting, and analyzing data. It also includes drawing reliable conclusions and making sound judgments based on this analysis. Statistics can refer to the data or numerical summaries derived from the data, such as averages. Examples of this include employment data, accident statistics, and more.

1.2 Population and sample

1.2.1 Population

In statistics, a population refers to the entire set of individuals, items, or data that we want to study or gather information about. It is the complete group that shares at least one characteristic that is of interest for our investigation.

Here, some examples of Population.

Example

- **All Students in the institution:** If we want to know the average height of students at our institution, we need to look at all the students who go there.
- **All Apples in an Orchard:** The population of apples cultivated in that orchard is what a farmer needs to know to find out the average weight of apples.

- **All Residents of a City:** The population of a city is everyone who lives there. This is how the government can figure out how much money people in that city make on average.
- **All Manufactured Cars in a Year:** The population for a vehicle firm assessing failure rates can be all the automobiles made in 2025.

Note

- The study's goal defines the population.
- It might be limited (like 500 pupils at a school) or unlimited (like all the potential tosses of a die).
- Before gathering data, the population should be well defined.
- The population is significant in statistics since it is the group we want to learn more about.
- Parameters are measurements or qualities that are determined from the population (for example, the population mean or the population variance).

1.2.2 Sample

Observing the entire group is often impractical when collecting data about the characteristics of individuals or objects. For example, this can include measuring the heights and weights of university students or counting the number of defective and non-defective bolts produced in a factory on a specific day. Instead of examining the entire group, known as the population or universe, researchers typically focus on a subset of the group, called a sample.

A statistical sample is a subset of individuals, items, or data points selected from a larger population, which is used to represent that population for analysis. The goal is to conclude the entire population without having to collect data from every member. Here we have three concepts.

- **Population:** The entire group you want to study.
- **Sample:** A smaller group chosen from the population.

- **Sampling:** The process of selecting this subset.

In general, there are four types of sampling.

- **Random Sampling:** Every member has an equal chance of being selected.
- **Stratified Sampling:** The population is divided into subgroups, and samples are taken from each.
- **Systematic Sampling:** Every n^{th} member is selected from a list.
- **Convenience Sampling:** Samples are chosen based on ease of access (less reliable).

The importance of the sample appears in:

- Saves time and resources.
- Makes data collection feasible.
- Allows for statistical inference about the population.

Example

Suppose we want to know the average height of students in a university with 10,000 students. Measuring all 10,000 would be difficult, so we select a sample of 200 students at random. The average height from this sample is then used to estimate the average height for all students.

1.3 Variable

A statistical variable is a trait or quality that can have different values for the things or people being studied. In statistics, variables are things that you look at, measure, or change in an experiment or survey.

A variable is a symbol, such as X, Y, x, y, A, or B, that can take on any value from a specific set known as the domain of the variable. On the other hand, a constant remains fixed and does not change; it can only assume a single value.

In general, there are two types of statistical variables.

- Quantitative variable
- Qualitative variable

1.3.1 Quantitative variable

A quantitative variable (also known as a numerical variable) is a type of variable in statistics that represents measurable quantities and is expressed as numbers. These variables can be used to perform mathematical operations, such as addition, subtraction, averaging, etc. For instance, a quantitative variable is established when a measurement is needed to describe this characteristic or when a count is required to quantify it.

Here are some examples of quantitative variables:

Example

- Age (in years)
- Salary (in dollars)
- Test scores
- Distance (in kilometers)
- Temperature (in degrees)

Quantitative variables are generally divided into two types:

- Continuous variable
- Discrete variable

a Continuous variable

A continuous variable is a type of quantitative variable that can take any value within a given range. Unlike discrete variables, which can only take specific, separate values (often counts), continuous variables can have an infinite number of possible values, often including fractions and decimals. This variable is characterized by:

- Can be measured, not counted (e.g., height, weight, temperature).
- Can assume any value within an interval (e.g., 2.5 kg, 37.8°C).
- Values are not restricted to whole numbers.

Here are some examples of Continuous Variables:

Example

- Height: 170.5 cm, 172.0 cm, 173.8 cm, etc.
- Weight: 65.2 kg, 70.75 kg, 80.0 kg, etc.
- Time: 2.03 $seconds$, 5.5 $minutes$, 1.25 $hours$, etc.
- Temperature: 36.6°C, 98.6°F, etc.
- Distance: 3.1 km, 5.85 $miles$, etc.

Data that a continuous variable can described is called continuous data.

Exemple

The heights of 100 university students are an example of continuous data.

b Discrete Variable

A discrete variable is a type of quantitative variable that can take only specific, distinct values—usually whole numbers. Discrete variables are often the result of counting something rather than measuring it. This variable is characterized by:

- Values are countable and separate (cannot take on every possible value within a range).
- No fractions or decimals between values (e.g., you cannot have 2.5 students).
- Usually represents counts or occurrences.

Here are some examples of Discrete Variables:

Example

- Number of students in a classroom (e.g., 20, 21, 22)
- Number of cars in a parking lot (e.g., 5, 10, 15)
- Number of books on a shelf
- Number of goals scored in a match
- Number of siblings a person has

Data that a discrete variable can described is called discrete data.

Example

The number of children in each of 1000 families is an example of discrete data.

1.3.2 Qualitative Variable

A qualitative variable (also known as a categorical variable) is a type of variable that describes or categorizes an attribute or quality, rather than measuring a numerical value. It is a characteristic of interest that results in a non-numeric value. The values of a qualitative variable are labels or categories rather than numbers. A qualitative variable is characterized by:

- Describes qualities or characteristics (not amounts or measurements).
- Values are categories or groups (e.g., color, type, brand).
- Cannot perform meaningful mathematical operations on the values.

The qualitative variable is usually divided into two types:

- **Nominal:** Categories with no natural order (e.g., gender, blood type, eye color).
- **Ordinal:** Categories with a meaningful order or ranking, but not measured (e.g., education level, satisfaction rating).

Here are some examples of qualitative variables:

- Gender: Male, Female, Other
- Marital status: Single, Married, Divorced
- Nationality: American, Indian, French
- Car brand: Toyota, Ford, BMW
- Eye color: Blue, Brown, Green

The available categories for qualitative variables are frequently coded for computerized statistical analysis. Marital status can be marked as 1, 2, or 3, where 1 denotes single, 2 denotes married, and 3 denotes divorced. Gender could be coded as 0 for females and 1 for males. Any qualitative variable's categories can be coded similarly. Despite the fact that numerical values are associated with the characteristic of interest after coding, the variable seems qualitative.

Chapter 2

One-variable statistical series

2.1 Statistical series

A statistical series is a set of pairs (x_i, n_i), where the x_i are the values taken from the character and the n_i is the number of times the value x_i appears.

2.2 Numerical representation

2.2.1 Frequency distribution

The frequency n_i of a particular value x_i is the number of times the value occurs in the data. The distribution of a variable is the pattern of frequencies, meaning the set of all possible values and the frequencies associated with these values.

The total number of categories is therefore:

$$n = \sum_i n_i$$

2.2.2 Relative frequency

The relative frequency f_i of a particular value x_i is obtained by dividing its frequency by the sum of all the frequencies.

$$f_i = \frac{n_i}{n}$$

Note

The sum of the relative frequencies will always equal one.

2.2.3 Percentage

The percentage of a given value x_i is obtained by multiplying its relative frequency by 100.

Note

The sum of the percentages for all the categories will always equal 100 percent.

Example

A statistical study was conducted on a group of Algerian university students about their hobbies, and the results were as follows :

sports music acting shaving reading music sports swimming shaving trade shaving acting trade music trade trade trade acting reading reading trade trade trade swimming swimming

The table below shows the frequency of each hobby for students, as well as the relative frequency and percentage.

Table 1. Example 1

Hobbies	Frequency	Relative frequency	Percentage
Sports	2	$2/25 = 0.08$	$100 \times 0.08 = 8$ %
Music	3	$3/25 = 0.12$	$100 \times 0.12 = 12$ %
Acting	3	$3/25 = 0.12$	$100 \times 0.12 = 12$ %
Shaving	3	$3/25 = 0.12$	$100 \times 0.12 = 12$ %
Reading	3	$3/25 = 0.12$	$100 \times 0.12 = 12$ %
Swimming	3	$3/25 = 0.12$	$100 \times 0.12 = 12$ %
Trade	8	$8/25 = 0.32$	$100 \times 0.32 = 32$ %

Example

We count the number of defects observed on a sample of a thousand pieces obtained from a factory's daily production :

Table 2. Example 2

Number of defects	Frequency
0	570
1	215
2	140
3	60
4	15

1. Determine the series (x_i, n_i)
2. Calculate the relative frequency of each category
3. Calculate the percentage of each category

2.2.4 Cumulative frequency distributions

A cumulative frequency distribution gives the total number of values that fall below various class boundaries of a frequency distribution.

Example

Table 3 shows the frequency distribution of the contents in milliliters of a sample of 25 one-liter bottles of soda.

Table 3. Example 3

Content in milliliters	Frequency
970 – 990	5
990˘1010	10
1010˘1030	5
1030˘1050	3
1050˘1070	2

2.2.5 Cumulative relative frequency distributions

Cumulative relative frequency is obtained by dividing the cumulative frequency by the total number of observations (N) in the data set.

$$\text{Cumulative Relative Frequency} = \frac{\text{Cumulative Frequency}}{\text{Total Number of Observations}}$$

This gives a running total of the proportion of observations up to each value, often expressed as a decimal or percentage.

Example

Table 4 presents the cumulative relative frequencies alongside the cumulative percentages corresponding to the frequency distribution outlined in Table 3.

Table 4. Example 4

Content less than	Cumulative frequency	Cumulative relative frequency	Cumulative percentage
970	0	$0/25 = 0$	0.00%
990	5	$5/25 = 0.2$	$20,00\%$
1010	$5 + 10 = 15$	$15/25 = 0.6$	$60,00\%$
1030	$15 + 5 = 20$	$20/25 = 0.8$	$80,00\%$
1050	$20 + 3 = 23$	$23/25 = 0.92$	$92,00\%$
1070	$23 + 2 = 25$	$25/25 = 1$	$100,00\%$

2.3 Characterization of a Class

A class is a grouping of data values or observations that fall within a specified range in a frequency distribution. Classes are commonly used when data are continuous or when the dataset is large, making it impractical to list every individual value. Each class represents an interval or category into which data points are sorted. Classes are characterized by their limits, boundaries, size, frequency, and midpoint, and should be mutually exclusive and collectively exhaustive.

2.3.1 Class Limits

Lower Class Limit: The smallest value that can belong to a class.

Upper Class Limit: The largest value that can belong to a class.

2.3.2 Class Boundaries

To keep the categories from having gaps, the real borders that divide them are generally figured out. We can tell each class to separate itself.

- Lower class boundary.
- Upper class boundary.

In practice, the class boundaries are obtained by adding the upper limit of one class interval to the lower limit of the next-higher class interval and dividing by 2.

2.3.3 Class Interval (size)

The size l of the class is defined as the difference between the upper and lower boundaries of a class.

$$l = U - L$$

Where :

- L : is the lower class boundary.
- U : is the upper class boundary.

2.3.4 Class Mark (Midpoint)

The value midway between the upper and lower class limits, calculated as:

$$\text{Class Mark} = \frac{L + U}{2}$$

For purposes of further mathematical analysis, all observations belonging to a given class interval are assumed to coincide with the class mark.

2.3.5 Mutual Exclusivity and Exhaustiveness

- Each data value should belong to one and only one class.
- The class should cover all possible data values.

Example

Consider the following exam scores, measured on a scale of 100:

Table 5. Example 5

Exam scores	40–49	50–59	60–69	70–79	80–89
Frequency	3	7	9	7	4

- Each range (e.g., 50-59) is a class.
- 50 and 59 are called class limits.
- 50 is the lower class limit.
- 59 is the upper class limit.
- The class interval 50-59 theoretically includes all exam scores from 49.5 to 59.5. These numbers are called class boundaries.
- 49.5 is the lower class boundary.
- 59.5 is the upper class boundary.

2.4 Graphical representation

There are indeed several ways to present statistical series, which help in organizing, analyzing, and visually interpreting data. The most common methods include:

1. **Tables :**
    - Simple tables: Present raw data in rows and columns.
    - Frequency tables: Show how often each value occurs.
    - Contingency tables: Display the relationship between two or more variables.

