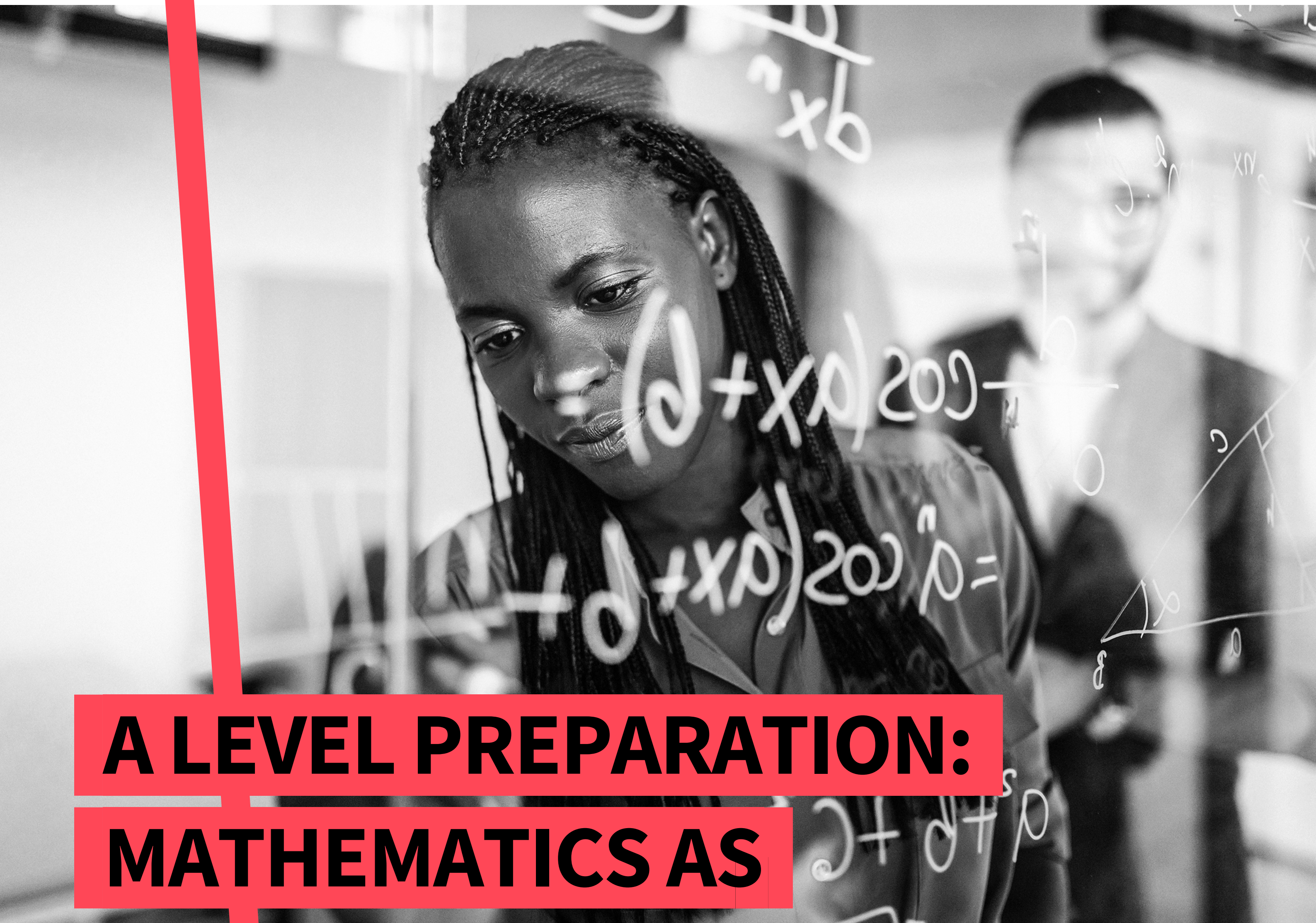
Course Book



A LEVEL PREPARATION: MATHEMATICS AS

DLAPMAT01

INTERNATIONAL
UNIVERSITY OF
APPLIED SCIENCES

A LEVEL PREPARATION:
MATHEMATICS AS

MASTHEAD

Publisher:
IU Internationale Hochschule GmbH
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt

Mailing address:
Albert-Proeller-Straße 15-19
D-86675 Buchdorf
media@iu.org
www.iu.de

DLAPMAT01
Version No.: 001-2022-0930
Dr. Florian Pausinger

MODULE DIRECTOR

DR. VERONICA MAS

Ms. Mas is a lecturer at IU International University of Applied Sciences. Her focus lies in chemical processes and renewable energy.

After completing her PhD, Ms. Mas worked several years as a consultant in the UK helping small and medium-sized enterprises become more energy-efficient and sustainable. She later moved to Germany to work as a researcher and manager, collaborating with the army on a project aiming to improve the detection of contaminants.

Additionally, Ms. Mas has extensive experience working in the higher education sector developing innovative approaches to deliver distance-learning work-based programs for companies such as Dyson, National Grid, and SSE.

TABLE OF CONTENTS

A LEVEL PREPARATION: MATHEMATICS AS

INTRODUCTION

WELCOME

SIGNPOSTS THROUGHOUT THE COURSE BOOK

This course book contains the core content for this course. Additional learning materials can be found on the learning platform, but this course book should form the basis for your learning.

The content of this course book is divided into units, which are divided further into sections. Each section contains only one new key concept to allow you to quickly and efficiently add new learning material to your existing knowledge.

At the end of each section of the digital course book, you will find self-check questions. These questions are designed to help you check whether you have understood the concepts in each section.

For all modules with a final exam, you must complete the knowledge tests on the learning platform. You will pass the knowledge test for each unit when you answer at least 80% of the questions correctly.

When you have passed the knowledge tests for all the units, the course is considered finished and you will be able to register for the final assessment. Please ensure that you complete the evaluation prior to registering for the assessment.

Good luck!

BASIC READING

Pemberton, S. (2018). *Cambridge international AS & A level mathematics: Pure mathematics 2 & 3 coursebook.* Cambridge University Press.

Goldie, S. (2018). *Cambridge International AS & A Level Mathematics Pure Mathematics 1* (2nd ed.). Hodder Education.

Goldie, S. & Gilbey, J. (Ed.). (2018). *Cambridge International AS & A Level Mathematics Pure Mathematics 2 and 3* (2nd ed.). Hodder Education.

Kranat, J. (2018). *Cambridge International AS & A Level Mathematics: Probability & Statistics 2 Coursebook Digital Edition*. Cambridge University Press.

FURTHER READING

UNIT 1

Hamming, R. W. (2004). Chapter 1: Prologue. In *Methods of Mathematics Applied to Calculus, Probability and Statistics*. Dover.

UNIT 2

Hamming, R. W. (2004). Chapter 4: Real Numbers, Functions and Philosophy. In *Methods of Mathematics Applied to Calculus, Probability and Statistics*. Dover.

UNIT 3

Kurgalin, S. & Borzunov, S. (2021). Chapter 7: Equation of a Straight Line on a Plane. In *Algebra and Geometry with Python*. Springer.

UNIT 4

Kurgalin, S. & Borzunov, S. (2021). Chapter 11: Curves of the second order. In *Algebra and Geometry with Python*. Springer.

UNIT 5

Rosenthal, D., Rosenthal, D. & Rosenthal, P. (2018). Chapter 11: Fundamentals of Euclidean Plane Geometry. In *A Readable Introduction to Real Mathematics*. Springer.

UNIT 6

Rosenthal, D., Rosenthal, D. & Rosenthal, P. (2018). Chapter 13: An Introduction to Infinite Series. In *A Readable Introduction to Real Mathematics*. Springer.

UNIT 7

Vince, J. (2019). Chapter 4: Derivatives and Antiderivatives. In *Calculus for Computer Graphics* (2nd edition). Springer.

UNIT 8

Vince, J. (2019). Chapter 7: Integral Calculus. In *Calculus for Computer Graphics* (2nd edition). Springer.

UNIT 9

Goh Ming Hui, E. (2019). Chapter 4: Descriptive Statistics. In *Learn R for Applied Statistics*. Apress, Springer.

UNIT 10

McShane-Vaughn, M. (2016). *Chapter 2: Learning to Count.* In *Probability Handbook.* American Society for Quality (ASQ).

UNIT 11

Hodges, J. L. & Lehmann, E. L. (2005). Chapter 4: Conditional Probability. In *Basic Concepts of Probability and Statistics* (2nd Edition). Society for Industrial and Applied Mathematics (SIAM).

UNIT 12

McShane-Vaughn, M. (2016). Chapter 4: Discrete Probability Distributions. In *Probability Handbook.* American Society for Quality (ASQ).

UNIT 13

McShane-Vaughn, M. (2016). Chapter 5: Continuous Probability Distributions. In *Probability Handbook.* American Society for Quality (ASQ).

LEARNING OBJECTIVES

The aim of this coursebook is to prepare students for Paper 1 (Pure Mathematics 1) and Paper 5 (Probability & Statistics 1) of the Cambridge International **AS Level Mathematics** Assessment.

Units 1 to 8 cover the material of Paper 1. The course starts with a gentle introduction to quadratics and functions. After the basics on coordinate geometry, the circular measure is introduced and problems in trigonometry are discussed. The first part is concluded with classical topics from calculus such as series, differentiation, and integration.

Units 9 to 13 prepare for Paper 5. The discussion starts by showing how to represent and analyze empirical data with basic statistical methods. In a next step, combinatorial concepts, such as permutations and combinations, are introduced. Such concepts play an important role in probability theory, which is a central topic of Paper 5. The last two units contain material on very important examples of discrete and continuous probability distributions, i.e., we discuss the binomial distribution as well as the normal distribution.

UNIT 1

QUADRATICS

STUDY GOALS

On completion of this unit, you will be able to ...

- define a quadratic expression.
- solve quadratic equations and inequalities.
- explain how to complete the square for a quadratic expression and use a completed square form.
- solve a pair of simultaneous equations in two variables of which one is linear, and one is quadratic.
- understand the relationship between a graph of a quadratic function and its associated algebraic equation.

1. QUADRATICS

Introduction

This unit is concerned with quadratic functions. A quadratic expression is of the form

$$ax^2 + bx + c$$

where $a,\ b,\ c \in \mathbb{R}$ with $a \neq 0$ and x is an independent variable. a is called the leading coefficient and c is the constant term.

A quadratic function has a maximum or a minimum value, and its graph has a curvature and interesting symmetry. These graphs are most familiar as the shape of the path of a ball or projectile as it travels through the air under the influence of gravity (known as the trajectory). Discovering that the trajectory is a quadratic was one of **Galileo**'s major achievements in the 17th century. Another occurrence of the quadratic function is explaining the path of planets in our solar system as they revolve around the sun.

Galileo
He was an Italian astronomer living in the 16th and 17th century.

Studying quadratics also offers a route to eventually thinking about more complicated functions such as cubic expressions (functions containing x^3) or even higher powers of variable x.

1.1 Solving Quadratic Functions by Factoring and Inequalities

In this unit we introduce the technique of factorizing a quadratic expression. This is an important method for solving quadratic equations as we will see.

Factorizing Quadratic Expressions

Factoring a quadratic is a method of expressing the function as a product of **linear polynomials**. Factorizing can be done via several different methods; in the first instance we shall explore the method of splitting the middle term.

Linear Polynomials
A linear polynomial (or a polynomial of order 1) in variable x is an expression of the form $ax + b$.

Example 1.1: Factorize $x^2 + 7x + 10$.

Solution. First observe that

$$x^2 + 7x + 10 = x^2 + 5x + 2x + 10$$

Here, the middle term of $7x$ has been split into two numbers whose product is 10. 10 has been chosen because this is the product of the coefficient of x^2 and the constant term. Now consider the first two termsx^2 and $5x$ which both have a common factor x of and factorize this.

$$= x(x + 5) + 2x + 10$$

Similarly to the last two terms $2x$ and 10, take out the common factor required to ensure it has the same linear factor as the first two terms

$$\begin{aligned} &= x(x + 5) + 2(x + 5) \\ &= (x + 2)(x + 5) \end{aligned}$$

Example 1.2: Factorize $12x^2 - 20x + 3$.

<u>Solution.</u> We proceed in a similar fashion.

$$\begin{aligned} &12x^2 - 20x + 3 \\ =&12x^2 - 18x - 2x + 3 \\ =&6x(2x - 3) - 1(2x - 3) \\ =&(6x - 1)(2x - 3) \end{aligned}$$

We are often asked to consider quadratics with no term. Specifically, if you are asked to factorize an expression which is one square number minus another you can factor it immediately as follows:

$$a^2 - b^2 = (a + b)(a - b)$$

This method is aptly called, the difference of two squares.

Example 1.3: Factorize $25 - x^2$.

<u>Solution.</u>

$$\begin{aligned} &25 - x^2 \\ =&5^2 - x^2 \\ =&(5 - x)(5 + x) \end{aligned}$$

Solving Quadratic Equations by Factorizing

To find the resulting answers for the unknown variable via factorizing in a quadratic equation, the first step is to ensure the equation has one side equal to zero. Then one can factor the expression as above into two linear polynomials. The key ingredient for the factorization method is to note that a product of two factors is zero if and only if one (or more) of the factors is zero itself. This mathematical fact allows us to separate the factored equation into simpler, more manageable linear equations to solve independently.

Example 1.4: Solve $x^2 - 3x - 10 = 0$.

Solution. Factorize the left-hand side,

$$(x-5)(x+2) = 0$$

now separate the factored equation into two smaller linear equations and solve for variable x,

$$x - 5 = 0 \implies x = 5$$

and

$$x + 2 = 0 \implies x = -2$$

Example 1.5: Solve the equation $2x^2 + 5x = 12$.

Solution. Make the right-hand side of the equation equal to zero before we begin,

$$2x^2 + 5x - 12 = 0$$

Then factorize by the method of splitting the middle term and proceed as in the previous example,

$$(2x-3)(x+4) = 0$$

hence

$$2x - 3 = 0 \implies 3x = 2 \implies x = \frac{2}{3}$$

and

$$x + 4 = 0 \implies x = -4$$

Example 1.6: Solve the equation $4y^2 - 1 = 0$.

Solution. Note that this expression is a difference of two squares and in that case can be factorized as follows

$$(2y-1)(2y+1) = 0$$

Therefore,

$$2y - 1 = 0 \implies y = \frac{1}{2}$$

and

$$2y + 1 = 0 \implies y = -\frac{1}{2}$$

Example 1.7: Solve the equation $x^4 - 5x^2 + 6 = 0$.

<u>Solution.</u> At first glance, this equation does not appear to have the same form of those we have studied so far. However, setting $y = x^2$ the equation becomes

$$y^2 - 5y + 6 = 0$$

Therefore, the equation is not a quadratic in variable x, but it is in variable y. Let's proceed to solve for y.

$$\begin{aligned} y^2 - 5y + 6 &= 0 \\ (y-3)(y-2) &= 0 \Longrightarrow y = 2,\ y = 3 \end{aligned}$$

We now have values for y. Recall however that our original equation was in variable x and therefore,

$$y = x^2 = 3 \Longrightarrow x = \pm\sqrt{3}$$

and

$$y = x^2 = 2 \Longrightarrow x = \pm\sqrt{2}$$

Hence, $x = \pm\sqrt{2},\ \pm\sqrt{3}$.

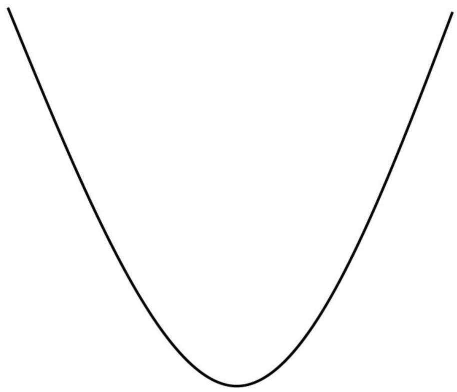
Solving Quadratic Inequalities

The reader should be familiar with the notation $a > b$ and $a \geq b$ meaning 'a is greater than b' and 'a is greater than, or equal to b'. Similarly, $a < b$ and $a \leq b$ will denote 'a is less than b' and 'a is less than, or equal to b'.

In contrast to the finite number of solutions we usually obtain when solving equations, inequalities often yield a range of possible values. Another important difference is that the inequality sign must be reversed if we multiply the inequality with or divide by a negative number. For example, the statement $2 < 3$ is true however $-2 < -3$ is false.

Solving quadratic inequalities can be done via sketching the graph and considering when the graph is above or below the x-axis. A satisfactory sketch of the curve can be made by knowing where the graph crosses the x-axis as well as the shape of the curve. That is, for a general quadratic function, the graph $y = ax^2 + bx + c$ is a parabola and if $a > 0$, the shape of the curve is:

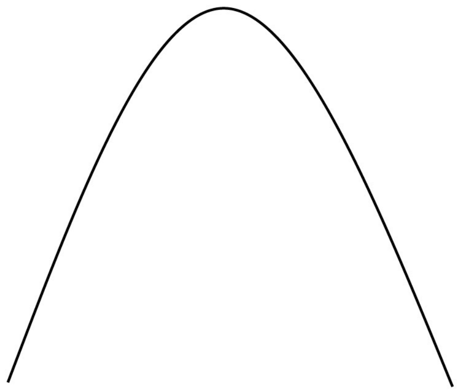
Figure 1: Parabola with Positive Leading Coefficient



Source: Florian Pausinger, (2022).

If $a < 0$, the shape of the curve is:

Figure 2: Parabola with Negative Leading Coefficient



Source: Florian Pausinger, (2022).

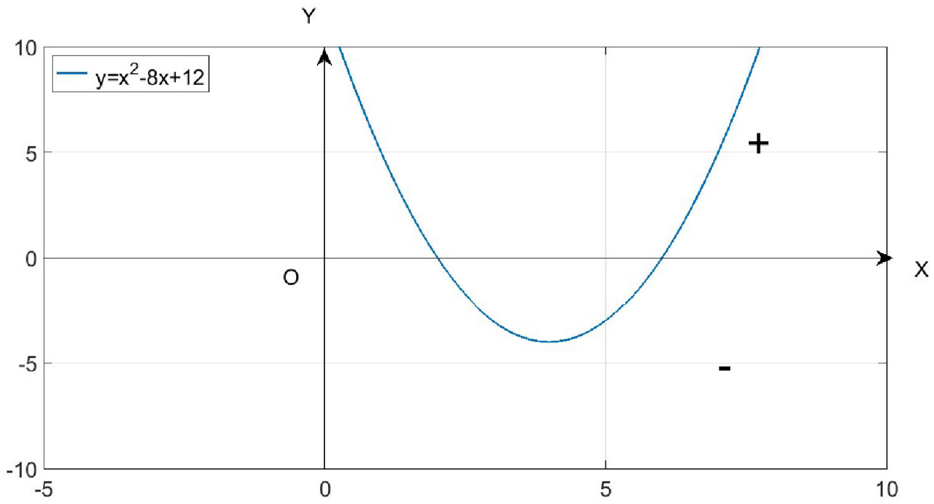
Example 1.8: Solve $x^2 - 5x - 14 > 0$.

Solution. Sketch the graph of $y = x^2 - 5x - 14$. To determine where the quadratic crosses the x-axis, we set $y = 0$. Hence

$$\begin{aligned} x^2 - 5x - 14 &= 0 \\ (x-7)(x+2) &= 0 \Longrightarrow x = 7,\ x = -2 \end{aligned}$$

So, the x-axis crossing points are $x = 7$ and $x = -2$.

Figure 3: Illustration of Example 1.8



Source: Florian Pausinger, (2022).

Now we need to determine the range of values of x for which the curve is positive (above the x-axis). By observation, the solution is $x < -2$ or $x > 7$.

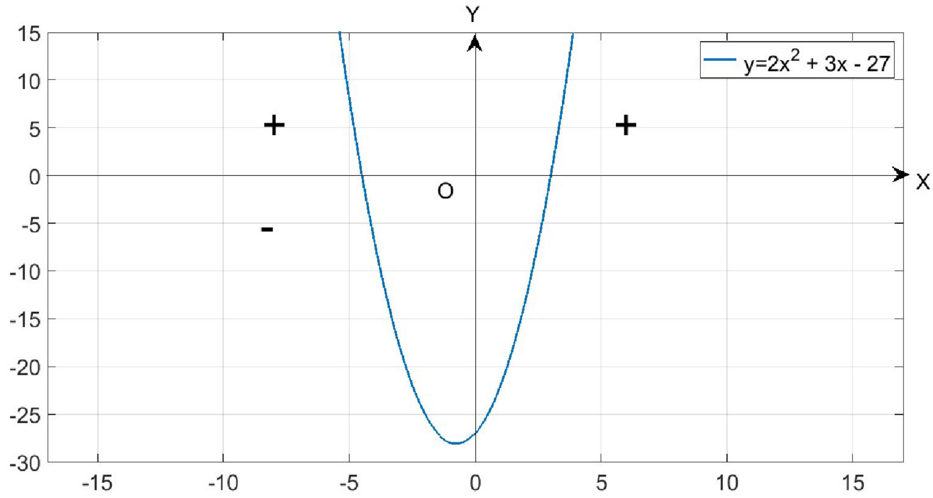
Example 1.9: Solve $2x^2 + 3x - 27 \leq 0$.

<u>Solution.</u> Sketch the graph of $y = 2x^2 + 3x - 27$. The curve crosses the x-axis when

$$2x^2 + 3x - 27 = 0$$
$$(2x + 9)(x - 3) = 0 \Longrightarrow x = 3, \ \ x = -4.5$$

Since the leading coefficient is positive, the graph of the curve is,

Figure 4: Illustration of Example 1.9



Source: Florian Pausinger, (2022).

By observing when the graph is below the x-axis in this case, the solution to the inequality is $2x^2 + 3x - 27 \leq 0$ is $-4.5 \leq x \leq 3$.

1.2 Completing the Square

Another method we can use to factorize and solve quadratic equations is completing the square. A perfect square is a particular kind of quadratic expression which contains two identical linear factors. For example,

$$(x + 6)^2$$

We can use the technique of completing the square for any quadratic expression and in return it will yield a perfect square term, like those examples above, and a single real number. For a quadratic expression with leading term equal to 1, i.e.,

$$x^2 + bx + c$$

the completed square form can be written as

$$\left(x + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}$$

Example 1.10: Complete the square on the expression $x^2 + 5x - 4$.

Solution. Following the completed square form as written above with $b = 5$ and $c = -4$, the expression $x^2 + 5x - 4$ becomes

$$\begin{aligned} &\left(x+\tfrac{5}{2}\right)^2 - 4 - \left(\tfrac{5}{2}\right)^2 \\ =&\left(x+\tfrac{5}{2}\right)^2 - 4 - \tfrac{25}{4} \\ =&\left(x+\tfrac{5}{2}\right)^2 - \tfrac{41}{4} \end{aligned}$$

One can also use the method of completing the square when the coefficient of x^2 is not equal to 1. However, one must be sure to factorize the leading coefficient before attempting to complete the square.

Example 1.11: Complete the square for the expression $2x^2 + 12x - 3$.

Solution.

$$\begin{aligned} &2x^2 + 12x - 3 \\ =&2\left(x^2 + 6x - \tfrac{3}{2}\right) \\ =&2\left((x+3)^2 - \tfrac{3}{2} - 9\right) \\ =&2\left((x+3)^2 - \tfrac{21}{2}\right) \\ =&2(x+3)^2 - 21 \end{aligned}$$

The vertex (extreme point) of a quadratic function can be easily read when the function is in completed square form. For a quadratic function in the form $k(x-l)^2 + m$, the vertex of the graph is given by $(l,\ m)$. In the last Example 1.11, the vertex of the graph is located at $(-3,\ -21)$.

Example 1.12: Find the minimum value of the function $y = 2x^2 + 8x - 3$, and the x value at which the function attains this minimum.

Solution. Let's begin by completing the square,

$$\begin{aligned} &= 2\left(x^2 + 4x - \tfrac{3}{2}\right) \\ &= 2\left((x+2)^2 - \tfrac{11}{2}\right) \\ &= 2(x+2)^2 - 11 \end{aligned}$$

Hence the minimum value of the function is -11 and is attained when $x = -2$.

Completing the square is a powerful technique when working with quadratic expressions and it is often employed when solving quadratic equations.

Example 1.13: Solve the quadratic equation $x^2 - 8x - 29 = 0$ by completing the square.

Solution. Complete the square on the quadratic function,

$$(x-4)^2 - 45 = 0$$

Then proceed in the usual manner to make x the subject of the equation. Be sure to remember that the square root produces a plus and minus answer.

$$\begin{aligned}(x-4)^2 &= 45 \\ x-4 &= \pm\sqrt{45} \\ x &= 4 \pm 3\sqrt{5}\end{aligned}$$

1.3 Discriminant

To give some motivation for this topic, we briefly note the solutions of several quadratic equations.

- $x^2 - 5x + 6 = 0 \implies x = 2$ and $x = 3$, hence there exist two real roots.
- $x^2 - 4x + 4 = 0 \implies x = 2$ and $x = 2$, there exists one repeated root.
- $x^2 + x + 1 = 0$ yields no real roots.

We observe that the set of solutions to any quadratic equation $ax^2 + bx + c = 0$ fits into one of three categories:

- two distinct real roots
- one real repeated root
- no real roots

Fortunately, we don't have to solve a quadratic equation to determine how many real roots it has. A quantity named the discriminant allows this to be done with ease. The discriminant of a quadratic function $ax^2 + bx + c$ is defined as,

$$b^2 - 4ac$$

The following table classifies the nature of the roots of a given quadratic equation.

Table 1: Classification of the Roots of a Quadratic Equation

$b^2 - 4ac$	
> 0	Two distinct real roots
$= 0$	One repeated root
< 0	No real roots

Source: Florian Pausinger, (2022).

Example 1.14: Find the discriminant of $4x^2 + 4x + 1$, and hence write down how many real roots exist. Confirm your answer by solving the equation $4x^2 + 4x + 1 = 0$ via any appropriate method.

Solution. Let's calculate the discriminant where $a = 4,\ b = 4,\ c = 1$:

$$b^2 - 4ac = 4^2 - 4(4)(1) = 16 - 16 = 0$$

Hence there exists one repeated root to the equation $4x^2 + 4x + 1 = 0$. Solving via the quadratic formula,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-4 \pm \sqrt{16 - 4(4)(1)}}{2(4)} = \frac{-4}{8} = -\frac{1}{2}$$

The discriminant can also be employed to find the value or values, that a particular parameter can take in a quadratic equation for there to exist a desired and specified number of roots.

Example 1.15: Find the values of k such that the following equation $x^2 + kx + k + 3 = 0$ has only one root.

Solution. For the quadratic listed above, $a = 1,\ b = k$ and $c = (k + 3)$. We also require only one root, therefore $b^2 - 4ac = 0$. Hence

$$k^2 - 4(1)(k + 3) = 0$$

which can be expanded and simplified to

$$k^2 - 4k - 12 = 0$$

Solving by the method of factorizing yields,

$$(k - 6)(k + 2) = 0 \implies k = 6,\ k = -2$$

That is, the quadratic equations $x^2 + 6x + 9 = 0$ and $x^2 - 2x + 1 = 0$ have one repeated root [Check this by solving!].

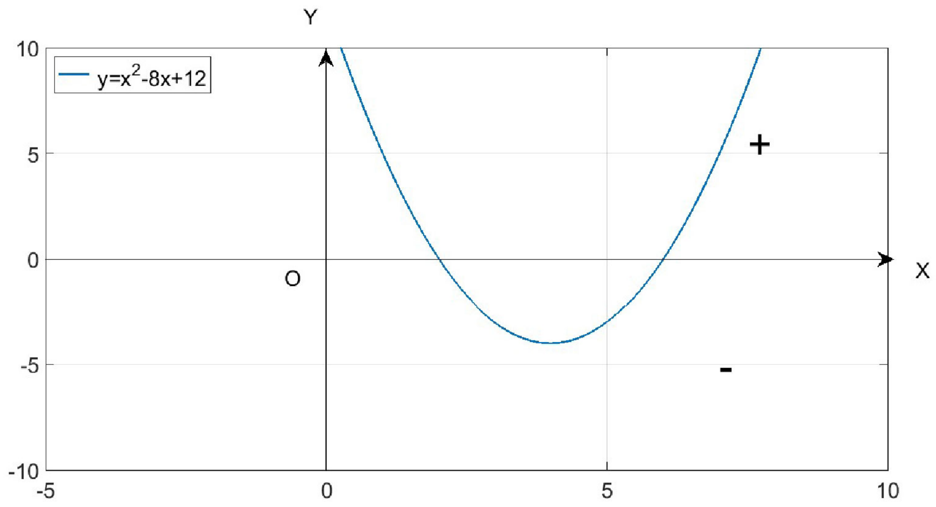
Example 1.16: Find the range of values that the constant m can take so that the equation $x^2 + (m + 2)x + 3m - 2 = 0$ has real roots.

Solution. For real roots, the discriminant must be greater than or equal to zero. Then for $a = 1,\ b = m + 2$ and $c = 3m - 2$ the discriminant comes up as

$$\begin{aligned} b^2 - 4ac &\geq 0 \\ (m + 2)^2 - 4(1)(3m - 2) &\geq 0 \\ m^2 - 8m + 12 &\geq 0 \end{aligned}$$

Notice that we are required to solve an inequality which is quadratic in variable m. From the previous section, we must sketch the graph of $y = m^2 - 8m + 12$ which crosses the m-axis at $m = 2$ and $m = 6$. Therefore, we have $m \leq 2$ or $m \geq 6$.

Figure 5: Illustration of Example 1.16



Source: Florian Pausinger, (2022).

The discriminant can be employed as an algebraic method to determine if lines and/or curves intersect.

Example 1.17: Does the line $y = 2x - 4$ intersect with the quadratic curve $y = x^2 + 2$?

Solution. First, we form the intersection equation by setting the equations for the line and the curve equal and by rearranging.

$$x^2 + 2 = y = 2x - 4$$
$$x^2 - 2x + 6 = 0$$

If the line and the curve intersect, this equation will have at least one real root. Conversely, if the line and curve don't intersect, there will be no real roots. The discriminant can inform us of which case we have,

$$b^2 - 4ac = (-2)^2 - 4(1)(6) = 4 - 24 = -20 < 0$$

Therefore, we are left with no real roots, and we can conclude that the line and the curve do not intersect.

Example 1.18: A curve has an equation $y = 3x^2 - 2$ and a straight line has the equation $y = mx - 5$. The line does not meet the curve. Find the range of possible values of m.

Solution. We must analyze the intersection equation of the line and the curve in a similar method to the previous example.

$$3x^2 - 2 = mx - 5$$
$$3x^2 - mx + 3 = 0$$

We are told that the line and curve do not meet, therefore

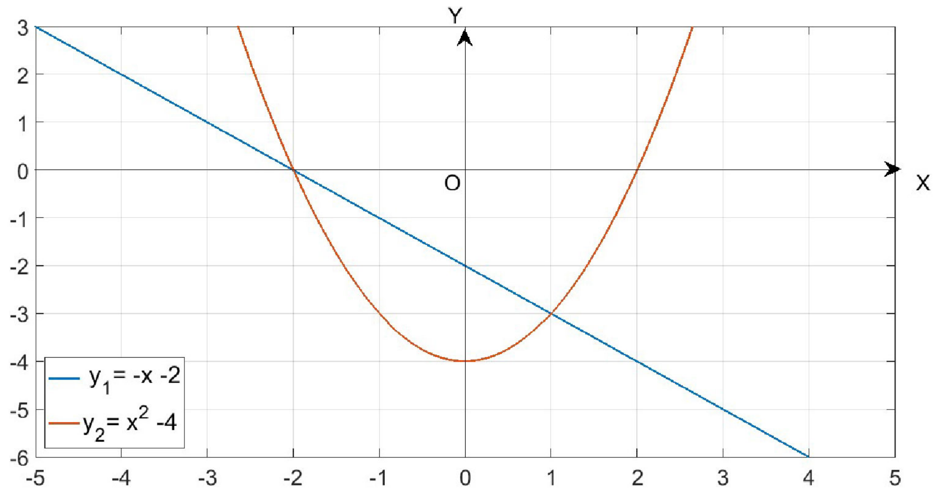
$$b^2 - 4ac = m^2 - 4(3)(3) < 0$$
$$m^2 - 36 < 0$$

Solving in the usual manner, we see that the curve $y = m^2 - 36 = (m-6)(m+6)$ crosses at $m = 6$ and $m = -6$. The curve is a ‘U’ shape positive curve, and we require the section below the x-axis. Hence $-6 < m < 6$.

1.4 Simultaneous Equations

In this section we shall demonstrate how to solve simultaneous equations where one equation is linear, and the second equation is quadratic.

Figure 6: Simultaneous Equations



Source: Florian Pausinger, (2022).

Above are the graphs of $y = -x - 2$ and $y = x^2 - 4$. The coordinates of the points of intersection of the two graphs are $(-2,\ 0)$ and $(1,\ -3)$. The solutions to the simultaneous equations $y = -x - 2$ and $y = x^2 - 4$ are therefore $x = -2,\ y = 0$ and $x = 1, y = -3$.

This simultaneous equation problem can also be completed algebraically. The general method that one adopts is substitution. That is, substituting the linear equation into the quadratic equation eliminating one of the variables. One can then solve this equation via any of the methods outlined in previous sections and the values substituted back into the linear equation to obtain the corresponding values for the other variable.

Example 1.19: Solve the simultaneous equations $y = -x - 2$ and $y = x^2 - 4$ algebraically.

Solution. Since both equations are equal to variable y, we can derive the quadratic equation

$$-x - 2 = x^2 - 4 \tag{1.1}$$

which can be rearranged to yield

$$x^2 + x - 2 = 0$$

This quadratic can be solved via factoring,

$$(x+2)(x-1) = 0 \Longrightarrow x = 1 \quad \text{and} \quad x = -2$$

Substituting back into the linear equation $y = -x - 2$,

$$y = -1 - 2 = -3 \quad \text{and} \quad y = -(-2) - 2 = 0$$

Hence the answers to the simultaneous equations are $x = -2,\ y = 0$ and $x = 1, y = -3$.

It is also required from time to time to form the simultaneous equations before attempting to solve.

Example 1.20: Find the two numbers whose sum is 26 while their product evaluates to 153.