2. **Graphics :**

- Bar Charts: Used to compare quantities across categories.
- Pie Charts: Illustrate proportions of a whole.
- Scatter Plots: Show relationships between two numerical variables.
- Dot Plots: Display individual data points.
- Histograms: Show the distribution of a continuous variable.
- Polygon : Display the distribution of a dataset.

2.4.1 Bar chart

A bar chart (or bar graph) is a type of graph that uses rectangular bars to represent and compare the values of different categories. The length or height of each bar is proportional to the value it represents. Bar charts are commonly used to display and compare discrete categories or groups.

- Bars can be vertical or horizontal.
- Each bar represents a category (e.g., different car colors).
- The height/length of the bar shows the value or frequency for that category.
- Categories are typically placed along one axis (often the x-axis), while the values are shown on the other axis.

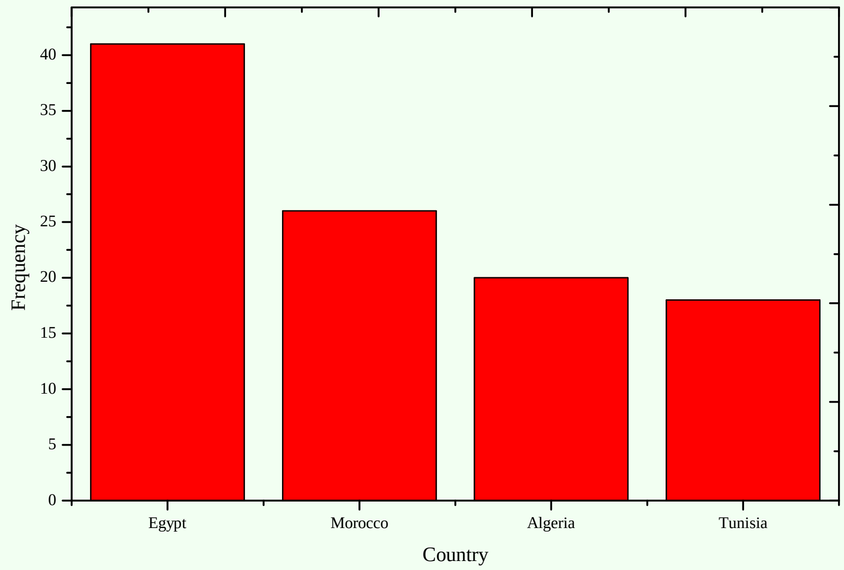
Example

Arab athletes from North African countries have won 105 medals throughout their participation in the Olympic Games, from 1896 to 2024 (Table 6).

Table 6. Example 6

Country	Egypt	Morocco	Algeria	Tunisia
Number of medals	41	26	20	18

To construct a bar graph, the categories are placed along the horizontal axis, and frequencies are marked along the vertical axis.



2.4.2 Pie chart

A pie chart is a circular graph used to represent data, typically qualitative (categorical) variables. It shows how a whole is divided into different parts, with each "slice" representing a category's proportion or percentage of the total. A pie chart illustrates the relative proportions of several categories, providing a swift visual representation of how a whole is divided among its components.

Key characteristics of a pie chart are as follows:

- Circle divided into slices: Each slice represents one category.
- Size of slices: The size (angle/area) of each slice is proportional to the category's value or percentage.
- Total equals 100%: All slices together make up the whole circle (100% of the data).

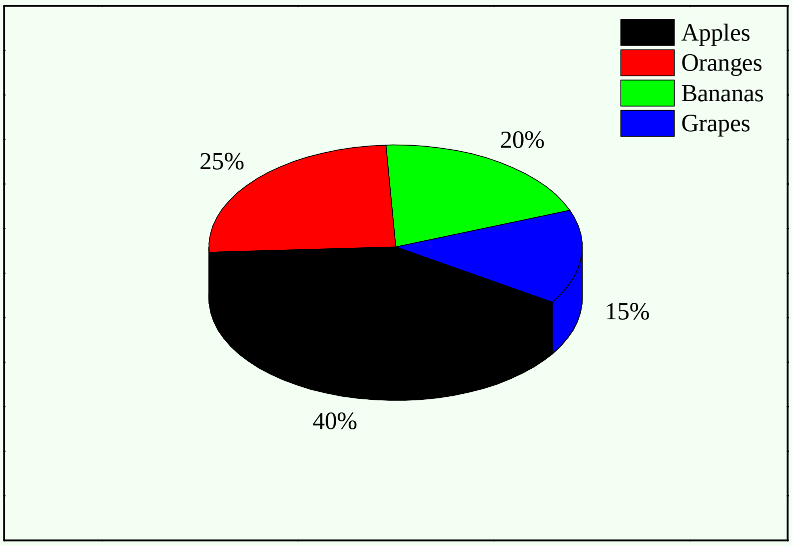
Example

Suppose we surveyed 100 people about their favorite fruit:

Table 7. Example 7

Fruit	Apples	Oranges	Bananas	Grapes
Frequency	40	25	20	15

A pie chart would have four slices, with the size of each slice showing the proportion of people who chose each fruit.



2.4.3 Scatter Plots

A scatter plot is a type of data visualization that uses dots to represent the values obtained for two different variables—one plotted along the x-axis and the other along the y-axis. Scatter plots are especially useful for identifying relationships, trends, or correlations between variables. Key Features of a Scatter Plot are as follows:

- X-axis and Y-axis: Represent two variables (e.g., height vs. weight).
- Dots: Each dot represents one data point.
- Trend Lines (optional): Sometimes a line (like a line of best fit) is added to show the overall direction of the data.

Here, some examples of use cases:

- Examining Correlation: For example, comparing hours studied and exam scores.
- Detecting Outliers: Identifying data points that do not follow the pattern.
- Visualizing Clusters: Observing groups or patterns within the data.

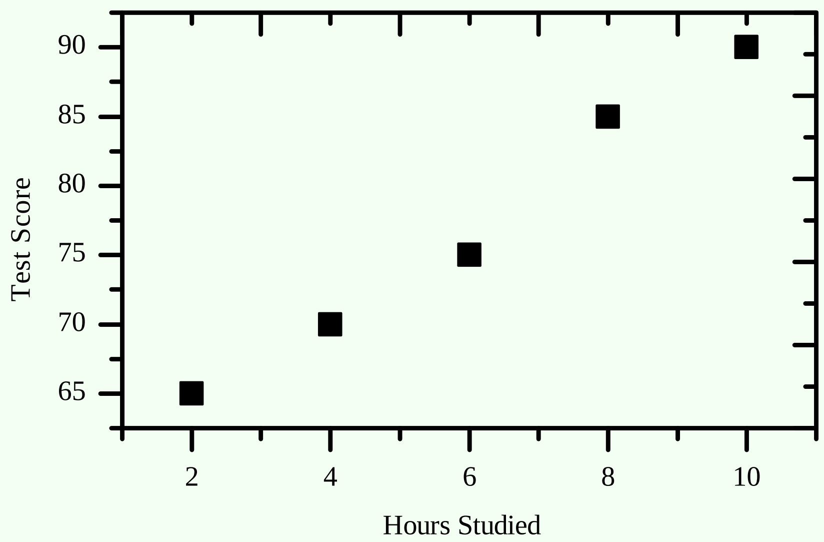
Example

Consider data reflecting the correlation between hours studied and the associated test results of many students:

Table 8. Example 8

Hours Studied	2	4	6	8	10
Test Score	65	70	75	85	90

In a scatter plot, "Hours Studied" should be shown on the x-axis and "Test Score" on the y-axis. Every student would represent a point on the graph.



2.4.4 Dot Plots

A dot plot is a simple and effective way to display small sets of numerical data using dots. Each dot represents one occurrence of a value. Dot plots are useful for visualizing the distribution and frequency of data points along a number line.

The primary characteristics of a dot plot can be described as follows:

- Number Line: The values of the variable are placed along a horizontal axis.
- Dots: Each dot represents one data point for a particular value. If a value occurs more than once, dots are stacked vertically.
- Simplicity: Best used for small data sets or data with limited, discrete values.

Dot plots are used :

- To show the frequency of small data sets.
- When you want a clear, simple visual of data distribution.
- To make it easy to compare counts for each category.

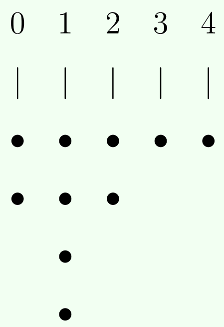
Example

Suppose you surveyed the number of pets owned by 10 students:

Table 9. Example 9

Number of Pets	0	1	2	3	4
Number of Students	2	4	2	1	1

The dot plot would look like this:



Each dot stands for one student.

2.4.5 Histogram

A histogram is a kind of bar graph that shows how often data points fall within certain ranges or intervals, which are called "bins." This shows how the data is spread out. Histograms are great for showing the shape, spread, and central tendency of data that changes over time.

The primary characteristics of a Histogram can be described as follows:

- X-axis (Horizontal): Shows the intervals (bins) into which the data is grouped.
- Y-axis (Vertical): Shows the frequency (number of data points) for each bin.

- Bars: The number of data points in a certain range is shown by each bar. The bars touch each other to show that the data is continuous.

Example

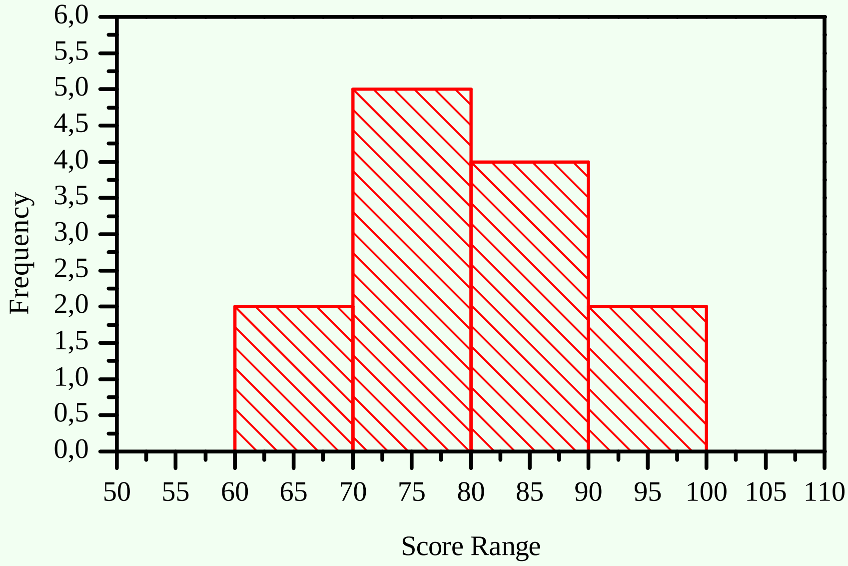
Let us say we have test scores for a group of students:

Score	62	67	70	72	73	75	77	80	82	85	87	90	92

We could divide scores into bins (e.g., 60–69, 70–79, 80–89, 90–99):

Score Range	60–69	70–79	80–89	90–99
Frequency	2	5	4	2

The histogram would have four bars, each one's height corresponding to the frequency of scores in that bin.



The histogram is used :

- To visualize the distribution of continuous or large discrete data.
- To identify patterns (like skewness, modality, or outliers) in the data.
- To compare the spread and central tendency of data sets.

Note

In histograms, the bars touch each other and are used for continuous data. This is different from bar charts. Bar charts are used for categorical data, but the bars do ont touch each other.

2.4.6 Polygon

In statistics, a polygon often refers to a frequency polygon, which is a graphical way to display the distribution of a dataset. A frequency polygon is similar to a histogram but uses points connected by straight lines instead of bars.

The main features of a Frequency Polygon are as follows:

- X-axis: Represents the class intervals (midpoints of bins).
- Y-axis: Represents the frequency (number of data points in each bin).
- Points: Plotted at the midpoint of each interval at a height corresponding to the frequency.
- Lines: Straight lines connect these points, forming a polygonal shape.

We usually use frequency polygons to compare distributions, which means putting a set of polygons along the same axes. This is to find out the shape of a distribution, or when we need a linear option instead of a histogram.

A frequency polygon shows frequency with connected dots instead of bars, which is different from a histogram.

Example

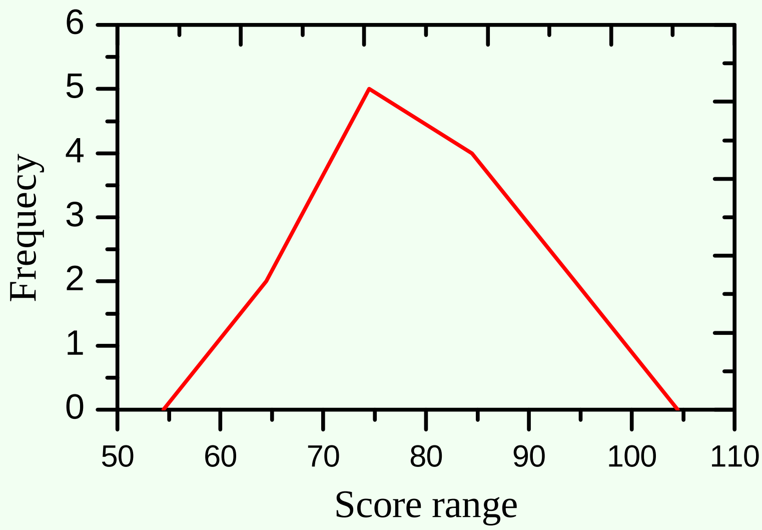
Suppose we have the following data on test scores:

Score Range	60–69	70–79	80–89	90–99
Frequency	2	5	4	2

First, we need to find the class Marks.

Score Range	60–69	70–79	80–89	90–99
Class Marks	64.5	74.5	84.5	94.5

The polygon frequency is illustrated in this figure.



2.5 Characteristics of position

There are many different measures of central tendency. The three most widely used measures of central tendency are

- The mean.
- The median.
- The mode.

2.5.1 The mean

The mean for a sample consisting of n observations is

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

Example

The mean of the numbers 8, 3, 5, 12, and 10 is

$$\overline{x} = \frac{8+3+5+12+10}{5} = 7.6$$

And the mean for a population consisting of N observations is

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum x}{N}$$

The mean for grouped data is given by

$$\overline{x} = \frac{\sum_{i=1}^{n} n_i x_i}{n \sum_{i=1}^{n} n_i}$$

Example

If 5, 8, 6, and 2 occur with frequencies 3, 2, 4, and 1, respectively, the mean is

$$\overline{x} = \frac{(3)(5) + (2)(8) + (4)(6) + (1)(2)}{3 + 2 + 4 + 1} = 5.7$$

2.5.2 The median

The median of a set of numbers arranged in order of magnitude (i.e., in an array) is either the middle value or the arithmetic mean of the two middle values.

Example

The set of numbers 3, 4, 4, 5, 6, 8, 8, 8, and 10 has a median of 6.

Example

The set of numbers 5, 5, 7, 9, 11, 12, 15, and 18 has a median $\frac{1}{2}(9 + 11) = 10$

For grouped data, the following formula is sometimes used to find the median.

$$Median = L + \frac{0.5N - c}{f_{median}} \times (U - L)$$

Where,

- L : the lower class boundary of the median class
- U : the upper class boundary of the median class
- N : the number of items in the data
- c : sum of frequencies of all classes lower than the median class
- f_{median} : frequency of the median class

2.5.3 The mode

The mode of a set of numbers is that value which occurs with the greatest frequency. The mode may not exist, and even if it does exist, it may not be unique.

Example

The set 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, and 18 has mode 9, and is called unimodal.

A statistical series may have no mode.

Example

The set 3, 5, 8, 10, 12, 15, and 16 has no mode.

A statistical series can also have more than one mode.

Example

The set 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, and 9 has two modes, 4 and 7, and is called bimodal

2.6 Dispersion characteristics

It is desirable to have numerical values to describe the spread or dispersion of a data set. The three most widely used measures of dispersion are

- The range
- The variance
- The standard deviation

2.6.1 The range

The range of a set of numbers is the difference between the largest and smallest numbers in the set.

Example

The range of the set 2, 3, 3, 5, 5, 5, 8, 10, 12 is $12 \div 2 = 10$. Sometimes the range is given by simply quoting the smallest and largest numbers; in the above set, for instance, the range could be indicated as 2 to 12, or $2 - 12$.

2.6.2 The variance

The variance of a population of size N is represented by σ^2 and is given by

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Example

The mean of the numbers 8, 3, 5, 12, and 10 is

$$\sigma^2 = \frac{(8-7.6)^2 + (3-7.6)^2 + (5-7.6)^2 + (12-7.6)^2 + (10-7.6)^2}{5} = 10.64$$

Shortcut formulas are useful in computing variances and standard deviations.

$$\sigma^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$

Example

The variance of the numbers 8, 3, 5, 12, and 10 is

$$\sigma^2 = \frac{8^2 + 3^2 + 5^2 + 12^2 + 10^2}{5} - 7.6^2 = 68.4 - 57.76 = 10.64$$

The variance for grouped data is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} n_i x_i^2}{\sum_{i=1}^{n} n_i} - \overline{x}^2$$

Example

If 5, 8, 6, and 2 occur with frequencies 3, 2, 4, and 1, respectively, the variance is

$$\sigma = \frac{(3)(5)^2 + (2)(8)^2 + (4)(6)^2 + (1)(2)^2}{3+2+4+1} - 5.7^2 = 35.1 - 32.49 = 2.61$$

2.6.3 The standard deviation

The standard deviation of a data set measures the spread of the data about the mean of the data set.

The standard deviation is

$$\sigma = \sqrt{\sigma^2}$$

2.6.4 Coefficient of variation

The coefficient of variation is equal to the standard deviation divided by the mean. The result is usually expressed as a percent.

The coefficient of variation for a population is given by

$$CV = \frac{\sigma}{\mu} \times 100\%$$

Application

This table gives the annual returns for 30 randomly selected mutual funds

10.5	12.5	14.5	22.0	12.5
-2.5	20.2	3.5	7.5	14.5
0	17.5	14.0	12.0	17.0
3	27.5	22.5	10.5	40.5
5	12.7	35.5	38.0	10.5
0	-5.5	19.0	14.5	10.5

1. Find the mean, the median, and the mode for the annual returns.
2. Find the range, the variance, and the standard deviation for the annual returns

Chapter 3

Two-variables statistical series

3.1 Definitions

A two-variable statistical series (also known as a bivariate data set or bivariate statistical series) is a collection of paired data points, where each pair consists of two variables measured for the same item or subject. It is a series of n measurements of two quantities $(x_i; y_j; n_{ij}; i = 1, \cdots, r; j = 1, \cdots, p$ where n_{ij} denotes the number of observations presenting both the modality x_i and the modality y_j.

The main purpose is to study the relationship or association between two quantitative variables.

Note

We are mainly interested in the question of

- Whether the measurements of the two values are independent or not?
- If they are not independent, how are they connected?

3.2 Presentation of bivariate statistical series

3.2.1 Contingency table

A contingency table (also known as a cross-tabulation or crosstab) is a type of table in statistics that displays the frequency distribution of variables. It is widely used to analyze the relationship between two or more categorical variables. It shows how the frequency of each modality of a character is distributed according to the modalities of the other.

	y_1	$\cdots$	y_j	$\cdots$	y_p	$n_{i\cdot}$
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1p}	$n_{1\cdot}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ip}	$n_{i\cdot}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_r	n_{r1}	$\cdots$	n_{rj}	$\cdots$	n_{rp}	$n_{r\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot p}$	$n_{\cdot\cdot}$

There is a structure for a contingency table that includes the components.

- Rows and columns correspond to the categories of the variables being analyzed.
- Cells contain the count of observations for each combination of categories.
- Marginal totals (row totals, column totals) and the total are typically included to provide summary counts.

Example

For instance, a study with 100 volunteers looked at handedness (right-handed vs. left-handed) based on gender (male vs. female):

	Right-handed	Left-handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100

This is a 2×2 contingency table. Both the marginal totals (for example, the total number of right-handed people) and the total (100) are presented in a comprehensible manner.

Contingency tables can contain more than two dimensions. For example, a 2×3 table can show the relationship between gender (male, female) and coffee preference (light, regular, dark). We can also create a 3×3 table, such as the table showing gender, coffee preference, and age group. For clarification, they are often displayed as several two-way tables.

3.2.2 Scatter diagram

A scatter diagram, scatter plot, or scatter graph is a graphical tool that shows how two quantitative variables are related. The Cartesian axes show each data point as an observation, with one variable on the x-axis and the other on the y-axis.

The scatter plot shows a bivariate series as follows.

- The x-axis usually shows the independent variable, and the y-axis shows the dependent variable.
- We can see patterns like positive or negative correlation, outliers, and the overall strength of the relationship in this picture.

A scatter plot can help us figure out:

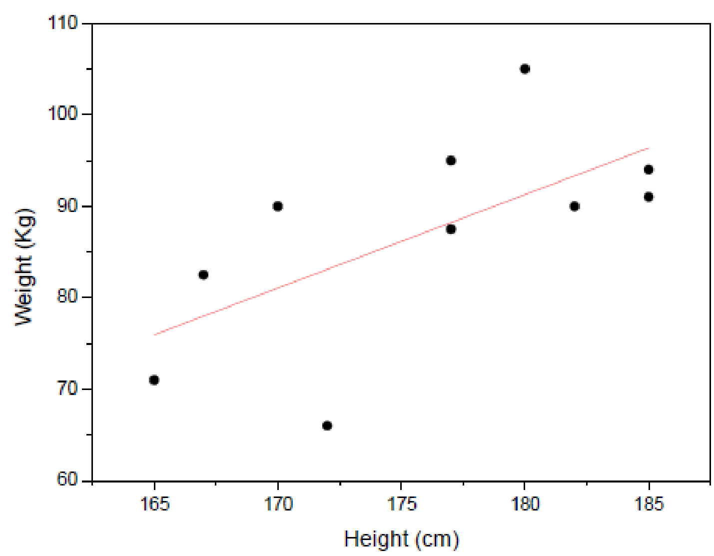
- **The direction of the relationship**
  - Positive correlation: the points go up from left to right.
  - Negative correlation: the points go down.
  - No correlation: points seem to be scattered randomly.
- **The strength of the relationship**
  - Strong correlation: when points are close to each other along a line.
  - Weak correlation: the points are more spread out.
  - Outliers: It is easy to spot points that do not fit the pattern.

Example

The table shows the weights (in kg) of adult males depend to some degree on their heights (in cm).

Heights	165	167	170	172	177	177	180	182	185	185
Weights	71	82.5	90	66	87.5	95	105	90	91	94

The next step is to plot the points on a rectangular coordinate system. The resulting set of points is called a scatter diagram.



The data appear to be approximated well by a straight line, and so we say that a linear relationship exists between the variables.

3.3 Bivariate statistical series parameters

In general, we can group the different types of parameters used to describe a bivariate statistical series into two groups:

- **Univariate-focused parameters:** The parameters that are focused on univariate analysis are those that characterize each variable on an individual basis using marginal and conditional distributions.

- **Bivariate-focused parameters:** Those that encapsulate the comprehensive joint distribution and the interrelationship between the two variables.

3.3.1 Univariate-focused parameters

a Marginal distributions

It describes each variable in isolation, disregarding the influence of the other variable. In the context of discrete cases, one computes the total by summing the joint probabilities.

Conversely, in the realm of continuous cases, the appropriate approach involves integrating over the other variable.

When we use marginal distributions, we can look at one variable at a time, which makes analysis easier. The sum of all the marginal probabilities must be 1 (or 100%). They are the starting point for looking into conditional distributions, independence, and how variables are related.

Marginal frequency Marginal frequency denotes the cumulative total of observations found within a specific row or column of a contingency table. This illustrates the distribution of a single variable without consideration of the influence or presence of another variable. The totals are located in the margins of the table, which gives rise to the designation "marginal frequency."

Joint frequencies represent the counts within each cell of the table, reflecting the frequency of specific combinations of categories.

The marginal frequency associated with x_i is given by

$$n_{i\cdot} = \sum_{j=1}^{p} n_{ij}$$

Likewise, the marginal frequency associated with y_j is given by

$$n_{\cdot j} = \sum_{i=1}^{r} n_{ij}$$

Marginal Relative Frequency The marginal relative frequency is expressed as a fraction or percentage of the whole.