Solution. Form the simultaneous equations by letting x and y be the two numbers we must find. Then “sum of two numbers is 26” implies $x + y = 26$, and “product of the two numbers is 153” implies $xy = 153$. Rearranging the linear equation, $y = 26 - x$ and we substitute this into the non-linear equation then expand,

$$\begin{aligned} x(26-x) &= 153 \\ x^2 - 26x + 153 &= 0 \end{aligned}$$

Solving this equation by completing the square,

$$\begin{aligned} (x-13)^2 + 153 - 169 &= 0 \\ (x-13)^2 &= 16 \end{aligned}$$

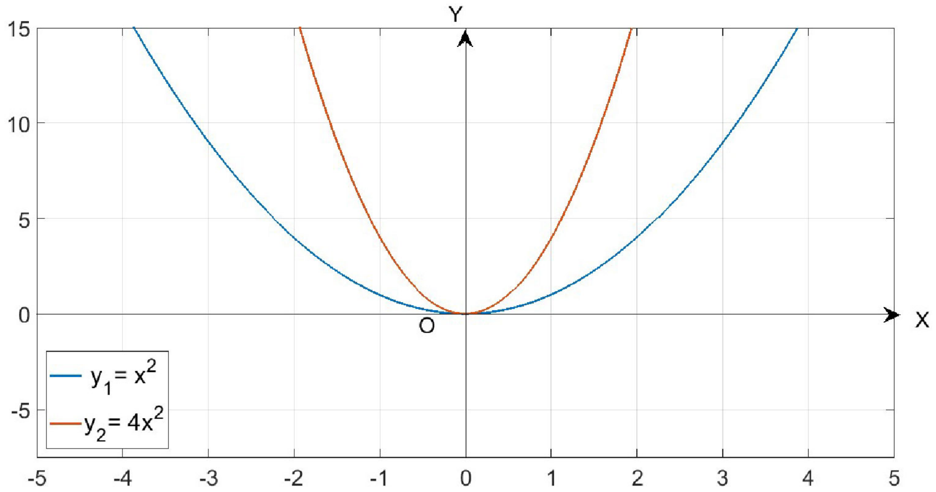
Leading to two solutions $x = 13 + 4 = 17$ and $x = 13 - 4 = 9$. Substituting these values back into the linear equation, $y = 26 - 9 = 17$ and $y = 26 - 17 = 9$. In any case, the two numbers are 9 and 17.

1.5 Graphs of Functions

It has already been discussed that a quadratic graph $y = ax^2 + bx + c$ has a general shape of a parabola. In this final section, we further investigate the effect that changing the values of coefficients a, b, and c has on the shape of the graph of a quadratic function.

The sign of the coefficient of x^2 determines whether the curve is a maximum or a minimum curve. However, the magnitude of the coefficient determines how wide or narrow the graph is. The greater the coefficient of x^2, the narrower the parabola. For example, the figure below shows the graphs of $y = x^2$ (red) and $y = 4x^2$ (blue).

Figure 7: Width of Parabola



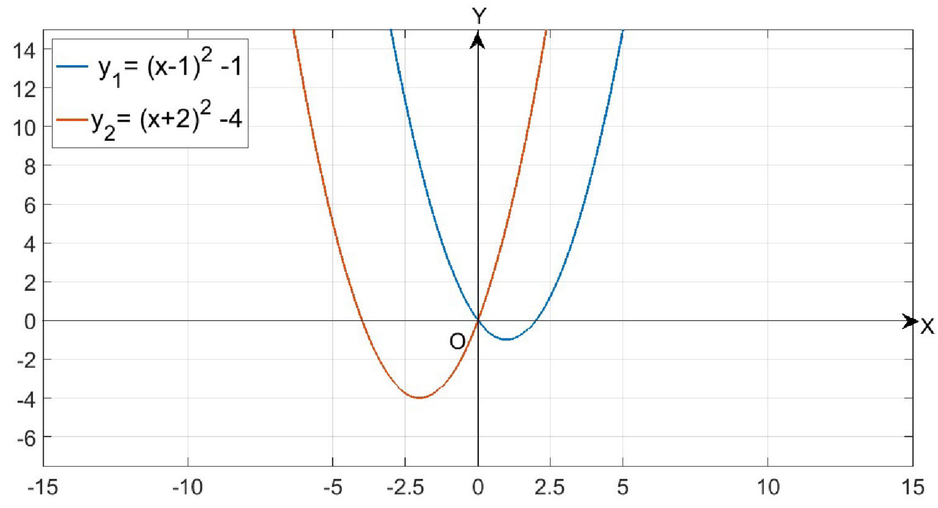
Source: Florian Pausinger, (2022).

The linear-term coefficient, b, shifts the axis of symmetry of the quadratic curve away from the y-axis to the left or to the right. Given a positive value of a, i.e., a minimum quadratic curve,

- the graph shifts to the right of the y-axis if b is negative, and
- the graph shifts to the left of the y-axis if$b$ is positive.

The reverse is true if we are given a negative value of a.

Figure 8: Quadratic Shift



Source: Florian Pausinger, (2022).

Lastly, the constant term c affects the y-intercept. The greater the number, the higher the intercept point on the y-axis.

From the previous figure we can also make another important observation. We can complete the square to bring the quadratic $ax^2 + bx + c$ into the form $a(x-h)^2 + k$. Now, if a is positive, then the function has a minimum at (h,k), while if a is negative, there is a maximum at (h,k). We can start with the function $f(x) = x^2$, which has a minimum at (0,0). The blue graph shows the function $f(x) = x^2 - 2x$ which can also be written as $f(x) = (x-1)^2 - 1$, indicating that there is a minimum at (1,-1). Similarly, the red graph represents the function $f(x) = x^2 + 4x$ which can also be written as $f(x) = (x+2)^2 - 4$, indicating that there is a minimum at (-2,-4).

SUMMARY

You have now seen three techniques for solving quadratic equations; factorization, completing the square and the quadratic formula. The second and third technique can be used in all cases no matter what the quadratic expression presented.

To solve a quadratic inequality, the answer will always be a range of values and it is a great idea to sketch the quadratic graph to determine what form the answer takes.

To solve pairs of simultaneous equations where one is quadratic and the other is linear, use the method of substitution and substitute the linear equation into the quadratic equation.

A quadratic equation can have a possible zero, one or two roots. Without the need to solve, one can employ the discriminant to determine how many solutions are to be expected.

UNIT 2

FUNCTIONS

STUDY GOALS

On completion of this unit, you will be able to ...

- explain the terms function, domain, range, one-one function, inverse function, and composition of functions.
- locate the range of a given function.
- describe the composition of two functions.
- recognize whether or not a given function is one-one and find the inverse of a one-one function.
- illustrate the relation between a one-one function and its inverse.
- explain and apply basic transformations of the graph of a function.

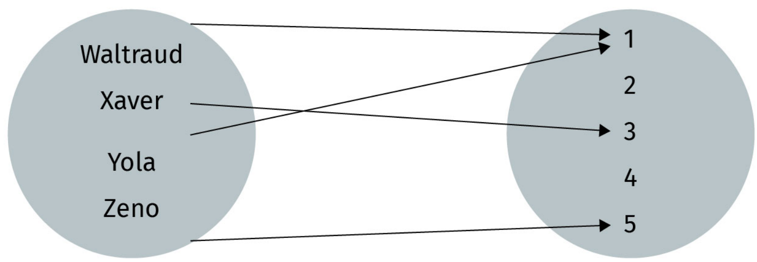
2. FUNCTIONS

Introduction

A function is a relation that uniquely associates members of one set with members of another set. Consider the non-mathematical example:

Four children Waltraud, Xaver, Yola, and Zeno are given a homework assignment which is marked out of a total of five. Their marks for the assignment are shown in the arrow diagram below.

Figure 9: Example Function



Source: Florian Pausinger, (2022).

By following the arrows from the set of names on the left to the set of marks on the right, we can find the mark that a student obtained. Formally, the first set is said to be the domain of the function and the second set is the range.

As an introductory example, suppose a function has the set X as the domain and maps any element $x \in X$ to twice its valueto give the corresponding element of the range. This would be written $f(x) = 2x$ for $x \in X$. As this function would map the number 2 to 4, we write this as $f(2) = 4$.

Other letters such as f, g or h are often used to represent an individual function.

2.1 Mapping

A mapping and a function are two different terms for the same mathematical object. Any relationship which takes one element of one set and assigns to it one and only one element of a second set is called a function (or a mapping).

Example 2.1: If $f(x) = 3x + 4$, find $f(5)$ and $f(x + 1)$.

Solution. Substituting $x = 5$, then $f(5) = 3(5) + 4 = 19$. Similarly, replacing $x + 1$ in the function we obtain,

$$f(x+1) = 3(x+1) + 4 = 3x + 7$$

Domain and Range

The domain of a function is the set of values which you are allowed to input to the function. When defining a function, it is extremely important to also specify the domain. The range of the function is the set of all output values of the function.

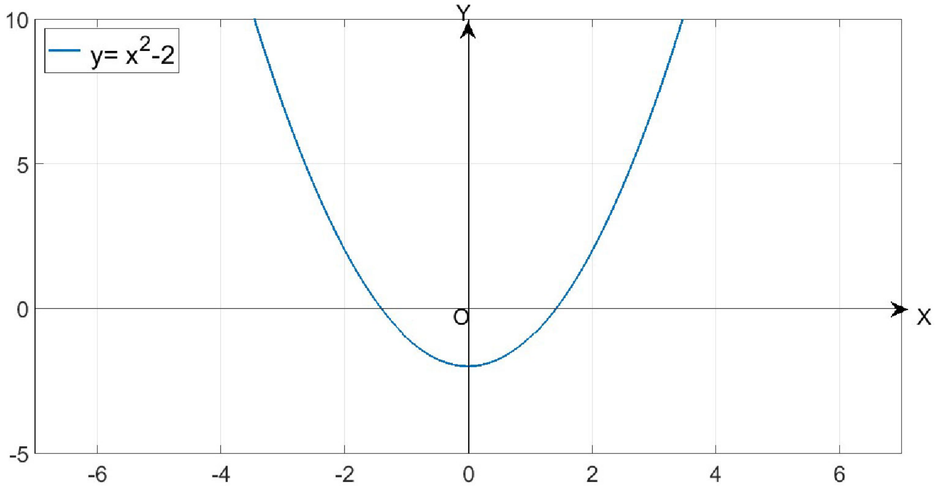
For example, take the function $f(x) = x^2$, and choose the domain to be all the real numbers. Then the range of $f(x)$ is all real numbers which are greater than or equal to zero. A standard method for determining the range of a given function and domain is to first draw a sketch of the function, then observe what range of $f(x)$ values are part of the output.

Example 2.2: Let $f(x) = 2x - 5$ for $x \in \mathbb{R}, \ -2 \leq x \leq 4$. Write down the range of the function f.

Solution. Since the graph is a linear straight line (it is of the form $y = mx + c$), let's calculate the function values at the end points of the domain. When $x = -2, f(-2) = 2(-2) - 5 = -9$. When $x = 4, f(4) = 2(4) - 5 = 3$. Therefore, the range of f is $-9 \leq f(x) \leq 3$.

Example 2.3: Sketch the function $g(x) = x^2 - 2$ for $x \in \mathbb{R}$. State the range of $g(x)$.

Figure 10: Illustration of Example 2.3



Source: Florian Pausinger, (2022).

Hence the range of the function g is $g(x) \geq -2$.

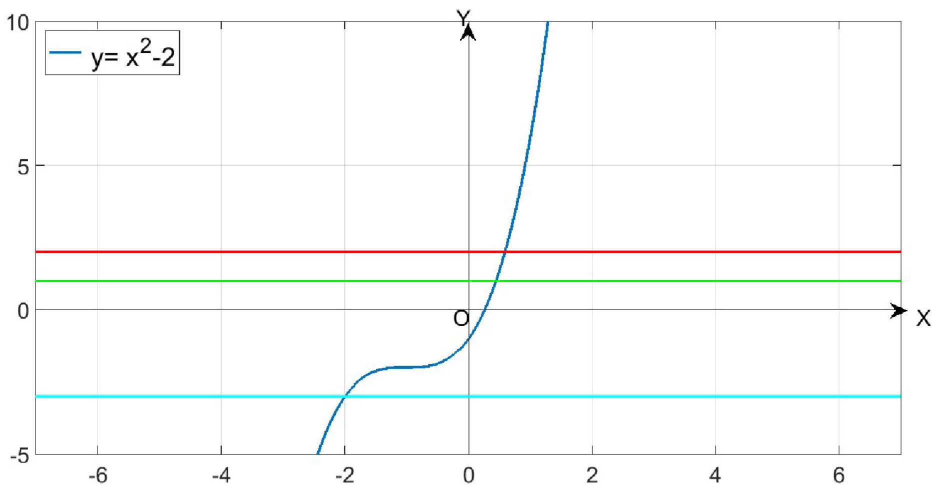
One-To-One or Many-To-One

Notice from the first few examples, that a function can map more than one element of the domain onto the same element of the range. For example, for $g(x) = x^2 - 2$ with $x \in \mathbb{R}$ we have that $g(2) = g(-2) = 2$. Such functions are said to be many-to-one. On the other hand, functions for which each element of the domain is mapped onto a different element of the range are said to be one-to-one. These functions are important to consider in a subsequent sect ion when studying inverse functions.

To test whether a function is one-to-one, we can use the horizonal line test. Given the graph of a function $y = f(x)$, we draw several horizonal lines through the graph and investigate where they intersect. If each of our lines intersect the graph at, at most one point, then we can conclude that the function is one-to-one.

Example 2.4: Sketch the graph of $y = (x+1)^3 - 2$ and determine via the horizontal line test if the function is one-to-one.

Figure 11: Illustration of Example 2.4



Source: Florian Pausinger, (2022).

<u>Solution.</u> We note that all horizontal lines drawn on the graph cross at only one point. Therefore, the function is indeed one-to-one.

It is worth noting that, relationships which are one-to-many can also occur. However, from our definition of a function, these relationships are not functions.

2.2 Composite Functions

When a function is followed by another function, the resulting function is called a composite function. Almost all functions we come across can be described as a composite function, even in the simplest examples. For example,$f(x) = 2x - 4$ is the combination of $g(x) = 2x$ and $h(x) = x - 4$. The obvious question to consider is, in which order should we compose the functions to form f correctly?

We must be careful of the order when composing functions, $fg(x)$, means "do $g(x)$ first, and then $f(x)$" and notice that in general $fg(x) \neq gf(x)$. Importantly, it is not always possible to compose functions. For functions $f\colon C \to D$ and $g\colon A \to B$, then $fg\colon A \to D$ exists if and only if B **is contained within** . That is, the range of $g(x)$ must be contained within the domain $f(x)$ for $fg(x)$ to exist.

Set B contained within set C
This implies that every element of B is also an element of C.

Example 2.5: Let $f(x) = 3x - 5$ and $h(x) = 3 - 2x$. Find both $fh(x)$ and $hf(x)$.

<u>Solution.</u> Firstly, $fh(x) = f(h(x)) = f(3 - 2x) = 3(3 - 2x) - 5 = 4 - 6x$. On the other hand, $hf(x) = h(3x - 5) = 3 - 2(3x - 5) = 13 - 6x$. Note that, as stated above $fh(x) \neq hf(x)$.

Example 2.6: Let $f(x) = (x - 4)^2 - 1$ for all $x \in \mathbb{R}$, and $g(x) = \frac{2x + 3}{x - 2}$ for$x \in \mathbb{R},\ x > 2$. Find $fg(4)$.

<u>Solution.</u> Note that the range of g is contained within the domain of f, hence fg exists. To find the composite function $fg(x)$,

$$\begin{aligned} fg(x) &= f(g(x)) = f\left(\frac{2x + 3}{x - 2}\right) = \left(\left(\frac{2x + 3}{x - 2}\right) - 4\right)^2 - 1 \\ &= \left(\frac{11 - 2x}{x - 2}\right)^2 - 1 \end{aligned}$$

Then $fg(4) = \left(\frac{11 - 2(4)}{4 - 2}\right)^2 - 1 = \left(\frac{3}{2}\right)^2 - 1 = \frac{9}{4} - 1 = \frac{5}{4}$.

Example 2.7: Let $f(x) = 2x + 3$ for all $x \in \mathbb{R},\ x \neq -1$, and $g(x) = \frac{12}{1 - x}$ for $x \in \mathbb{R},\ x \neq 1$. Find $gf(x)$ and hence solve $gf(x) = 2$.

<u>Solution.</u> The composite function is

$$gf(x) = g(f(x)) = \frac{12}{1 - (2x + 3)} = \frac{12}{-2x - 2} = -\frac{6}{x + 1}$$

Then

$$\begin{aligned} gf(x) &= -\frac{6}{x + 1} = 2 \\ -6 &= 2(x + 1) = 2x + 2 \\ 2x &= -8 \Longrightarrow x = -4 \end{aligned}$$

2.3 Inverse Functions

The inverse of a function $f(x)$ is the function that reverses what $f(x)$ has done. We write the inverse of a function $f(x)$ as $f^{-1}(x)$. We can immediately state that:

- the domain of $f^{-1}(x)$ is the range of $f(x)$, and
- the range of $f^{-1}(x)$ is the domain of $f(x)$.

Importantly, not every function has an inverse. For example, the inverse of a many-to-one function is a one-to-many relationship and we know that this is not a function (but a so-called relation). Therefore, a function has an inverse if and only if it is one-to-one.

To find an inverse function, we must first write $y = f(x)$ and rearrange this equation to make x the subject of the formula. The resulting expression will be $x = f^{-1}(y)$ and therefore to finish we must rewrite the right-hand side of the expression in variable .

Example 2.8: Let $f(x) = \sqrt{x+2} - 7$ for $x \in \mathbb{R},\ x \geq -2$. Find an expression for $f^{-1}(x)$.

Solution. The function $f(x)$ is one-to-one (this can be checked via the horizontal line test), and therefore we can find the inverse function. Set

$$\begin{aligned} y &= \sqrt{x+2} - 7 \\ y + 7 &= \sqrt{x+2} \end{aligned}$$

Hence,

$$x = (y+7)^2 - 2$$

and therefore

$$f^{-1}(x) = (x+7)^2 - 2$$

Example 2.9: Let $f(x) = 5 - (x-2)^2$ for $x \in \mathbb{R},\ x \geq k$. State the smallest value of k for which the function f has an inverse. With this value of k, find an expression for f^{-1} stating the domain and the range.

Solution. The function $f(x)$ is in completed square form. We can deduce that the curve has a maximum vertex at the coordinate $(2,\ 5)$ and is a quadratic curve. Therefore, restricting the domain from $x \in \mathbb{R}$ to $x \geq 2$, the function becomes one-to-one and thus has an inverse. To find the inverse function,

$$\begin{aligned} y &= 5 - (x-2)^2 \\ (x-2)^2 &= 5 - y \\ x &= 2 + \sqrt{5-y} \end{aligned}$$

Therefore, $f^{-1}(x) = 2 + \sqrt{5 - x}$ with domain $x \leq 5$ and range $f^{-1}(x) \geq 2$.

The Graph of a Function and its Inverse

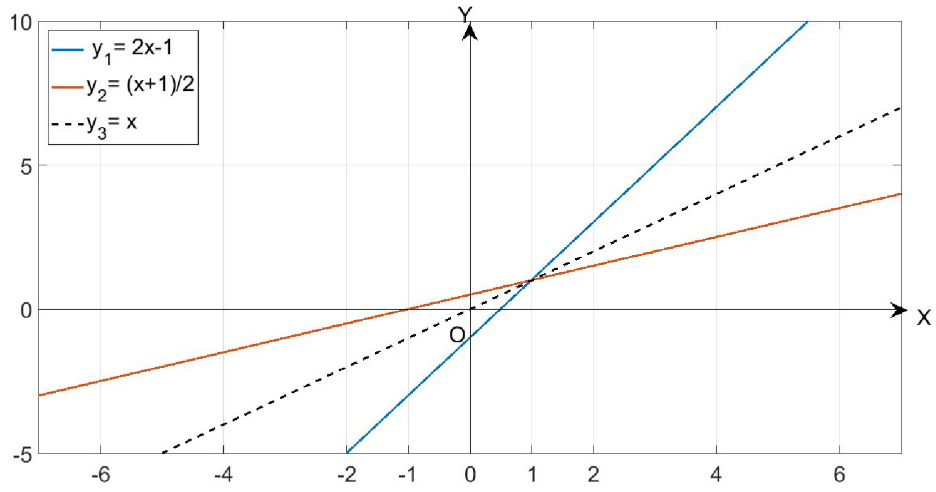
A function and its inverse have a geometrical relationship when plotted. The graphs of f and f^{-1} are **reflections** of each other in the line $y = x$.

Reflection
A reflection of a shape A is its mirror image through a so-called axis of reflection.

Example 2.10: Let $f(x) = 2x - 1$ for $x \in \mathbb{R}, \ -1 \leq x \leq 3$. Find an expression for $f^{-1}(x)$, stating the domain and range of f^{-1}. Sketch on the same axes the graphs of f and f^{-1}, making clear the relationship between the graphs.

Solution. Note in the first instance that since $f(-1) = -3$ and $f(3) = 5$, the range of f is $-3 \leq f(x) \leq 5$. Then proceeding like in the previous examples, we find that $f^{-1}(x) = \frac{x+1}{2}$ with domain $-3 \leq x \leq 5$ and range $-1 \leq f^{-1}(x) \leq 3$.

Figure 12: Illustration of Example 2.10



Source: Florian Pausinger, (2022).

The graphs of $f(x) = 2x - 1$ (red) and $f^{-1}(x) = \frac{x+1}{2}$ (green) with symmetry in the main diagonal line $y = x$ (black dotted).

2.4 Transformations

We study three types of transformations of the graph of a function $y = f(x)$: translations, reflections and stretches. We begin by considering the transformations as single transformations and then combinations of these transformations are discussed. We write the results using the same function notation as has been used so far.

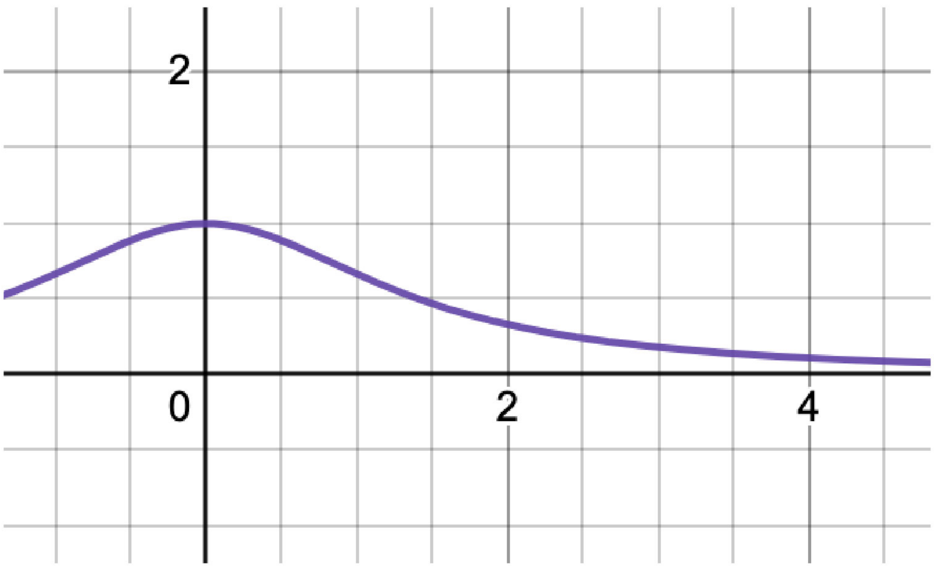
Translations

- The graph of $y = f(x) + a$ is a translation of the graph $y = f(x)$ by the vector $\begin{pmatrix} 0 \\ a \end{pmatrix}$.
- The graph $y = f(x - a)$ of is a translation of the graph$y = f(x)$ by the vector $\begin{pmatrix} a \\ 0 \end{pmatrix}$.

Therefore, as a combination of these first two results, the graph of $y = f(x - a) + b$ is a translation of the graph $y = f(x)$ by the vector $\begin{pmatrix} a \\ b \end{pmatrix}$.

Example 2.11: The diagram shows the graph of $y = f(x)$.

Figure 13: Illustration of Example 2.11, Part 1

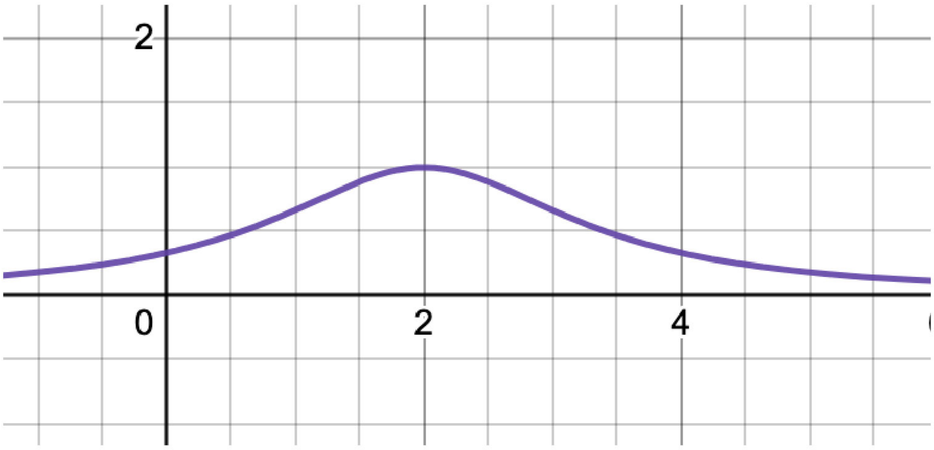


Source: Florian Pausinger, (2022).

Sketch the graph of the two functions $y = f(x - 2)$ and $y = f(x + 1) - 5$.

Solution. The transformation $y = f(x - 2)$ will translate the graph by the vector $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$, i.e., two units to the right.

Figure 14: Illustration of Example 2.11, Part 2



Source: Florian Pausinger, (2022).

Finally, the transformation $y = f(x+1) - 5$ will translate the graph by the vector $\begin{pmatrix} -1 \\ -5 \end{pmatrix}$. One unit left and five units down. The vertex of the graph is now at the coordinate $(-1, -4)$.

Figure 15: Illustration of Example 2.11, Part 3

Source: Florian Pausinger, (2022).

Example 2.12: The graph of $y = \sqrt{2x}$ is translated by the vector $\begin{pmatrix} -5 \\ 3 \end{pmatrix}$. Find the equation of the resulting graph.

<u>Solution.</u> Let $f(x) = \sqrt{2x}$, then from the remarks above,

$$f(x+5)+3=\sqrt{2(x+5)}+3=\sqrt{2x+10}+3$$

Therefore, the resulting graph is $y=\sqrt{2x+10}+3$.

Reflections

- The graph of $y=f(-x)$ is a reflection of the graph $y=f(x)$ in the y-axis.
- The graph of $y=-f(x)$ is a reflection of the graph $y=f(x)$ in the x-axis.

Example 2.13: The graph of $y=f(x)$ is a quadratic curve with minimum at $(5,-7)$. Find the coordinate of the vertex and state whether it is a maximum or a minimum for the graph $y=-f(x)$ and $y=f(-x)$.

Solution. Firstly, the transformation $y=-f(x)$ reflects the original graph in the x-axis. Therefore, the vertex will now be at the coordinate $(5{,}7)$. It is a maximum point. On the other hand, $y=f(-x)$ reflects the graph in the y-axis and hence the vertex is at $(-5,-7)$ and remains a minimum point.

Stretches

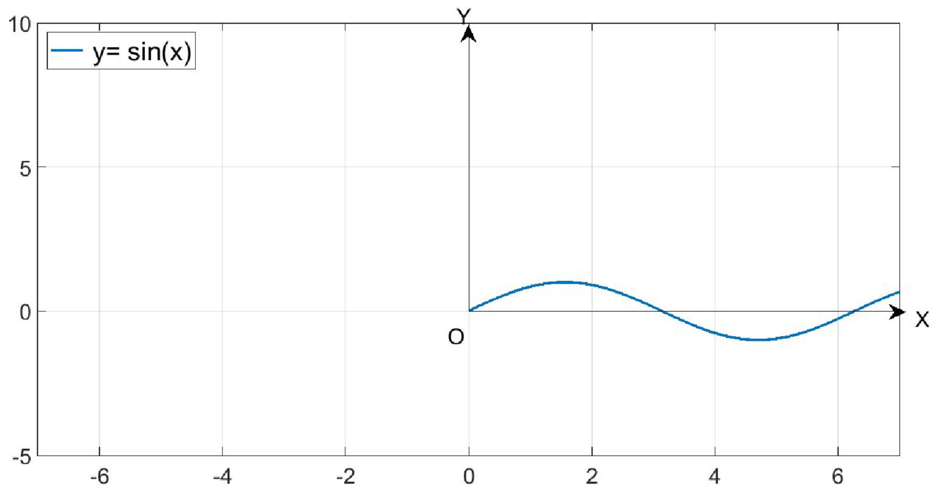
- For $a>0$, the graph of $y=af(x)$ is a stretch of the graph $y=f(x)$ with a stretch factor of a parallel to the y-axis.
- For $a>0$, the graph of $y=f(ax)$ is a stretch of the graph $y=f(x)$ with stretch factor of $\frac{1}{a}$ parallel to the x-axis.

Example 2.14: The graph of $y=4-\frac{1}{2}x^2$ is stretched with stretch factor of 3 parallel to the y-axis. Find the equation of the resulting graph.

Solution. Let $f(x)=4-\frac{1}{2}x^2$. Then $3f(x)=3\left(4-\frac{1}{2}x^2\right)=12-1.5x^2$. Therefore, the resulting graph is $y=12-1.5x^2$.

Example 2.15: The graph below shows $y=g(x)$.

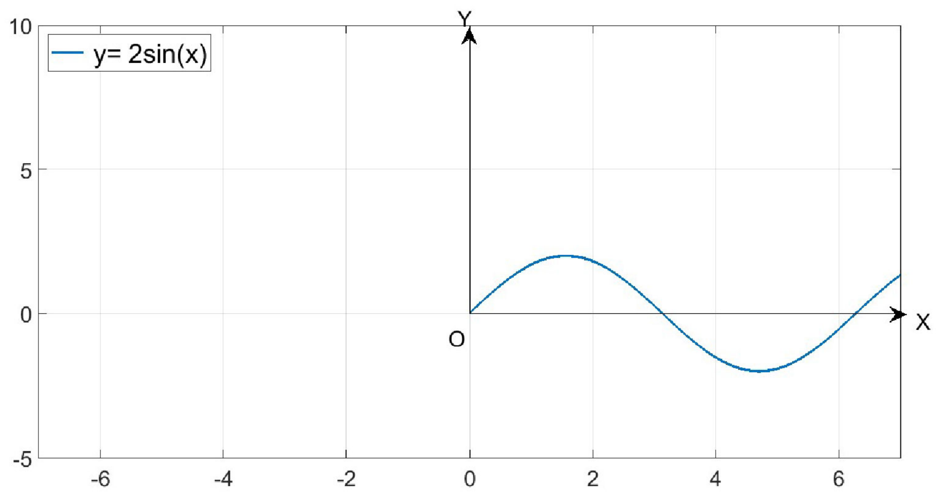
Figure 16: Illustration of Example 2.15, Part 1



Source: Florian Pausinger, (2022).

Draw the graph of $y = 2g(x)$ on a separate grid.

Figure 17: Illustration of Example 2.15, Part 2



Source: Florian Pausinger, (2022).

Combinations of Transformations

Any transformation of the graph $y = f(x)$ that is of interest to us in this course has the form,

$$y = Cf(Bx + A) + D$$

for numbers A, B, C and D. Horizontal transformations must be applied before vertical transformations. More specifically, transformations must be applied in the order A then B then C then D. That is, the horizontal translation must be applied before the horizontal stretch/reflection, thereafter the vertical stretch/reflection must be applied before the vertical translation.

Example 2.16: Determine the sequence of transformations that maps $y = f(x)$ to $y = \frac{1}{2}f(x) + 2$.

Solution. We must first stretch by a factor of $\frac{1}{2}$ parallel to the y-axis and then translate the graph by the vector $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

Example 2.17: Determine the sequence of transformations that maps $y = f(x)$ to $y = f(2x - 6) + 2$.

Solution. Following the order of transformation rules above, we first translate the graph by the vector $\begin{pmatrix} 6 \\ 0 \end{pmatrix}$ and then stretch by a factor of $\frac{1}{2}$ parallel to the x-axis to complete the horizontal transformations. To finish, we translate the vector by vector $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

SUMMARY

A function is a procedure that assigns to each member x of a set another value y from a second set. The first set is the domain of the function, the second set is the range of the function.

A function is said to be one-to-one if for every point y in the range of the function, there is only one value of x in the domain such that $y = f(x)$. A function is one-to-one if any horizontal line cuts the graph only once.

We can apply a function to the output of another function. This is written as $fg(x)$ which means that we carry out the function g first on x, and then f. It is important to remember that $fg(x)$ does not equal $gf(x)$.

The inverse, $f^{-1}(x)$, of a one-to-one function, $f(x)$, is the function which reverses the effect of the original function. Remember that inverse functions only exist for one-to-one functions.

Finally, we have seen basic transformations. The graph of $y = f(x) + a$ is a translation of the graph $y = f(x)$ by the vector $\begin{pmatrix} 0 \\ a \end{pmatrix}$. The graph of $y = f(x - a)$ is a translation of the graph $y = f(x)$ by the vector $\begin{pmatrix} a \\ 0 \end{pmatrix}$. The graph of $y = f(-x)$ is the reflection of the graph $y = f(x)$ in the y-axis. The graph of $y = -f(x)$ is the reflection of the graph $y = f(x)$ in the x-axis. Furthermore, for $a > 0$, the graph of $y = af(x)$ is a stretch of the graph $y = f(x)$ with a stretch factor of a parallel to the y-axis. For $a > 0$, the graph of $y = f(ax)$ is a stretch of the graph $y = f(x)$ with a stretch factor of$\frac{1}{a}$ parallel to the a-axis.

UNIT 3

COORDINATE GEOMETRY

STUDY GOALS

On completion of this unit, you will be able to ...

- determine the equation of a straight line when given sufficient information.
- understand how to represent a circle algebraically when given its radius and center.
- use different forms for straight lines and circles in solving problems.
- apply algebraic methods to find points of intersections between lines and circles.