The marginal relative frequency associated with x_i is given by

$$f_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \quad where\ i = 1, \cdots, r$$

Likewise, the marginal relative frequency associated with y_j is given by

$$f_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}} \quad where\ j = 1, \cdots, p$$

Marginal distributions table A marginal distribution table is typically derived from a two-way contingency table.

Marginal X distributions table

X	x_1	x_2	$\cdots$	x_i	$\cdots$	x_r	$\sum$
$n_{i\cdot}$	$n_{1\cdot}$	$n_{2\cdot}$	$\cdots$	$n_{i\cdot}$	$\cdots$	$n_{r\cdot}$	$n_{\cdot\cdot}$
$f_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}}$	$f_{1\cdot}$	$f_{2\cdot}$	$\cdots$	$f_{i\cdot}$	$\cdots$	$f_{r\cdot}$	1

Marginal Y distributions table

Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_p	$\sum$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot p}$	$n_{\cdot\cdot}$
$f_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}}$	$f_{\cdot 1}$	$f_{\cdot 2}$	$\cdots$	$f_{\cdot j}$	$\cdots$	$f_{\cdot p}$	1

Marginal means Marginal means are the averages of one variable, which are found by taking the mean of all levels of that variable. The means mentioned above are on the edges or margins of the table.

For X we have

$$\overline{x} = \frac{1}{n_{\cdot\cdot}} \sum_{i=1}^{r} \left(\sum_{j=1}^{p} (n_{ij}) \times x_i \right)$$

For Y we have

$$\overline{y} = \frac{1}{n_{\cdot\cdot}} \sum_{j=1}^{p} \left(\sum_{i=1}^{r} (n_{ij}) \times y_j \right)$$

Marginal variances It refers to the variance of a single variable within a joint (bivariate) distribution, excluding the effects of the other variable. This metric measures how spread out or dispersed the variable is on its own, without taking into account any correlation with other variables.

For X we have

$$\sigma_x^2 = \sum_{i=1}^{r} \frac{n_{i\cdot}}{n_{\cdot\cdot}} \times (x_i - \overline{x})^2$$

For Y we have

$$\sigma_y^2 = \sum_{j=1}^{p} \frac{n_{\cdot j}}{n_{\cdot\cdot}} \times (y_j - \overline{y})^2$$

b Conditional distributions

A conditional distribution delineates the probability distribution of a specific variable, dependent on the known value of another variable. This shows how the distribution of (Y) changes when (X) stays the same at a specific value.

The conditional distribution of X given that $Y = y_j$, is given by

$$X/Y = y_j = \left\{\left(x_i; f_i^j\right);\quad i = 1, \cdots, r\right\} \quad where \quad f_i^j = \frac{n_{ij}}{n_{.j}}$$

Likewise, The conditional distribution of Y given that $X = x_i$, is given by

$$Y/X = x_i = \left\{\left(y_j; f_j^i\right);\quad j = 1, \cdots, p\right\} \quad where \quad f_j^i = \frac{n_{ij}}{n_{i.}}$$

Conditional means The conditional mean, or conditional expectation, of a variable (Y) in relation to another variable (X) indicates the expected value of (Y) when (X) is fixed or set.

For $X\,/Y = y_j$ we have

$$\overline{x_j} = \sum_{i=1}^{r} f_i^j x_i$$

For $Y\,/X = x_i$ we have

$$\overline{y_i} = \sum_{j=1}^{p} f_j^i y_j$$

Conditional variances The conditional variance of a random variable (Y) concerning another variable (X) measures the degree of variability of (Y) in relation to its conditional mean, given the information about (X).

For $X\,/Y = y_j$ we have

$$V_j(x) = \sum_{i=1}^{r} f_i^j \left(x_i - \overline{x_j}\right)^2 \quad with \quad j = 1, \cdots, p$$

For $Y\,/X = x_i$ we have

$$V_i(y) = \sum_{j=1}^{p} f_j^i \left(y_j - \overline{y_i}\right)^2 \quad with \quad i = 1, \cdots, r$$

Example

Let us say we have information on 200 people with two quantities that can be measured:

- Age (in years)

- Daily Screen Time (hours per day)

We can make four bins (intervals) for each variable:

- Age bins (X):
  - A1: 0–19 years
  - A2: 20–39 years
  - A3: 40–59 years
  - A4: 60+ years
- Screen Time bins (Y):
  - S1: 0–2 hours
  - S2: 2–4 hours
  - S3: 4–6 hours
  - S4: 6+ hours

Now we make a 4×4 contingency table where each cell shows how many people fall into the right combination of intervals:

Age \ Screen Time	$0-2\ h$	$2-4\ h$	$4-6\ h$	$> 6\ h$
$0-19$ years	10	15	5	0
$20-39$ years	8	25	20	7
$40-59$ years	5	10	15	10
> 60 years	2	8	10	50

1. Construct the Marginal distributions table for X and Y.
2. Find the marginal means and variances.
3. Suppose that we want to study the distribution of Daily Screen Time for individuals aged 35, the variable studied is the conditional variable $X/Y = 35$.
   (a) Construct the conditional distributions table of this variable.

(b) Find the conditional mean and variance of this variable.

4. Let us now study the age distribution of Daily Screen Time equal to 5 h. The variable studied is the conditional variable $Y/X = [4, 6\ h[$.

   (a) Construct the conditional distributions table of this variable.

   (b) Find the conditional mean and variance of this variable.

Statistical independence For each value y_i of Y, the conditional distribution of $X/Y = y_i$ is identical to the marginal distribution of X; we say that X is statistically independent of Y.

X is statistically independent of Y if and only if :

$$\frac{n_{ij}}{n_{.j}} = \frac{n_{i.}}{n_{..}} \Leftrightarrow f_i^j = f_{i.} \quad \forall(i,j)$$

Properties

If X is statistically independent of Y , then:

- Y is also independent of X : we say that X and Y are independent.
- $n_{ij} \times n_{..} = n_{.j} \times n_{i.} \quad \forall(i,j) or f_{ij} = f_{.j} \times f_{i.}$
- $\frac{n_{(i-1)(j-1)}}{n_{i(j-1)}} = \frac{n_{(i-1)j}}{n_{ij)}} \forall(i,j)$

3.4 Covariance and linear correlation coefficient

We clarify covariance and the linear correlation coefficient, including their meaning and fundamental attributes.

3.4.1 Covariance

Covariance measures how much two variables change at the same time. It shows which way their linear relationship goes:

- Positive covariance means that the two variables tend to go up (or down) at the same time.
- Negative covariance means that when one variable goes up, the other one goes down.
- A covariance of 0 signifies a lack of linear correlation; however, it does not always imply independence.

The covariance of the variables X and Y is represented by $cov(X, Y)$, which is equal to

$$cov(X, Y) = \sum_{i=1}^{r}\sum_{j=1}^{p} f_{ij}(x_i - \overline{x})(y_j - \overline{y})$$

This can also be expressed as:

$$cov(X, Y) = \sum_{i=1}^{r}\sum_{j=1}^{p} f_{ij}x_i y_j - \overline{x}\overline{y}$$

3.4.2 Correlation coefficient

The correlation coefficient is a statistical measure that shows how strong and in what direction two variables are related in a linear way. The values go from -1 to $+1$.

- A value of $+1$ means that there is a perfect positive linear relationship.
- A value of -1 means that there is a perfect negative linear correlation.
- A value of 0 means that there is no linear relationship.

The Pearson Product-Moment Correlation Coefficient (r) is the most common type. It is the covariance of the two variables divided by the product of their standard deviations:

$$r(x, y) = \frac{cov(X, Y)}{\sigma_y \sigma_x}$$

Properties

- Symmetry: $r(X, Y) = r(Y, X)$.
- Scale-invariance: Unaffected by changes in units or shifts in mean.
- Sensitive to outliers: Extreme values can distort r significantly.

- Positive r: Both variables increase together.
- Negative r: One variable increases as the other decreases.
- Magnitude indicates strength:
    - $0.7 \leq r \leq 1.0$: Very strong correlation.
    - $0.3 \leq r \leq 0.7$: Moderate correlation.
    - $r < 0.3$: Weak or negligible correlation.

3.5 Regression curves and regression lines

Regression curves and regression lines are both concepts from statistics used to describe the relationship between variables.

3.5.1 Regression curves

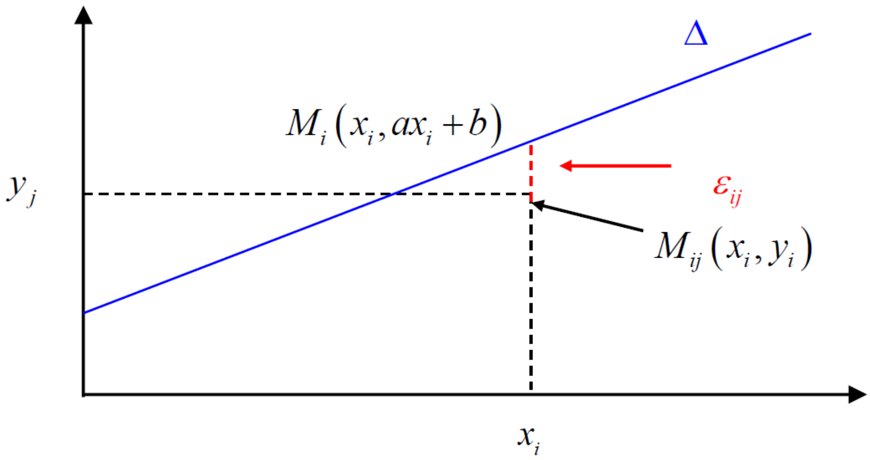
A regression curve is a curve that best fits the data points. It is used when the relationship between variables is not linear. It can take many forms—quadratic, cubic, exponential, logarithmic, etc. These curves are used when data points follow a pattern that is curved rather than straight, indicating a more complex relationship.

We want to give a graphic interpretation to the notion of correlation between 2 variables X and Y.

- The regression curve of Y in X is the curve $C_{Y/X}$ of points $M_i(x_i, \overline{y_j})$.
- The regression curve of X in Y is the curve $C_{X/Y}$ of points $M_j(\overline{x_i}, y_j)$.
- The areas of the points M_i and M_j are proportional respectively to $n_{i\cdot}$ and $n_{\cdot j}$.

3.5.2 Regression lines

A regression line is a straight line that best fits the data points on a scatter plot. The linear regression line models the relationship between two variables. It is used when the relationship between X and Y is approximately linear. We seek to summarize the cloud of points M_{ij} as best as possible by a straight line. The criterion used is that of "**least-squares**". The problem of its determination concerns the so-called linear adjustment method.



Δ is a straight line with equation :

$$y = ax + b$$

and

$$\varepsilon_{ij} = (y_j - ax - b)$$

The problem is how to identify a Δ line by minimizing the following quantity :

$$\sum_{i=1}^{r}\sum_{j=1}^{p}\frac{n_{ij}}{n_{..}} \times (\varepsilon_{ij})^2 = \sum_{i=1}^{r}\sum_{j=1}^{p}\frac{n_{ij}}{n_{..}} \times (y_j - ax_i - b)^2$$

The solution to the problem that ε_{ij} is counted parallel to the axis Y is the regression line of y on x, denoted by $D_{Y/X}$.

3.5.3 Equation of $D_{y/x}$

We will denote this equation by $y = \hat{a}x + \hat{b}$, because the exact values of a and b will never be known. After all, in statistical problems, we mainly work with samples.