3. COORDINATE GEOMETRY

Introduction

To help navigation, every point on the earth has unique coordinates so that it can be easily located. The coordinate system of the earth is based on imaginary lines called latitudes and longitudes, with two special lines, i.e., the Greenwich Longitude and the Equator Latitude denoting the reference axes of this coordinate system. In a similar fashion, we can locate points in a two-dimensional plane (or on a piece of paper). Here, we use the well-known horizontal and vertical coordinate axes, usually denoted as x-axis and y-axis.

Coordinate geometry is the study of geometric figures by representing them using a given coordinate system. Geometric objects such as straight lines, curves, circles, ellipses, hyperbolas, and parabolas can be drawn and transformed using their algebraic representations.

In this unit we study different ways of representing straight lines and circles in the two-dimensional plane. Circles are an example of a collection of mathematical shapes called **conic sections** which are fascinating mathematical objects that have been studied for thousands of years, especially in the context of astronomy.

Conic Sections
A conic section is a curve obtained as the intersection of a plane with a cone.

We will see how to use the algebraic representations and properties of line segments, straight lines, and circles to solve different geometry problems.

3.1 Line Segment

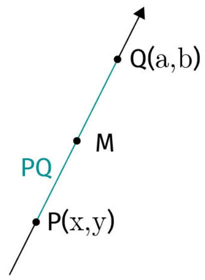
In this section we define the straight-line segment between two points in the plane and calculate its midpoint and length. Given two points $P(x, y)$ and $Q(a, b)$ in the plane, we can always find a line l which contains P and Q. This is one of the axioms of Euclidean geometry. The line segment PQ is defined as the part of the line l which starts in P and ends in Q. Consequently, every line segment has finite length. The length of the line segment PQ can be calculated with the formula:

$$\text{length } PQ = \sqrt{(a-x)^2 + (b-y)^2}$$

Moreover, the midpoint M of the ling segment PQ is given as

$$M = \left(\frac{x+a}{2}, \frac{y+b}{2}\right)$$

Figure 18: A Line Segment and Its Midpoint



Source: Florian Pausinger, (2022).

Example 3.1: The point $M(7, -10)$ is the midpoint of the line segment PQ with $P(3,2)$ and $Q(a, b)$. Find the value of a and the value of b.

Solution. We can solve this question in two different ways. First, we derive an algebraic answer. We use the midpoint formula and set $x = 3$ and $y = 2$. Then we have

$$\left(\frac{3+a}{2}, \frac{2+b}{2}\right) = (7, -10)$$

To find we equate the first coordinate on the left-hand side with the first coordinate on the right-hand side. We find that

$$\frac{3+a}{2} = 7 \Leftrightarrow a = 14 - 3 \Leftrightarrow a = 11$$

Similarly, we equate the second coordinate on the left-hand side with the second coordinate on the right-hand side and find

$$\frac{2+b}{2} = -10 \ \Leftrightarrow b = -20 - 2 \ \Leftrightarrow b = -22$$

Therefore, we conclude that $Q(11, -22)$.

Alternatively, we can also argue that since M is the midpoint of PQ the two-line segments PM and MQ have to be the same. We have that

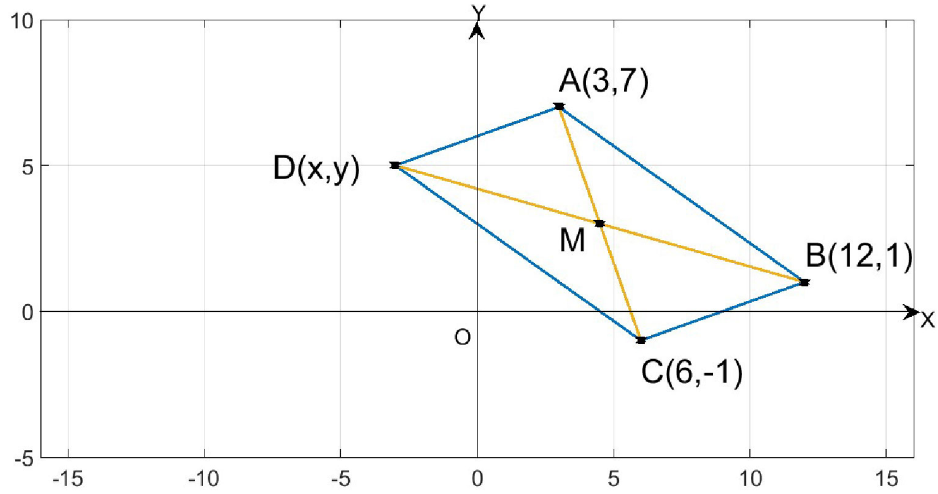
$$PM = \begin{pmatrix} 7 \\ -10 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ -12 \end{pmatrix}$$

and

$$MQ = \begin{pmatrix} 4 \\ -12 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} 7 \\ -10 \end{pmatrix}$$

From this it is easy to see that $a = 11$ and $b = -22$ as required.

Figure 19: Illustration of Exercise 3.1



Source: Florian Pausinger, (2022).

Example 3.2: We are given three of the four vertices of a parallelogram ABCD, i.e., $A(3{,}7)$, $B(12{,}1)$ and $C(6,\ -1)$. Find the coordinates of $D(x,y)$.

Solution. We can solve this problem in two steps. First, we calculate the midpoint M of AC. Next, we use the assumption that $ABCD$ is a parallelogram. Importantly, this assumption implies that M is also the midpoint of BD, and we can proceed as in Example 3.1 to find $D(x,y)$.

We have that

$$M = \left(\frac{3+6}{2}, \frac{7-1}{2}\right) = \begin{pmatrix} 4.5 \\ 3 \end{pmatrix}$$

Furthermore,

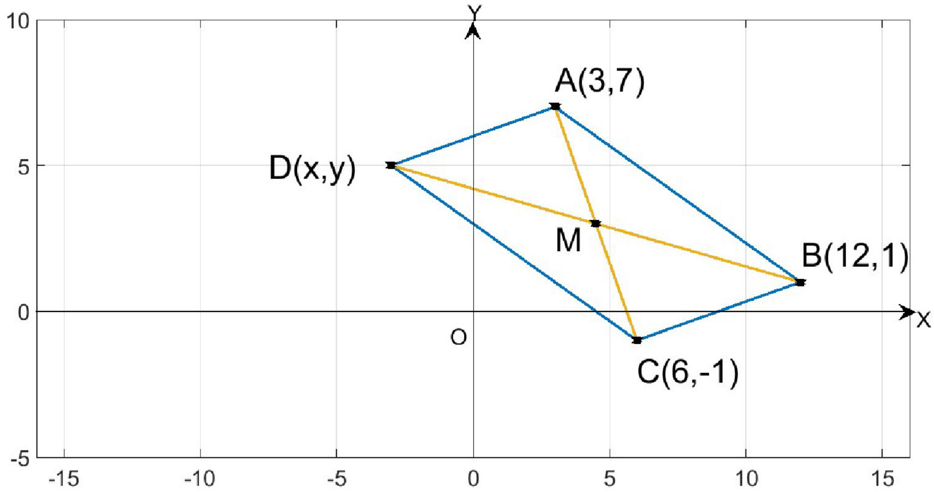
$$BM = \begin{pmatrix} 4.5 \\ 3 \end{pmatrix} - \begin{pmatrix} 12 \\ 1 \end{pmatrix} = \begin{pmatrix} -7.5 \\ 2 \end{pmatrix}$$

such that

$$MD = \begin{pmatrix} -7.5 \\ 2 \end{pmatrix} = \begin{pmatrix} x - 4.5 \\ y - 3 \end{pmatrix}$$

Consequently, we see that $x = \ -3$ and $y = 5$.

Figure 20: Illustration of Exercise 3.2



Source: Florian Pausinger, (2022).

Example 3.3: Assume that the distance between two points $P(7, -a)$ and $Q(a-3,1)$ is $\sqrt{85}$. Find the two possible values of a.

Solution. We use the formula for the length of the line segment and solve for a. We have that

$$\text{length of } PQ = \sqrt{(a-3-7)^2 + (1+a)^2} = \sqrt{(a-10)^2 + (1+a)^2}$$
$$= \sqrt{85}$$

Hence, we see that

$$\begin{aligned} 85 &= a^2 - 20a + 100 + 1 + 2a + a^2 \\ 0 &= 2a^2 - 18a + 101 - 85 \\ 0 &= a^2 - 9a + 8 = (a-8)(a-1) \end{aligned}$$

Therefore, the two possible values are $a = 1$ and $a = 8$.

3.2 Parallels and Perpendicular Lines

The aim of this section is to introduce and apply the concept of the gradient of a line segment. We can use the gradient to determine whether two lines are parallel or perpendicular.

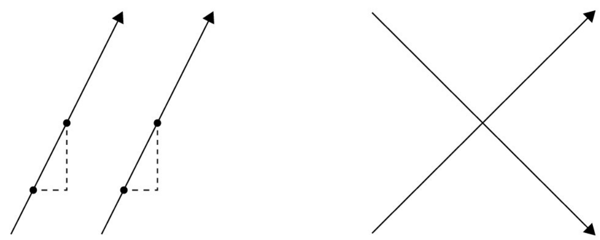
The gradient of a line segmentPQ measures how steep a line is. As a rule of thumb, the larger the gradient, the steeper the line. On the other hand, a horizontal line has a gradient of zero. Hence, the smaller the gradient, the shallower the line. The gradient of a line segment PQ is formally defined as the change in the y coordinates divided by the change in the x coordinates, i.e.

$$\text{grad } PQ = \frac{\text{change in } y}{\text{change in } x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Note that if there is no change in the x coordinates, i.e., if the line is vertical, then the gradient is not defined. If two lines have the same gradient, then they are parallel. Furthermore, if a line l has gradient $m \neq 0$, then every line with gradient $-\frac{1}{m}$ is said to be perpendicular to l.

Finally, three points are collinear if they lie on the same line. In this case, all pairwise line segments of the three points will have the same gradient.

Figure 21: Two Parallel Lines (Left) And Two Perpendicular Lines (Right)



Source: Florian Pausinger, (2022).

Example 3.4: The coordinates of three points are $A(5-k, -10-2k)$, $B(5, -k)$ and $C(7{,}1)$. Assume that A, B and C are collinear and find the two possible values k.

<u>Solution.</u> By the definition, if the three points are collinear, then the line segments AB and BC have the same gradient. Therefore,

$$\begin{aligned} \text{grad } AB = &\text{grad } BC \Rightarrow \frac{-k+10+2k}{5-5+k} = \frac{1+k}{7-5} \Rightarrow \frac{10+k}{k} = \frac{1+k}{2} \\ 2(10+k) = &k(k+1) \Rightarrow 0 = k^2 - k - 20 \\ 0 = &(k+4)(k-5) \end{aligned}$$

It follows that the two possible values are $k = -4$ and $k = 5$.

To check our result, we can calculate the gradient of A and C and compare it to the gradients of AB and BC. We have that

$$\text{grad } AC = \frac{1+10+2k}{7-5+k} \Longrightarrow -\frac{3}{2} \text{ or } 3$$

Substituting the values of k into the above gradients, we see that all the gradients are indeed the same.

Example 3.5: The vertices of the triangle ABC are $A(4{,}3)$, $B(1,\,-2)$ and $C(2k,\,-k)$. Assume that the angle ABC is a right angle. Find the possible values of k and draw a diagram to visualize the solutions.

<u>Solution.</u> Since the angle ABC is a right angle, we have that $\mathrm{grad}\ AC \cdot \mathrm{grad}\ CB = -1$. Therefore,

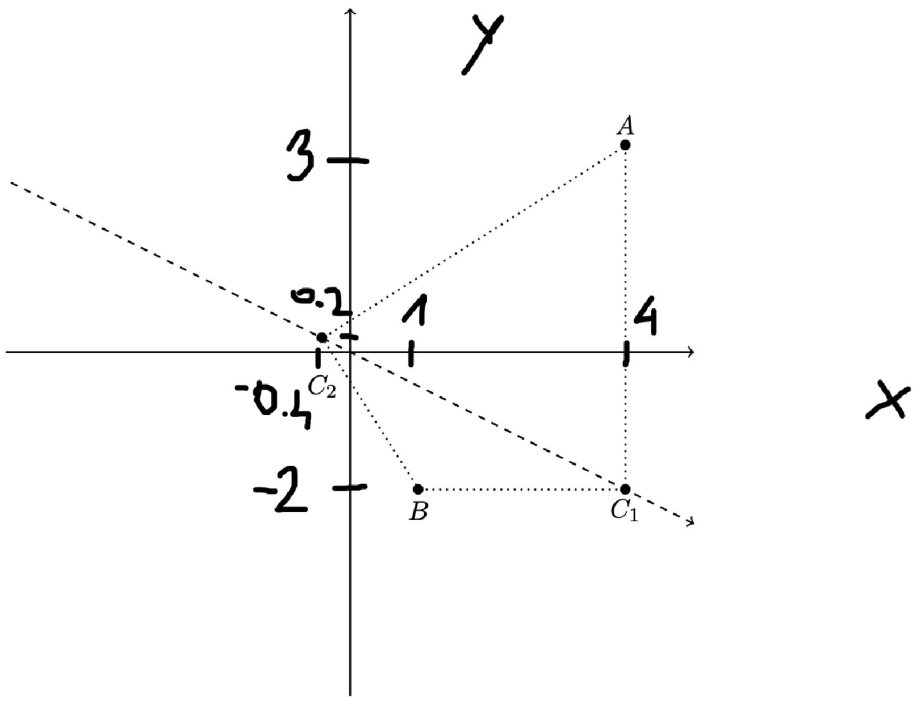
$$\begin{aligned} -1 &= \frac{-k-3}{2k-4} \cdot \frac{-2+k}{1-\ 2k} \\ (2k-4)(2k-1) &= (k-2)(-3-k) \\ 5k^2 - 9\ k - 2 &= 0 \end{aligned}$$

Here, we use the quadratic formula to get the solutions, i.e.,

$$k_{1,2} = \frac{9 \pm \sqrt{81 - 4 \cdot 5 \cdot (-2)}}{2 \cdot 5} = \frac{9 \pm \sqrt{121}}{10} = \frac{9 \pm 11}{10}$$

From this we see that $k = 2$ or $k = -0.2$. Hence, we get $C(4,\,-2)$ and $C(-0.4{,}0.2)$.

Figure 22: Illustration of Exercise 3.5



Source: Florian Pausinger, (2022).

3.3 Straight Line and Circle Equation

In the following we see different ways of representing lines before we introduce the completed square form and the general form as two ways of representing a circle. Mathematically, we think of lines and circles as the set of points in the plane whose coordinates satisfy a given equation. In this way, it is easy to check whether a given point lies on a particular line or circle: We just need to check whether the point satisfies the equation.

Equations of Straight Lines

The equation of a non-vertical straight line is

$$y = mx + c$$

in which m is the gradient of the line, c is the y-intercept of the line and $x,\ y$ are the coordinates of an arbitrary point on this line. If the line is vertical, the formula reduces to

$$x = b$$

in which b is the x-intercept and y is arbitrary, i.e., every point of the form $P(b, y)$ lies on the vertical line with x-intercept b.

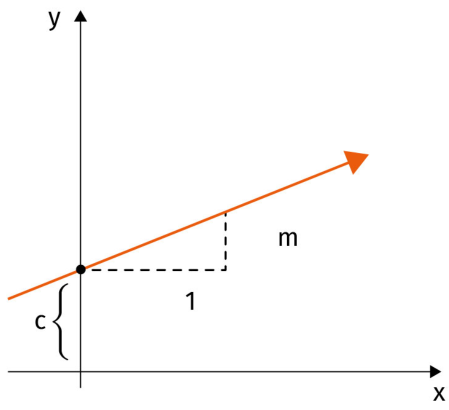
Importantly, there is an alternative representation if we know one point on the line as well as the gradient of the line. Consider a line with gradient m that passes through the known point $A(a, b)$ and whose general point is $P(x, y)$. Then

$$\text{grad } AP = m$$

Hence,

$$\frac{y-b}{x-a} = m \Rightarrow (y - b) = m(x - a)$$

Figure 23: Illustration of Equation of a Straight Line



Source: Florian Pausinger, (2022).

Example 3.6: Find the equation of the straight line with gradient 3 that passes through the point $A(3{,}2)$.

Solution. We use the formula $(y-b)=m(x-a)$ with $m=3$, $a=3$ and $b=2$. Therefore,

$$(y-2)=3(x-3)\Rightarrow 3x-y=7$$

Example 3.7: Find the equation of the straight line passing through the points $P(-4{,}13)$ and $Q(1,-2)$.

Solution. We solve this task in two steps. First, we calculate the gradient of the line segment PQ, then we follow the lines of Example 3.6 to obtain the equation.

$$\text{grad } PQ=\frac{-2-13}{1+4}=-\frac{15}{5}=-3$$

We use the formula $(y-b)=m(x-a)$ with $m=-3$, $a=-4$ and $b=13$:

$$(y-13)=-3(x+4)\Rightarrow 3x+y=1$$

Example 3.8: Given two points $P(3,-1)$ and $Q(8{,}9)$. Find the equation of the perpendicular bisector of the line segment PQ.

Solution. To answer this question, we first calculate the midpoint M of the line segment PQ since the bisector has to pass through M. Next, we calculate the gradient of PQ and invert it to obtain the gradient m of any perpendicular line. Finally, we again follow Example 3.6 to obtain the desired equation.

$$\begin{aligned} M&=\left(\frac{3+8}{2},\frac{-1+9}{2}\right)=\left(\frac{11}{2},4\right)\\ \text{grad } PQ&=\frac{9-3}{8+1}=\frac{2}{3}\Rightarrow m=-\frac{3}{2}\end{aligned} \tag{3.1}$$

We use the formula $(y-b)=m(x-a)$ with $m=\ -\frac{3}{2}$, $a=\frac{11}{2}$ and $b=4$:

$$(y-4)=-\frac{3}{2}\left(x-\frac{11}{2}\right)\Rightarrow\frac{3x}{2}+y=\frac{33}{4}+4\Rightarrow 6x+4y$$
$$=33+16\Rightarrow 6x+4y=49$$

Equations of Circles

A circle is defined as the set of all points in the plane that are a fixed distance (the radius) from a given point (the center). To find the equation of a circle, let $P(x,y)$ be any point on a circle with center $C(a,b)$ and radius r. We write $|AB|$ for the length of the line segment from A to B in the following. Using the Theorem of Pythagoras on the triangle PCQ gives:

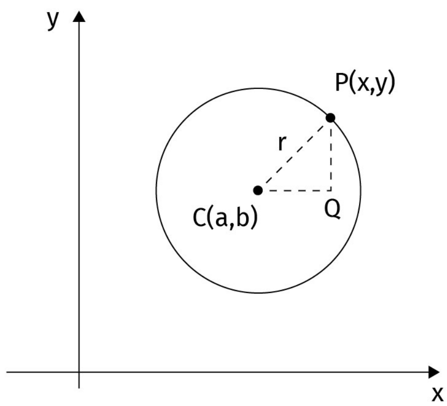
$$|CQ|^2+|PQ|^2=r^2$$

We can substitute $|CQ|=x-a$ and $|PQ|=y-b$ to get

$$(x-a)^2+(y-b)^2=r^2\,.$$

This equation is known as the completed square form of a circle with center $C(a,b)$ and radius r.

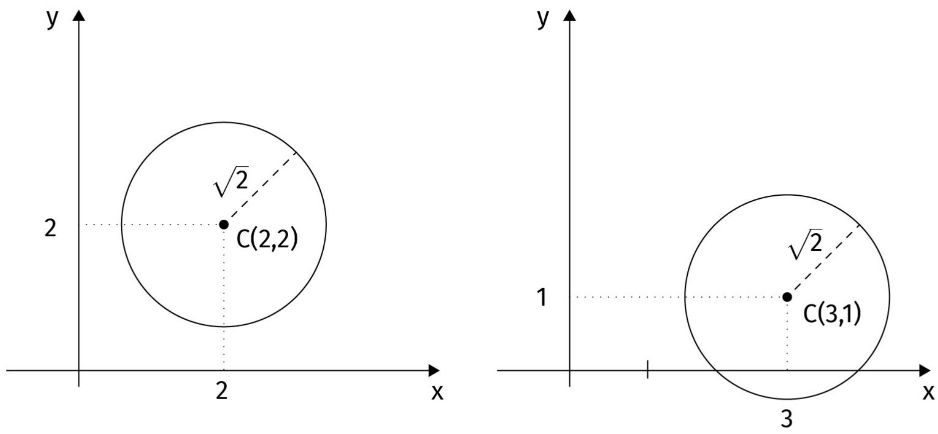
Figure 24: Illustration of the Equation of a Circle



Source: Florian Pausinger, (2022).

We can explore the completed square form graphically by changing the different parameters. In the following figure, we keep the radius fixed and change $a=2$ to $a=3$ and $b=2$ to $b=1$ resulting in a translation of the given circle.

Figure 25: Explore Completed Square Form



Source: Florian Pausinger, (2022).

Example 3.9: Find the equation of the circle centered at $C(3{,}7)$ with radius 4.

Solution. We can use the completed square form with $a = 3$, $b = 7$ and $r = 4$. Hence, we get

$$(x-3)^2 + (y-7)^2 = 16$$

Example 3.10: Given two points $A(1{,}5)$ and $B(3,\,-3)$. Find the equation of the circle having AB as a diameter.

Solution. First, note that the center C of the circle is the midpoint M of the line segment AB. Moreover, the radius of the circle is the length of the line segment AC. Thus, we get

$$C{=}M = \left(\frac{1+3}{2}, \frac{5-3}{2}\right) = (2{,}1)$$
$$\text{length } AC{=}\sqrt{(2-1)^2 + (1-5)^2} = \sqrt{17}$$

Therefore, the equation of the circle is

$$(x-2)^2 + (y-1)^2 = 17$$

When we expand the equation of a circle, we can make some interesting observations:

$$(x-a)^2 + (y-b)^2 = r^2 \Rightarrow x^2 + y^2 - 2ax - 2by \; + \left(a^2 + b^2 - r^2\right) = 0\,.$$

Note that the coefficients of x^2 and y^2 are the same. Moreover, there is no mixed xy term in this equation. This expanded form of a circle is known as the general form. It is often written as:

$$x^2 + y^2 + 2gx + 2fy + c = 0$$

Comparing with the formula above we see that the circle has center $(-g, -f)$ and radius $\sqrt{g^2 + f^2 - c}$.

Example 3.11: Find the center and the radius of the circle

$$x^2 + y^2 + 14x - 6y \ - 20 = 0$$

Solution. We can answer this question by completing the square.

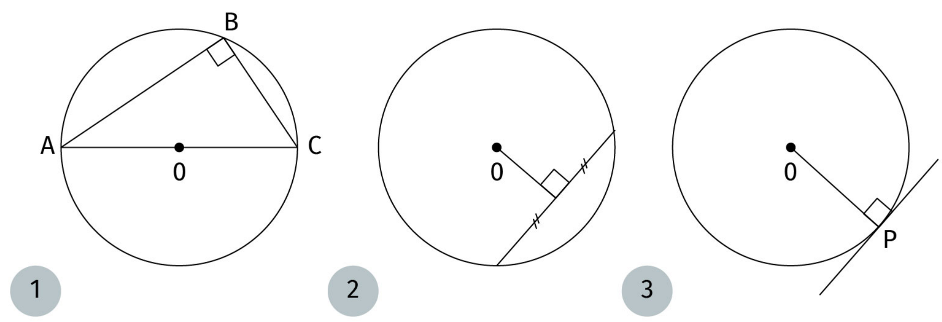
$$\begin{aligned} x^2 + 14x + 49 \ - \ 49 + y^2 - \ 6y \ + 9 \ - \ 9 \ - 20 &= 0 \\ (x+7)^2 + (y-3)^2 &= 78 \end{aligned}$$

Consequently, the center is at $C(-7{,}3)$ and the radius is $\sqrt{78}$.

We close this section by recalling three important facts about right angles and circles.

1. The angle in a semicircle is a right angle.
2. The perpendicular from the center of a circle to the chord bisects the chord.
3. The tangent to a circle at a point is perpendicular to the radius at that point.

Figure 26: Illustration of the Important Facts



Source: Florian Pausinger, (2022).

We can conclude the following fact from the above:

Given a right-angled triangle ABC such that the right angle is at B. Then the points $A,\ B,\ C$ lie on a circle with diameter AC.

Example 3.12: A circle passes through the points $D(-2{,}4),\ E(2{,}7),\ F(6{,}3)$. Find the equation of the circle.

Solution. We know that all three points lie on the circle and, hence, must satisfy the equation of the circle. This gives us a system of three equations in three unknowns which we can solve. Using the coordinates of the points we get:

$$\begin{aligned}(-2-a)^2+(4-b)^2 &= r^2 \Rightarrow a^2+b^2+4a-8b+20=r^2\\ (2-a)^2+(7-b)^2 &= r^2 \Rightarrow a^2+b^2-4a-14b+53=r^2\\ (6-a)^2+(3-b)^2 &= r^2 \Rightarrow a^2+b^2-12a-6b+45=r^2\end{aligned}$$

We can now subtract the third equation from the first, and the third from the second to obtain two simultaneous equations for a and b, i.e.,

$$\begin{aligned}4a+12a-8b+6b+20-45 &= 0\\ -4a+12a-14b+6b+53-45 &= 0\end{aligned}$$

This implies that $a=\frac{27}{14}$ and $b=\frac{41}{14}$ from which we get that $r=5\sqrt{\frac{65}{7}}$.

3.4 Points of Intersection

In this final section we solve problems about the intersection of lines and circles. Therefore, it is useful to recall that points of intersection can be found by solving the equations of the line and circle simultaneously. Note that if the resulting equation is of the form

$$ax^2+bx+c=0$$

then the sign of the discriminant b^2-4ac tells us whether there are two (discriminant positive), there is one (discriminant zero) or whether there are no real solutions to the equation (discriminant negative).

Example 3.13: The line $x=3y+10$ intersects the circle $x^2+y^2=30$ at the points A and B. Find the coordinates of the points A and B.

Solution. We substitute $3y+10$ for x in the equation of the circle to get

$$\begin{aligned}(3y+10)^2+y^2 &= 30\\ y^2+6y\ +\ 7 &= 0\end{aligned}$$

Using the quadratic formula, we get $y_{1,2}=\ -3\pm\sqrt{2}$. It follows that the two points of intersection have the coordinates:

$$\left(1+3\sqrt{2},\ -3+\sqrt{2}\right) \text{ and } \left(1-3\sqrt{2},\ -3-\sqrt{2}\right)$$

Example 3.14: Show that the line $y=x-2$ is a tangent to the circle

$$x^2+y^2-12x-2y\ +\frac{65}{2}=0$$

Solution. We substitute $x-2$ for y and expand the equation of the circle.

$$\begin{aligned} x^2+(x-2)^2-12x-2(x-2)+\tfrac{65}{2}&=0 \\ 4x^2-36x+81&=0 \\ (2x-9)^2&=0 \end{aligned}$$

This equation has one repeated root, i.e., $x=4.5$, and consequently $y=x-2$ is indeed a tangent.

SUMMARY

In this unit we have seen how to calculate the midpoint, gradient and length of a line segment. The gradient, m, of a line segment can be used to determine whether two given lines are parallel or perpendicular. It can also be used to find the equation of a straight line via

$$y-b=m(x-a)$$

in which (a,b) is a known point on the line. Alternatively, we can write

$$y=mx+c$$

in which m is the gradient of the line, c is the y-intercept of the line and $x,\ y$ are the coordinates of an arbitrary point on this line. Finally, we have seen how to represent a circle with a given center $C(a,b)$ and a given radius r algebraically in completed square form as

$$(x-a)^2+(y-b)^2=r^2$$

Furthermore, we can expand this equation to obtain the general form often written as

$$x^2+y^2+2gx+2fy+c=0$$

Here, the circle has center $(-g,\ -f)$ and radius $\sqrt{g^2+f^2-c}$.

UNIT 4

CIRCULAR MEASURE

STUDY GOALS

On completion of this unit, you will be able to ...

- explain the definition of a radian.
- convert between degrees and radian measure.
- apply the formulae for the arc length and area of a circle to solve geometric problems.

4. CIRCULAR MEASURE

Introduction

So far when studying circles, you may have only considered measuring angles in degrees. However, why are there $360°$ in one full revolution? The original reason for choosing degrees as a unit of circular measurement is unknown; one possible theory is that **ancient astronomers** assumed that the sun moves one degree each day on its trajectory around the earth while believing that a solar year lasts 360 days.

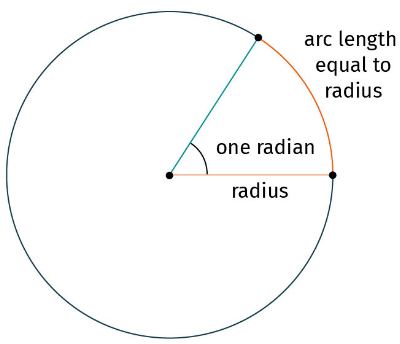
Ancient astronomers
The division of the circle by ancient astronomers into is contained in Babylonian texts dating back to 2000 BC.

Degrees are not the only way in which angles can be measured. There also exists the radian measure that we use extensively in mathematics problems. It involves the number π, which in turn often simplifies our formulae and calculations involving arc length and area of a sector. Hence radian measure is sometimes referred to as the natural unit of circular measurement.

4.1 Radians and Arc Length

One radian is approximately equal to $57.3°$, and is formally defined as the angle subtended at the centre by an arc equal in length to the radius.

Figure 27: Radian Measure



Source: Florian Pausinger, (2022).

This formal definition implies that

- 2π radians $= 360°$
- π radians $= 180°$

1 radian or 1 rad is used to denote angles measured in radians, however often no symbol at all is used after the numerical value. For example, $90° = \frac{\pi}{2}$. We can convert arbitrary angles that are written in degrees into radians, and vice versa via the following rules:

- To change from degrees into radians, multiply by $\frac{\pi}{180}$
- To change from radians into degrees, multiply by $\frac{180}{\pi}$

Example 4.1: Convert $30°$ into radians.

Solution. Using the above rule, $30 \cdot \frac{\pi}{180} = \frac{\pi}{6}$ rad.

Example 4.2: Convert $\frac{5\pi}{8}$ into degrees.

Solution. To calculate the angle in radians, we compute $\frac{5\pi}{8} \cdot \frac{180}{\pi} = 112.5°$.

Example 4.3: Write $28°$ in radians, correct to 3 significant figures.

Solution. Using the standard rule, $28 \cdot \frac{\pi}{180} = 0.489$.

Given a circle of radius r and recalling the definition of a radian, we conclude that an **arc** subtending an angle of 1 radian at the center of the circle is of length r. Therefore, if an arc subtends an angle of θ rad at the center, then the length of the arc (which is usually denoted as s) shall be equal to $r \cdot \theta$. That is,

Arc
An arc of a circle is defined as a connected subset of the circumference of a circle.

$$s = r\theta \,.$$

Example 4.4: An arc subtends an angle of $\frac{\pi}{3}$ at the center of a circle, with radius 15 cm. What is the arc length? Write your answer in terms of π.

Solution. Arc length $= r\theta = 15 \cdot \frac{\pi}{3} = 5\pi \text{ cm}$.

4.2 Sector Area

When the angle subtended at the center of a sector is measured in radians, one can utilize a simple formula to determine the area of the sector.

We have the following ratio (Why is this correct? Try to find an intuitive geometric argument!):

$$\frac{\text{area of sector}}{\text{area of circle}} = \frac{\text{angle in the sector}}{\text{complete angle at the centre}}$$

Assuming that θ is measured in radians, the ratio becomes

$$\frac{\text{area of sector}}{\pi r^2} = \frac{\theta}{2\pi}$$

which can be rearranged to give

$$\text{Area of sector} = \frac{1}{2}r^2\theta$$

Example 4.5: Find the area of the sector with radius $5\ \text{cm}$ and subtended angle of $1\frac{3}{5}\ rad$.

Solution. Using the formula derived above,

$$\text{Area} = \frac{1}{2}r^2\theta = \frac{1}{2} \cdot 5^2 \cdot \frac{8}{5} = 20\text{cm}^2$$

Example 4.6: AOB is a sector of a circle, center O, with radius $8\ \text{cm}$. The length of the arc AB is $12\ \text{cm}$. Find the area of the sector AOB.

Solution. To be able to use the area of a sector formula, we need to know the angle at the center of the circle. We can determine this value via the radius and arc length using $s = r\theta$,

$$s = r\theta \Longrightarrow 12 = 8 \cdot \theta \Longrightarrow \theta = \frac{3}{2}$$

Therefore,

$$A = \frac{1}{2}r^2\theta \implies A = \frac{1}{2} \cdot 8^2 \cdot \frac{3}{2} = 48\text{cm}^2 \tag{4.1}$$

4.3 Problems

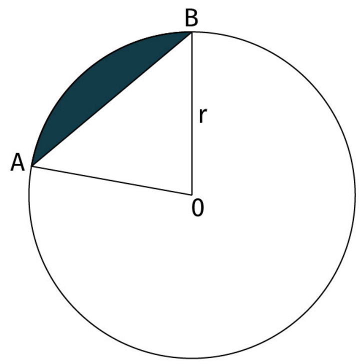
We now use the results derived in the previous subsections to consider more complex problems.

Knowledge of the area of a segment of a circle will be useful in what follows. Note that the area of a sector is given by $A = \frac{1}{2}r^2\theta$ and the area of a triangle is given by $A = \frac{1}{2}ab\sin C$ in which C denotes the angle enclosed by the two sides of the triangle of length a and length b. Hence, we can derive an expression for the area of a segment of a circle.

$$\text{Area of segment} = \frac{1}{2}r^2\theta - \frac{1}{2}r^2\sin\theta = \frac{1}{2}r^2(\theta - \sin\theta)$$

Example 4.7: The circle has radius $8\ \text{cm}$ and center O. AB is a chord and angle $AOB = 1.4$ rad. Find a) the area of the sector AOB, b) the area of triangle AOB and c) the area of the shaded segment.

Figure 28: Illustration of Example 4.7



Source: Florian Pausinger, (2022).