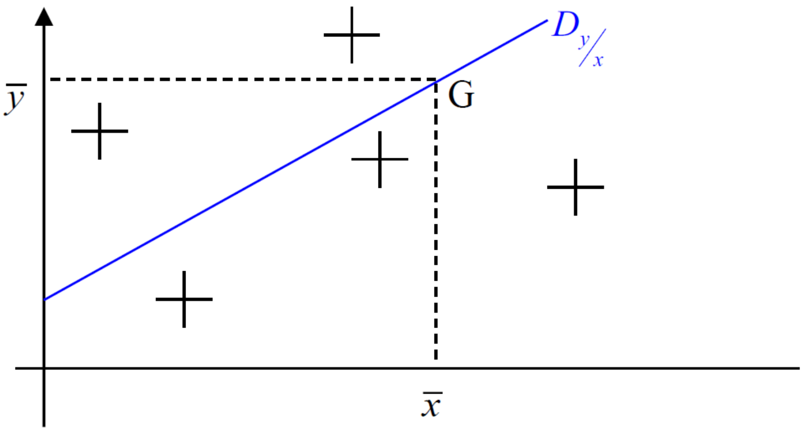
- $\hat{a}$ and $\hat{b}$ are therefore, for a given sample, approximate values of a and b.
- $\hat{a}$ is the slope of $D_{y/x}$ given by:

$$\hat{a} = \frac{cov(X, Y)}{\sigma_x^2}$$

- $\hat{b}$ represents the y-intercept of $D_{y/x}$, defined as follows:

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

The equation $\hat{b} = \bar{y} - \hat{a}\bar{x}$ translates to $D_{y/x}$ passes through $G(\bar{x}, \bar{y})$, which is the center of gravity of the point cloud.



3.5.4 Equation of $D_{x/y}$

The straight line $D_{x/y}$ is the line that minimizes the following quantity:

$$\sum_{i=1}^{r}\sum_{j=1}^{p}\frac{n_{ij}}{n_{..}} \times (e_{ij})^2 = \sum_{i=1}^{r}\sum_{j=1}^{p}\frac{n_{ij}}{n_{..}} \times \left(x_i - \tilde{a}y_j - \tilde{b}\right)^2$$

The equation of $D_{x/y}$ is given by :

$$x = \tilde{a}y + \tilde{b}$$

and

$$e_{ij} = (x_i - \tilde{a}y_j - \tilde{b})$$

The error e_{ij} is a deviation counted parallel to the axis (Ox). In the same way as for $D_{y/x}$, we obtain:

- $\tilde{a}$ is the slope of $D_{x/y}$ given by:

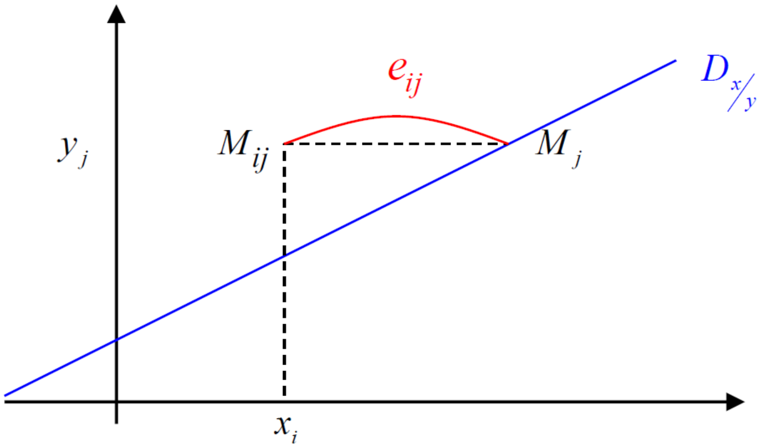
$$\tilde{a} = \frac{cov(X, Y)}{\sigma_y^2}$$

- $\tilde{b}$ represents the y-intercept of $D_{x/y}$, defined as follows:

$$\tilde{b} = \bar{x} - \tilde{a}\bar{y}$$

The final equation of $D_{x/y}$ is given by :

$$y = \frac{1}{\tilde{a}}x - \frac{\tilde{b}}{\tilde{a}}$$



3.5.5 Coefficient of determination

To assess the quality of a linear adjustment, we use the following indicator:

$$R^2 = \frac{[cov(X,Y)]^2}{\sigma_x^2 \sigma_y^2}$$

R^2 is the coefficient of determination for the linear regression; a higher value indicates a better fit.

Properties

- $R^2 = |r^2(x,y)|$: is the square of linear correlation coefficient
- $0 \leq R^2 \leq 1$, car $|r(x,y)| \leq 1$
- If $R^2 = 1$, there exists an affine relation between X and Y, the points M_{ij} are aligned.
- If $R^2 = 1$, the two adjustment lines are parallel to the axes.

3.6 Application

Weather stations have been established at various elevations to examine the length of the vegetative cycle in mountainous regions. The table below shows the average temperature (variable Y in degrees Celsius) and the altitude (variable X in thousands of meters) for each station.

X	1.1	1.2	1.5	1.6	1.7	1.9	2.2	2.5	2.8	3.1
Y	7.4	6	4.5	3.8	2.9	1.9	1	-1.2	-1.5	-4.5

What average temperature do you predict at 1400 m?

Chapter 4

Combinatory analysis

4.1 Sets and elements

A set is a fundamental concept in mathematics, defined as any well-defined collection of objects, which are called elements or members of the set. The "well-defined" aspect means that it is always clear whether an object belongs to the set or not. One usually uses :

- A, B, X, Y, $\cdots$ to denote sets.
- a, b, x, y, $\cdots$ to denote elements of sets.
- $a \in S$ denotes that a belongs to a set S.
- $\in$ means "is an element of".
- Sets are usually denoted using curly braces $\{ \ \}$.
- For example, the set of natural numbers less than 5 can be written as:

$$A = \{1, 2, 3, 4\}$$

- The objects within the set $(1, 2, 3, 4)$ are its elements. The order of elements and repetition do not matter in sets. For example, 1, 2, 3 is the same set as 3, 2, 1, 2.

Here we mention some important notes and points.

- Sets are usually denoted using curly braces $\{ \ \}$.

- For example, the set of natural numbers less than 5 can be written as:

$$A = \{1, 2, 3, 4\}$$

- The objects within the set $(1, 2, 3, 4)$ are its elements.
- The order of elements and repetition do not matter in sets. For example, $\{1, 2, 3\}$ is the same set as $\{3, 2, 1, 2\}$.

There are two ways to specify a particular set :

- $A = \{1, 3, 5, 7, 9\}$
- $B = \{x | x \textit{ is an even integer}, \quad x > 0\}$

Note that :

- The set A can also be written as $A = \{x | x \textit{ is an odd integer}, \quad x < 10\}$
- We can not list all the elements of B although we specify the set by $B = \{2, 4, 6, \cdots\}$

Let us have the following three different sets.

- $E = \{x | x^2 - 7x + 12 = 0\}$.
- $F = \{3, 4\}$.
- $G = \{3, 4, 4, 3\}$.

Then

$$E = F = G$$

4.1.1 Subsets

Suppose every element in a set A is also an element of a set B, that is, suppose $a \in A$ implies $a \in B$. Then A is called a subset of B. We also say that A is contained in B or that B contains A. This relationship is written.

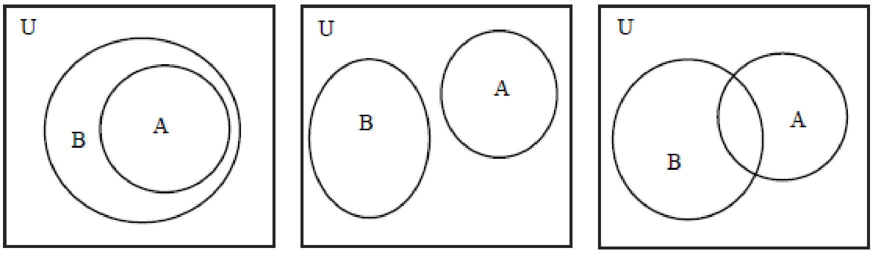
$$A \subset B \text{ or } B \supset A$$

Two sets are equal if they both have the same elements. That is

$$A = B \text{ if and only if } A \subset B \text{ and } B \subset A$$

4.1.2 Venn diagrams

A Venn diagram is a pictorial representation of sets in which enclosed areas in the plane represent sets. It shows logical links between sets. Closed curves—usually circles—represent each set, with crossings representing common components and non-overlapping portions representing unique elements.



(a) $A \subset B$ (b) *A and B* are disjoint (c) In general

Title

Here we find : The interior of a rectangle represents the Universal set U. If $A \subset B$, then the disk of A will be entirely within the disk representing B (Figure 1.a) If A and B are disjoint, then the disk representing A will be separated from the disk representing B (Figure 1.b) Hence in general we represent A and B as in (Figure 1.c)

4.2 Basic counting principles

This chapter employs two fundamental counting principles. The first includes addition, whereas the second involves multiplication.

- **Sum rule principle:** Assume that one event E can occur in m ways and another event F can occur in n ways, and that both events cannot occur at the same time. Then E or F can happen in $m + n$ different ways.

- **Product rule principle:** Assume there is an event E that can occur in m ways and, independent of this event, there is a second event F that can occur in n ways. Then combinations of E and F can occur $m \times n$ ways.

4.3 Mathematical functions

4.3.1 Factorial function

$n!$ denotes the product of positive numbers from 1 to n inclusive, often referred to as **"n factorial"**. Namely :

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-2)(n-1)n$$

Accordingly,

- $1! = 1$
- It is also convenient to define $0! = 1$.
- $n! = n(n-1)!$

Example

- $3! = 3 \times 2 \times 1 = 6,$
- $4! = 4 \times 3! = 24,$
- $5! = 5 \times 4! = 5(24) = 120,$

4.3.2 Binomial coefficients

The symbol $\binom{n}{r}$, read **"nCr"** where r and n are positive integers with $r \leq n$, is defined as follows :

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

Note that $n - (n - r) = r$, which yields the following important relation.

$$\binom{n}{n-r} = \binom{n}{r}$$

Example

- $\binom{8}{2} = \frac{8!}{2!(6)!} = \frac{8\times 7}{2\times 1} = 28$

- $\binom{10}{7} = \binom{10}{3} = \frac{10\times9\times8}{3\times2\times1} = 120$

4.4 Permutations

A permutation of the objects is any arrangement of a set of n objects in a given order. An "r-permutation" is any arrangement of any $r \leq n$ of these items in a given order.

Example

Consider the set of letters A, B, C, and D. Then :

- $BDCA$, $DCBA$, and $ACDB$ are permutations of the four letters (taken all at a time).
- BDC, DBA, and ADB are permutations of the four letters (taken three at a time).
- BC, BA, and AD are permutations of the four letters (taken two at a time).

The number of permutations of n objects taken r at time will be:

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

4.4.1 Permutations with repetitions

We frequently want to know the number of permutations of a multiset, which is a collection of items, some of which are similar. We will let

$$P(n; n_1, n_2, \cdots, n_r)$$

denote the number of permutations of n objects of which n_1 are alike, n_2 are alike,$\cdots$, n_r are alike. The general formula follows :

$$P(n; n_1, n_2, \cdots, n_r) = \frac{n!}{n_1!n_2!\cdots n_r!}$$

Example

Determine the number m of eight-letter words that may be created from the letters in the word "**NINETEEN**". We seek the number of permutations of 8 objects, of which 3 are alike (the three N's), and 3 are alike (the three E's). By equation

$$m = P(8; 3, 3) = \frac{8!}{3!3!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = 1120$$

4.4.2 Ordered samples

Many problems involve selecting an element from a set S of n elements. When we select one element after another, say r times, we refer to the selection as an ordered sample of size r. We look at two scenarios.

a Sampling with replacement

Before selecting the next element, the element in the set S is replaced. As a result, each time there are n possibilities to select an element (repetition is permitted). According to the Product rule, the number of such samples is:

$$n \cdot n \cdot n \cdots n \cdot n \cdot = n^r$$

b Sampling without replacement

In this case, the element in the set S is not altered before the next element is chosen. As a result, the ordered sample contains no repeats. This type of sample is simply an r-permutation. As a result, the number of such samples is:

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

Example

Three cards are drawn sequentially from a 52-card deck. Determine the number m of possible solutions:

- (a) with replacement;
- (b) without replacement.

Answer:

- (a) Each card can be selected in 52 different ways. Thus

$$m = 52 \cdot 52 \cdot 52 = 140608.$$

- (b) There is no substitute in this case. Thus, the first card can be chosen in 52 different ways, the second in 51 different ways, and the third in 50 different ways. Therefore:

$$m = P(52, 3) = 52 \cdot 51 \cdot 50 = 132600.$$

4.5 Combinations

Assume S is a set with n items. Any selection of r of the items where order does not matter is a mixture of these n elements taken r at a time. An r-combination is a selection that is just a subset of S with r members. The total number of such combinations will be represented by

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

Example

An engineer wanted to purchase equipment consisting of 3 computers, 2 scanners, and 4 printers from a man who owns 6 computers, 5 scanners, and 8 printers. Determine the number m of options available to the farmer.