Solution.

a) Begin by calculating the area of the sector AOB,

$$A = \frac{1}{2} \cdot 8^2 \cdot 1.4 = 44.8\text{cm}^2$$

b) Using the standard formula for the area of a triangle $A = \frac{1}{2}ab\sin C$, we calculate the area of the triangle AOB

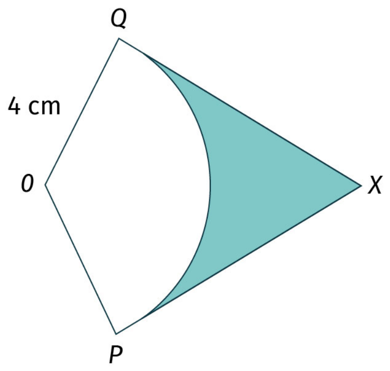
$$A = \frac{1}{2}(8)(8)\sin 1.4 = 31.5\text{cm}^2$$

c) The desired area is therefore the difference between the area of the sector and the area of the triangle,

$$\text{Area of segment} = 44.8 - 31.5 = 13.3\text{cm}^2$$

Example 4.8: The diagram shows a sector POQ of a circle, center O, with radius $4\ \text{cm}$. The length of the arc PQ is $7\ \text{cm}$. The lines PX and QX are tangents to the circle at P and Q, respectively. Find the area of the shaded section (Pemberton, 2018, p. 109: Example 4C.5)

Figure 29: Illustration of Example 4.8



Source: Florian Pausinger, (2022), based on Pemberton (2018).

<u>Solution.</u> The arc length PQ is equal to 7 cm and the radius of the sector POQ is 4 cm, therefore the angle POQ can be calculated as follows.

$$s = r\theta \Longrightarrow 7 = 4 \cdot \theta \Longrightarrow \theta = \frac{7}{4}$$

Notice next that POX is a right-angled triangle since PX is the tangent to the circle at point P . The anglePOX is equal to $\frac{7}{8}$ radians since the line OX bisects the angle POQ. Hence, using a standard trigonometric relation we can calculate the length of PX

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} \Longrightarrow \tan\left(\frac{7}{8}\right) = \frac{PX}{4}$$

Thus

$$PX = 4 \cdot \tan\left(\frac{7}{8}\right) = 4.79 \text{ cm}$$

Once again, because POX is a right-angled triangle we can calculate the area of $PXOQ$

$$\text{Area of PXOQ} = 4 \cdot 4.79 = 19.16 \text{ cm}^2$$

The area of the sector is

$$A = \frac{1}{2} \cdot 4^2 \cdot \frac{7}{4} = 14 \text{ cm}^2$$

The desired region is clearly the difference between these calculated values,

$$\text{Area of shaded region} = 19.16 - 14 = 5.16 \text{ cm}^2$$

SUMMARY

The radian measure is used extensively in difficult circle geometry problems because it is intuitively more suited to measuring angles than degrees as it simplifies many calculations.

To convert between degrees and radians, use: π radians $= 180°$.

When the angle subtended at the center of a circle by a chord is measured in radians, we can use the following formula:

- The area of a triangle $= \frac{1}{2}ab\sin C$
- The length of an arc $= r\theta$
- The area of a sector $= \frac{1}{2}r^2\theta$
- The area of a segment $= \frac{1}{2}r^2(\theta - \sin\theta)$

UNIT 5

TRIGONOMETRY

STUDY GOALS

On completion of this unit, you will be able to ...

- sketch the graphs of , and for a given range of values of .
- understand and use the notation of , and to denote the inverse trigonometric relations.
- apply the relations and to prove trigonometric identities.
- use identities to solve simple trigonometric equations for a given interval.

5. TRIGONOMETRY

Introduction

Trigonometry is concerned with basic geometric objects such as triangles or circles, with functions associated to it and with their application to geometric problems involving oscillations and waves; hence, it is important for many areas of geometry and physics.

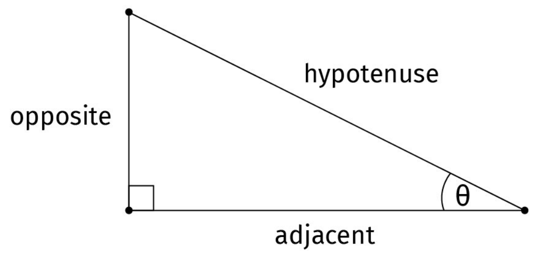
Trigonometry is often described as if it is all about triangles, but it is a lot more interesting than that. For one thing, it describes the behavior of waves and resonance. Thus, trigonometric functions play a crucial role in many engineering problems and can be used to describe how sound or light is moving.

Apart from that, they are also interesting from the point of view of pure mathematics as they often miraculously appear as solutions to complicated equations and, as such, indicate that a given problem must have some intricate, underlying geometric structure.

5.1 Trigonometric Functions and Graphs

Trigonometric functions are often used in right-angled triangles to calculate missing angles and sides. To define the three elementary trigonometric ratios, we must first label the sides in relation to the angle θ as follows.

Figure 30: Right-angled Triangle



Source: Florian Pausinger, (2022).

Therefore, we have

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$$

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$

Exact Values of Trigonometric Functions

Motivated from the angles inside a right-angled triangle, we can tabulate the exact values of the three trigonometric functions for $30°$, $45°$ and $60°$.

Table 2: Exact Values of Functions

	$\sin x$	$\cos x$	$\tan x$
$x = 30° = \frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{3}}$
$x = 45° = \frac{\pi}{4}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	1
$x = 60° = \frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$

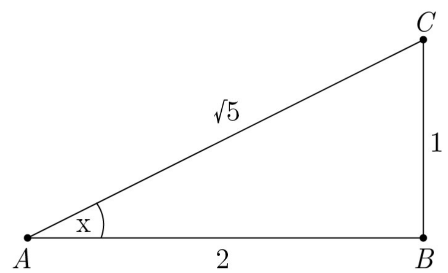
Source: Florian Pausinger, (2022).

Example 5.1: Assume that we have a right-handed triangle such that $\cos x = \frac{2}{\sqrt{5}}$ with x being acute, find the exact value of $\sin x$ and $\tan x$.

Solution. From the definition of the cosine, we can conclude that the third side of the triangle has length 1 using the Theorem of Pythagoras (Note that there are in principle infinitely many triangles satisfying our assumption. However, assuming that the first two sides are of length 2 and $\sqrt{5}$ automatically implies that the third side has length 1.)

Using the basic ratios above, we can draw the following triangle with angle $B\widehat{A}C = x$.

Figure 31: Illustration of Example 5.1



Source: Florian Pausinger, (2022).

Therefore, $\sin x = \frac{opp}{hyp} = \frac{1}{\sqrt{5}} = \frac{\sqrt{5}}{5}$ and $\tan x = \frac{opp}{adj} = \frac{1}{2}$.

Example 5.2: Solve the equation $8x^2 + 14x - 16 = -\tan\frac{\pi}{4}$.

Solution. From the table above, $\tan\frac{\pi}{4} = 1$ therefore we can rewrite the equation to give the quadratic,

$$8x^2 + 14x - 16 = -1 \Longrightarrow 8x^2 + 14x - 15 = 0$$

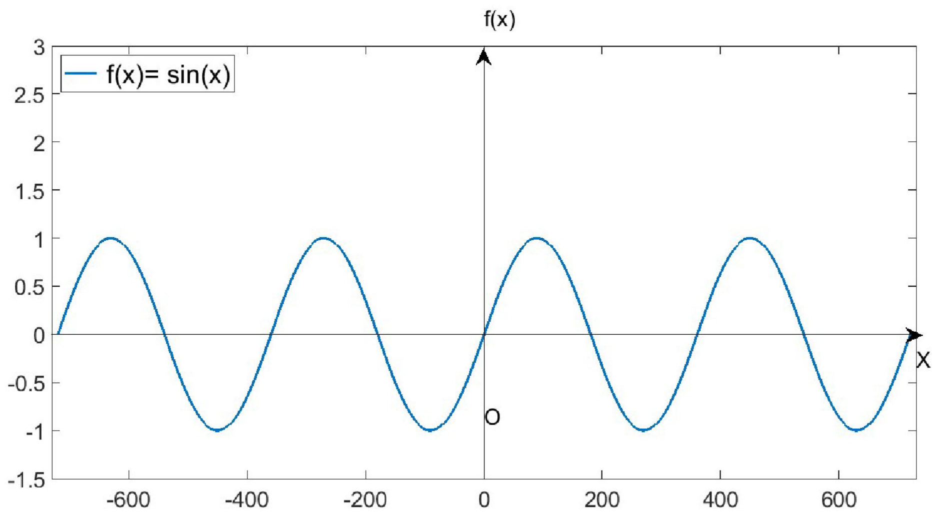
Factorizing and solving gives,

$$(4x - 3)(2x + 5) = 0 \Longrightarrow x = \frac{3}{4}, -\frac{5}{2}$$

Trigonometric Graphs

You are required to be familiar with the key features of the sine, cosine, and tangent graphs. In what follows, we plot the graphs with the x-axis, however, you should also be familiar with plotting in radians. Let's first look at $y = \sin x$.

Figure 32: Sine Curve

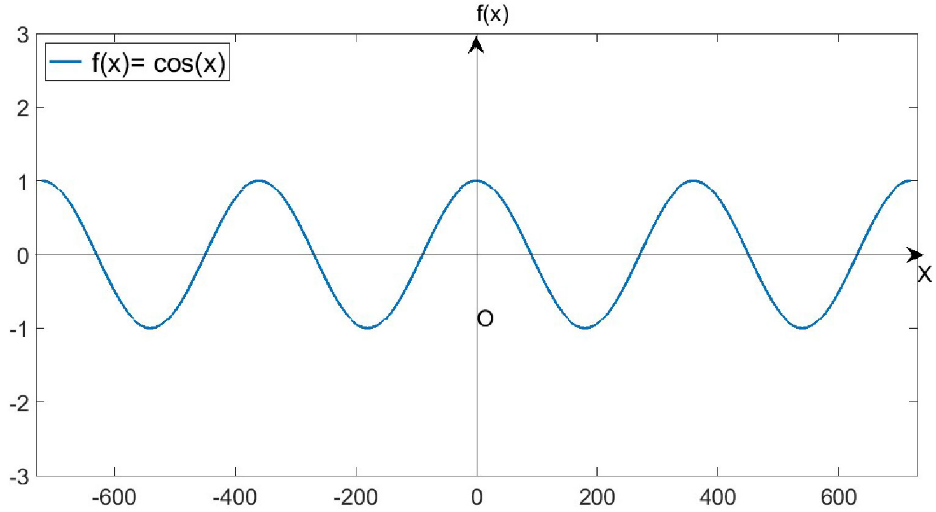


Source: Florian Pausinger, (2022).

- The curve $y = \sin x$ passes through the origin.
- It reaches a maximum value of 1 when $x = 90°$, and a minimum value of -1 when $x = 270°$.
- It has a period of $360°$.
- $y = 0$ for every integer multiple of π.
- Due to the periodicity, $\sin(x - 360°) = \sin x = \sin(x + 360°)$.
- Finally, we have that $\sin(-x) = -\sin(x)$.

Now we draw the curve of $y = \cos x$.

Figure 33: Cosine Curve

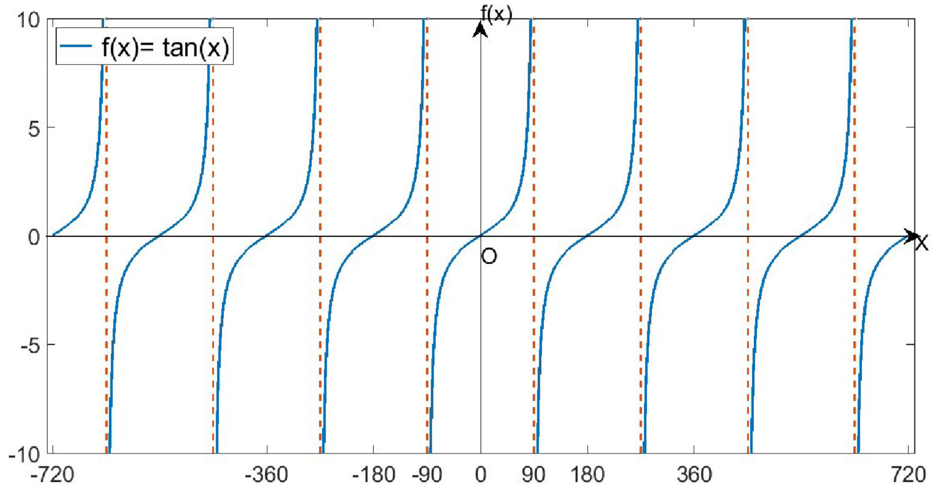


Source: Florian Pausinger, (2022).

- The curve $y = \cos x$ has a maximum value of 1 when $x = 0°$, and a minimum value of -1 when $x = 180°$.
- It has a period of $360°$.
- Due to the periodicity, $\cos(x - 360°) = \cos x = \cos(x + 360°)$.
- The cosine curve has symmetry about the line $x = 0°$, meaning that $\cos(x) = \cos(-x)$.

Lastly, consider the graph of $y = \tan x$.

Figure 34: Tangent Curve



Source: Florian Pausinger, (2022).

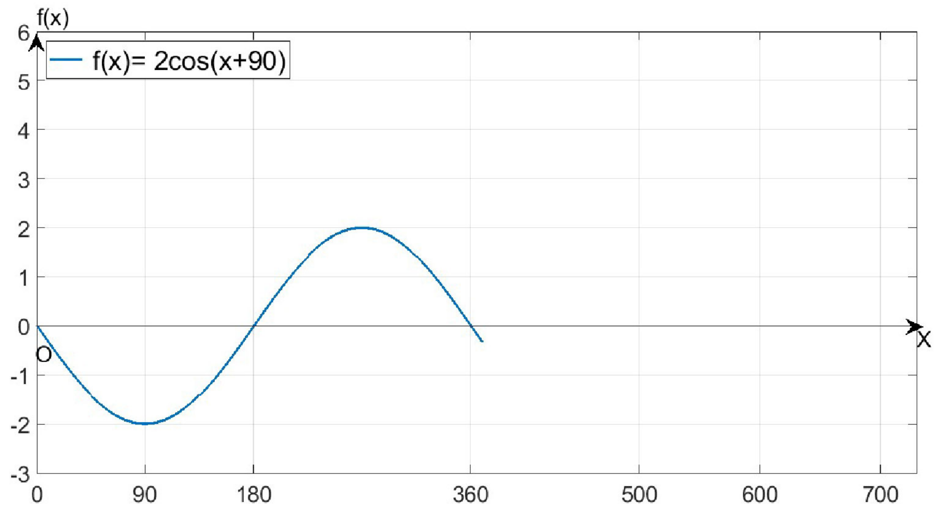
- The curve $y = \tan x$ has period of $180°$, therefore, $\tan(x) = \tan(x + 180°)$.
- $y = 0$ when $x = ..., \; -180°, \; 0°, \; 180°, \; ...$ or at every integer multiple of π.
- There are no maxima or minima. Instead, the curve approaches asymptotes at $x = ..., \; -270°, \; -90°, \; 90°, \; 270°, \; ...$. We write $y \to \infty$ as $x \to 90°$.
- The tangent curve does not have any lines of symmetry.
- Finally, we have that $\tan(-x) = -\tan(x) and \tan(90° + x) = -\tan(90° - x)$.

We can apply transformations on the trigonometric graphs.

Example 5.3: Sketch the graph $y = 2\cos(x + 90)$ of for the range $0° \leq x \leq 360°$.

Solution. We must perform the translation of 90 units left first and thereafter we can perform the stretch by scale factor 2 parallel to the y-axis. Therefore,

Figure 35: Illustration of Example 5.3



Source: Florian Pausinger, (2022), based on Pemberton (2018).

5.2 Inverse Trigonometric Functions

The functions of $\sin x$, $\cos x$ and $\tan x$ for $x \in \mathbb{R}$ are many-to-one functions. A many-to-one function does not have an inverse function (since it would be a one-to-many relation, which is not a function). However, it is possible to restrict the domain of each of these functions to obtain a one-to-one function leading to restricted inverse functions.

Note that:

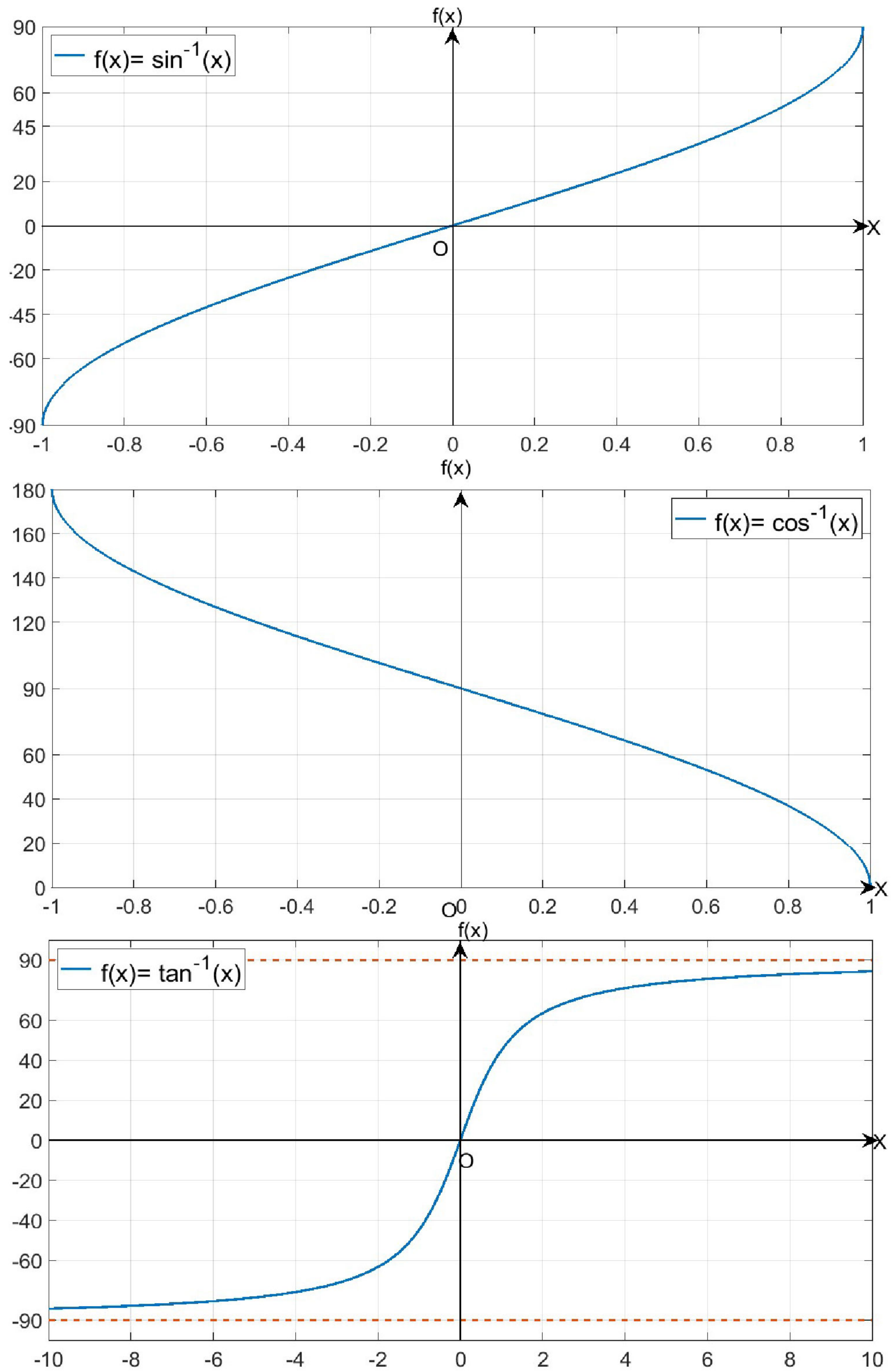
- $y = \sin x$ is one-to-one for $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$
- $y = \cos x$ is one-to-one for $0 \leq x \leq \pi$
- $y = \tan x$ is one-to-one for $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$

The inverse trigonometric functions are hence defined as:

- $y = \sin^{-1} x$ for domain $-1 \leq x \leq 1$ and range $-\frac{\pi}{2} \leq \sin^{-1} x \leq \frac{\pi}{2}$
- $y = \cos^{-1} x$ for domain $-1 \leq x \leq 1$ and range $0 \leq \cos^{-1} x \leq \pi$
- $y = \tan^{-1} x$ for domain $x \in \mathbb{R}$ and range $-\frac{\pi}{2} \leq \tan^{-1} x \leq \frac{\pi}{2}$

These are sometimes referred to as $\arcsin$, $\arccos$ and $\arctan$ respectively. The graphs of each inverse function are as follows.

Figure 36: Inverse Trig Graphs



Source: Florian Pausinger, (2022).

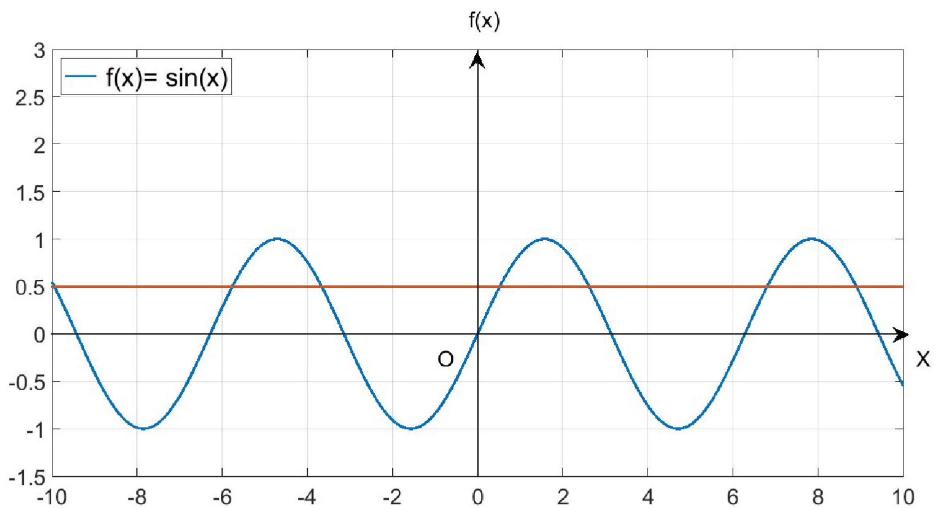
Example 5.4: The function $f(x) = 3\sin x - 4$ is defined for the domain $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. Find the range of $f(x)$. Does the function $f(x)$ have an inverse? If yes, find $f^{-1}(x)$.

Solution. Note that $f(x)$ is one-to-one in the domain $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$, therefore find end point values $f\left(-\frac{\pi}{2}\right) = -7$ and $f\left(\frac{\pi}{2}\right) = -1$. The range of $f(x)$ is $-7 \leq f(x) \leq -1$. Since the $f(x)$ is one-to-one, the inverse exists and can be written as $f^{-1}(x) = \sin^{-1}\left(\frac{x+4}{3}\right)$ with domain $-7 \leq x \leq -1$.

5.3 Trigonometric Equations and Identities

Consider the solution x to the equation $\sin x = 0.5$. Using a calculator we are certain that one solution is $x = \sin^{-1}(0.5) = \frac{\pi}{6}$ (or $30°$). However, if we plot the graphs of $y = \sin x$ and $y = 0.5$, we can see that there are multiple (actually infinite) solutions to the above equation.

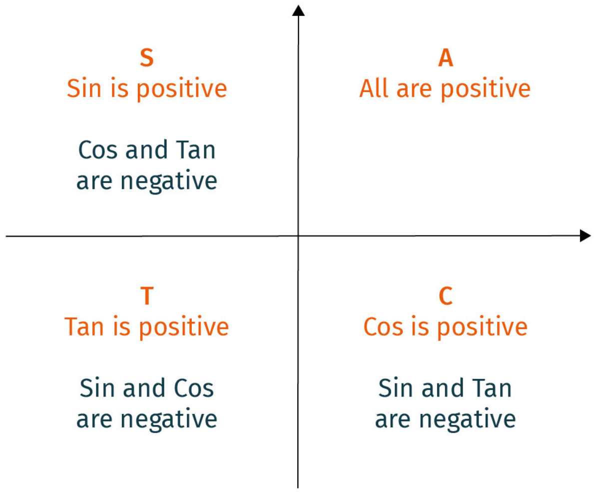
Figure 37: Infinite Sine Solutions



Source: Florian Pausinger, (2022).

In order to aid solving trigonometric equations like the one above, we introduce a tool called the CAST diagram. The CAST diagram helps you to find the sine, cosine, and tangent of angles outside of the range $0°$ to $90°$. It is often used between $0°$ to $360°$ or between $-180°$ to $180°$.

Figure 38: CAST Diagram



Source: Florian Pausinger, (2022).

To use a CAST diagram …

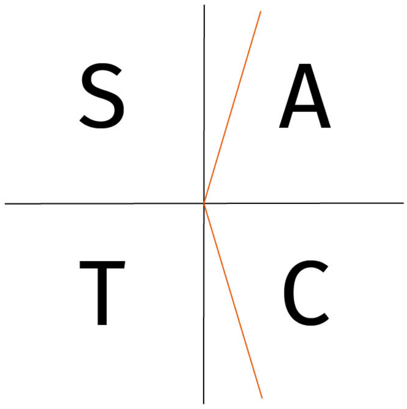
- … use the calculator to take the inverse trigonometric function and find an answer to the problem (this is called the ‘Principal Value’ or PV).
    - Ignore any minus signs if they are present.
- … depending on which function was used and whether there was a minus sign, we plot the PV in the appropriate quadrants in the CAST diagram.
    - Measure from the horizontal always.

- … measure in a circular manner, collecting all answers you come across, dependent on the range given in the question.
    - Moving anti-clockwise results in positive answers.
    - Moving clockwise results in negative answers.

Example 5.5: Solve $\cos x = 0.2$ for $0 \leq x \leq 2\pi$.

Solution. Find the principal value $x = \cos^{-1}(0.2) = 1.37$ rad. Place this angle in the CAST diagram in the C and A quadrants because these are the quadrants in which the cosine is positive.

Figure 39: Illustration of Example 5.5



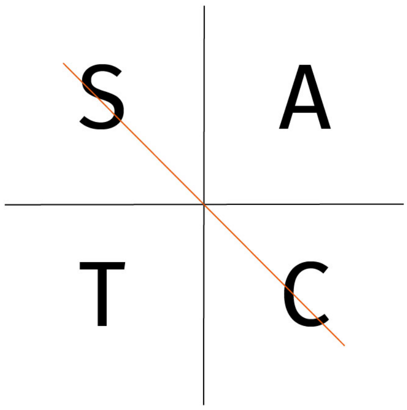
Source: Florian Pausinger, (2022).

Hence, $x = 1.37$ and $x = 2\pi - 1.37 = 4.91$.

Example 5.6: Solve $\tan x = -1.2$ for $-180° \leq x \leq 180°$.

<u>Solution.</u> We find the principal value by taking inverse tan of positive 1.2. We have $x = \tan^{-1}(1.2) = 50.42°$. We input this to the CAST diagram in the S and C quadrants since this is exactly where tangent is negative.

Figure 40: Illustration of Example 5.6



Source: Florian Pausinger, (2022).

Reading from the CAST diagram in the range $-180° \leq x \leq 180°$, i.e., one half turn clockwise and one half turn anti-clockwise, we get $x = 129.58°, \ -50.42°$.

Trigonometric Identities

Throughout this section, we use the symbol $\equiv$ in place of $=$. This is the identity (or sometimes called, equivalence) symbol. It indicates that something is always true, not just for particular values of a variable.

The first useful identity involves all three of the trigonometric functions used above and gives some insight on how they are linked (and why the graph of $\tan x$ is so different to those of $\sin x$ and $\cos x$).

$$\tan x \equiv \frac{\sin x}{\cos x}$$

This follows immediately from the elementary trigonometric ratios given in the first section. A second important formula is:

$$\sin^2 x + \cos^2 x \equiv 1$$

These two identities can be used to simplify many other trigonometric expressions and to prove many other identities. When proving an identity, it is usual to start with the more complicated side of the identity and prove that it simplifies to the less complicated side.

Example 5.7: Prove the identity $\tan x + \frac{1}{\tan x} \equiv \frac{1}{\sin x} + \frac{1}{\cos x}$

Solution. Begin with the LHS (left-hand side)

$$\begin{aligned} LHS &= \frac{\sin x}{\cos x} + \frac{1}{\left(\frac{\sin x}{\cos x}\right)} = \frac{\sin x}{\cos x} + \frac{\cos x}{\sin x} \\ &= \frac{\sin x + \cos x}{\sin x \cos x} \\ &= \frac{1}{\cos x} + \frac{1}{\sin x} = RHS \end{aligned}$$

Example 5.8: Prove $\frac{\cos^4 x - \sin^4 x}{\cos^2 x} \equiv 1 - \tan^2 x$.

Solution. Once again, begin with the LHS and perform the difference of two squares factorization on the numerator

$$LHS = \frac{\cos^4 x - \sin^4 x}{\cos^2 x} = \frac{\left(\cos^2 x + \sin^2 x\right)\left(\cos^2 x - \sin^2 x\right)}{\cos^2 x}$$

Notice now that $\cos^2 x + \sin^2 x \equiv 1$, hence

$$= \frac{\cos^2 x - \sin^2 x}{\cos^2 x} = \frac{\cos^2 x}{\cos^2 x} - \frac{\sin^2 x}{\cos^2 x} = 1 - \tan^2 x$$

as required.

Furthermore, we can also use these original identities to help solve more complicated trigonometric equations.

Example 5.9: Solve $2\sin^2 x + 3\cos x - 3 = 0$ for $0 \leq x \leq 2\pi$.

Solution. Replace $\sin^2 x$ with $1 - \cos^2 x$.

$$2\left(1 - \cos^2 x\right) + 3\cos x - 3 = 0$$

Rearranging gives the equation,

$$2\cos^2 x - 3\cos x + 1 = 0$$

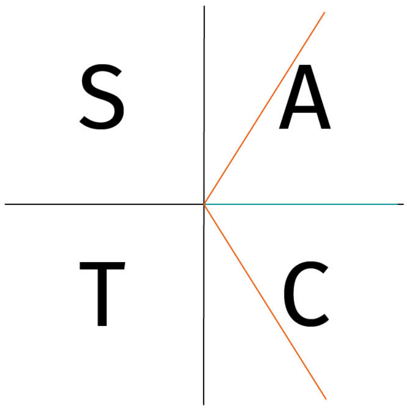
Then factorize and separate,

$$(2\cos x - 1)(\cos x - 1) = 0 \Longrightarrow \cos x = \frac{1}{2} \quad \text{and} \quad \cos x = 1$$

Therefore, we have that either $x = \cos^{-1}(0.5) = \frac{\pi}{3}$ or $x = \cos^{-1}(1) = 0$. Following the same method as outlined previously, we plot the two principal values above into a **CAST diagram**.

CAST diagram
One can use two separate CAST diagrams here, or input both principal values into the same CAST diagram.

Figure 41: Illustration of Example 5.9



Source: Florian Pausinger, (2022).

Hence $x = 0,\ \frac{\pi}{3},\ -\frac{\pi}{3},\ 2\pi$ in the range $0 \leq x \leq 2\pi$.

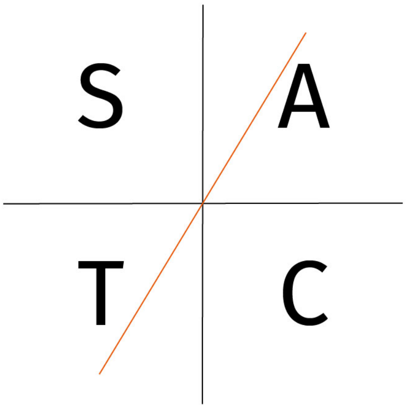
Example 5.10: Solve $\sqrt{3}\cos\theta = \sin\theta$ for $-\pi \leq \theta \leq \pi$.

Solution. Divide by $\cos\theta$ to obtain,

$$\tan\theta = \sqrt{3} \Longrightarrow \theta = \tan^{-1}\left(\sqrt{3}\right) = \frac{\pi}{3}$$

Hence $\theta = \frac{\pi}{3}$ is the principal value. We put this into the CAST diagram,

Figure 42: Illustration of Example 5.10



Source: Florian Pausinger, (2022).

Therefore, reading from the CAST diagram $\theta = \frac{\pi}{3}, \quad -\frac{2\pi}{3}$ in the desired range.

SUMMARY

The three basic trigonometric functions of $\sin x$, $\cos x$ and $\tan x$ can be defined as the ratio of sides of a right-angled triangle.

$$\sin\theta = \frac{opp}{hyp}$$
$$\cos\theta = \frac{adj}{hyp}$$
$$\tan\theta = \frac{opp}{adj}$$

The trigonometric functions can be inversed (to obtain $\sin^{-1} x$, $\cos^{-1} x$ and $\tan^{-1} x$) if one restricts the domain so that the functions become one-to-one.

- $y = \sin x$ is one-to-one for $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$
- $y = \cos x$ is one-to-one for $0 \leq x \leq \pi$
- $y = \tan x$ is one-to-one for $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$

The graphs of $\sin x$ and $\cos x$ are translated versions of each other with periodic maxima and minima. However, $\tan x$ follows a different pattern due to the identity $\tan x \equiv \frac{\sin x}{\cos x}$. We also have another important identity $\sin^2 x + \cos^2 x \equiv 1$. These identities are used to simplify and solve simple and complex trigonometric equations. All solutions are found via plotting the principal value on a CAST diagram.

UNIT 6

SERIES

STUDY GOALS

On completion of this unit, you will be able to ...

- apply the expansion of $(a+b)^n$, where n is a positive integer.
- understand the role of Pascal's triangle in the expansion of $(a+b)^n$.
- identify arithmetic and geometric progressions and calculate the sum of the first n terms of such progressions.
- solve problems involving arithmetic and geometric progressions.
- apply the formula for the sum to infinity for a geometric progression.

6. SERIES

Introduction

A set of ordered numbers, such that each number can be obtained from the previous number by a certain rule, is called a sequence. Each number of the sequence is called a term. When terms of a sequence are added together, we obtain a series.

Mathematical sequences crop up in surprising places, particularly and often in the natural world. There exists a kind of shellfish called the nautilus, which is one of the oldest known animals on earth. The shell of a nautilus grows in a spiral shape and each time a new chamber is grown, it is larger than the last. Comparing the size of each chamber with the last, a very consistent mathematical sequence is formed which can be described as a geometric progression.

6.1 Pascal’s Triangle

We already know how to expand the simple expressions

$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

Blaise Pascal
He was a 17th century mathematician and played a key role in the invention of the first mechanical calculators.

However, for any power larger than 3, things begin to get a little tricky and tedious in the calculation. Notice in the first instance that if we extract the coefficients of the above expansions, we get the pattern as detailed in the following figure. This is called **Pascal**’s triangle. Pascal's triangle can be easily constructed from scratch. Note that each value in row j, is the sum of the two values diagonally to the left and diagonally to the right in row j-1. This is illustrated in the below figure for the value 10 in row 5, which is the sum of 4 and 6 coming from row 4. Similarly, the value 35 in row 7, is obtained by adding 10 and 15 coming from row 6.

Figure 43: Pascal's Triangle

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

Source: Florian Pausinger, (2022).

Pascal's triangle can be used to expand expressions in the form $(a+b)^n$.

Example 6.1: Expand $(x+2)^5$.

<u>Solution.</u> From Pascal's triangle above, take the coefficients from the line that begins with 1 and 5. That is, the coefficients of the expansion will be $1,\ 5,\ 10,\ 10,\ 5,\ 1$. The powers of x will decrease from 5 to zero as the powers of the second term 2 will increase from zero to 5.

$$(x+2)^5 = 1x^5 + 5x^4(2) + 10x^3\left(2^2\right) + 10x^2\left(2^3\right) + 5x\left(2^4\right) + 1\left(2^5\right)$$

Hence,

$$(x+2)^5 = x^5 + 10x^4 + 40x^3 + 80x^2 + 80x + 32$$

6.2 Binomial Notation and Expansion

Pascal's triangle provides a tool to expand expressions of the form $(a+b)^n$, however, with large n, this task of finding the 'correct line' of the triangle to determine the coefficient is cumbersome. For this reason, it is almost always more convenient to use the binomial theorem. The coefficients in any binomial expansion are called binomial coefficients and we can write them as $\binom{n}{r}$ (or sometimes written as nCr) for positive integers n and r with r less than or equal to n. These coefficients can be calculated as

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

where $n! = n \cdot (n-1) \cdot (n-2) ... \cdot 3 \cdot 2 \cdot 1$ denotes n factorial. If n is a positive integer, then

$$(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n$$

Example 6.2: Find the first four terms of the expansion of $(1+2x)^{10}$.

<u>Solution.</u> Using the first expansion above,

$$\begin{aligned}(1+2x)^{10} &= 1 + \binom{10}{1}(2x) + \binom{10}{2}(2x)^2 + \binom{10}{3}(2x)^3 + \dots \\ &= 1 + 10(2x) + 45(4x^2) + 120(8x^3) + \dots \\ &= 1 + 20x + 180x^2 + 960x^3 + \dots\end{aligned}$$

Example 6.3: Expand $(1+x)^4$. Using the expansion, find the exact value of $\left(1+\sqrt{3}\right)^4$.

<u>Solution.</u> With $n = 4$,

$$(1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4$$

To find the exact value of $\left(1+\sqrt{3}\right)^4$, set $x = \sqrt{3}$. Therefore,

$$\begin{aligned}&= 1 + 4\left(\sqrt{3}\right) + 6\left(\sqrt{3}\right)^2 + 4\left(\sqrt{3}\right)^3 + \left(\sqrt{3}\right)^4 \\ &= 1 + 4\sqrt{3} + 6(3) + 4(3)\sqrt{3} + 9 \\ &= 1 + 18 + 9 + 4\sqrt{3} + 12\sqrt{3} \\ &= 28 + 16\sqrt{3}\end{aligned}$$

6.3 More Complicated Expansions

The examples in this last subsection of the form $(1+x)^n$ are among the easiest to expand. The binomial expansion theorem can also be used to expand more complex expressions of the form $(a+b)^n$ since

$$(a+b)^n = a^n\left(1+\frac{b}{a}\right)^n$$

Alternatively, one can use the following expansion formula

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{n}b^n$$

Example 6.4: Expand $(5-2x)^5$.

Solution. Let's factorize $(5-2x)^5 = 5^5\left(1-\frac{2x}{5}\right)^5$ and then expand using the original expansion,

$$(5-2x)^5 = 3125\left(1+\binom{5}{1}\left(-\frac{2x}{5}\right)+\binom{5}{2}\left(-\frac{2x}{5}\right)^2+\binom{5}{3}\left(-\frac{2x}{5}\right)^3+\binom{5}{4}\left(-\frac{2x}{5}\right)^4+\binom{5}{5}\left(-\frac{2x}{5}\right)^5\right)$$
$$(5-2x)^5 = 3125\left(1-2x+\frac{8x^2}{5}-\frac{16x^3}{25}+\frac{16x^4}{125}-\frac{32x^5}{3125}\right)$$

Hence,

$$(5-2x)^5 = 3125-6250x+5000x^2-2000x^3+400x^4-32x^5$$

Example 6.5: Find the term which is independent of x in the expansion of $\left(x+\frac{5}{x^2}\right)^9$.

Solution. The term independent of x is the term which when simplified does not contain x. We use the alternative binomial expansion this time, with $a=x$, $b=\frac{5}{x^2}$ and $n=9$. Hence,

$$\left(x+\frac{5}{x^2}\right)^9 = x^9+9x^8\left(\frac{5}{x^2}\right)+36x^7\left(\frac{5}{x^2}\right)^2+84x^6\left(\frac{5}{x^2}\right)^3+\ldots$$

Taking the last displayed term in the above expansion, we obtain

$$84x^6\left(\frac{5}{x^2}\right)^3 = 84x^6\left(\frac{125}{x^6}\right) = 84(125) = 10500$$

6.4 Sequences

This section goes slightly beyond the material of AS Level Mathematics. Note that you are not expected to be able to calculate limits of sequences in the exam other than infinite sums of geometric progressions as discussed in the next section. For more information on limits of sequence see Abbott (2016).

As defined in the introduction, a sequence is a set of ordered numbers, such that each subsequent number can be obtained from the previous number by a certain rule. For example, one of the most famous sequences in called the **Fibonacci sequence.**

Fibonacci Sequence
This sequence can describe the growth of a population of rabbits.

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ \ldots$$

In this sequence, the rule we must follow is that the next term is determined by adding together the two previous terms. This kind of sequence is also called a recurrence relation. If F_n denotes the n^{th} term of the Fibonacci sequence, then we can write $F_n = F_{n-1}+F_{n-2}$ to express that we obtain the next term F_n by adding together the two previous terms F_{n-1} and F_{n-2}.

We classify sequences depending on their behavior.

Convergent Sequences

A sequence is said to be convergent if it eventually approaches some finite limit. For example, take the sequence defined as$u_n = \frac{1}{2^n}$ for $n \geq 1$, i.e.

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32} \cdots$$

This is a convergent sequence with limit of 0. More formally, we say that the sequence is tending towards 0 as n tends to infinity. We write this $\frac{1}{2^n} \longrightarrow 0$ as $n \longrightarrow \infty$.

Example 6.6: A sequence is defined by the recurrence relation $u_{n+1} = a - \frac{u_n}{3}$ where $u_1 = 12,\ n \geq 1$ and a is a constant. Given that $u_2 = 32$, find a. This sequence is convergent and therefore tends towards a limit L, find L.

Solution. Using the information given, substitute into the recurrence relation and solve to obtain

$$u_2 = a - \frac{u_1}{3} \Longrightarrow 32 = a - \frac{12}{3} \Longrightarrow a = 36$$

Now to analyze the behavior of the sequence when n is large, let $n \longrightarrow \infty$ .. In this limit, make the observation that term u_n and term u_{n+1} will be almost indistinguishable from the limit L, and each other. Therefore, we can write

$$L = 36 - \frac{L}{3}$$

which implies $L = 27$.

Divergent Sequences

A sequence is called divergent if it does not converge. For example, the sequence of powers of two

$$2,\ 4,\ 8,\ 16,\ 32,\ \ldots$$

clearly diverges, since the next element is always twice as big as the previous one.

Example 6.7: Find the first four terms of the sequence with n^{th} term defined as $u_n = n^2 + 3$ for $n \geq 1$. State whether the sequence converges or diverges.

Solution. By simple calculation,

$$u_1 = 1 + 3 = 4$$
$$u_2 = 2^2 + 3 = 7$$
$$u_3 = 3^2 + 3 = 12$$
$$u_4 = 4^2 + 3 = 19$$

Hence the sequence $4,\ 7,\ 12,\ 19,\ \ldots$ clearly diverges because the sequence of squares of consecutive integers grows without bounds.

6.5 Arithmetic and Geometric Progressions

Arithmetic and geometric progressions are basic but important mathematical structures that find many applications. Both types of progression start from a given start value. The next values are then calculated by either adding the same fixed number repeatedly (arithmetic progressions) or by repeated multiplication with a fixed number (geometric progression).

Arithmetic Progression

Consider the sequence $3,\ 5,\ 7,\ 9,\ 11,\ \ldots$ The first term 3 is and each subsequent term is obtained by adding 2 to the previous term. In the general case, we denote the first term by a and a common difference d and we call this kind of sequence an arithmetic progression (shortened to A.P.). A general infinite A.P. would be:

$$a,\ a+d,\ a+2d,\ \ \ldots$$

From this pattern, we can write down an expression for the n^{th} term of an A.P.:ht

$$u_n = a + (n-1)d$$

Example 6.8: How many terms are there in the sequence $-2,\ 3,\ 8,\ 13,\ 18, \ldots,\ 208$?

Solution. Take the n^{th} term formula for an A.P. with $a = -2$ and $d = 5$. We would like to know what numbered term is 208.

$$208 = -2 + (n-1)5$$
$$n - 1 = 42 \Longrightarrow n = 43$$

There are 43 terms in the sequence.

Example 6.9: The sum of the first two terms of an arithmetic series is 2. The 41st term is equal to 475. Show that the first term and common difference are -5 and 12 respectively.

Solution. Use the expression for the n^{th} term. The sum of the first two terms can be written as $a+(a+d)=2a+d$, and the 41st term is written as $a+40d$. Hence we are required to solve the simultaneous equations,

$$\begin{aligned} 2a+d &= 2 \\ a+40d &= 475 \end{aligned}$$

By the algebraic method, we obtain $d=12$ and therefore $a=-5$.

As mentioned in the introduction, one can add together the terms of a sequence to obtain a series. For an arithmetic progression with first term a and common difference d, the sum of the first n terms denoted by S_n can be written as

$$S_n=\frac{n}{2}[2a+(n-1)d]$$

Example 6.10: The first term of an arithmetic series is 1. The common difference is 6. The sum of the first n terms of the series is 7400. Find the value of n.

Solution. Substituting the known values into the S_n formula,

$$\begin{aligned} S_n &= \frac{n}{2}[2a+(n-1)d] \\ 7400 &= \frac{n}{2}[2(1)+(n-1)6] \\ 14800 &= 2n+6n^2-6n \Longrightarrow 6n^2-4n-14800=0 \end{aligned}$$

Implying the answers, $n=50$ and $n=-\frac{148}{3}$. The correct answer in our context is $n=50$ since we are looking for a number of terms and this can only be a positive integer.

Geometric Progression

Consider the sequence $2,\ 6,\ 18,\ 54,\ 162,\ ...$. Each term of the series can be obtained by multiplying the previous term by 3. In general, and in a similar fashion to the arithmetic progression, if the first term is denoted by a and the common ratio r then a geometric progression (shortened to G.P.) can be written as

$$a,\quad ar,\quad ar^2,\quad ar^3,\ ...$$

From this pattern, we can write down an expression for the n^{th} term of a G.P.:

$$u_n=ar^{n-1}$$

Example 6.11: The second term of a G.P. is 48 and the fourth term is 3. What are the possible values of the common ratio r?

Solution. From the n^{th} term formula, we can write

$$u_2=ar=48 \quad \text{and} \quad u_4=ar^3=3$$

Since $a \neq 0$, divide these equations by one another

$$\frac{ar^3}{ar} = \frac{3}{48} \Longrightarrow r^2 = \frac{1}{16} \Longrightarrow r = \pm\frac{1}{4}$$

We can add together the terms of a geometric sequence to obtain a series. The sum of the first n terms can be calculated by

$$S_n = \frac{a(1-r^n)}{1-r}$$

Geometric progressions are more interesting than arithmetic progressions due to the following example. Consider the infinite geometric series

$$2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + ...$$

The first term is 2 and the common ratio is $\frac{1}{2}$. For this G.P., we have the following partial summations

$$\begin{aligned}S_2&=3\\S_3&=3.5\\S_4&=3.75\\S_5&=3.875\\S_6&=3.9375\\S_7&=3.96875\end{aligned}$$

It sure looks like the sum S_n is approaching the number 4. In fact, by taking a sufficiently large enough value of n, we can make S_n as near to 4 as we wish. In this case, we say that S_n tends towards a limiting value of 4 as n approaches infinity. This can be written as $S_n \to 4$ as $n \to \infty$. The number 4 is called the sum to infinity of the above G.P.

The sum to infinity, denoted by S_∞ of a G.P. with first term a and common ratio r, can be written in general as

$$S_\infty = \frac{a}{1-r}$$

Notice that indeed for the above example with $a = 2$ and $r = \frac{1}{2}$,

$$S_\infty = \frac{2}{1-\frac{1}{2}} = 4$$

However, it is extremely important to remember that S_∞ only exists for an infinite geometric progression in which $|r| < 1$. In this case, the terms of the series will decrease in magnitude and the series is said to be convergent. Any series whose sum does not approach some finite limit as $n \to \infty$ is said to be divergent, i.e., for a G.P. with $|r| \geq 1$, the series is divergent.

Example 6.12: The first three terms of a geometric series are $(k+4)$, k and $(2k-15)$ respectively where $k > 0$. Show that $k = 12$ and hence calculate S_∞.

Solution. We must form an equation using the information given in the question. Notice that the ratios $\frac{k}{k+4}$ and $\frac{2k-15}{k}$ must be equal to the common ratio r and, consequently, equal to one another. Hence

$$\frac{k}{k+4} = \frac{2k-15}{k}$$

which can be rearranged to yield the quadratic equation

$$k^2 - 7k - 60 = 0 \Longrightarrow k = 12 \quad \text{or} \quad k = -5$$

Clearly $k = 12$ from the condition that $k > 0$. Now, the common ratio can be determined as

$$\frac{k}{k+4} = \frac{12}{16} = \frac{3}{4} = r$$

Since $|r| = \frac{3}{4} < 1$, the sum to infinity exists and is therefore equal to

$$S_\infty = \frac{a}{1-r} = \frac{16}{1-\frac{3}{4}} = \frac{16}{\frac{1}{4}} = 64$$

Example 6.13: Express the recurring decimal $0.2\dot{3}\dot{5}$ as a fraction in the lowest terms.

Solution. By definition $0.2\dot{3}\dot{5} = 0.2353535... = 0.2 + 0.035 + 0.00035 + ...$ which is 0.2 plus an infinite geometric series with $a = 0.035$ and $r = \frac{1}{100}$. Calculate the sum to infinity of this G.P.,

$$S_\infty = \frac{0.035}{1-\frac{1}{100}} = \frac{7}{198}$$

Therefore,

$$0.2\dot{3}\dot{5} = 0.2 + \frac{7}{198} = \frac{233}{990}$$

SUMMARY

The binomial expansion theorem provides a way to expand brackets of two terms quickly

$$(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + ... + \binom{n}{n}x^n$$

or the more general,

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n}b^n$$

The numbers $\binom{n}{r}$ can be determined via Pascal's triangle or by the closed formula

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

A sequence is an ordered list of numbers, generated by a certain rule. A series is the sum of the terms of a sequence. Sequences can be formed by a formula for the n^{th} term, or by a recurrence relation involving previous terms.

An arithmetic progression is a sequence in which successive terms differ by a constant, called the common difference. An A.P. has general term

$$u_n = a + (n-1)d$$

and the sum to n terms can be written as

$$S_n = \frac{n}{2}[2a + (n-1)d]$$

A geometric progression is a sequence in which the ratio of successive terms is a constant, called the common ratio. The general term of a G.P. is

$$u_n = ar^{n-1}$$

and the sum to n terms can be written as

$$S_n = \frac{a(1-r^n)}{1-r}$$

For an infinite geometric progression which satisfies the condition $|r| < 1$, the sum to infinity can be calculated as

$$S_\infty = \frac{a}{1-r}$$

UNIT 7

DIFFERENTIATION

STUDY GOALS

On completion of this unit, you will be able to ...

- understand the meaning of the gradient of a curve at a point.
- use the notations $f'(x)$, $f''(x)$ and $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$ to denote the first and second derivatives.
- understand how to find the derivative of x^n (where n is rational) and the derivative of composite functions using the chain rule.
- apply differentiation to find gradients of tangents.
- find the equation of tangents for a given point.

7. DIFFERENTIATION

Introduction

Sir Isaac Newton
He was a 17th and 18th century mathematician, physicist, astronomer, and inventor.

Calculus is the study of how things change, and it has two main aspects: differentiation and integration. Modern calculus was developed by **Isaac Newton** in the early 17th century. In this unit, we will study the first of these.

Differentiation studies the *rate* at which quantities change and has many applications in different scientific disciplines. For example, in physics the derivative of the displacement of a moving object with respect to time is the velocity and differentiating a second time yields an expression for the acceleration of the object. In operations research, derivatives determine the most efficient ways to transport materials from one location to another.

Our study in this unit is confined to determining the gradient of a curve at a given point and the applications to equations of linear lines.

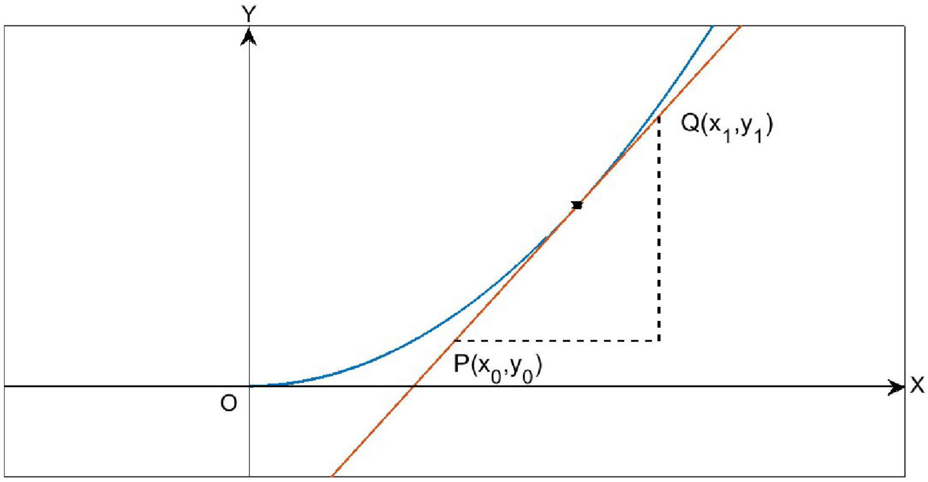
7.1 Differentiation from Definition

The gradient of a straight line is the same at all points on the line. With a curve however, the gradient will depend upon where we are on the curve. The gradient on a curve at a given point P is defined as the gradient of the tangent drawn to the curve at point P.

If we want to find the gradient of a curve at a given point, an accurate drawing of the curve, the tangent to the curve at the required point, and choosing two arbitrary points on the tangent $(x_0,\ y_0)$ and $(x_1,\ y_1)$ would allow the measurement of the gradient via the formula

$$\text{gradient} = \frac{\text{change in } y}{\text{change in } x} = \frac{y_1 - y_0}{x_1 - x_0}$$

Figure 44: Gradient of a Curve



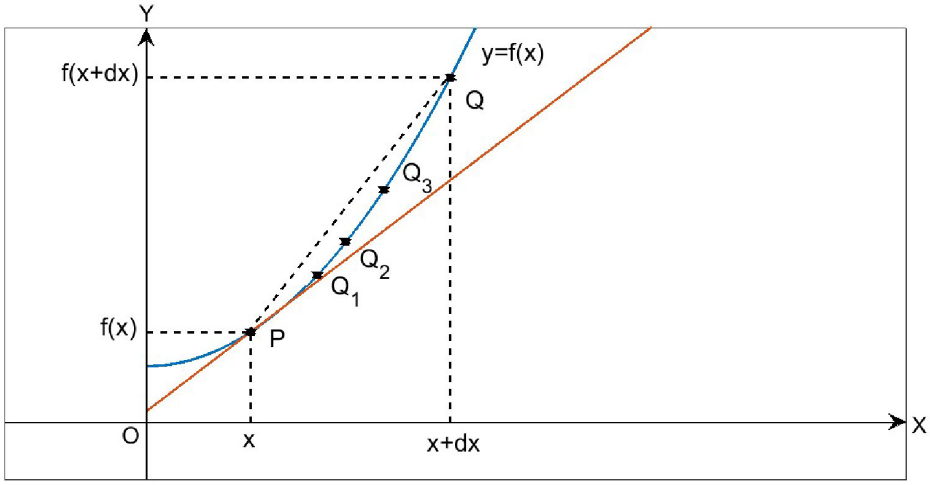
Source: Florian Pausinger, (2022).

Clearly, the described method is impractical due to the difficulty of drawing these curves and tangents to a high enough degree of accuracy. We must have an alternative method for finding the gradient accurately at a given point on a curve. This process is called differentiation. The value of the derivative of the equation of the curve at a given point (x, y) is how one defines and determines the gradient of a curve at this point. There are many rules and 'tricks' that allow us to carry out the differentiation process quickly with little effort. However, before we present these rules, it will be useful to discuss differentiation from first principles.

Suppose that we want to find the gradient of a curve $y = f(x)$ at a point $P(x,\ y)$ on the curve. We consider a second point Q lying on the curve near to point P with the x-coordinate given by $x + \delta x$ where δx is used to denote a small length in the x-direction. From the above paragraph, the gradient of the chord PQ is then given by

$$\text{gradient of chord } PQ = \frac{f(x + \delta x) - f(x)}{\delta x} \,.$$

Figure 45: Differentiation from Frist Principles



Source: Florian Pausinger, (2022).

Consider next moving the point Q closer P and notice as we pass points $Q_1,\ Q_2,\ Q_3 \ldots$ the gradients of the chords $PQ_1,\ PQ_2,\ \ldots$ give better approximations for the gradient of the curve at point P. Therefore, the gradient of the curve $y = f(x)$ at point $P(x,\ y)$ is denoted by $f'(x)$ or $\frac{dy}{dx}$, and is given by

$$f'(x) = \lim_{\delta x \to 0}\left[\frac{f(x+\delta x) - f(x)}{\delta x}\right]$$

Example 7.1: Find, from first principles, the expression for $\frac{\mathrm{dy}}{\mathrm{dx}}$ for the curve $y = x^2$.

<u>Solution.</u> Let $f(x) = x^2$. Then

$$\frac{\mathrm{dy}}{\mathrm{dx}} = \lim_{\delta x \to 0}\left[\frac{f(x+\delta x) - f(x)}{\delta x}\right]$$

$$\frac{dy}{dx} = \lim_{\delta x \to 0}\left[\frac{(x+\delta x)^2 - x^2}{\delta x}\right] = \lim_{\delta x \to 0}\left[\frac{2x\delta x + (\delta x)^2}{\delta x}\right] = \lim_{\delta x \to 0}[2x + \delta x] = 2x\,.$$

Therefore, $\frac{dy}{dx} = 2x$.

Example 7.2: Evaluate the derivative of $y = x^n$ at the point $x = 2$ from first principles.

<u>Solution.</u> Let $f(x) = x^n$. Then

$$\frac{\mathrm{dy}}{\mathrm{dx}} = \lim_{\delta x \to 0}\left[\frac{f(x+\delta x) - f(x)}{\delta x}\right]$$

$$\frac{dy}{dx} = \lim_{\delta x \to 0}\left[\frac{(x+\delta x)^n - x^n}{\delta x}\right]$$

Using the binomial expansion formula for the first term on the numerator, we obtain

$$\frac{dy}{dx} = \lim_{\delta x \to 0}\left[\frac{x^n + \binom{n}{1}x^{n-1}(\delta x)^1 + \ldots + \binom{n}{n-1}x^1(\delta x)^{n-1} + (\delta x)^n - x^n}{\delta x}\right]$$

which can be simplified to

$$\frac{dy}{dx} = \lim_{\delta x \to 0}\left[\binom{n}{1}x^{n-1} + \ldots + \binom{n}{n-1}x^1(\delta x)^{n-2} + (\delta x)^{n-1}\right]$$

by dividing each term by δx. Then let $\delta x \to 0$, and we obtain

$$\frac{dy}{dx} = nx^{n-1}$$

Hence at $x = 2$, $\frac{dy}{dx} = n2^{n-1}$.

7.2 Differentiation Rules

We would like to avoid differentiating from first principles in favor of certain fixed rules. In the following we introduce important rules for the differentiation of particular types of functions.

Differentiation of Power Functions

It can be proved from first principles (but is not part of this course book) that for $y = ax^n$ for any constant a, we have

$$\frac{dy}{dx} = anx^{n-1}$$

One usually remembers this rule in words as: 'multiply by the power and decrease the power by one'. We often need to recall standard index rules to rewrite the function in the form $y = ax^n$.

Example 7.3: Find the derivative of $\frac{1}{x^2}$.

<u>Solution.</u> Let $y = \frac{1}{x^2}$ and rewrite the function $y = x^{-2}$. By the simple rule above, we have

$$\frac{dy}{dx} = -2 \cdot x^{-2-1} = -2x^{-3}$$

Hence the derivative of $\frac{1}{x^2}$ is $-\frac{2}{x^3}$.

Example 7.4: Find the derivative of $\sqrt{x}$.

Solution. Let $y = \sqrt{x}$. Using the index laws, we once again rewrite the expression as $y = x^{\frac{1}{2}}$. Hence

$$\frac{dy}{dx} = \frac{1}{2}x^{-\frac{1}{2}}$$

or

$$\frac{dy}{dx} = \frac{1}{2\sqrt{x}}\,. \tag{7.1}$$

The established rule for power functions above combined with the following two rules allow us to differentiate many polynomial functions.

- If k is a constant and $f(x)$ is a function, then $\frac{d}{dx}[kf(x)] = k\frac{d}{dx}[f(x)]$ and
- if $f(x)$ and $g(x)$ are functions, then $\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}[f(x)] \pm \frac{d}{dx}[g(x)]$.

Example 7.5: Find the expression for $\frac{dy}{dx}$ for $y = 3x - \frac{5}{x} + \frac{6}{x^2} + 6.$

Solution. Let $y = 3x - \frac{5}{x} + \frac{6}{x^2} + 6$. Write the function as

$$y = 3x - 5x^{-1} + 6x^{-2} + 6x^0\,.$$

Then

$$\frac{dy}{dx} = 3x^0 + 5x^{-2} - 12x^{-3} + 0$$

and hence the derivative is equal to

$$\frac{dy}{dx} = 3 + \frac{5}{x^2} - \frac{12}{x^3}$$

The Chain Rule

When $y = x + 4$, we say that y is a function of x. If $y = (x+4)^3$, we say that y is a function of a function of x (add 4 to x, and then cube the resulting function).

One way to calculate the derivative of a function like $y = (x+4)^3$ is to rewrite the expression via the binomial theorem, and then differentiate term by term as shown earlier in this unit. However, one can imagine if the power on the bracket increases (to say, above 5) then this task would be a little tiresome.

Suppose that $y = u^3$ and $u = x + 4$, then using this substitution and another variable u leads to $y = (x + 4)^3$. The motivation for introducing the second variable u, which links y and x, is to allow us to differentiate these simple expressions $y = u^3$ and $u = x + 4$ in the first instance and then compose the expressions in a certain manner to yield the desired derivative $\frac{dy}{dx}$. We do this via the chain rule which can be stated for $y = f(u(x))$ as

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Example 7.6: Find $\frac{dy}{dx}$ if $y = \left(3x^2 - 3\right)^4$.

Solution. For $y = \left(3x^2 - 3\right)^4$, let $u = 3x^2 - 3$ and hence $y = u^4$. Therefore,

$$\frac{dy}{du} = 4u^3 \quad \text{and} \quad \frac{du}{dx} = 6x$$

Using the chain rule above,

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = 4u^3 \cdot 6x = 24x\left(3x^2 - 3\right)^3$$

The final answer should always be given in terms of the variable used in the question and not the introduced third variable u.

Example 7.7: Find$\frac{dy}{dx}$ if $y = \frac{1}{\sqrt{x^2 - 1}}$.

Solution. For $y = \frac{1}{\sqrt{x^2 - 1}}$, rewrite the expression as $y = \left(x^2 - 1\right)^{-\frac{1}{2}}$. Then let $u = x^2 - 1$ and hence $y = u^{-\frac{1}{2}}$. Therefore,

$$\frac{dy}{du} = -\frac{1}{2}u^{-\frac{3}{2}} \quad \text{and} \quad \frac{du}{dx} = 2x$$

Using the chain rule,

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = -\frac{1}{2}u^{-\frac{3}{2}} \cdot 2x = -\frac{x}{u^{\frac{3}{2}}} = -\frac{x}{\left(x^2 - 1\right)^{\frac{3}{2}}}$$

7.3 Equation of a Tangent

The gradient of a tangent of a curve at a given point was already discussed at the beginning of this unit. We wish to extend this discussion and determine the full equation of the tangent to the curve given a point on the curve. In elementary geometry, the equation of a straight line is given by

$$y - y_1 = m(x - x_1)$$

where m is the gradient of the line and $(x_1,\ y_1)$ is a point lying on the straight line. From the work carried out above given a point (x, y) lying on a curve, the gradient of the curve at this point will be the value of $\frac{dy}{dx}$ evaluated at this point.

Example 7.8: Find the equation of the tangent at the point $(2{,}8)$ on the curve $y = x^2 - x + 6$.

Solution. First, find an expression for $\frac{dy}{dx}$. By the rules above

$$\frac{dy}{dx} = 2x - 1$$

Therefore, at the point $x = 2$, $\frac{dy}{dx} = 3$. Hence

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - 8 &= 3(x - 2) \end{aligned}$$

which simplifies to

$$y = 3x + 2$$

Example 7.9: Find the equation of the tangent to the curve$y = \frac{10}{(x-1)^3}$ at the point $(2{,}10)$.

Solution. Following the same method as the previous example, we first need an expression for the first derivative. However, let's first rewrite the equation of the curve as $y = 10(x-1)^{-3}$ and using the chain rule (with $u = x - 1$),

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ \frac{dy}{dx} &= -30u^{-4} \cdot 1 \\ \frac{dy}{dx} &= \frac{-30}{(x-1)^4} \end{aligned}$$

Therefore, evaluating the last expression at the point $(2{,}10)$ yields $\frac{dy}{dx} = -30$. Hence

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - 10 &= -30(x - 2) \\ y &= -30x + 70 \end{aligned}$$

It is worth noting that the normal of the curve at a given point is the line perpendicular to the tangent. Therefore, if the gradient of the tangent is equal to $\frac{dy}{dx} = m$, then the equation of the normal at the point $(x_1,\ y_1)$ lying on the curve can be given as

$$y - y_1 = -\frac{1}{m}(x - x_1)$$

7.4 Second Derivative

So far, we have only spoken of 'the' derivative of a function, however we should really be more specific.

For a function $y = f(x)$, $\frac{dy}{dx}$ is called the first derivative of y and $\frac{dy}{dx}$ is another function of x . Therefore, if we differentiate $\frac{dy}{dx}$ again, we obtain the second derivative which is written as $\frac{d^2y}{dx^2}$. The application of the second derivative will be explored in an upcoming section, we first give examples on the calculation of $\frac{d^2y}{dx^2}$.

Example 7.10: A curve has the equation $y = 2x^3 + 3x^2 - 9x + 2$. Find expressions for $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$.

<u>Solution.</u> Let $y = 2x^3 + 3x^2 - 9x + 2$. Using the rules for power function and addition and subtraction, then

$$\frac{dy}{dx} = 6x^2 + 6x - 9$$

and differentiating the first derivative expression again

$$\frac{d^2y}{dx^2} = 12x + 6$$

7.5 Stationary Points and Concavity

Stationary points on a curve are those points where the gradient is equal to zero, i.e., where $\frac{dy}{dx} = 0$. The method to determine the location of a stationary point involves differentiating the curve to find $\frac{dy}{dx}$, set the expression equal to zero and solve for x. The corresponding y coordinates of the stationary points can be found by substituting the value of x back into the original equation.

Stationary Points
In a differentiable function, a point on the graph of the function where the function's derivative is zero.

Example 7.11: Find the location of the stationary points on the curve $y = x^3 - \frac{3}{2}x^2 + 2$.

Solution. Following the method laid out above, let's find an expression for the first derivative and solve when the equation equals zero. That is, $\frac{dy}{dx} = 3x^2 - 3x = 0 \Longrightarrow 3x(x-1) = 0 \Longrightarrow x = 0$ and $x = 1$. The corresponding y values are therefore

$$y = 0^3 - \tfrac{3}{2}\left(0^2\right) + 2 = 2$$

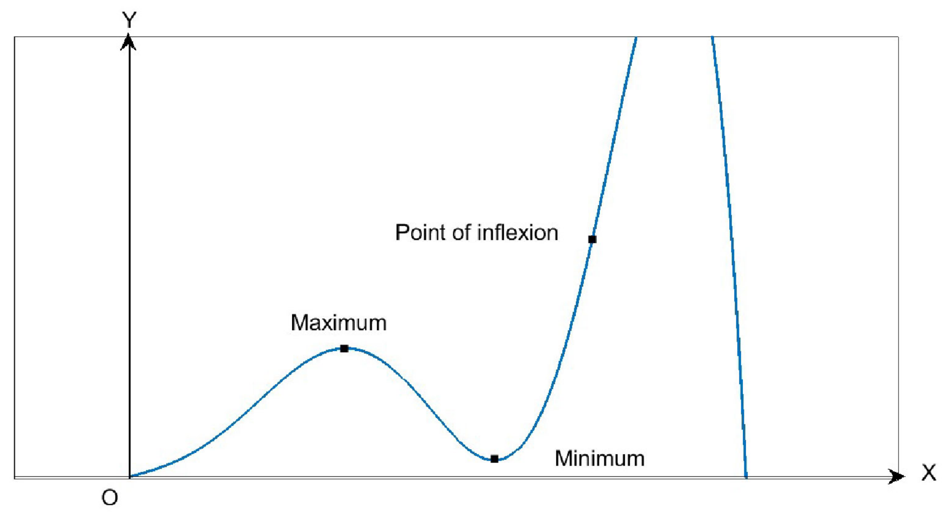
and

$$y = 1^3 - \tfrac{3}{2}\left(1^2\right) + 2 = \tfrac{3}{2}$$

Hence, the turning points are located at $(0{,}2)$ and $\left(1, \frac{3}{2}\right)$.