The engineer has $C(6, 3)$ options for computers, $C(5, 2)$ options for scanners, and $C(8, 4)$ options for printers.

Thus the number m of choices follows:

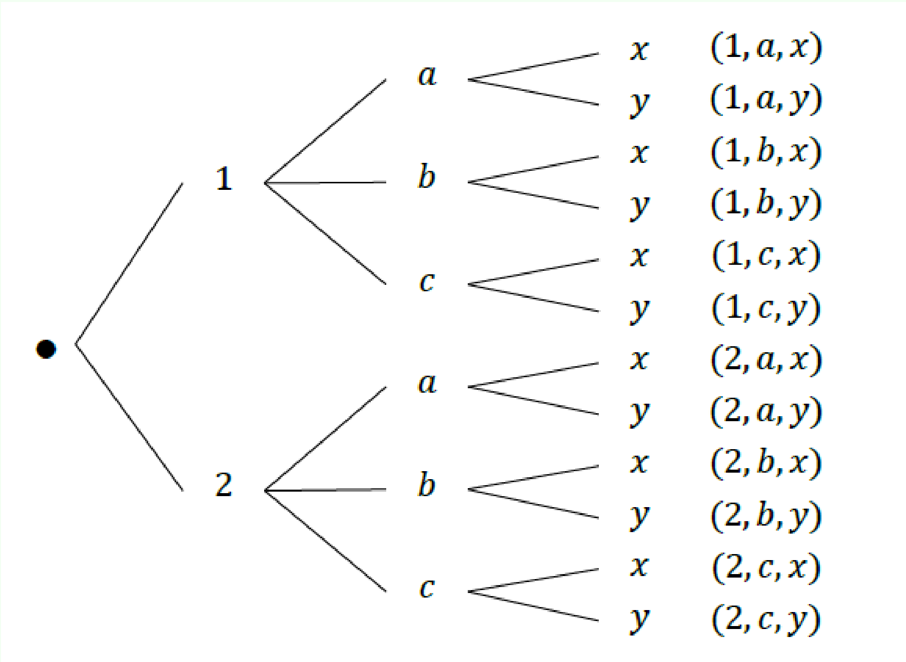
$$m = C(6, 3)C(5, 2)C(8, 4) = \frac{6!}{3!3!} \cdot \frac{5!}{2!3!} \cdot \frac{8!}{4!4!} = 14000$$

4.6 Tree Diagrams

A tree diagram is a tool for listing all of the possible outcomes of a series of events, where each event can occur in a limited number of ways. The following example shows how to make a tree diagram.

Example

Calculate the product set $A \times B \times C$, where $A = \{1, 2\}$, $B = \{a, b, c\}$, and $C = \{x, y\}$.



The tree is built from left to right here, with the number of branches at each point representing the various outcomes of the next event. Each tree endpoint (leaf) is labeled with the corresponding element of $A \times B \times C$. As previously stated, $A \times B \times C$ contains

$$m = 2(3)(2) = 12$$

Chapter 5

Probability, conditioning and independence

5.1 Probability and events

Probability theory began with the study of chance games like roulette and cards. The probability p of an event A was defined as follows : if A can happen in s of n equally likely ways, then

$$p = P(A) = \frac{s}{n}$$

The **sample space** is defined as the set S of all possible results of a given experiment. A **sample point** or **sample** is a specific outcome (an element in S).

- An event A is a collection of outcomes, or a subset of the sample space S.
- An **elementary event** is one that consists of $\{a\}$ single sample $a \in S$.
- The empty $\emptyset$ and S itself are events.
- $\emptyset$ set is frequently referred to as the **impossible** event, and S is referred to as the **certain** or **sure** event.

Using the various set operations, we may combine events to create new events:

- $A \cup B$ is the event that occurs if A or B (or both) occurs.
- $A \cap B$ is the event that occurs if A and B both occur.

- A^c, the complement of A, is the event that occurs if A does not occur.
- Two events A and B are called **mutually exclusive** if they are disjoint $(A \cap B) = \emptyset$.

Example

Toss a die and note the number that comes up on top. The sample space then consists of the following six numbers:

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

Let A represent the occurrence of an even number, B the occurrence of an odd number, and C the occurrence of a prime number:

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

Then :

- $A \cup C = \{2, 3, 4, 5, 6\}$
- $B \cap C = \{3, 5\}$
- $C^c = \{1, 4, 6\}$
- $A \cap B = \emptyset$: A and B are mutually exclusive.

5.2 Probability rules

Let S represent a sample space, ε represent a class of events, and P represent a real-valued function defined on ε. Then P is referred to as a probability function, and $P(A)$ is referred to as the probability of the occurrence A if the following rules are true.

- **R1:** For each event A, $0 \leq P(A) \leq 1$.

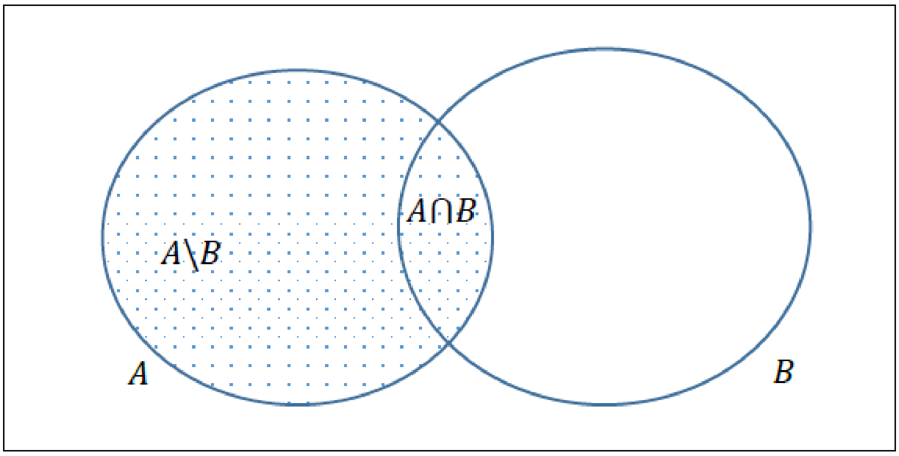
- **R2:** $P(S) = 1$.

- **R3:** $P(\emptyset) = 0$.

- **R4:** $P(A^c) = 1 - P(A)$.

- **R5:** If $A \subset B$, then $P(A) \leq P(B)$

- **R6:** If A and B are any two events, then



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

- **R7:** If A and B are mutually exclusive events, then

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

- **R8:** If A and B are any two events, then

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

- **R9:** If $A_1, A_2, \cdots,$ is a sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

5.3 Probabilistic spaces

Let $S = \{a_1, a_2, \cdots, a_n\}$ be a finite sample space. A finite probability space is obtained by assigning to each point $a_i \in S$ a real number p_i, called the probability of a_i, satisfying the following properties:

i) each p_i is nonnegative, $p_i \geq 0$

ii) the sum of the p_i is one,

$$p_1 + p_2 + \cdots + p_n = 1$$

The probability $P(A)$ of any event A is then defined to be the sum of the probabilities of the points in A. For notational convenience we write $P(a_i)$ for $P(\{a_i\})$.

Example

Let three coins be tossed and the number of heads observed; then the sample space is

$$S = \{0, 1, 2, 3\}$$

We obtain a probability space by the following assignment

- $P(0) = \frac{1}{8}$
- $P(1) = \frac{3}{8}$
- $P(2) = \frac{3}{8}$
- $P(3) = \frac{1}{8}$

Since each probability is nonnegative and the sum of the probabilities is 1. Let A be the event that at least **one head** appears and let B be the event that **all heads or all tails appear**:

- $A = \{1, 2, 3\}$
- $B = \{0, 3\}$

Then, by definition

$$P(A) = P(1) + P(2) + P(3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}$$

And

$$P(B) = P(0) + P(3) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

5.3.1 Equiprobable spaces

Frequently, the physical characteristics of an experiment suggest that the various outcomes of the sample space be assigned equal probabilities. Such a finite probability space S, where each sample point has the same probability, will be called an **equiprobable** or **uniform** space. In particular, if S contains n points, then the probability of each point is $\frac{1}{n}$. Furthermore, if an event A contains r points then its probability is

$$r \cdot \frac{1}{n} = \frac{r}{n}$$

In other words,

$$P(A) = \frac{\text{number of ways that the event } A \text{ can occur}}{\text{number of ways that the sample space } S \text{ can occur}}$$

Example

Let 2 items be chosen at random from a lot containing 12 items, of which 4 are defective. Let

- $A = \{\text{both items are defective}\}$
- $B = \{\text{both items are non-defective}\}$

Find $P(A)$ and $P(B)$. Now

- S can occur in $C(12, 2) = 66$ ways.
- A can occur in $C(4, 2) = 6$ ways.
- B can occur in $C(8, 2) = 28$ ways.

Accordingly

- $P(A) = \frac{6}{66} = \frac{1}{11}$

- $P(B) = \frac{28}{66} = \frac{14}{33}$

The probability that at least one item is defective. It is the complement of B, $C = B^c$. Thus,

$$P(C) = P(B^c) = 1 - P(B) = 1 - \frac{14}{33} = \frac{19}{33}$$

5.3.2 Infinite sample spaces

Now suppose $S = \{a_1, a_2, \cdots\}$ is a countably infinite sample space. As in the finite case, we obtain a probability space by assigning to each $a \in S$ a real number p_i, called its probability, such that.

i) $p_i \geq 0$

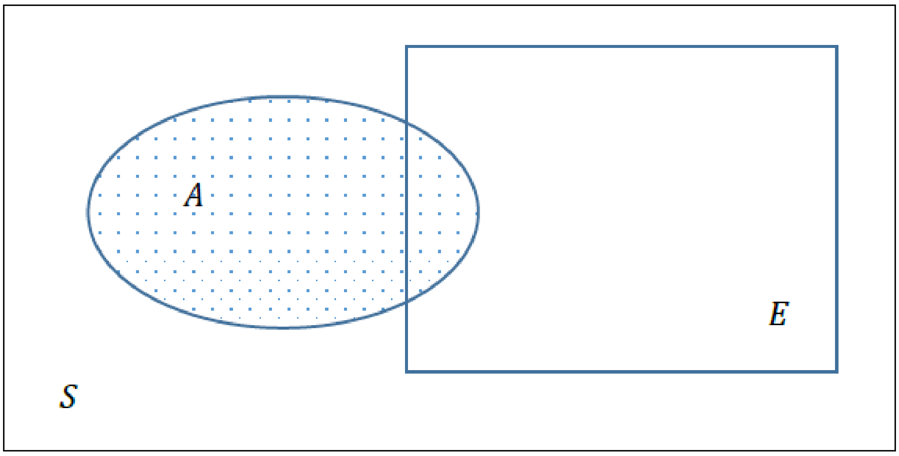
ii) $p_1 + p_2 + \cdots = \sum_{i=1}^{\infty} p_i = 1$

The probability $P(A)$ of any event A is then the sum of the probabilities of its points.

5.4 Conditional probability

Let E be an arbitrary event in a sample space S with $P(E) > 0$. The probability that an event A occurs once E has occurred is called the conditional probability of A given E, written $P(A|E)$, where :

$$P(A|E) = \frac{P(A \cap E)}{P(E)}$$



As seen in the adjoining Venn diagram, $P(A|E)$ in a certain sense measures the relative probability of A with respect to the reduced space E.