There are three kinds of stationary point: maxima, minima, and points of inflexion. We refer to the figure below for visual aid. It is clear to see from the image that the gradient is equal to zero (or the tangent is a horizontal line) for each of these points.

Figure 46: Stationary Points



Source: Florian Pausinger, (2022).

How do we determine which of these is the case? One approach is to manually consider the gradient each side of the location of the turning point. This method is shown in the next example.

Example 7.12: For the curve $y = x^2 + 4$, there exists a stationary point at the point $x = 0$. Classify the turning point as a maximum, minimum, or point of inflexion.

Solution. For $y = x^2 + 4 \Longrightarrow \frac{dy}{dx} = 2x$. Clearly this confirms that there is a stationary point at $x = 0$. Consider the gradient function either side of $x = 0$, i.e. at $x = -1$ and $x = 1$. Therefore

$$\frac{dy}{dx} = 2(-1) = -2 < 0$$

and

$$\frac{dy}{dx} = 2(1) = 2 > 0$$

Since the gradient is negative before the stationary point and positive after the stationary point, we can conclude that the stationary point is a minimum.

Alternatively, one can use the second derivative to determine the nature of a stationary point. We do not care about the exact value of the second derivative when it is evaluated at the point on the curve, rather the sign of the value. For a curve $y = f(x)$ and a point on the curve (x, y), we have that:

Table 3: Stationary points

$\frac{dy}{dx}$ **evaluated at** (x, y)	**Nature of stationary point**
> 0	Minimum
< 0	Maximum
$= 0$	Point of inflexion

Source: Florian Pausinger, (2022).

Note that this classification is on the premise that the third derivative is not equal to zero – however, in general, we shall not encounter such cases in this course.

Example 7.13: Find and classify the stationary points of the curve given by $y = x^3 - 3x - 2$.

Solution. To determine the location of the stationary points, we must compute the first derivative and explore when the function is equal to zero. That is,

$$\frac{dy}{dx} = 3x^2 - 3 = 0 \Longrightarrow x = 1, \ -1$$

Therefore, the stationary points are located at $(1, \ -2)$ and (-1, 0). Let's differentiate again to obtain

$$\frac{d^2y}{dx^2} = 6x$$

Then:

- At $x = 1,\ \frac{d^2y}{dx^2} = 6(1) = 6 > 0$. Hence, the stationary point$(1,\ \ -2)$ is a minimum.
- At $x = \ -1,\ \frac{d^2y}{dx^2} = 6(-1) = \ -6 < 0$. Hence, the stationary point$(-1,\ 0)$ is a maximum.

Finally, more generally and related to the second derivative we can discuss the concavity of a curve at any point. We define the concavity of a curve as the rate of change of a curve's gradient function. There are two kinds of concavity: concave upward and concave downward.

- If the second derivative of a curve is increasing, $\frac{d^2y}{dx^2} > 0$, then the curve is said to be concave upward.
- If the second derivative of a curve is decreasing, $\frac{d^2y}{dx^2} < 0$, then the curve is said to be concave downward.

It can be easily seen that at a minimum point, a curve can also be described as concave upward and similarly, at a maximum point a curve can be described as concave downward.

SUMMARY

The gradient or rate of change of a curve $y = f(x)$ is denoted by $\frac{dy}{dx}$ or $f'(x)$. The process of finding the rate of change is called differentiation.

In this unit we saw how to differentiate monomials of the form ax^n, in which a is a constant. We can use this rule together with two general rules for scalar multiplication and addition of functions to find the derivative of arbitrary polynomials, i.e.,

$$\frac{dy}{dx}\left(\sum a_n x^n\right) = \sum a_n n x^{n-1}$$

Furthermore, to differentiate a function of a function, we introduced a third variable $u(x)$ so that $y = f(u(x))$. Then we can differentiate this via the so-called chain rule:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

We also saw how to obtain the equation of the tangent and of the normal at an arbitrary point of a curve using the gradient of the curve, i.e., if the gradient of the tangent is equal to $\frac{dy}{dx} = m$, then the equation of the tangent at the point (a, b) lying on the curve is $y - b = m(x - a)$ and the normal is given as

$$y - b = -\frac{1}{m}(x - a)$$

Importantly, the first derivative can be used to locate stationary points on a curve. Stationary points are those points where the gradient is equal to zero, and there are three variants – maxima, minima, and points of inflexion.

Finally, the second derivative of a function, denoted by $\frac{d^2y}{dx^2}$ can be found by differentiating the first derivative again (or similarly, differentiating the original function two times). It can be used to find the nature of a stationary points.

UNIT 8

INTEGRATION

STUDY GOALS

On completion of this unit, you will be able to ...

- explain the relation between integration and differentiation.
- integrate basic functions and find the value of a constant of integration if given sufficient information.
- understand the difference between a definite, an indefinite, and an improper integral.
- use definite integration to calculate the area of a region between two curves or between a curve and a straight line.
- calculate volumes of revolution about a fixed coordinate axis.

8. INTEGRATION

Introduction

One main goal of integration is to describe areas or volumes enclosed by curves or surfaces. The ancient Greek astronomer Euxodus (about 370BC) provides the first documented and systematic technique to determine the value of an integral using the so-called **method of exhaustion,** a method which was later refined by Archimedes. Similar methods have been developed independently in China (around 300 AC). However, it took until the 17th century before integration was formalized in a way that allowed the discovery of one of the most important theorems of modern calculus.

Method of Exhaustion
The main idea of this method is to approximate a given area by inscribing a sequence of polygons whose areas converge to the area of interest.

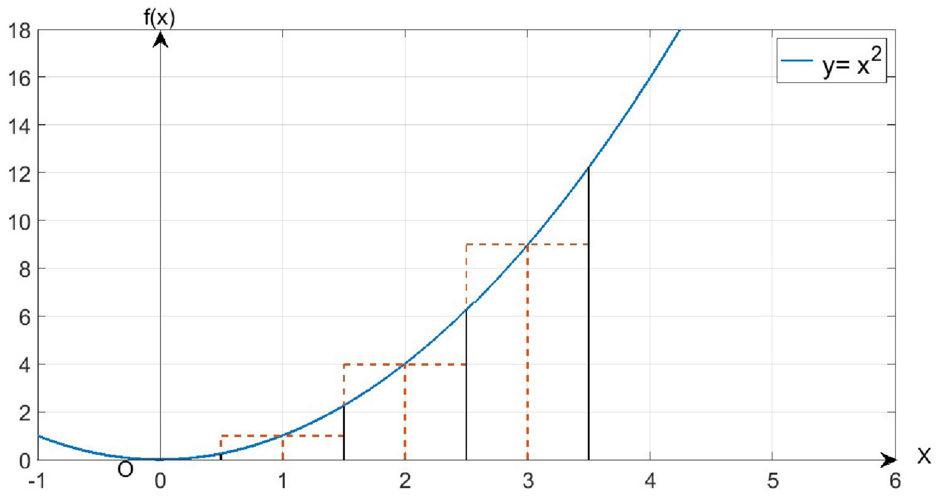
The Fundamental Theorem of Calculus was independently discovered by Leibniz and Newton. This theorem relates integration and differentiation and, in particular, can be exploited to calculate many integrals in a straightforward manner. The main discovery was that integration is, in a certain way, the reverse process of differentiation, for which it is sometimes called antidifferentiation. See (Stillwell, 1989) for more information on the history of integration.

We explore this relationship in the following, study how to evaluate basic integrals, and how to calculate areas enclosed by curves or volumes of revolution, i.e., volumes generated by rotating a curve around a fixed coordinate axis.

8.1 Integration

Imagine we are interested in the area bounded by the curve $y = x^2$, the x-axis as well as the two lines $x = 0.5$ and $x = 3.5$. A natural idea is to approximate the area by a set of rectangular strips of width δx and height corresponding to the function value $f(m)$ at the midpoint m of a given strip. Note that the area of each such strip is simply $\delta x \cdot f(m)$ and the sum of all such areas approximates the area under the curve.

Figure 47: Approximation of the Area Under a Curve



Source: Florian Pausinger, (2022).

In our concrete example we can subdivide the interval $[0.5,3.5]$ into 3 subintervals by setting $\delta x = 1$. The corresponding midpoints are then 1, 2 and 3, and the approximation $A_{\delta x}$ to the area is

$$A_1 = \sum \delta x \cdot f(m_i) = 1 \cdot f(1) + 1 \cdot f(2) + 1 \cdot f(3) = 1 + 4 + 9 = 14$$

By inspection, it is easy to see that the smaller δx, the better is the approximation of the area by the rectangles. For example, we could set $\delta x = 0.5$ giving midpoints $0.75,\ 1.25,\ 1.75,\ 2.25,\ 2.75,\ 3.25$. Evaluating the sum as before gives $A_{0.5} = 14.1875$. As $\delta x \longrightarrow 0$ this approximation will get better and better. The limit A_0 of this process is called the integral of the function $f(x) = x^2$ which is denoted with $\int$; i.e., if $\delta x \longrightarrow 0$, then

$$A \longrightarrow \int_{0.5}^{3.5} x^2 \; dx$$

in which the numbers 0.5 and 3.5 indicate that we are interested in the area between $x = 0.5$ and $x = 3.5$.

In general, if $y = f(x)$ is a function with $y \geq 0$ and if $x = a$ and $x = b$ are two lines, then the area between the graph of the function, the x-axis, and the lines is denoted as

$$\int_a^b f(x) d(x)$$

The main question becomes how to evaluate such integrals.

To illustrate the process, we present an exemplary calculation in the following. Note that you are not expected to perform such computations yourself. The following is rather meant to motivate an alternative way of calculating integrals, which will be discussed in the forthcoming sections of this unit.

In our concrete example we can proceed as follows. Let $f(x) = x^2$ and $[a, b] = [0.5{,}3.5]$. Now subdivide the interval of length 3 into n strips, i.e., $\delta x = \frac{3}{n}$. The midpoint of the k^th rectangular strip can then be obtained from the arithmetic progression

$$\frac{1}{2} - \frac{\delta x}{2} + k \cdot \delta x = \frac{1}{2} + \delta x\left(k - \frac{1}{2}\right)$$

and we get that

$$A_{\delta x} = \sum_{k=1}^{n-1} \delta x\left(\frac{1}{2} + \delta x\left(k - \frac{1}{2}\right)\right)^2 = \frac{3}{n}\sum_{k=1}^{n-1} \left(\frac{1}{4} + \frac{3}{n}\left(k - \frac{1}{2}\right) + \frac{9}{n^2}\left(k - \frac{1}{2}\right)^2\right)$$

With the help of a computer algebra system, this can be simplified to

$$\frac{3 \cdot 19}{4} - \frac{3 \cdot 49}{4n} + \frac{3 \cdot 39}{4n^2} - \frac{3 \cdot 9}{4n^3}$$

Letting $\delta x \longrightarrow 0$, which corresponds to letting $n \longrightarrow \infty$, we obtain $\frac{3 \cdot 19}{4} = 14.25$ as the value of our integral, which also fits well with our two initial approximations.

However, as you can see, it is very tedious to explicitly calculate the limit of the sum as $\delta x \longrightarrow 0$ for every single integral of interest. Fortunately, there is a powerful theorem that simplifies such calculations tremendously.

8.2 Antidifferentiation and the Indefinite Integral

The process of differentiation is concerned with understanding the rate of change of a given function $y = f(x)$. In particular, there is a streamlined method to obtain $\frac{dy}{dx}$ when y is known. Recall the rule for power functions:

$$\text{If } y = x^n, \text{ then } \frac{dy}{dx} = n \cdot x^{n-1}$$

Now, let's apply this rule (recalling that it can be applied to each summand separately in a given sum) and study the derivatives of different functions:

$$y = \frac{x^3}{3} - 3 \Longrightarrow \frac{dy}{dx} = x^2$$
$$y = \frac{x^3}{3} + 24 \Longrightarrow \frac{dy}{dx} = x^2$$
$$y = \frac{x^5}{5} + 32 \Longrightarrow \frac{dy}{dx} = x^4$$
$$y = -\frac{x^{-2}}{2} + 1 \Longrightarrow \frac{dy}{dx} = x^{-3}$$

These examples suggest the following observation: If $\frac{dy}{dx} = x^n$, then $y = \frac{1}{n+1}x^{n+1} + c$ in which c is an arbitrary constant and $n \neq 1$. Note that this implies that there are infinitely many different functions y such that $\frac{dy}{dx} = x^n$, i.e., one function for every possible value of the constant c.

The main discovery of Newton and Leibniz was to **formally prove** that antidifferentiation is virtually the same as integration. Therefore, we will not use the term antidifferentiation in the following but only speak of integration. The most important observation for us is that the process of calculating integrals can be streamlined and does not require us to explicitly calculate limits any longer.

Formal Proof
An axiom is a statement that is taken to be true. A formal proof is a finite sequence of statements such that each statement is either an axiom or follows from the previous statement via a rule of inference.

We have already seen that the symbol $\int$ is usually used to denote integration. Using our observation from above we can write

$$\int x^2 dx = \frac{1}{3}x^3 + c$$

Here, the expression $\int x^2 dx$ is called the indefinite integral of x^2 with respect to x. The integral is called indefinite since it has infinitely many solutions. We can use this new notation to rewrite the above observation as $\int x^n dx = \frac{1}{n+1}x^{n+1} + c$ in which c is a constant and $n \neq 1$. Moreover, there are two general rules for the integral of a function that is multiplied by a constant c as well as for the integration of a sum of two functions:

$$\int c \cdot f(x) dx = c \int f(x) dx$$
$$\int (f(x) \pm g(x)) dx = \int f(x) dx \pm \int g(x) dx$$

Finding the Constant of Integration

In case we have additional information about a function of interest, we can determine the constant of integration as shown in the following example.

Example 8.1: Assuming $f(-1) = 1$ and $f'(x) = 10x^4 - 4x$, find $f(x)$.

Solution. We first integrate $f'(x)$ to obtain $f(x) = 2x^5 - 2x^2 + c$. Next, we use the extra information $f(-1) = 1$ to calculate c:

$$f(-1) = 1 \Rightarrow 1 = -2 - 2 + c \Rightarrow c = 5$$

Hence, we have that $f(x) = 2x^5 - 2x^2 + 5$.

Integration of $(ax + b)^n$

Recall that

$$\frac{d}{dx}\left[\frac{1}{2 \cdot 6}(2x - 3)^6\right] = (2x - 3)^5$$

Therefore, we have that

$$\int (2x - 3)^5 dx = \frac{1}{2 \cdot 6}(2x - 3)^6 + c$$

We can generalize this observation to obtain the following rule: Let $n \neq 1$, $a \neq 0$ and let c be an arbitrary constant, then

$$\int (a\ x + b)^n dx = \frac{1}{a(n+1)}(ax + b)^{n+1} + c$$

Note that this only works for powers of affine functions of the form $a\ x + b$!

Example 8.2: Find the indefinite integral $\int \frac{20}{(1-4x)^6} dx$.

<u>Solution.</u> We have that

$$\int \frac{20}{(1-4x)^6} dx = 20 \int (1-4x)^{-6} dx = \frac{20}{-4 \cdot (-6+1)}(1-4x)^{-5} + c$$
$$= \frac{1}{(1-4x)^5} + c$$

8.3 The Definite Integral

In this section we return to our example from the first section. We calculated that the area bounded by the curve $y = x^2$, the x-axis as well as the two lines $x = 0.5$ and $x = 3.5$ is 14.25. Furthermore, we know that

$$\int x^2 dx = \frac{1}{3}x^3 + c$$

The main question is how can we relate indefinite integrals to the concrete calculation of areas.

The answer is so-called *definite integration*. Let us first observe the following:

$$\frac{1}{3}3.5^3 - \frac{1}{3}0.5^3 = 14.25,$$

i.e., evaluating the function, we calculated at the two interval bounds and then taking the difference of these two values miraculously gives the same number that we computed for the area! It turns out that this works in general. If we want to emphasize that a function should be only integrated over a given interval $[a, b]$, we write

$$\int_a^b f(x)dx$$

and call this expression the definite integral of $f(x)$ with respect to x between the limits a and b. As a general rule, definite integrals can be evaluated via

$$\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a)$$

Here the capital F denotes the antiderivative of f. This is the usual notation in this context. In analogy to indefinite integrals, we have the following rules for definite integrals:

$$\int_a^b c \cdot f(x)dx = c \int_a^b f(x)dx$$
$$\int_a^b (f(x) \pm g(x))dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx$$

Furthermore, there are two new rules to keep in mind:

$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$
$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$$

Example 8.3: Evaluate the definite integral $\int_{-1}^1 \frac{4}{(5-2x)^2}dx$.

Solution.

$$\int_{-1}^1 \frac{4}{(5-2x)^2}dx = \int_{-1}^1 4(5-2x)^{-2}dx = \left[\frac{4}{(-2)\cdot(-1)}(5-2x)^{-1}\right]_{-1}^1$$
$$= \frac{2}{3} - \frac{2}{7} = \frac{8}{21}$$

8.4 Finding Area Using Integration

As already mentioned, one main goal of integration is to calculate the area under a curve, or the area bounded by two curves.

Area Bounded by a Curve and a Coordinate Axis

Historically, the area under a curve was approximated by a series of rectangular strips. With the discovery that antidifferentiation is basically the same as integration, this procedure was much simplified. We have the general rule that:

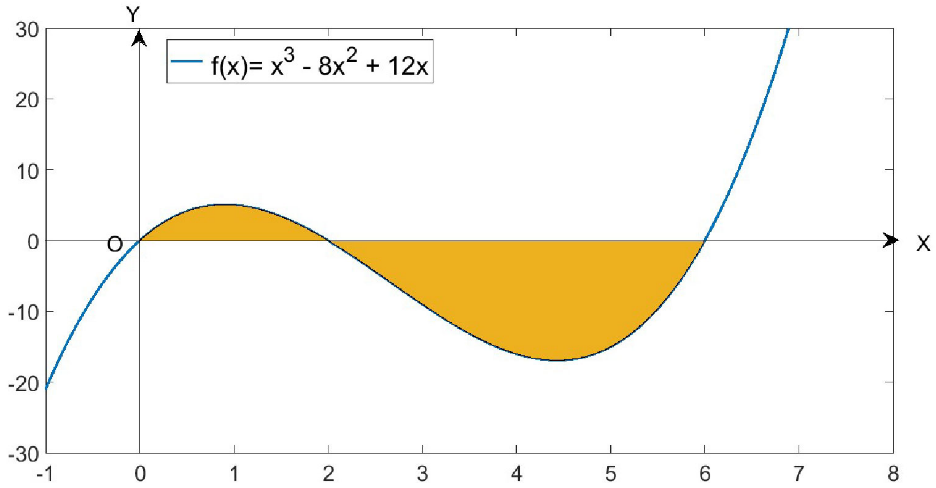
If $y = f(x)$ is a non-negative function, i.e., $y \geq 0$, then the area, A, bounded by the graph of the function as well as the x-axis and the lines $x = a$ and $x = b$ is given as

$$A = \int_a^b f(x)dx$$

Note that this formula returns a positive value, since the function values are all positive, i.e., the height values of the rectangular strips. If $y = f(x)$ is a non-positive function, i.e., $y \leq 0$, the value of the integral is negative. We have to keep this in mind when we are interested in the area of a region with a section above the x-axis and a section below the x-axis. In this case we have to evaluate each area separately as the following example shows.

Example 8.4: Find the area of the shaded region enclosed by $f(x) = x^3 - 8x^2 + 12x$ and the x-axis.

Figure 48: Illustration of Example 8.4



Source: Florian Pausinger, (2022).

Solution. We can factor the polynomial to get $x^3 - 8x^2 + 12x = x(x-2)(x-6)$. This representation shows that the function has zeros at $x = 0$, $x = 2$ and $x = 6$. The function is non-negative in the interval $[0{,}2]$ and non-positive in the interval $[2{,}6]$. Hence, in order to calculate the total area of the shaded region, we have to evaluate two integrals separately.

$$A = \int_0^2 (x^3 - 8x^2 + 12x)dx \; - \int_2^6 (x^3 - 8x^2 + 12x)dx$$

We can first compute the definite integral and then evaluate it at the boundary points:

$$\int (x^3 - 8x^2 + 12x)dx = \frac{1}{4}x^4 - \frac{8}{2}x^3 + 6x^2 + c$$

From which we get

$$A = \left[\frac{1}{4}x^4 - \frac{8}{2}x^3 + 6x^2\right]_0^2 - \left[\frac{1}{4}x^4 - \frac{8}{2}x^3 + 6x^2\right]_2^6 = \frac{20}{3} - \ 0 \ + 36 + \frac{20}{3}$$
$$= 49 + \frac{1}{3}$$

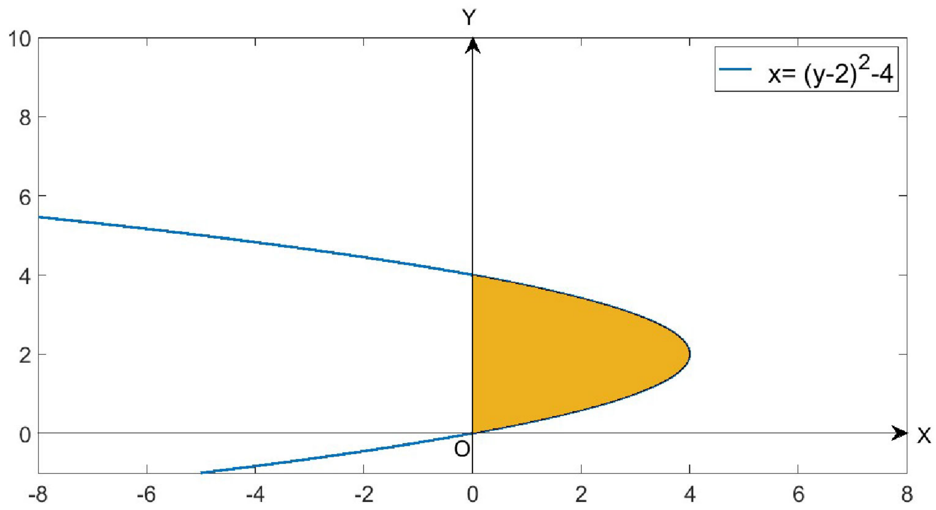
Hence, the total area of the shaded region is$49 + \frac{1}{3}$ units².

Similarly, it can be that we would like to calculate the area enclosed by a curve and the y-axis. We can proceed in an analogous fashion as the following general rule indicates:

If $x = f(y)$ is a non-negative function, i.e., $x \geq 0$, then the area, A, bounded by the graph of the function as well as the y-axis and the lines $y = a$ and $y = b$ is given as

$$A = \int_a^b f(y)dy$$

Figure 49: Area Enclosed by a Curve and y-axis



Source: Florian Pausinger, (2022).

Example 8.5: Find the area of the region bounded by the y-axis, the line $y = 9$ and by the curve $y = 2x^2 + 1$.

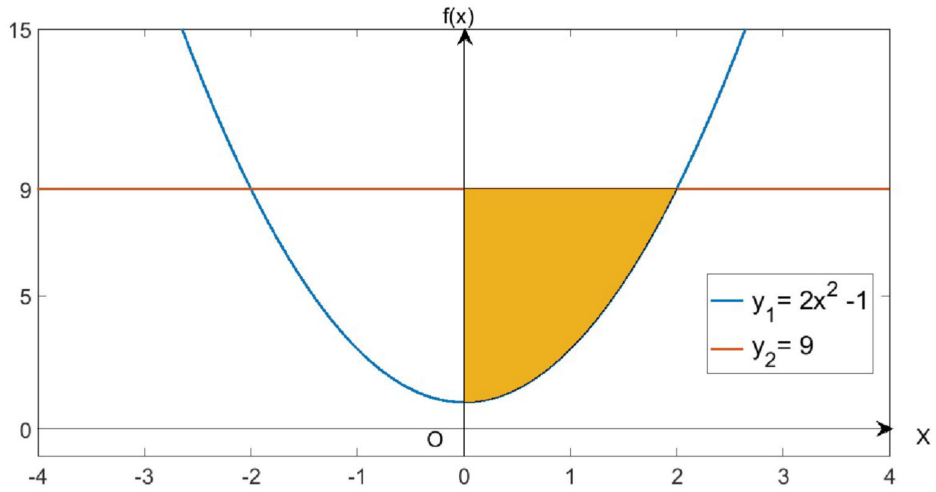
<u>Solution.</u> First, we note that $y = 1$ if $x = 0$. Next, we see that

$$y = 2x^2 + 1 \Rightarrow x = \sqrt{\frac{y-1}{2}}$$

From this we get that

$$\int_1^9 \sqrt{\frac{y-1}{2}}dy = \left[\frac{\sqrt{2(y-1)^{\frac{3}{2}}}}{3}\right]_1^9 = \frac{2^{\frac{1}{2}}8^{\frac{3}{2}}}{3} - 0 = \frac{2^{\frac{1}{2}}2^{\frac{9}{2}}}{3} = \frac{2^5}{3} = \frac{32}{3}$$

Figure 50: Illustration of Example 8.5



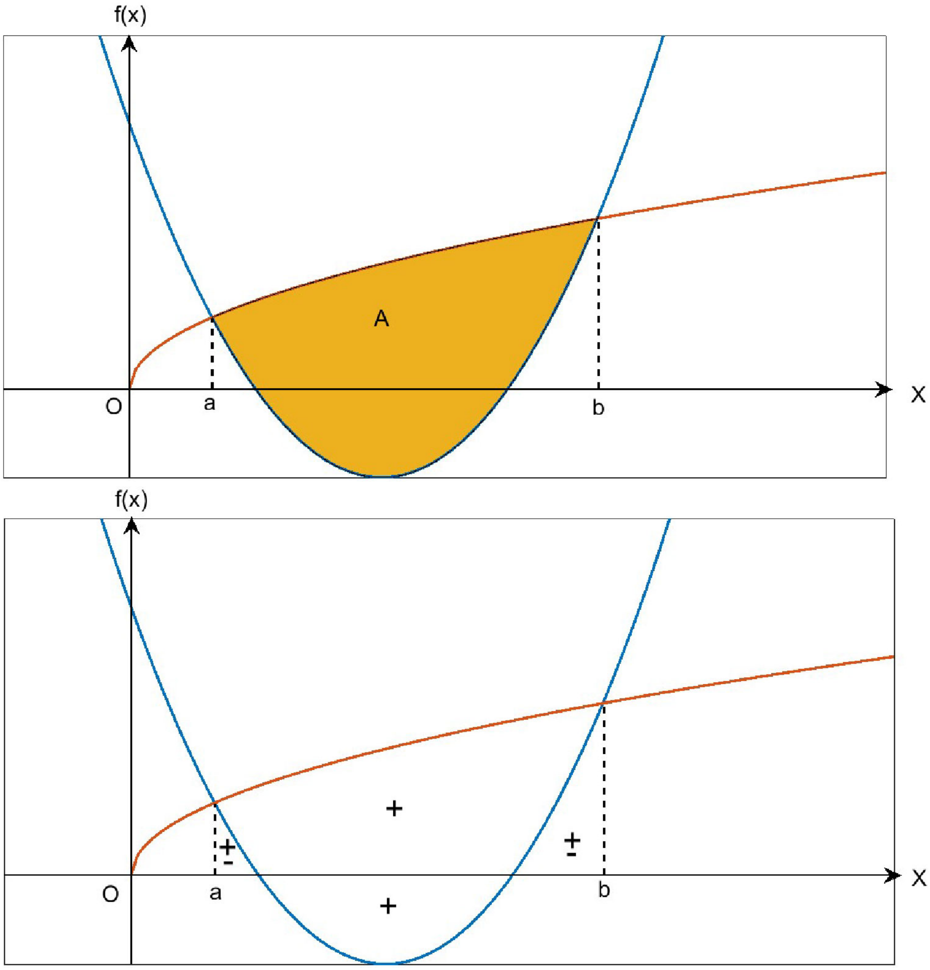
Source: Florian Pausinger, (2022).

Area Bounded by Two Curves

We can extend the method of calculating areas under curves to more complicated situations. If two functions $f(x)$ and $g(x)$ intersect in two points, say $x = a$ and $x = b$, then the area, A, enclosed between the two curves is given by

$$A = \int_a^b f(x)dx - \int_a^b g(x)dx$$

Figure 51: Illustration of Area Formula



Source: Florian Pausinger, (2022).

The figure above illustrates how the formula works. In the left image we see the area A enclosed by the red curve (corresponds to $f(x)$) and the black curve (corresponds to $g(x)$). On the right, the signs in the 4 different bounded regions correspond to the signs of the areas as calculated in the given formula. Note that the $\pm$ indicates that this area is calculated twice, once with positive sign and once with negative sign, hence it cancels and all what remains are the two regions with a positive sign which give exactly the shaded area on the left.

Example 8.6: Calculate the area enclosed by the graph of $f(x) = x^2 - 6x - 1$ and the line $g(x) = 2x - 8$.

<u>Solution.</u> First, we calculate the two points of intersection. We have that

$$f(x)=g(x) \Longrightarrow x^2-6x-1=2x-8 \Rightarrow x^2-8x+7=0 \Longrightarrow (x-1)(x-7)$$
$$=0$$

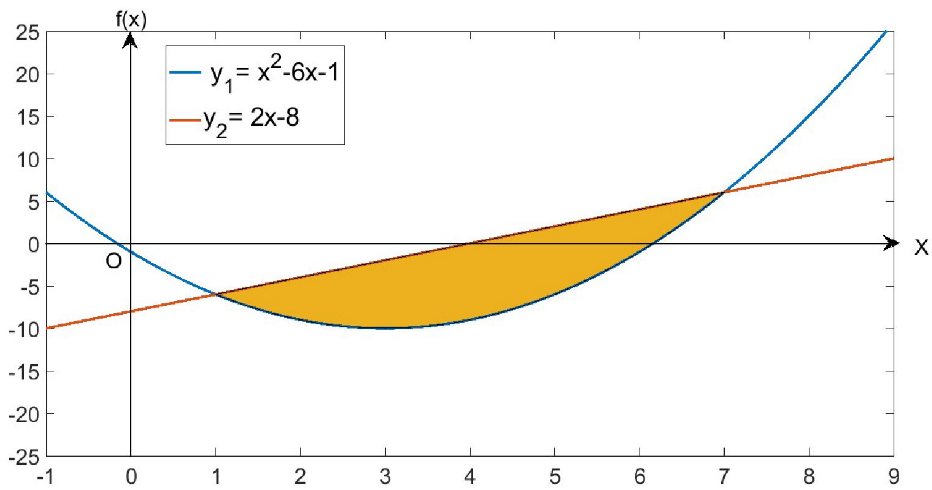
Hence, the two graphs intersect at $x=1$ and $x=7$. We apply our formula and obtain

$$A=\int_1^7 2x-8dx-\int_1^7 x^2-6x-1dx=\int_1^7 -x^2+8x-7dx$$
$$=\left[-\frac{1}{3}x^3+4x^2-7x\right]_1^7=36$$

Note that the result would be the same but negative in case we calculated $\int_1^7 x^2-6x-1dx-\int_1^7 2x-8dx$ instead. Hence, it is not a problem if we do not know whether to consider $f-g$ or $g-f$. We simply report the absolute value of the difference.

Therefore, the area enclosed by the curve and the line is 36 units[2].

Figure 52: Illustration of Example 8.6



Source: Florian Pausinger, (2022).

8.5 Improper Integrals

There is an important extension to definite integrals. If one part of a definite integral becomes infinite, we speak of an improper integral. In this section we discuss two different types of improper integrals.

Type 1

The first type of an improper integral is a definite integral in which one or both integration limits are infinite. We can evaluate integrals of the form

$$\int_{-\infty}^{a} f(x)dx,\ \int_{a}^{\infty} f(x)dx\ or \int_{-\infty}^{\infty} f(x)dx$$

by replacing the limits that are $\pm\infty$ with a finite value such as W, then evaluate the integral and finally take the limit $W \to \pm\infty$ if it exists.

Example 8.7: Evaluate the improper integral $\int_2^\infty \frac{1}{x^2}dx$.

Solution. First, we replace the upper integration bound with W and evaluate the resulting definite integral:

$$\int_2^W \frac{1}{x^2}dx = \left[-x^{-1}\right]_2^W = -\frac{1}{W} + \frac{1}{2}$$

As $W \to \infty$, we have that $-\frac{1}{W} \to 0$. Therefore, we have that

$$\int_2^\infty \frac{1}{x^2}dx = \frac{1}{2}$$

Type 2

In contrast to the first type, it can also happen that instead of the integration limits, the function to be integrated approaches $\pm\infty$ when approaching one or both integration limits. Importantly, we have to distinguish between the situation when a function is not defined at a certain value and the situation when it approaches infinity. For example, $\int_{-2}^{2} \frac{1}{x^4}dx$ is not a valid integral because $\frac{1}{x^4}$ is not defined at $x = 0$. On the other hand, $\int_0^2 \frac{1}{x^4}dx$ is an improper integral because $\frac{1}{x^4}$ approaches infinity as $x \to 0$ and the function is well-defined everywhere else in the interval over which it is being integrated. As a general rule, we can evaluate integrals of the form

$$\int_a^b f(x)dx$$

in which $f(x)$ is not defined for either $x = a$ or $x = b$ by replacing the boundary value a (or b) with W, evaluate the integral and then take the limit $W \to a$ (or $W \to b$), if this limit exists. (If the limit does not exist, then we say that the integral is undefined.)

Example 8.8: Evaluate the integral $\int_0^3 \frac{3}{x^2}dx$.

Solution. We first note that the integrand is not defined for $x = 0$. Hence, we replace the lower integration limit with W and evaluate the resulting definite integral:

$$\int_W^3 \frac{3}{x^2}dx = \left[-3x^{-1}\right]_W^3 = -\frac{3}{3} + \frac{3}{W} = \frac{3}{W} - 1$$

We see that $\frac{3}{W}$ tends to infinity for $W \to 0$, hence the integral is undefined.

Example 8.9: Evaluate the integral $\int_0^2 \frac{5}{\sqrt{2-x}} dx$.

Solution. We note again that the integrand is not defined for $x = 2$. We replace this integral limit with W and evaluate the resulting integral

$$\int_0^W \frac{5}{\sqrt{2-x}} dx = \int_0^W 5(2-x)^{-\frac{1}{2}} dx = \left[-10(2-x)^{\frac{1}{2}} \right]_0^W = -10\sqrt{2-W}$$

$$+ 10\sqrt{2}$$

We see that$= -10\sqrt{2-W} \rightarrow 0$ when $W \rightarrow 2$. Therefore, we conclude that

$$\int_0^2 \frac{5}{\sqrt{2-x}} dx = 10\sqrt{2}$$

8.6 Volumes of Revolution

In this final section, we take a brief look at volumes in three dimensions that are generated from curves that rotate about a coordinate axis. For example, think (again) of the area bounded function $f(x) = x^2$, the x-axis and the lines $x = 0.5$ and $x = 3.5$. If this area is rotated about the x-axis (here we assume a full rotation of $360°$) then a so-called solid of revolution is formed. The volume of such a solid is known as volume of revolution. The volume, V, can be approximated based on an idea that is similar to the approximation of areas in two dimensions. We approximate V by a series of cylindrical discs of thickness (or height) δx and radius y which corresponds to value of the function $f(x)$. The volume of each cylindrical disc is given by the intuitive formula

$$\text{area of disc} \cdot \text{thickness},$$

i.e., $\pi y^2 \delta x$. Hence, if we sum over all discs indexed with i then we get the following approximation

$$V = \sum \pi {y_i}^2 \delta x$$

It can be shown that this leads to a general formula for the volume: The volume, V, which is bounded between the lines $x = a$ and $x = b$, as well as the x-axis and the graph of the function $f(x)$ and which is obtained by a full rotation of the graph of the function around the x-axis, is given by the formula

$$V = \int_a^b \pi f(x)^2 dx$$

A similar formula holds when we obtain a volume by a rotation around the y-axis instead of a rotation around the x-axis. That is, the volume, V, which is bounded between the lines $y = a$ and $y = b$, as well as the y-axis and the graph of the function $f(y)$ and which is obtained by a full rotation of the graph of the function around the y-axis, is given by the formula

$$V = \int_a^b \pi f(y)^2 dy$$

Example 8.10: The graph of the function $f(x) = \frac{9}{3x+2}$ is rotated about the x-axis. Calculate the volume of revolution obtained assuming that only the piece of $f(x)$ bounded by the lines $x = 1$ and $x = 2$ is considered.

Solution. According to the above formula we have that

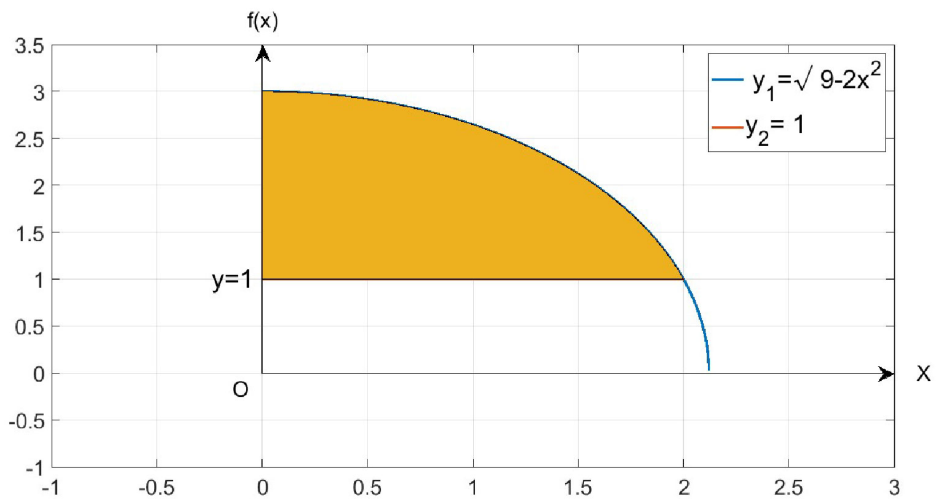
$$V = \int_1^2 \pi \left(\frac{9}{3x+2}\right)^2 dx = \int_1^2 \pi\ 81(3x+2)^{-2} dx = \pi\left[\frac{81}{-3}(3x+2)^{-1}\right]_1^2$$
$$= \pi\left(-\frac{27}{8} + \frac{27}{5}\right) = \frac{81\pi}{40}$$

Hence, the volume of revolution is $\frac{81\pi}{40}$ units³.

Example 8.11: The graph of the function $f(x) = \sqrt{9 - 2x^2}$ is rotated about the x-axis.

a) Calculate the volume of revolution obtained assuming that only the piece of $f(x)$ bounded by the lines $x = 0$ and $x = 2$ is considered.
b) Now consider that only the shaded region is rotated.

Figure 53: Illustration of Example 8.11.



Source: Florian Pausinger, (2022).

Solution. (a) First, we apply our formula to obtain

$$V = \int_0^2 \pi(9 - 2x^2)dx = \pi\left[9x - \frac{2x^3}{3}\right]_0^2 = 18\pi - \frac{16\pi}{3} = \frac{38\pi}{3}$$

(b) In a second step we observe that rotating the line $y = 1$ about the x-axis generates a cylinder. We can calculate the volume either using our formula or apply the formula for the volume of a cylinder directly. We have for a cylinder or length 2 and radius 1 that

$$\text{volume cylinder} = \pi \cdot 1^2 \cdot 2 = 2\pi$$

Therefore, the volume of the object obtained by rotating the shaded area is

$$\frac{38\pi}{3} - 2\pi = \frac{32\pi}{3}$$

SUMMARY

One of the main goals of integration is to calculate areas and volumes. In this unit we saw that integration is the reverse of differentiation which simplifies the calculation of many integrals. We can summarize this observation as

$$\text{If } \frac{d}{dx}[F(x)] = f(x), \text{ then } \int f(x)dx = F(x) + c$$

for a constant c. Based on this fundamental insight, we saw how to integrate functions of the form $(ax+b)^n$. Furthermore, we introduced so-called definite integration as an important tool for the calculation of areas, i.e.,

$$\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a)$$

Importantly, there are basic rules for indefinite and definite integration of sums of two functions as well as of a product of a constant with a function.

It can happen that either one of the integration limits is ∞ or that the function is ∞ when evaluated at one of the integration limits. In this case we speak of an improper integral and saw various ways to evaluate such integrals.

Finally, we saw how to calculate volumes of revolution obtained via rotating a curve around the x- or y-axis. For example, the volume, V, which is bounded between the lines $x = a$ and $x = b$, as well as the x-axis and the graph of the function $f(x)$ and which is obtained by a full rotation of the graph of the function around the x-axis, is given by the formula

$$V = \int_a^b \pi f(x)^2 dx$$

UNIT 9

REPRESENTATION OF DATA

STUDY GOALS

On completion of this unit, you will be able to ...

- illustrate and interpret numerical data.
- choose a method for displaying different kind of data.
- determine different measures of central tendency such as the mode, the mean, or the median.
- explain and use different measures of variation such as the range, the interquartile range, or the standard deviation.
- calculate and interpret the variance of a given data set.

9. REPRESENTATION OF DATA

Introduction

There are many ways in which data is being collected nowadays. Understanding data is crucial to understanding questions whose answer is not immediately obvious. The methods presented in this unit aim to improve our data-handling skills.

Importantly, there are two fundamentally different types of data. Categorical data are generally non-numerical and described by words, for example nationalities. Quantitative data on the other hand take numerical values and are further distinguished into discrete and continuous data. The important distinction is that discrete data are in general counted while continuous data stem from measurements that are given to a certain degree of accuracy.

As an example of discrete data, we can ask for the number of appearances of a certain letter in an English text. Or for the values of the coins in our wallet. Note that Euro coins can have a value of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1 and 2 Euros. Even if the individual values are rational numbers rather than integers, we still speak of discrete data as there are only 8 different possibilities for the value of any given coin.

An example of continuous data is measurements such as body height, which generally falls into the interval from 50–250 cm, or time needed to finish a task, such as a marathon. Here, depending on the accuracy of the measure, we can, in principle, measure every possible value within the (limited) range of the data.

The aim of this unit is to introduce different methods to visualize and interpret discrete and continuous data.

9.1 Stem-and-Leaf Diagrams

The most important aim in the representation of data is to present the data in a way that reveals certain properties that are otherwise hard to grasp. A stem-and-leaf diagram is a type of table that can be used to display small amounts of discrete data.

Each row in the table of a stem-and-leaf diagram corresponds to one stem together with its leaves. In other words, each row forms a class of values. In addition, the diagram comes with a key that explains what the values in the diagram mean.

As an example, assume that we wish to visualize the following set of 12 numbers which could correspond to the number of rainy days per month in an arbitrary city:

$$21, 15, 12, 9, 5, 4, 1, 3, 7, 10, 12, 17$$

For ease of representation, we decide that numbers 0, …, 9 should be represented with a leading 0. Next, we need to specify what the stems and leaves represent. In our example, the leaf should contain the last digit of a number, while the stem contains its leading digit. This implies that we group the data into classes of width 10. The corresponding stem-and-leaf diagram looks as follows:

Figure 54: A Stem-and-Leaf Diagram

Stem	Leaf
0	1,3,4,5,7.9
1	0,2,2,5,5
2	1

Key: 1|2
represents a month with 12 rainy days

Source: Florian Pausinger, (2022).

Back-to-back stem-and-leaf diagrams are important extensions of stem-and-leaf diagrams such as the one above. They are useful if we want to compare two data sets with each other. Imagine we count the rainy days in another city as well and obtain the following numbers

$$9{,}10{,}13{,}12{,}14{,}15{,}15{,}13{,}12{,}10{,}8,\ 11$$

With the same assumptions as above, we can now extend the diagram to the right. While the left column contains the leaves corresponding to the rainy days in the first city, we can now write the leaves corresponding to the rainy days in the second city into a new column on the right:

Table 4: A Back-to-back Stem-and-Leaf Diagram

City 2	Stem	City 1
9,8	0	1,3,4,5,7,9
5,5,4,3,3,2,2,1,0,0	1	0,2,2,5,7
	2	1

Source: Florian Pausinger, (2022).

This second diagram suggests the conclusion that the number of rainy days is relatively constant throughout the year in City 2, while City 1 seems to experience dry weather most of the year with a few, very rainy months.

Example 9.1: There are two bridges over the river that passes through a city. Over a 10-day period data were collected on the number of cars driving over the two bridges. The results are presented in the following back-to-back stem-and-leaf diagram.

1. How many more cars used Bridge 1 than Bridge 2?
2. The toll on Bridge 1 is 1.80 Euros, the toll on Bridge 2 is x Euros. Overall, the collected tolls on Bridge 2 were 79.4 Euros more than on Bridge 1. Find the value of x.

Table 5: Back-to-back Stem-and-Leaf Diagram for Example 9.1

Bridge 2	stem	Bridge 1
9,8	3	2,3,4,5,
5,4,3,3,2,2,1,0,0	4	0,1,2,5,7
4,2,0	5	1,3,4
	6	3,7

Source: Florian Pausinger, (2022).

Solution. If we sum all values on the right, we get that 637 cars used Bridge 1, while only 613 used Bridge 2. Hence, there were 24 cars more driving over Bridge 1 than over Bridge 2.

On Bridge 1,

$$637 \cdot 1.8 = 1146.6$$

Euros of tolls were collected. Hence, we get that

$$x = \frac{1146.6 + 79.4}{613} = 2$$

Hence, the toll on Bridge 2 is 2 Euros.

9.2 Histograms and Cumulative Frequency Graphs

In contrast to discrete data, continuous data is not limited to a finite number of values but can, in principle, take every value in an interval. However, note that in practice every measurement device has limited accuracy, hence continuous data will usually be rounded to a pre-specified number of decimal places etc.

Different classes of values are defined via their lower and upper boundary values and are often identified with their class mid-value. The width of a class is defined as the difference between the upper and lower boundary value of the class.

As an example, think of measuring the temperature on different days of the year. Assume that we measure temperature with a thermometer that displays the temperature in degree Celsius to one decimal place. We can now define different classes. Note that the underlying data is continuous, and that the thermometer is rounding its results. Hence, we have to make sure we do not leave gaps between our classes. For a given temperature t, we can define classes such as $15 \leq t < 20$ and $20 \leq t < 25$, i.e., we avoid gaps of length 0.1 between classes corresponding to the given degree of accuracy of our measurements, even if we cannot measure a temperature of 19.96 degrees.

Histograms

We often use a histogram to display continuous data. Histograms are among the **seven basic tools of quality control**. Given unstructured continuous data, we first group the data into a so-called grouped frequency table. In other words, we define different classes and count how many values fall into each class.

Seven Basic Tools of Quality Control
These tools refer to a set of seven graphical methods which are believed to be most useful in troubleshooting issues related to quality control.

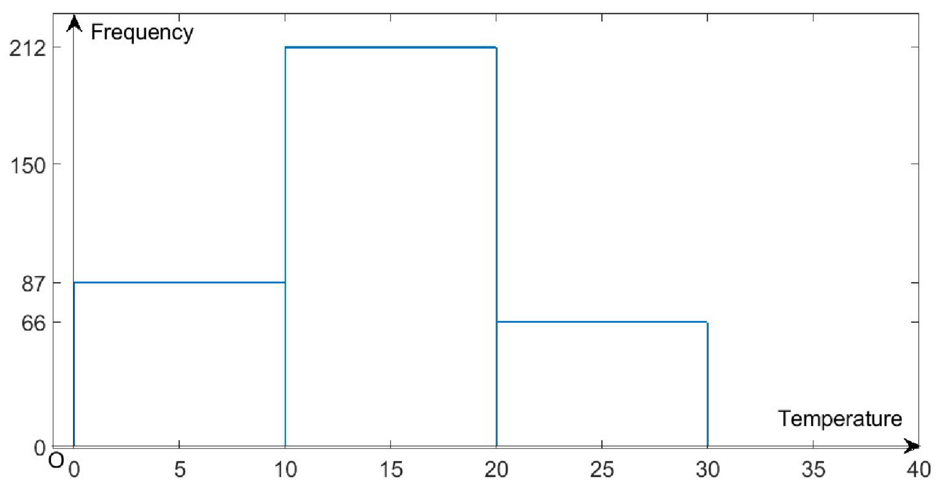
Table 6: Table: A grouped frequency table

Temperature (degrees Celsius)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$
No. days (frequency f)	87	212	66

Source: Florian Pausinger, (2022).

In a next step we can illustrate the data in a frequency diagram. The diagram consists of three columns of equal width corresponding to 10 degrees while the height of the columns is equal to the class frequency.

Figure 55: A Frequency Diagram



Source: Florian Pausinger, (2022).

The grouped frequency table allows us to calculate the areas of the columns in the frequency diagram. We simply multiply the frequency with the width to obtain the area. Importantly, since the width is the same for all columns, we see that the ratio between areas of the columns is the same as the ratio of the frequencies. This is an important property of every histogram: The ratio of the areas of the columns is the same as the ratio of the frequencies.

It may happen, that the classes do not have equal widths. However, we would like to preserve the above property. In this case, we need to apply an extra trick. In this case we introduce the so-called frequency density, which measures frequency per unit interval. Very often this unit interval is assumed to be of length 1; in such cases we get that

$$\text{Frequency density} = \frac{\text{class frequency}}{\text{class width}}$$

In a histogram, the vertical axis is typically labelled frequency density, and we can compare the relative frequencies of classes by comparing column areas.

Example 9.2: Different types of pineapple are grouped according to their masses into two classes.

1. Please illustrate the data in a histogram.
2. Estimate the number of pineapples with masses between 1 and 2 kg.

Table 7: Masses of Pineapples

Mass (in kg)	$1 \leq m < 1.5$	$1.5 \leq m < 3$
No. fruits (frequency f)	34	81

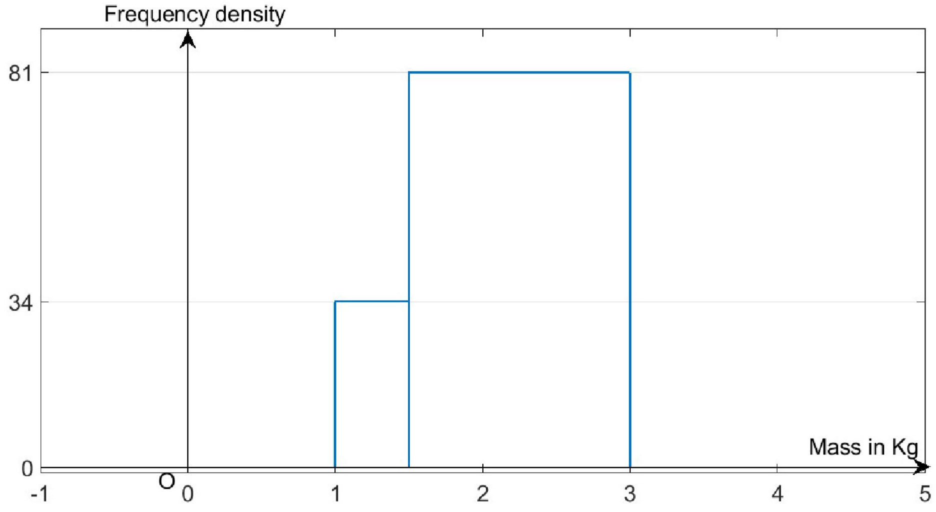
Source: Florian Pausinger, (2022).

Solution. First, we calculate the frequency densities for the two classes. We get that

$$\frac{34}{0.5} = 68, \ \frac{81}{1.5} = 54$$

Hence, we can draw the following histogram:

Figure 56: Histogram for Example 9.2



Source: Florian Pausinger, (2022).

For the second part, we note that all pineapples in the first class have mass between 1 and 2 kg, while (probably) only a fraction of the pineapples in the second class fall into this interval. Our best guess is to estimate the number of pineapples from the second class by splitting the class into two subclasses, i.e., $1.5 \leq m < 2$ and $2 \leq m < 3$ and calculating the frequency of the first subclass:

$$\text{Frequency} = \text{width} \cdot \text{frequency density} = 0.5 \cdot 54 = 27$$

Hence, our estimate is that in total 34+27=61 pineapples have a mass between 1 and 2 kg.

Cumulative Frequency Graphs

Another common way to represent continuous data is to use a cumulative frequency graph. As the name suggests, cumulative frequency is the total frequency of all values less than a given value. When given grouped data, we plot cumulative frequencies (denoted as cf) against the upper class boundaries for all intervals. The points are usually joined by straight line segments to give the cumulative frequency graph.

For example, assume that we have 10 values below 5, 31 values below 10 and 40 values below 15. We would then plot the three points (5,10), (10,31) and (15,40).

Note that we can also join the points with other curves than straight lines. However, the convention to use straight lines has the advantage to be well defined (or unambiguous) and makes sure everyone can produce the same graph from the same data.

Example 9.3: The following table shows the number of days in which the temperature at noon in a fantasy city falls into a certain interval.

1. Draw the cumulative frequency polygon.
2. Use the diagram to estimate the number of days in which the temperature is less than 15 degrees.

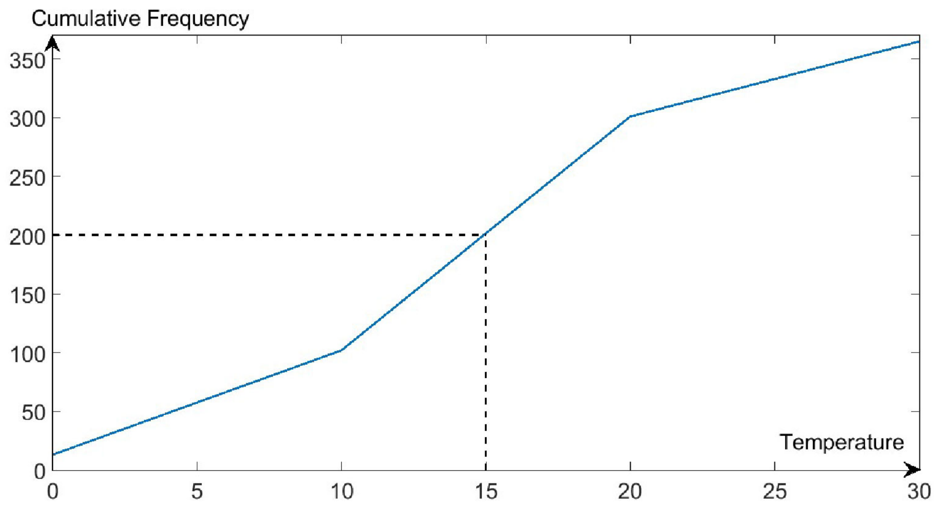
Table 8: A Grouped Frequency Table for Example 9.3

Temperature (degrees Celsius)	$-10 \leq t < 0$	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$
No. days (frequency f)	13	89	200	63

Source: Florian Pausinger, (2022).

<u>Solution.</u> To obtain the frequency polygon we have to add the frequencies. There are 13 days with temperature < 0, there are 102 days with $t < 10$, 302 days with $t < 20$ and 365 days with $t < 30$. Therefore, we have to plot the points (0,13), (10,102), (20,301) and (30,365).

Figure 57: Cumulative Frequency Polygon for Example 9.3



Source: Florian Pausinger, (2022).

To answer the second question, we use the diagram to check the cumulative frequency at temperature 15, which is 202.

9.3 Measures of Central Tendency

When given a data set, we are often interested to describe this data set with one significant value. Different so-called measures of central tendency have been developed to give an idea of the typical or average value of a given data set. In this section we introduce and discuss three such measures, i.e., the mode, the mean, and the median, which can all be used as representatives of a typical value in a data set.

The Mode and the Modal Class

The mode of a set of data is the most common value of this data set. Similarly, if we are given a set of grouped data, then the modal class is the class with the highest frequency density. Note that mode and modal class are neither unique, nor is there any reason why they have to exist at all for a given data set.

Example 9.4.: Find the modal class of the 120 exam results given in the following table.

Table 9: Frequency Table for Example 9.4

Mark range	100-70	69-60	59-50	49-40	39-0
No. of students (frequency)	17	32	29	23	19

Source: Florian Pausinger, (2022).

<u>Solution.</u> We have to find the class with the highest frequency density. We calculate the frequency densities as follows:

Table 10: Intervall Table for Example 9.4

Interval	No. of students	Width	Frequency Density
$100 \geq x > 69.5$	17	30.5	$17/30.5 \approx 0.56$
$69.5 \geq x > 59.5$	32	10	$32/10 = 3.2$
$59.5 \geq x > 49.5$	29	10	$29/10 = 2.9$
$49.5 \geq x > 39.5$	23	10	$23/10 = 2.3$
$39.5 \geq x > 0$	19	39.5	$19/39.5 \approx 0.48$

Source: Florian Pausinger, (2022).

Hence, we see that the modal class is 69-60.

The Arithmetic Mean

Arithmetic Mean
Apart from the arithmetic mean, the geometric mean is a second important measure that is often considered in a mathematical context. It is defined as the **th** root of the product of the n numbers, i.e., $\sqrt[n]{x_1 \cdot \dots \cdot x_n}$

The **arithmetic mean**, often simply referred to as the mean or the average, is defined as the sum of values divided by the number of values. In mathematics, the upper-case Greek letter 'sigma' is usually used to write sums, i.e.,

$$x_1 + x_2 + \dots + x_n = \sum_{i=1}^{n} x_i$$

If it is clear from the context that we sum over a given set of data x, we may simply write $\sum x$ to denote the sum over all elements of the set x. The mean of the data set is often denoted as $\overline{x}$ and, therefore, we can write the mean as

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

This works in a similar fashion for grouped data. However, here we assume that all data points in a given class have the same data value. Hence, the sum of data values is the frequency of a given class times the representative data value of members of that class, while the number of data values is the sum of frequencies, i.e.,

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

Example 9.5: Students sitting an exam got the following marks:

$$88,\ 43,\ 56,\ 65,\ 71,\ 55,\ 61,\ 37,\ 45,\ 60,\ 50,\ 69$$

Find the average mark.

<u>Solution.</u> We have to sum all the values and divide by the number of values to obtain the average. The sum of the 12 values is 700, so the average is

$$\overline{x} = \frac{700}{12} \approx 58$$

Example 9.6: In Austria, the exams of students are marked from 1 to 5 in which 1 is the best mark and 5 is the fail mark. Calculate the average mark for the following mark distribution:

Table 11: Frequency Table for Example 9.6

Mark	1	2	3	4	5
No. of students (frequency)	12	31	53	17	8

Source: Florian Pausinger, (2022).

Solution. To get the number of students who sat the exam, we have to sum all frequencies. We get $\sum f = 12 + 31 + 53 + 17 + 8 = 121$. The sum of data values is obtained by first multiplying each mark with its frequency and then summing over the five values:

$$\sum fx = 12 \cdot 1 + 31 \cdot 2 + 53 \cdot 3 + 17 \cdot 4 + 8 \cdot 5 = 341$$

Hence, the average mark is $\overline{x} = \frac{341}{121} \approx 2.8$.

Importantly, in general, if data is represented in grouped frequency tables, or illustrated in a histogram, we may lose information about the raw data. For example, it is not possible to calculate the mean exactly for data given as in the frequency table in Example 9.4. The best thing we can do is to use mid-values to represent the values in each class and calculate an approximation for the mean. However, note that this approximation can be far from the true average, for example, if many marks are close to the lower class boundaries (because professors tried to assign the better of two marks to students wherever possible).

Finally, it sometimes makes sense to transform a data set before calculating its mean. Adding a fixed positive or negative constant to all elements of a data set produces a set of so-called coded data. Note that

$$\begin{aligned}\sum (x-b) &= (x_1 - b) + (x_2 - b) + \ldots + (x_n - b) = x_1 + x_2 + \ldots + x_n - n \cdot b \\ &= \sum x - \sum b\end{aligned}$$

With this insight, it is easy to see that the following formulas hold:

$$\begin{aligned}\overline{x} &= \frac{\sum (x-b)}{n} + b \\ \overline{x} &= \frac{\sum (x-b)f}{\sum f} + b\end{aligned}$$

Example 9.7.: The age of an individual man is denoted by m, while the age of an individual woman is denoted with w. Exactly 3 years ago, the sum of ages of the 11 male players of a volleyball club was 243, i.e., $\sum (m-3) = 243$. On the other hand, in exactly 4 years, the sum of the ages of the 7 female members of the club will be 303, i.e., $\sum (w+4) = 303$. Find the mean age today of the 11 male players, the 7 female players as well as the combined mean age of both male and female players.

Solution. To get the mean age of the male players we have to add 3 to the known mean:

$$\overline{m} = \frac{243}{11} + 3 = \frac{276}{11} \approx 25$$

Similarly, to get the mean age of the female players, we have to subtract 4 from the known mean:

$$\overline{f} = \frac{303}{7} - 4 = \frac{275}{7} \approx 39$$

To get the combined mean it is important to note that it is not enough to simply add the two means and divide by 2. The reason is that there is a different number of players in each class. The correct way is to combine the two sets into a new set with 11+7=18 elements and to compute the mean of this new set:

$$\frac{276+275}{11+7} = \frac{551}{18} \approx 31$$

The Median

The third measure of central tendency to be discussed here is the median. The median is defined as the value in the middle of an ordered set of data. More precisely, if the data set contains n values, the median is the $\left(\frac{n+1}{2}\right)th$ value of the ordered set. If n is an odd number, then it is clear which value is the median. If n is an even number, however, we calculate the median as the mean of the two elements with index closest to $\left(\frac{n+1}{2}\right)$.

Importantly, in large data sets, when the values are grouped, we often cannot access individual data values. In this case we can at least estimate the median by reading its value from a cumulative frequency graph. The median can be estimated by the value whose cumulative frequency is equal to half of the total frequency.

When to Choose which Average

The different measures introduced in this section all come with different properties that make them more useful in certain situations than others. Before deciding which average to choose to describe a data set it is good to reflect on the different features of these measures.

Finally, if a data set is not symmetrical, it is called skewed. It can be positively skewed (if the curve's longer tail points towards the right on a standard coordinate axis) or negatively skewed (if the curve's longer tail points towards the left on a standard coordinate axis). As a rule of thumb, mode<median<mean if the data are positively skewed with reversed greater signs if the data are negatively skewed.

9.4 Measures of Variation

It is not too difficult to construct two data sets such that their mode, mean, and median are the same, while their individual values are very different. Therefore, the measures of central tendency only give a partial description of a data set, but do not say much about the spread of the values. For this reason, data sets are often described by a measure of central tendency together with a measure of variation. In this section we introduce two common measures of variation, i.e., the range and the interquartile range, while the next section contains another such measure, i.e., the standard deviation.

The Range

The range is defined as the numerical difference between the two extreme values of a data set. Depending on the context, it may be more informative to state the minimal and maximal value, than their difference. In the context of grouped data, we can approximate the range by using the lower and upper boundary values of the data.

The Interquartile Range and Percentiles

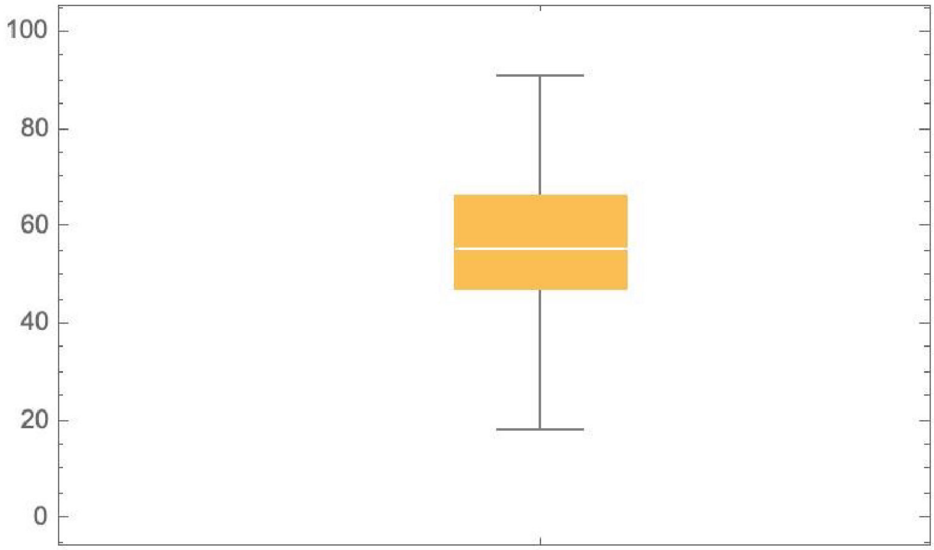
While the median splits a data set into two parts, the lower quartile and upper quartile together with the median split a data set into four parts, such that each part contains an equal number of data points. Usually, these three separators are denoted with Q1 (lower quartile), Q2 (median) and Q3 (upper quartile). Importantly, the interquartile range is the difference between the upper and lower quartile and gives the range of the middle half of the values. The advantage of the interquartile range over the range is that it is unaffected by outliers (very small or large values compared to the bulk of values). For ungrouped data can determine the quartiles by finding the median first, because then the quartiles are simply the medians of the first and second half of the data. For grouped data we can use the cumulative frequency graph to estimate the lower and upper quartiles similarly to how we estimated the median.

Box-and-Whisker Diagrams

The box-and-whisker diagram or box-and-whisker plot is a method to illustrate key features of a data set such as spread and skewness using the quartiles of the data. Given a data set, the box-and-whisker diagram consists of two parts as the name suggests. The box is an empty rectangle drawn from Q1 to Q3 which contains one bar indicating Q2. In addition, there are lines (the whiskers) that connect the smallest value of the data set with the side of the rectangle corresponding to Q1 and the largest value with the side corresponding to Q3. Thus, a box-and-whisker plot displays the range of the data, as well as its quartiles and its mean. This approach also gives a good idea about the skewness as it can be seen from any asymmetry of the box around the bar of the median.

The following plot summarizes the exam result of a particular exam in which 40 students participated. The average mark was 54, while the median is 55.5 as indicated by the white line in the box. The range is from 91 (highest mark) to 18 (lowest mark) as indicated by the two black whiskers.

Figure 58: A Box-and-Whisker Plot



Source: Florian Pausinger, (2022).

9.5 Variance

Imagine we have a data set of mean 0 containing two values -10 and 10. Intuitively, this set has a larger spread than a set of two points consisting of -1 and 1 which also has mean 0. It is often important to quantify the spread of data around its mean. To do this we introduce a new quantity, called the variance.

The idea is to calculate the difference of each data value and the mean, then take the square of it and divide by the number of elements in the set. Apart from algebraic reasons, taking the square also has the advantage that the outcome is always positive and, thus, ensures, that we do not have unwanted cancellations in our measure of the spread. We have that

$$Var(x) = \frac{\sum (x - \overline{x})^2}{n}$$

Now imagine we take measurements in a certain unit such as meters, the variance returns a value in square meters. Thus, one often considers the square root of the variance as a measure of the spread. This notion is known as standard deviation, i.e.,

$$\text{Standard deviation of } x = \sqrt{Var(x)} = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

The reason why the variance and standard deviation are popular measures of the spread stems (partly) from the fact that there is a simplified formula which makes the calculation much easier than it seems at first. It can be shown that:

$$Var(x) = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

Consequently, we can find the variance and standard deviation from n, $\sum x$ and $\sum x^2$, i.e., the number of values, the sum of the values and the sum of the squares of the values.

A small standard deviation means that most of the values are close to the mean, whereas a large standard deviation means that the values are widely spread. However, note that taking the square of the differences takes away any information about the skewness of the data. Hence, we do not know whether most of the data is smaller or larger than the mean or whether the data set is perfectly symmetric.

In comparison to the interquartile range and a box-and-whisker plot, we note that the standard deviation is much simpler to calculate, the data values do not have to be ordered and we take all values into consideration. However, in contrast to the interquartile range, the standard deviation is much more affected by extreme values and, because of the properties of taking squares, it gives greater emphasis to large deviations than to small deviations.

Example 9.8: Find the standard deviation for the seven numbers 2, 8, 13, 15, 21, 24, 29. Which of the seven numbers are more than one standard deviation away from the mean?