In particular, if S is a finite equiprobable space and $|A|$ denotes the number of elements in an event A, then

- $P(A \cap E) = \frac{|A \cap E|}{|S|}$
- $P(E) = \frac{|E|}{|S|}$

Hence,

$$P(A|E) = \frac{P(A \cap E)}{P(E)} = \frac{|A \cap E|}{|E|}$$

That is,

$$P(A|E) = \frac{\text{number of ways } A \text{and} E \text{ can occur}}{\text{number of ways } E \text{ can occur}}$$

Example

Let a pair of fair dice be tossed. If the sum is 6, find the probability that one of the dice is a 2. In other words, if

- $E = \{(l, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$
- $A = \{\text{a 2 appears on at least one die}\}$

Now E consists of five elements and two of them, $(2, 4)$ and $(4, 2)$, belong to A :

$$A \cap E = \{(2, 4), (4, 2)\}$$

Then,

$$AP(A|E) = \frac{2}{5}$$

On the other hand, since A consists of eleven elements,

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

and S consists of 36 elements,

$$P(A) = \frac{11}{36}$$

5.4.1 Multiplication theorem

If we cross-multiply the above equation defining conditional probability and use the fact that $A \cap E = E \cap A$, we obtain the following useful formula.

$$P(E \cap A) = P(E)P(A|E)$$

This theorem can be extended by induction. For any events $A_1, A_2, \cdots, A_n$. we have

$$P(A_1 \cap A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n_1})$$

This theorem is called the multiplication theorem.

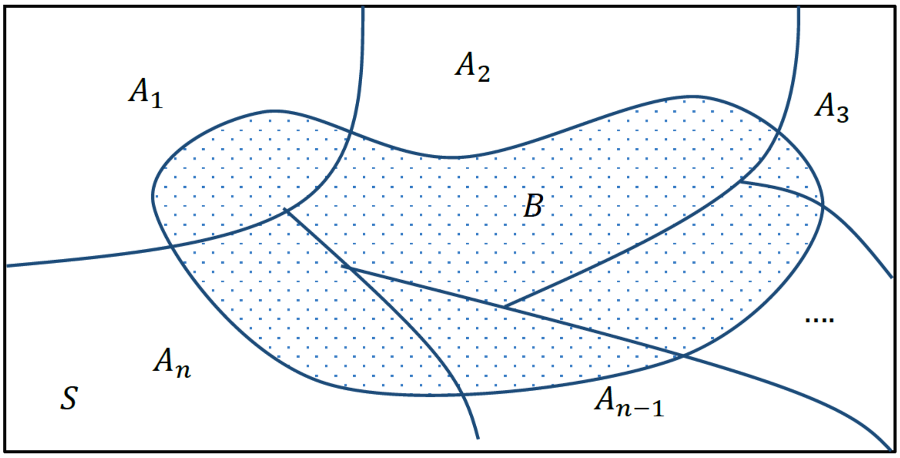
5.4.2 Finite stochastic processes and tree diagrams

A (finite) stochastic process is a (finite) set of trials in which each experiment has a finite number of outcomes with given probabilities. A tree diagram is a convenient means of expressing such a process and determining the likelihood of each occurrence; the preceding section's multiplication theorem is used to compute the chance that the result represented by any given path of the tree occurs.

5.5 Bayes formula

Suppose the events $A_1, A_2, \cdots, A_n$ form a partition of a sample space S; that is, the events A_i are mutually exclusive and their union is S. Now let B be any other event. Then

$$B = S \cap B = (A_1 \cup A_2 \cup \cdots \cup A_n) \cap B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$



Where the $A_i \backslash B$ are also mutually exclusive. Accordingly,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

Thus, by the multiplication theorem,

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$$

On the other hand, for any i, the conditional probability of A_i given B is defined by

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

Thus,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

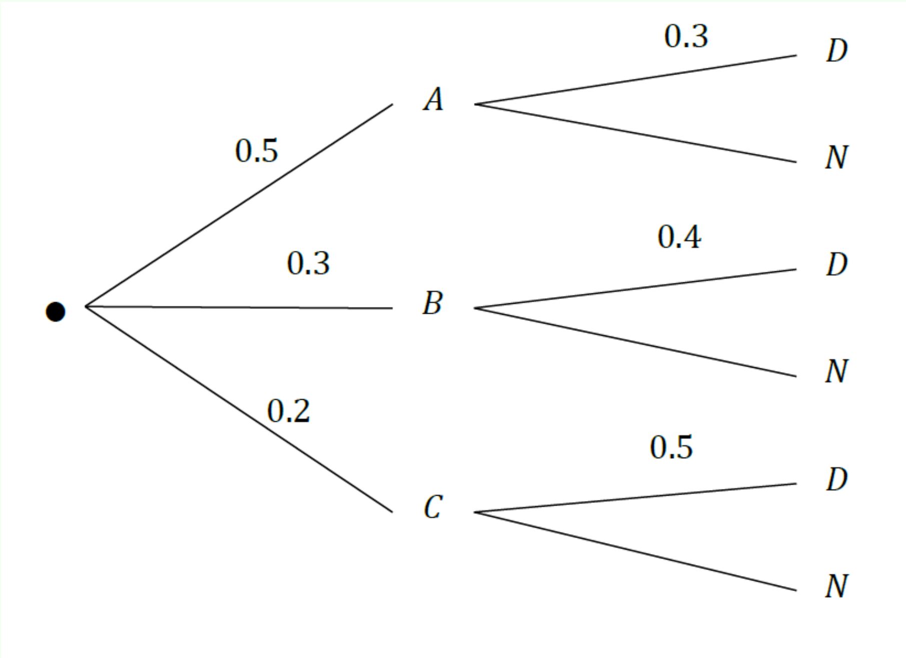
Example

Three machines A, B, and C produce respectively 50%, 30%, and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4%, and 5%. If an item is selected at random, find the probability that the item is defective.

Let X be the event that an item is defective. Then by (1) above,

$$\begin{aligned} P(X) &= P(A)P(X|A) + P(B)P(X|B) + P(C)P(X|C) \\ &= (.50)(.03) + (.30)(.04) + (.20)(.05) \\ &= .037 \end{aligned}$$

Observe that we can also consider this problem as a stochastic process having this tree diagram.



5.6 Independence

An event B is said to be independent of an event A if the probability of B occurring is unaffected by whether or not A has occurred. In other words, if the probability of B equals the conditional probability of B given A:

$$P(B) = P(B|A)$$

Now substituting $P(B)$ for $P(B|A)$ in the multiplications theorem we obtain

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

So events A and B are independent if

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

Otherwise, they are dependent.

Example

Let a fair coin be tossed three times; we obtain the equiprobable space.

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Consider the events

- $A = \{\text{first toss is heads}\}$
- $B = \{\text{second toss is heads}\}$
- $C = \{\text{exactly two heads are tossed in a row}\}$

Clearly, A and B are independent events; this fact is verified below. We claim that A and C are independent, but that B and C are dependent. We have then

$$\begin{aligned} P(A) &= P(\{HHH, HHT, HTH, HTT\}) = \frac{4}{8} = \frac{1}{2} \\ P(B) &= P(\{HHH, HHT, THH, THT\}) = \frac{4}{8} = \frac{1}{2} \\ P(C) &= P(\{HHT, THH\}) = \frac{2}{8} = \frac{1}{4} \end{aligned}$$

Then

$$\begin{aligned} P(A \cap B) &= P(\{HHH, HHT\}) = \frac{1}{4} \\ P(A \cap C) &= P(\{HHT\}) = \frac{1}{8} \\ P(B \cap C) &= P(\{HHT, THH\}) = \frac{1}{4} \end{aligned}$$

Accordingly,

$$\begin{aligned}
P(A)P(B) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} = P(A\cap B), \text{ and so } A \text{ and } B \text{ are independent}\\
P(A)P(C) &= \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8} = P(A\cap C), \text{ and so } A \text{ and } C \text{ are independent}\\
P(B)P(C) &= \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8} \neq P(B\cap C), \text{ and so } B \text{ and } C \text{ are dependent}
\end{aligned}$$

Three events A, B and C are independent if:

i) $P(A\cap B) = P(A)P(B)$, $P(A\cap C) = P(A)P(C)$ and $P(B\cap C) = P(B)P(C)$

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

Note

Three events may be pairwise independent but not independent themselves.

Example

Let a pair of fair coins be tossed; here

$$S = \{HH, HT, TH, TT\}$$

is an equiprobable space. Consider the events

- $A = \{\text{heads on the first coin}\} = \{HH, HT\}$
- $B = \{\text{heads on the second coin}\} = \{HH, TH\}$
- $C = \{\text{heads on exactly one coin}\} = \{HT, TH\}$

Then $P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$ and

$$\begin{aligned}
P(A\cap B) &= P(\{HH\}) = \frac{1}{4}\\
P(A\cap C) &= P(\{HT\}) = \frac{1}{4}\\
P(B\cap C) &= P(\{TH\}) = \frac{1}{4}
\end{aligned}$$

Thus, the events are pairwise independent. However,

$$\begin{aligned}
A\cap B\cap C &= \emptyset \text{ and so}\\
P(A\cap B\cap C) &= P(\emptyset) = 0 \neq P(A)P(B)P(C)
\end{aligned}$$

In other words, the three events are not independent.

Chapter 6

Random variables and usual probability laws

6.1 Definitions and properties

A random variable X on a sample space S is a function from S into the real number set R, such that each interval in R has a preimage that is an event in S.

We express the probability of the events "X maps into a" and "X maps into the interval $[a, b]$" using the abbreviations $P(X = a)$ and $P(a \leq X \leq b)$. In other words,

$$P(X = a) = P(\{s \in S : X(s) = a\})$$

and

$$P(a \leq X \leq b) = P(\{s \in S : a \leq X(s) \leq b\})$$

6.2 Distribution function

Let X be a random variable on a sample space S with a finite image set; say,

$$X(S) = \{x_1, x_2, \cdots, x_n\}$$

By defining the probability of x_i, which is $P(X = x_i)$, we turn $X(S)$ into a probability space, and we write $f(x_i)$. The distribution or probability function of X is denoted by this function f on $X(S)$, and it is typically provided as a table:

x_1	x_2	$\cdots$	x_n
$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$

The distribution f satisfies the conditions :

- $f(x_i) \geq 0$
- $\sum_{i=1}^{n} f(x_i) = 1$

6.3 Mean (expectation)

Now, if X is a random variable with the above distribution, then the mean or the expectation of X, denoted by $E(X)$ or μ_x, is defined by:

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

That is, $E(X)$ is the weighted average of the possible values of X, each value weighted by its probability.

Example

A pair of fair dice is tossed. We obtain the finite equiprobable space S consisting of the 36 ordered pairs of numbers between 1 and 6a:

$$S = \{(1,1), (1,2), ..., (6,6)\}$$

Let X assign to each point (a, b) in S the maximum of its numbers, i.e. $X(a, b) = \max(a, b)$. Then X is a random variable with image set.

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

We compute the distribution f of X:

- $P(X = 1) = P(\{(1,1)\}) = \frac{1}{36}$
- $P(X = 2) = P(\{(1,2), (2,2), (2,1)\}) = \frac{3}{36}$
- $P(X = 3) = P(\{(1,3), (2,3), (3,3), (3,2), (3,1)\}) = \frac{5}{36}$
- $P(X = 4) = P(\{(1,4), (2,4), (3,4), (4,4), (4,3), (4,2), (4,1)\}) = \frac{7}{36}$
- $P(X = 5) = P(\{(1,5), (2,5), \cdots, (5,5), \cdots, (5,2), (5,1)\}) = \frac{9}{36}$

- $P(X=6)=P(\{(1,6),(2,6),\cdots,(6,6),\cdots,(6,2),(6,1)\})=\frac{11}{36}$

This information is put in the form of a table as follows:

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

We next compute the mean of X :

$$E(x)=\sum_{i=1}^{6} x_i f(x_i)=1\cdot\frac{1}{36}+2\cdot\frac{3}{36}+3\cdot\frac{5}{36}+4\cdot\frac{7}{36}+5\cdot\frac{9}{36}+6\cdot\frac{11}{36}=4.47$$

When it comes to gambling games, the expected value E of the game is regarded as the player's value. If E is positive, the game is considered advantageous to the player; if E is negative, it is considered unfavorable. It is a fair game if $E=0$.

Example

A player tosses a fair die. If a prime number occurs, he wins that number of dollars, but if a non-prime number occurs, he loses that number of dollars. The possible outcomes x_i of the game with their respective probabilities $f(x_i)$ are as follows:

x_i	2	3	5	-1	-4	-6
$f(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The expected value of the game is

$$E(x)=\sum_{i=1}^{6} x_i f(x_i)=2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+5\cdot\frac{1}{6}-1\cdot\frac{1}{6}-4\cdot\frac{1}{6}-6\cdot\frac{1}{6}=-\frac{1}{6}$$

Thus, the game is **unfavorable** to the player since the expected value is negative.

Proprieties

- Let X be a random variable and k a real number. Then

$$E(kX)=kE(X)E(X+k)=E(X)+k$$

- Let X and Y be random variables on the same sample space S. Then

$$E(X+Y)=E(X)+E(Y)$$

- Let $X_1, X_2, \cdots, X_n$ be random variables on S. Then

$$E\left(X_1 + \cdots + X_n\right) = E(X_1) + \cdots + E(X_n)$$

6.4 Variance and standard deviation

The mean of a random variable X measures the "average" value of X. Moreover, the variance of X measures the "dispersion" of X. Let X be a random variable with the following distribution:

x_1	x_2	$\cdots$	x_n
$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$

Then the variance of X, denoted by $Var(X)$, is defined by

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = E\left((X - \mu)^2\right)$$

where μ is the mean of X. The standard deviation of X, denoted by σ_x, is the (nonnegative) square root of $Var(X)$:

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

There is an alternate formula for calculating the variance of the random variable X.

$$Var(X) = \sum_{i=1}^{n} x_i^2 f\left(x_i\right) - \mu^2 = E\left((X)^2\right) - \mu^2$$

6.5 Joint distribution

Let X and Y be random variables on a sample space S with respective image sets

$$X(S) = \{x_1, x_2, \cdots, x_n\} Y(S) = \{y_1, y_2, \cdots, y_m\}$$

We make the product set

$$X(S) \times Y(S) = \{(x_1, y_1), (x_1, y_2), \cdots, (x_n, y_m)\}$$

into a probability space by defining the joint probability function of X and Y of the ordered pair (x_i, y_j) to be $P(X = x_i, Y = y_j)$, which we write $h(x_i, y_j)$.

	y_1	y_2	$\cdots$	y_m	$\sum$
x_1	$h(x_1, y_1)$	$h(x_1, y_2)$	$\cdots$	$h(x_1, y_m)$	$f(x_1)$
x_2	$h(x_2, y_1)$	$h(x_2, y_2)$	$\cdots$	$h(x_2, y_m)$	$f(x_2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$h(x_n, y_1)$	$h(x_n, y_2)$	$\cdots$	$h(x_n, y_m)$	$f(x_n)$
$\sum$	$g(y_1)$	$g(y_2)$	$\cdots$	$g(y_m)$	1

The above functions f and g are defined by

$$f(x_i) = \sum_{j=1}^{m} h(x_i, y_j)g(y_j) = \sum_{i=1}^{n} h(x_i, y_j)$$

$f(x_i)$ and $g(y_j)$ are called the marginal distributions. The joint distribution h satisfies the conditions

- $h(x_i, y_j) \geq 0$
- $\sum_{i=1}^{n} \sum_{j=1}^{m} h(x_i, y_j) = 1$

The covariance of X and Y, denoted by $Cov(X, Y)$ is defined by

$$Cov(X, Y) = \sum_{i,j} (x_i - \mu_x)(y_j - \mu_y) = E\left[(X - \mu_i)(Y - \mu_j)\right]$$

The correlation of X and Y, denoted by $\rho(X, Y)$, is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

The correlation ρ is dimensionless and has the following properties:

- $\rho(X, Y) = \rho(Y, X)$
- $\rho(X, X) = 1$ and $\rho(X, -X) = -1$
- $-1 \leq \rho \leq 1$
- $\rho(aX + b, cY + d) = \rho(X, Y)$, if $a, c \neq 0$

6.6 Independent random variables

A finite number of random variables $X, Y, \cdots, Z$ on a sample space S are said to be independent if

$$P(X = x_i, Y = y_j, \cdots, Z = z_k) = P(X = x_i)P(Y = y_j)\cdots)P(Z = z_k)$$

for any values $x_i, y_j \cdots, z_k$. In particular, X and Y are independent if

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

Now if X and Y have respective distributions f and g, and joint distribution h, then the above equation can be written as

$$h(x_i, y_j) = f(x_i)g(y_j)$$

In other words, X and Y are independent if each entry $h(x_i, y_i)$ is the product of its marginal entries.

6.6.1 Proprieties

We establish some important properties of independent random variables that do not hold in general.

Let X and Y be independent random variables. Then:

- $E(XY) = E(X)E(Y)$
- $Var(X + Y) = Var(X) + Var(Y)$
- $Cov(X, Y) = 0$

Example

Let X and Y be random variables with the following joint distribution:

	2	3	4	$\sum$
1	0.6	0.15	0.09	0.30
2	0.14	0.35	0.21	0.70
$\sum$	0.20	0.50	0.30	1

Thus, the marginal distributions of X and Y are as follows:

x	1	3
f(x)	0.30	0.70

y	2	3	4
g(y)	0.20	0.50	0.30

X and Y are independent random variables since each entry of the joint distribution can be obtained by multiplying its marginal entries.

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

for each i and each j.

6.7 Probability Laws

The objective of this section is to understand the fundamental laws of probability, their properties, and how to use them in concrete contexts.

6.7.1 Discrete Laws

a Bernoulli's Law

A discrete random variable X is said to adhere to Bernoulli's law if it assumes only two possible values: $X \in \{0, 1\}$ where $p = P(X = 1)$ is the probability of success.

$$P(X = x) = \begin{cases} 1 - p & \text{si } x = 0, \\ p & \text{si } x = 1. \end{cases}$$

Properties

- **The expectation :** $E[X] = p$
- **The variance :** $Var(X) = p(1 - p)$

Among the Applications of Bernoulli's law, we can cite, as an example, coin tosses (heads = 1, tails = 0) and the success/failure of an experiment.

b Binomial Law

The binomial distribution models the number of successes in a fixed number of independent trials, each with the same probability of success, p. If a random variable X follows this distribution with parameters n (number of trials) and p, it is denoted as $X \sim \text{Binomial}(n, p)$.

The probability of observing exactly k successes is given as:

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

Where, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$) is the binomial coefficient.

Properties

- **Expectation:** $E[X] = np$
- **Variance:** $Var(X) = np(1-p)$
- **Symmetry:** The distribution is symmetric when $p = 0.5$.
- **Median:** The median is usually between $\lfloor np \rfloor$ and $\lceil np \rceil$, although this depends on whether np is an integer.

c Poisson's law

A random variable X follows a Poisson distribution with parameter $\lambda > 0$ (denoted $X \sim Poisson(\lambda)$) if it satisfies the following conditions:

- Events occur independently.
- The average rate of events in the interval is λ.
- Two events cannot occur simultaneously.
- The Poisson distribution is often used to model rare events over continuous domains.

The probability that exactly k events occur is given by:

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

Here, λ is both the expected value and the variance of the distribution.

Properties

- **Expectation:** $E[X] = \lambda$
- **Variance:** $Var(X) = \lambda$
- **Law of Rare Events:** The Poisson distribution is the limit of the binomial distribution with $n \to \infty, p \to 0$, and $np = \lambda$ fixed.
- **Skewness:** The distribution is positively skewed, but becomes more symmetric as λ increases.

d Geometric Law

The geometry distribution describes the number of separate Bernoulli trials with the same chance of success p needed to get the first success. There are two common parameterizations:

1) Number of trials until first success:

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

.

2) Number of failures before the first success:

$$P(Y = k) = (1-p)^{k}p) \quad \text{for} \quad k = 0, 1, 2, \ldots$$

.

Properties

- **Trials until success** (X)**:** ◦ **Expectation:** $E[X] = \frac{1}{p}$
  - ◦ **Variance:** $Var(X) = \frac{1-p}{p^2}$
- **Failures before success** (Y)**:** ◦ **Expectation:** $E[Y] = \frac{1-p}{p}$
  - ◦ **Variance:** $Var(Y) = \frac{1-p}{p^2}$

e Hypergeometric Law

A hypergeometric distribution explains the odds of achieving exactly k successes from n draws without replacement in a finite population of size N with K successes. The binomial distribution treats each draw independently, but the hypergeometric distribution adjusts probability when items are removed.

The probability distribution is given by:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Where: N represents the total population size, K denotes the number of successes in the population, n is the number of draws, and k is the number of observed successes in the sample.

Properties

- **Expectation:** $E[X] = n \cdot \frac{K}{N}$
- **Variance:** $Var(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$
- **Dependent Trials:** Each trial affects the next in sampling without replacement, unlike binomial distribution trials.

6.7.2 Continues Laws

a Uniform Law

In the continuous uniform distribution, which is also called the rectangle distribution, every value in a closed interval $[a, b]$ has the same chance of happening. It is shown as $X \sim U(a, b)$. When only the range is given, this distribution has the most uncertainty (entropy).

The probability density function of a continuous uniform distribution is:

$$\begin{aligned} f(x) &= \tfrac{1}{b-a} \qquad && a \leq x \leq b \\ f(x) &= 0 && \text{otherwise} \end{aligned}$$

Properties

- **Expectation:** $E[X] = \frac{a+b}{2}$
- **Variance:** $Var(X) = \frac{(b-a)^2}{12}$
- **Symmetry:** The distribution is symmetric around its mean.
- **Equal-likelihood:** The probability is proportional to the length of any subinterval of fixed length within $[a, b]$, regardless of position.
- **Maximum entropy:** The uniform distribution has the maximum entropy of all distributions with a given support, making it the least informative given merely the range.

b Normal (Gaussian) Law

The Gaussian distribution, or normal distribution, is a continuous probability distribution for real-valued random variables. It has two parameters: mean μ and variance σ^2. Its probability density function is:

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

The Central Limit Theorem asserts that the sum (or average) of many independent, identically distributed variables tends toward a normal distribution under specific conditions, making this distribution important in statistics.

Properties

- **Expectation:** As the center of the distribution, the mean μ also represents the median and mode.
- **Variance:** $Var(X) = \sigma^2$
- **Symmetry:** The normal distribution is symmetric around its mean, with the mean, median, and mode all equal

c Exponential Law

The exponential distribution characterizes the time between events in a Poisson point process, where events occur independently and continuously at a constant average rate λ. Continuous, specific gamma distribution. It is memoryless and continuous like the geometric distribution.

The probability density function is given by:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for} \quad x \geq 0 \\ 0 \quad \text{for} \quad x < 0 \end{cases}$$

Where $\lambda > 0$ is the rate parameter

Properties

- **Expectation:** $E[X] = \frac{1}{\lambda}$
- **Variance:** $\text{Var}(X) = \frac{1}{\lambda^2}$
- **Higher Moments:** $E[X^n] = \frac{n!}{\lambda^n}$

d Gamma Law

The Gamma distribution is a continuous probability distribution with two positive parameters: shape α or k and scale (θ) or rate $(\beta = \frac{1}{\theta})$. It is typically used to model Poisson process waiting times till k events.

A gamma-distributed random variable X with shape α and rate β has the PDF:

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

Here, $\Gamma(\alpha)$ is the gamma function, a continuous extension of the factorial function.

Properties

- **Expectation:** $E[X] = \frac{\alpha}{\beta}$
- **Variance:** $\text{Var}(X) = \frac{\alpha}{\beta^2}$

These are fundamental properties, essential for understanding the distribution's behavior.

Bibliography

[1] D. Dacunha-Castelle and M. Duflo, *Probabilités et statistiques : Problèmes à temps fixe*, Masson, 1982.

[2] Murray R. Spiegel, Larry J. Stephens, *Statistics*, Schaum's Outline, 6th edition, McGraw-Hill Companies,2017.

[3] Larry J. Stephens, *Beginning Statistics*, Schaum's Outline, McGraw-Hill,1997.

[4] Seymour Lipschutz and John J. Schiller, *Introduction to Probability and Statistics*, Schaum's Outline, McGraw-Hill, 1998.

[5] J.F. Delmas *Introduction au calcul des probabilités et à la statistique*, Polycopié ENSTA, 2008.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley & Sons, Inc., vol. 1, 3rd edition, 1968.

[7] G. Grimmett, D. Stirzaker, *Probability and Random Processes*,2nd edition, Oxford University Press, 1992.

[8] J. Jacod and P. Protter, *Probability Essentials*, Springer, 2000.

[9] A. Montfort, *Cours de statistique mathématique*, Economica, 1988.

[10] A. Montfort, *Introduction à la statistique*, Ecole Polytechnique, 1991.