<u>Solution.</u> We first calculate the variance of the numbers. For this we calculate the sum of all values as well as the sum of squares:

$$\sum x = 2 + 8 + 13 + 15 + 21 + 24 + 29 = 112$$
$$\sum x^2 = 4 + 64 + 169 + 225 + 441 + 576 + 841 = 2320$$

Hence, we have that

$$Var(x) = \frac{2320}{7} - \frac{112^2}{49} = \frac{528}{7}$$

and, therefore, we get the standard deviation $\sqrt{\frac{528}{7}} \approx 8.68$, while the mean of the data is $\frac{112}{7} = 16$. We have that $16 - 8.68 = 7.31$ and $16 + 8.68 = 24.68$. Hence, 2 and 29 are more than one standard deviation from the mean.

Coded Data

We remember that we had the following formula for the mean of coded data

$$\overline{x} = \frac{\sum(x-b)}{n} + b$$

The next natural question is, what effect does addition of a constant have on the variation of the data?

It is a good exercise to verify the following

$$\frac{\sum(x-b)^2}{n} - \left(\frac{\sum(x-b)}{n}\right)^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

In particular this means that the variance is not affected by adding a constant to all data values. Intuitively, this makes of course sense as we just translate all point by a fixed number. We can summarize the above as

$$Var(x) = Var(x-b)$$

SUMMARY

In this unit we have seen how to represent different types of data in a graphical manner in stem-and-leave diagrams, histograms, cumulative frequency graphs and box-and-whisker plots. These different representations support our understanding of a given data set by highlighting structures that might otherwise be difficult to grasp from the raw data.

Furthermore, we introduced various measures of central tendency, such as the mode, the mean, and the median as a way to describe the average value of a data set. In short, the mode is the most commonly occurring value, the mean is obtained by summing all values and dividing by the number of added values and the median is the value in the middle of an ordered set of data. We discussed the different features of each measure to facilitate the choice of the right measure when choosing how to summarize a given set of data.

Finally, we noticed that the measures of central tendency do not convey any information about the spread of the data, for which reason such a measure is often complemented by an appropriate measure of variation such as the standard deviation. The spread and the skewness of data can also be illustrated using the range together with the interquartile range of data based on its median and quartiles.

UNIT 10

COMBINATORIAL STRUCTURES

STUDY GOALS

On completion of this unit, you will be able to ...

- explain the difference between a combination and a permutation.
- apply these concepts to simple selection problems.
- combine the two concepts to solve counting problems that involve the repetition of objects as well as restrictions to subsets of the objects.

10. COMBINATORIAL STRUCTURES

Introduction

Combinatorics
Methods of combinatorics can be traced throughout the history of mathematics. However, the current importance and rapid growth of combinatorics in the second half of the 20th century is mostly due to the abundance of demand in modern computer science and probability theory.

Combinatorics is concerned with counting objects or possibilities in a systematic way and provides a wealth of powerful methods. Especially when we are interested in how likely a certain event is, we often need to count the favorable outcomes of an experiment and compare it with all possible outcomes. Hence, combinatorics plays a crucial role for probability theory.

Counting problems can quickly become very intricate and combinatorics tries to provide a systematic way to analyze a given problem. In this unit we introduce elementary concepts on which much of the advanced theory is built, i.e., combinations and permutations. Given n objects, a permutation is a way to shuffle them. Imagine all your n objects are in a bag, and you take out one after the other and write down the order in which they were selected. The particular order is called a permutation of the n elements and we will see how to solve problems about the number of such permutations which can be subject to various restrictions. A combination on the other hand is a way of selecting a certain number of elements from a larger set. Here the order does not matter.

To illustrate the difference, imagine you have three fruits at home, a banana (B), an apple (A) and a pineapple (P) and you wish to make juice with two of them. Then there are three possible combinations of two fruits, i.e., AB, AP, BP. Note that the order does not matter for the final juice. However, if your two children cannot decide who should get which fruit to eat, you may suggest taking two random picks out of a bag containing the fruits. Here, the order matters and there are six possible outcomes of the selection, i.e., AB, BA, AP, PA, BP and PB.

10.1 Factorials

Imagine you want to multiply the first five positive integers. You can simply write

$$1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

However, this quickly gets tedious if you wish to multiply larger sets of numbers. The factorial function offers a way of expression of such products and is of central importance in many counting problems. It takes an integer n as an input and outputs the product of the first positive integers. Because it appears so frequently, we often write simply $n!$ for the factorial function, i.e.,

$$n! = 1 \cdot 2 \cdot \dots \cdot n$$

for any integer $n > 0$. In addition, we define $0! = 1$ which turns out to be very convenient in many calculations.

The following examples should give an idea of how convenient the factorial function can be for the calculation of products:

$$\begin{aligned}\frac{7!}{4!}&=\frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1}=7\cdot6\cdot5=210\\ \frac{7!}{5!}-\frac{3!}{0!}&=\frac{7!}{5!}-\frac{3!}{1}=7\cdot6-3\cdot2=36\\ 7!-5!&=5!(7\cdot6-1)=120\cdot41=4920\end{aligned}$$

10.2 Permutations

Recall that a permutation is a particular arrangement of objects on a line. In the following we will see how to count the number of all permutations and, importantly, also how to count the number of permutations that have a certain property of interest that distinguishes them from other permutations of the same number of objects.

Permuting n Distinct bjects

Given n distinct objects, we denote the number of all permutations of these n objects with ${}^{n}P_{n}$. Imagine you are given these n objects in a bag and you pick one after the other and arrange them in a line. You have n choices to pick the first object, you have $(n-1)$ choices for the second object and so on until there is only one object left that you necessarily must pick in the last round. In total you have $n!$ many different choices to pick n elements. Hence, we have that

$${}^{n}P_{n}=n!$$

for any integer $n>0$.

This formula is important when considering questions like how many three-digit numbers can be built from the given digits 2, 3, 4. Of course, we can just make a list, but this will not be practicable for larger sets of digits, i.e., we ask for all eight-digit numbers built from the list of digits 0, 1, 4, 5, 6, 7, 8, 9. The only information that we need here is the number of objects. We can immediately conclude that the number of such eight-digit numbers is

$${}^{8}P_{8}=8!=40320\,.$$

Permuting n Objects with Repetitions

Things get a bit more interesting when our set of objects contains repeated objects. For example, we ask for all four-digit numbers that we can build from 2, 3, 3, 4. For ease of illustration, we replace one of the threes with $\widehat{3}$ and observe that the arrangements $\widehat{3},\ 3,\ 2,\ 4$ and $3,\widehat{3},\ 2,\ 4$ are actually the same arrangement. To be more precise, in each arrangement we can swap the two threes to obtain an arrangement that appears to be the same. Now, recall that if the 4 digits were distinct, we would have $4!=24$ different arrangements. However, we have two reduce this number by one half because the repeated digit can be placed in $2! = 2$ different ways. From here, it is easy to deduce the following general rule:

Given n objects, of which n_1 are of one type, n_2 are of another type and so on up to n_k, with

$$n = n_1 + n_2 + \cdots n_k$$

then the number of different permutations is

$$\frac{^nP_n}{n_1! \cdot n_2! \cdot \cdots \cdot \; n_k!} = \frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot \; n_k!}$$

Example 10.1: Consider the district of Istanbul called GALATASARAY. In how many different ways can all the letters be arranged?

Solution. This word has 11 letters, five As, and one G, L, T, S, R and Y. Hence, we get that

$$\frac{11!}{5! \cdot 1! \cdot 1! \cdot 1! \cdot 1 \cdot 1! \cdot 1!} = 332640\,.$$

Permuting n Distinct Objects with Restrictions

A natural next step is to ask for the number of possible permutations of n objects when certain restrictions are in place. We use an example to illustrate this situation.

Example 10.2: How many even four-digit numbers, with distinct digits and greater than 2000 can be built from 1, 2, 3, 6?

Solution. There are two restrictions in place. The leftmost digit can only be 2, 3, 6. while the rightmost digit can only be 2 or 6. Note that 2 and 6 can be used at either of the restricted positions, so we consider the four-digit numbers starting with 2 and 6 separately.

Case 1: Start with 2. In this case we must place 6 at the right most point to obtain an even number, so we can only vary 1 and 3 at the two middle positions, i.e.,

$$1 \cdot 2 \cdot 1 \cdot 1 = 2 \text{ choices}$$

Case 2: Start with 6. This case is analogous to Case 1; i.e., again 2 choices.

Case 3: Start with 3. In this case, we can place 2 or 6 at the far right, while the other two positions can be filled with the remaining two elements, i.e.,

$$1 \cdot 2 \cdot 1 \cdot 2 = 4 \text{ choices}$$

Altogether, we see that $2 + 2 + 4 = 8$ even numbers with distinct digits and greater than 2000 can be made.

As a general rule, we note that we should always start to investigate the number of choices for the restricted positions first, and only in a second step, investigate the unrestricted positions.

Permuting r out of n Objects

Finally, we go one step further and consider permutations in which only a subset of the original n objects are selected and shuffled. Choosing r objects from a set of n objects and arranging them in a particular form is called a permutation of r from n.

This is certainly the most interesting case we have studied so far. We denote the number of permutations of r from n as ${}^{n}P_{r}$. To obtain a formula for this number, observe that we have n choices to select the first object, $n-1$ choices for the second and so on, until we are left with $n-r+1$ choices to pick the rth object. Hence, we have that

$${}^{n}P_{r} = \frac{n!}{(n-r)!} = n \cdot (n-1) \cdot \dots \cdot (n-r+1) \, .$$

Example 10.3: How many four-digit numbers can be made from $0{,}1, \ \cdots, \ 9$ if each digit can be used only once?

Solution. We ask for all permutations of 4 from 10. According to the above formula we have

$${}^{10}P_{4} = \frac{10!}{(10-4)!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

Example 10.4: In how many ways can 23 people be seated on the 15 chairs in a train car if the elderly man and the pregnant lady among them must be offered a seat?

Solution. The elderly man and the pregnant lady can sit on any of the 15 chairs. Hence, there are $15 \cdot 14$ many ways to sit them. The remaining 21 people are then placed on the remaining 13 chairs. Depending on where the elderly man and the pregnant lady sit, we have ${}^{21}P_{13}$ choices to place the remaining passengers. So, overall, we have

$$15 \cdot 14 \cdot {}^{21}P_{13} = 266\ 098\ 657\ 144\ 320\ 000 \ \text{possibilites}$$

This example shows that the number of possibilities can get very large, very quickly in case the order matters. In the next section, we will see that numbers might be much smaller if the order does not matter. For example, we could simply ask, what are my chances of being one of the 8 unlucky passengers who do not get a seat? Here the order does not matter.

10.3 Combinations and the Binomial Coefficient

In contrast to a permutation, the order of the selection is not important in a combination. For example, when you can choose three flavors of ice cream out of 10 possibilities, the order in which you choose does not matter. More abstractly, a combination refers to choosing r objects in no particular order from a set of n objects.

We write ${}^{n}C_{r}$ for a combination of r objects from n. Note that there are ${}^{r}P_{r} = r!$ ways to arrange the selected objects. Hence, we can write

$$ {}^{n}C_{r} \cdot {}^{r}P_{r} = {}^{n}P_{r} $$

and from this we can easily deduce the formula

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

Importantly, note the symmetry in this formula. It implies that selecting r objects and ignoring $(n-r)$ is the same as selecting $(n-r)$ objects and ignoring r. Hence, we have

$$ {}^{n}C_{r} = {}^{n}C_{n-r} $$

It is customary to say we choose r objects from n and there is the following alternative notation:

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

Binomial Coefficient The integers that appear as the coefficients in the so-called Binomial Theorem are called **binomial coefficients**. As such, binomial coefficients are directly related to Pascal's triangle.

This is usually read as "n choose r" and is also referred to as **binomial coefficient**. A way to remember this rule is to memorize

$$ \binom{n}{r} = \frac{(\text{Number we select from})!}{(\text{Number of selected})!\ (\text{Number of not selected})!} $$

Example 10.5: Assume that there are 9 books and 10 magazines in your bookshelf. In how many ways can you select three magazines and four books?

Solution. We can select three from 10 magazines and four from 9 books. Hence,

$$ {}^{10}C_{3} = \frac{10!}{3!7!} = 10 \cdot 3 \cdot 4 = 120 $$
$$ {}^{9}C_{4} = \frac{9!}{4!5!} = 9 \cdot 2 \cdot 7 = 126 $$

Overall, we have ${}^{10}C_{3} \cdot {}^{9}C_{4} = 120 \cdot 126 = 15120$ possibilities.

Note that the numbers of possible selections are multiplied since books and magazines are selected independently!

Example 10.6: A bouquet of 5 flowers is to be chosen from 6 roses and 5 sunflowers. What is the number of bouquets in which there are more roses than sunflowers?

<u>Solution.</u> There are 3 cases in which a bouquet contains more roses than sunflowers, i.e., bouquets with 5, 4 or 3 roses. The possible make-up of a bouquet is then

$$^6C_5 \cdot {^5C_0} = 6$$
$$^6C_4 \cdot {^5C_1} = 75$$
$$^6C_3 \cdot {^5C_2} = 200$$

So, in total we have $6 + 75 + 200 = 281$ different bouquets with more roses than sunflowers.

Example 10.7: Assume that you are given 5 cards each of which has one digit printed on it. How many different three-digit numbers can be made from your cards if the digits on the cards are 1, 1, 3, 4 and 5?

<u>Solution.</u> Since there is one repeated digit, we have to distinguish three different cases.

Case 1. No 1 is selected. This means 3, 4 and 5 are selected and arranged: $3! = 6$.

Case 2. One 1 is selected. Here, two of the three digits 3, 4, 5 are selected in addition to the 1: $^3C_2 \cdot 3! = 18$.

Case 3. Both 1s are selected. Then, one additional digit is selected and arranged with the two 1s: $^3C_1 \cdot \frac{3!}{2!} = 9$. Note that we divide by 2! because the digit 1 is repeated.

In total this gives us $6 + 18 + 9 = 33$ different three-digit numbers.

Basic Probabilities

An important application of permutations and combinations is to find the probability of certain events. The basic counting techniques outlined in this section are often used to determine the number of favorable outcomes among all possible outcomes of a random experiment. In particular, if an event consists of a number of favorable permutations that are equiprobable, or a number of favorable such combinations, then

$$P(event) = \frac{\text{Number of favorable permutations}}{\text{Number of possible permutations}}$$
$$P(event) = \frac{\text{Number of favorable combinations}}{\text{Number of possible combinations}}$$

Example 10.8: There are 15 identical wine bottles in the cellar. They are all without a label, but your grandfather remembers that eight contain red wine (R), four contain white wine (W) and three contain cider (C). If seven bottles are randomly selected without replacement, find the probability that exactly 5 of them contain red wine.

Solution. A favorable selection occurs when five bottles of red wine and two bottles of another liquid are selected. Note that it does not matter whether the two bottles contain white wine or cider. Hence, we denote the 15 bottles by 8R and 7R', in which R' means “not red wine.”

We have in total ${}^{15}C_7$ possible combinations, of which ${}^{8}C_5 \cdot {}^{7}C_2$ are favorable combinations. Therefore,

$$P(\text{select 5 bottles of red wine}) = \frac{{}^{8}C_5 \cdot {}^{7}C_2}{{}^{15}C_7} = \frac{56 \cdot 21}{6435} = 0.183\,.$$

Thus, the probability of selecting 5 bottles of red wine is roughly 18%.

SUMMARY

In this unit the basic tools of combinatorics are introduced. The product of the first n consecutive and non-negative integers is denoted as $n!$ and it is referred to as the factorial function. This function plays a key role in many counting problems.

A permutation is an arrangement of n objects in which the order matters. We have seen that n objects can be arranged in $n!$ different ways on a line and, hence, the number of distinct permutations of n objects is $n!$.

An arrangement of n objects in which the order does not matter is called a combination. We write ${}^{n}C_r$ for a combination of objects r from n and have the formula

$${}^{n}C_r = \frac{n!}{r!\,(n-r)!}$$

Based on the concepts of permutation and combination, we addressed different counting problems. We saw how to determine the number of permutations when some objects appear more than once, or how to determine the number of permutations subject to certain types of restrictions. Finally, we considered permutation of r from n as well as combination of r objects from n.

UNIT 11

PROBABILITY

STUDY GOALS

On completion of this unit, you will be able to ...

- determine probabilities of events by counting equally likely events and by using addition and multiplication of probabilities.
- understand the concepts of mutually exclusive, independent, and dependent events.
- explain and use conditional probabilities.

11. PROBABILITY

Introduction

Probability Theory
This theory is very frequently applied, important, and at the same time a rather young subdiscipline of mathematics. The roots of modern probability theory can be traced back to the 16^{th} and 17^{th} century before Andrei N. Kolmogorov formalized modern probability theory in 1933.

In **probability theory** we are always concerned with the probability of an event. A single random experiment and its result are referred to as an elementary event, e.g., rolling a die once. A combination of elementary events is simply called an event, e.g., obtaining a total of 6 when scores on two fair dice are added together. In this example, the event is a combination of two elementary events.

The likelihood of an event, its probability, is measured on a scale from 1, which means the event is certain, to 0 which means the event is impossible. For example, the event to have a score of 7 when rolling an ordinary die is 0, since there is no 7 on a normal die.

The main goal of probability theory is to provide a set of tools and methods to calculate the probability of more complicated events. In this unit we introduce important terminology and rules of probability theory, which are necessary to consider interesting applications as we will see.

Probability theory plays a key role in our modern world. For example, the whole insurance industry is based on assessing risk and estimating probabilities of certain events. Similarly, online betting or gaming crucially depends on assessing probabilities. Finally, also government spending is often decided on the basis of how likely it will bring benefits to society.

11.1 Events and their Outcomes

A classic example of an elementary event is tossing a fair coin which has two possible outcomes, i.e., heads or tails. Another example is rolling an ordinary, fair die which has six possible outcomes, i.e., 1, 2, 3, 4, 5 or 6. Asking for the probability of obtaining an even number when rolling a die, is an event that has three favorable outcomes, i.e., 2, 4 or 6.

To evaluate probabilities of events, we introduce the notion of selecting an object at random. In this context this means that each object has the same chance of being selected in a single experiment. Selecting an object at random is also called fair or unbiased and the selection of any object is said to be equiprobable. To introduce this standard notation, we reformulate the above as follows.

Given n objects: The probability, P, of selecting one object at random is equiprobable if

$$P(\text{selecting an arbitrary object}) = \frac{1}{n}$$

Importantly, when probabilities of all possible events are equiprobable, the probability that an event occurs is equal to the ratio between all possible favorable outcomes and all possible outcomes, i.e.,

$$P(event) = \frac{\text{Number of favourable and equiprobable outcomes}}{\text{Number of all equiprobable outcomes}}.$$

As an example, consider randomly selecting an easter egg from a basket with 18 eggs in which 12 are red and 6 are blue. Then there are 18 possible outcomes of which 12 are favorable for the event $\text{selecting a red egg}$ and 6 are favorable for the event $\text{selecting a blue egg}$. In other words, the probability to select any particular egg is

$$P(\text{select any particular egg}) = \frac{1}{18}$$

while the probabilities of selecting a red egg or a blue egg are

$$P(\text{Selecting a red egg}) = \frac{12}{18}, \quad P(\text{Selecting a blue egg}) = \frac{6}{18}$$

Exhaustive Events and Trials

A set of events is exhaustive if it contains all the possible outcomes of an experiment. The simplest example of an exhaustive set of events is an event A together with its complement $not \quad A$ also denoted as A'. The sum of the probabilities of these two events is 1 because one of them is certain to occur. For example, we either obtain a 3 when we roll a die, or we do not get a 3. More formally, we can state a first rule, i.e,

$$P(A) + P(not\ A) = 1$$

Obviously, **random experiments** can also be repeated, i.e., we can roll the same fair die more than once and write down the results. A trial is one repetition of a random experiment. We call the proportion of trials in which a particular event occurs its relative frequency. Importantly, the relative frequency of events is often used to estimate the probability of events! On the other hand, if we know the probability of an event, we can estimate how often it is likely to occur in a series of trials. This number is then referred to as the expectation. We can summarize this in a second rule:

Random experiments
In Probability Theory, a random experiment is defined as an experiment that can be performed multiple times under the same conditions. However, their outcomes cannot be predicted in advance.

$$\text{An event } A \text{ is expected to occur } n \cdot P(A) \text{times in } n \text{ trials}.$$

Example 11.1: The probability of sunshine on any particular day in Belfast is 0.57. On how many days of the year are we safe from the sun?

<u>Solution.</u> We have that $n = 365$ and we have two possible outcomes for the experiment, i.e., day with sunshine and its complement day without sunshine. Hence,

$$P(\text{no sunhine}) = 1 - P(\text{sunshine}) = 1 - 0.57 = 0.43$$

Consequently, we have $365 \cdot 0.43 \approx 157$ days without sunshine.

Example 11.2: A player randomly picks one of the following ten letters:

$$A, c, B, C, b, a, C, c, B, C$$

If the player repositions the letters and repeats the experiment 40 times, how often can they be expected to pick a lower-case letter that is not a b?

Solution. To answer this question, we define an appropriate event. We call the event $\text{picking a lower-case letter that is not a b}$ and we see that there are 4 lower-case letters of which three are different from b, i.e., there are three favorable outcomes. Hence, the probability of our event is

$$P(\text{picking a lower-case letter that is not b}) = \frac{3}{10}$$

Now, if we repeat the experiment 40 times, i.e., if we have 40 trials, we expect to pick a lower-case letter that is not b

$$40 \cdot P(\text{picking a lower-case letter that is not b}) = 40 \cdot \frac{3}{10} = 12 \text{ times}$$

11.2 The Addition Law

Two events A and B are $\text{mutually exclusive}$ if they do not share any favorable outcomes. More formally, we have that $P(A \quad and \quad B) = 0$. In this case, we can simply add the probabilities of the events, if want to calculate the probability that event A or event B occurs. As an example, consider rolling an ordinary die. The events selecting a prime, i.e., 2, 3, 5 and $\text{selecting a square number}$, i.e., 1, 4 are mutually exclusive, because they have no favorable outcome in common. Consequently,

$$P(\text{selecting prime or square number}) = P(\text{selecting prime}) + P(\text{selecting square})$$

since there is no number that is a prime and a square number. On the other hand, the events selecting a prime, i.e., 2, 3, 5 and selecting an odd number, i.e., 1, 3, 5 are not mutually exclusive since they share some favorable outcomes, i.e., 3, 5. In this case

$$P(\text{selecting an odd prime number}) \neq P(\text{selecting prime}) + P(\text{selecting odd number})$$

In general, we have an addition law for mutually exclusive events A, B, i.e.,

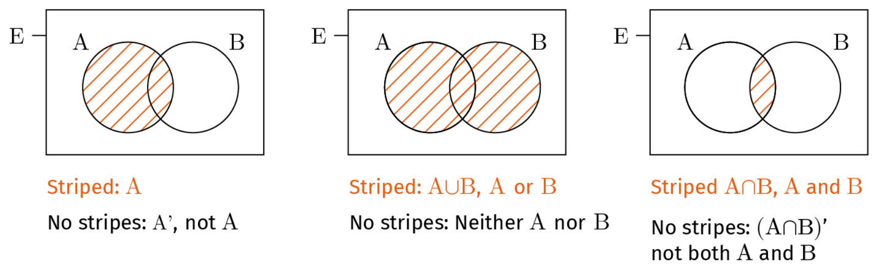
$$P(A \text{ or } B) = P(A) + P(B)$$

Note that this can be extended to any number of mutually exclusive events.

Venn diagrams can be useful when solving basic problems in probability. The universal set E represents the set of all possible outcomes of an experiment and is called the possibility space. In a Venn diagram, the universal set is usually depicted as the frame of the diagram, which contains all possible outcomes. Subsets of E correspond to different events and are depicted as circles. If events are mutually exclusive the circles are disjoint. If they share favorable outcomes, then the circles intersect. In this way, different relations between sets can be illustrated with Venn diagrams as the following figure shows.

Venn diagrams
John Venn (1834-1923) was an English mathematician, logician and philosopher who developed these diagrams when lecturing in Cambridge around 1869.

Figure 59: Examples of Venn Diagrams



Source: Florian Pausinger, (2022).

As the figure indicates, for non-mutually exclusive events, the probability $P(A \cup B)$ can be found by counting all favorable and equally likely outcomes and making sure not to count any such outcomes twice. We have the general rule

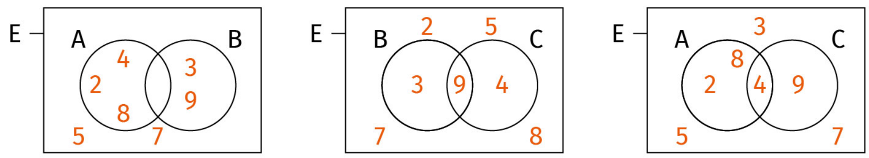
$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example 11.3: We select a digit randomly from 2, 3, 4, 5, 7, 8, 9 and consider the following events: (A) a multiple of 2 is selected, (B) a multiple of 3 is selected, (C) a square number is selected.

a) Use Venn diagrams to illustrate the pairwise relationships of the three sets.
b) Calculate the probabilities $P(A \cup B),\ P(A \cup C),\ P(B \cup C)$ and $P(A \cup B \cup C)$.

<u>Solution.</u> (a) We have that $E = \{2{,}3,4{,}5,7{,}8,9\},\ A = \{2{,}4,8\},\ B = \{3{,}9\},\ C = \{4{,}9\}$. This can be illustrated as follows.

Figure 60: Illustration of Example 11.3



Source: Florian Pausinger, (2022).

Therefore, we see that only A and B are mutually exclusive.

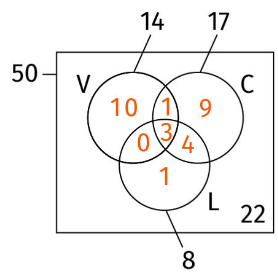
(b) We have that $P(A \cup B) = P(A) + P(B) = \frac{3}{7} + \frac{2}{7} = \frac{5}{7}$. In the other two cases, we have to make sure not to double count any event. Let $n(\,.\,)$ denote the number of elements in a set. Then,

$$\begin{aligned} n(A \cup C) = n(A) + n(C) - n(A \cap C) = 3 + 2 - 1 = 4 \\ n(B \cup C) = n(B) + n(C) - n(B \cap C) = 2 + 2 - 1 = 3 \end{aligned}$$

Consequently, we have that $P(A \cup C) = \frac{4}{7},\ P(B \cup C) = \frac{3}{7}$.

Example 11.4: In a survey, a group of 50 students is asked which ice cream flavor they like from a choice of vanilla (V), chocolate (C), and lemon (L). The depicted Venn diagram shows the outcome of the survey. Find the probability that a randomly selected student likes lemon and chocolate ice cream.

Figure 61: Illustration of Example 11.4.



Source: Florian Pausinger, (2022).

Solution. We need to find $P(C \cup L)$. First, we calculate the number of favorable outcomes: $n(C \cup L) = n(C) + n(L) - n(C \cap L) = 17 + 8 - 7 = 18$. Hence, the probability is $\frac{18}{50}$.

11.3 The Multiplication Law

If two events can occur without either being affected by the occurrence of the other, then we call the two events independent. The most basic examples are selections with replacement, i.e., you select one out of n objects and put it back before selecting the next object or rolling two identical dice.

If A and B are two independent events, then

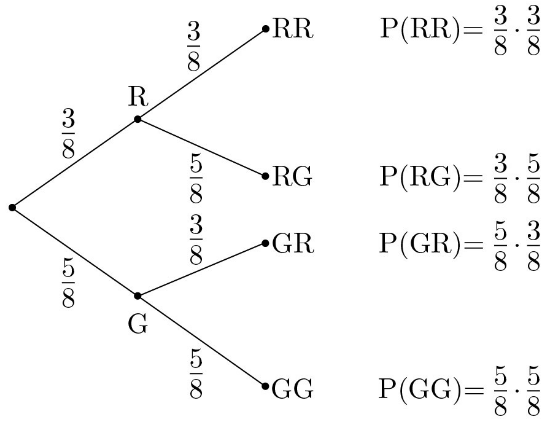
$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B)$$

This is called the multiplication law for independent events and can be extended to any number of independent events.

As an example, consider a bag with three red balls (R) and five green balls (G). Assume that we select one ball at random, put it back into the bag, and then select a second ball. First, we want to find the probabilities of the four possible outcomes, i.e., RR, RG, GR, and GG, of this experiment.

Note that the two selections are independent of each other. For the first selection we have $P(R) = \frac{3}{8}$ and $P(G) = \frac{5}{8}$, the same probabilities hold for the second selection. We can illustrate the experiment in a tree diagram, which shows how to use the multiplication law to determine the probabilities of the different outcomes.

Figure 62: A Tree Diagram



Source: Florian Pausinger, (2022).

Note that the multiplication of the probabilities of independent events is always performed from left to right following the branches. This can be combined with addition of mutually exclusive events, i.e., all events that we find on the same vertical level of the tree are mutually exclusive and can be added. In this way, we can also compute the probabilities of events, such as $\text{selecting balls of different colors}$ or $\text{selecting balls of the same color}$ or $\text{selecting a red ball first}$.

$$
\begin{aligned}
P(\text{different colour}) &= P(RG \text{ or } GR) = P(RG) + P(GR) \\
&= \left(\frac{3}{8} \cdot \frac{5}{8}\right) + \left(\frac{5}{8} \cdot \frac{3}{8}\right) = \frac{30}{64} \\
P(\text{same colour}) &= P(RR \text{ or } GG) = P(RR) + P(GG) \\
&= \left(\frac{3}{8} \cdot \frac{3}{8}\right) + \left(\frac{5}{8} \cdot \frac{5}{8}\right) = \frac{34}{64} \\
P(\text{red first}) &= P(RG \text{ or } RR) = P(RG) + P(RR) \\
&= \left(\frac{3}{8} \cdot \frac{5}{8}\right) + \left(\frac{3}{8} \cdot \frac{3}{8}\right) = \frac{24}{64}
\end{aligned}
$$

Alternatively, we can use a possibility diagram to illustrate the experiment. The following diagram shows the$8 \cdot 8 = 64$ equally likely outcomes of the experiment. We can simply count the favorable outcomes for each event in order to determine its probability.

Figure 63: A Possibility Diagram

Second selection	R	RR	RR	RR	GR	GR	GR	GR	GR
	R	RR	RR	RR	GR	GR	GR	GR	GR
	R	RR	RR	RR	GR	GR	GR	GR	GR
	G	RG	RG	RG	GG	GG	GG	GG	GG
	G	RG	RG	RG	GG	GG	GG	GG	GG
	G	RG	RG	RG	GG	GG	GG	GG	GG
	G	RG	RG	RG	GG	GG	GG	GG	GG
	G	RG	RG	RG	GG	GG	GG	GG	GG
		R	R	R	G	G	G	G	G
		First selection							

Source: Florian Pausinger, (2022).

Example 11.5: Given three ordinary dice each of which is rolled once. Find the probability that the sum of the scores of the three dice is more than 15.

<u>Solution.</u> We have three possible scores larger than 15, i.e., 16, 17 and 18. Hence,

$$P(sum > 15) = P(sum = 16) + P(sum = 17) + P(sum = 18)$$

Now, each combined roll has $216 = 6 \cdot 6 \cdot 6$ equiprobable outcomes because each individual die has 6 such outcomes. There is only one way to obtain a sum of 18, when all dice show 6; i.e., $P(sum = 18) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$. There are three possible ways to get a score of 17, when two dice show 6 and one die shows 5; i.e., $P(sum = 17) = 3 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$. Finally, a score of 16 can be obtained either with two 6s and a 4 or with one 6 and two 5s, hence we have six possible ways, i.e., $P(sum = 16) = 3 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$. In summary,

$$P(sum = 18) = \frac{1}{216}, \;\; P(sum = 17) = \frac{3}{216}, \;\; P(sum = 16) = \frac{6}{216}$$

Therefore, we have that

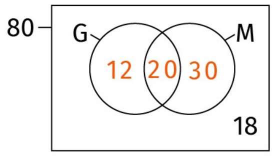
$$P(sum > 15) = \frac{1}{216} + \frac{3}{216} + \frac{6}{216} = \frac{10}{216} = \frac{5}{108}$$

An important application of the multiplication law is to show whether or not two events are independent. If $P(A \cap B) = P(A) \cdot P(B),$, then A and B are independent and vice versa.

Example 11.6: Of the 80 students staying in a dormitory, 50 are male (M) and 32 study German literature (G), 20 are male and study German literature and 18 belong to neither group. Determine whether the two events M and G are independent.

Solution. First, we can draw a Venn diagram.

Figure 64: Illustration of Example 11.6



Source: Florian Pausinger, (2022).

We see that $P(M) = \frac{50}{80},\ P(G) = \frac{32}{80},\ P(M \cap G) = \frac{20}{80}$. Now

$$P(M) \cdot P(G) = \frac{50}{80} \cdot \frac{32}{80} = \frac{5}{8} \cdot \frac{4}{10} = \frac{20}{80} = P(M \cap G)$$

Hence, the two events are independent.

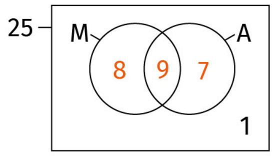
11.4 Conditional Probability

Often, we do not only want to know the probability of an event, but we want to evaluate the probability of an event dependent on additional information. We speak of conditional probability, if we want to emphasize that the probability of an event depends on information given about the outcome of another event.

For example, if there are two bananas, an apple, and three carrots in a bag and an item is selected at random, the probability of selecting the apple is $\frac{1}{6}$. However, if we are told that a fruit is selected, then the probability to select the apple is $\frac{1}{3}$ conditional on the information that a fruit was selected. Conditional probability is usually denoted with a vertical bar, i.e., $P(\text{apple}|\text{fruit}) = \frac{1}{3}$. We read $P(A|B)$, as “the probability that event A occurs, given that B occurs.”

Example 11.7: The following Venn diagram shows the number of students who study astronomy (A) and who study mathematics (M). Find the probability that a randomly selected student (a) studies mathematics, given that they study astronomy, and (b) does not study astronomy, given that they do not study mathematics. (c) Compare the probabilities with the probability of selecting a student that studies mathematics and of selecting a student that does not study astronomy.

Figure 65: Illustration of Example 11.7



Source: Florian Pausinger, (2022).

Solution. (a) We see that there are 16 students who study astronomy, nine of which study mathematics as well, hence $P(math\ |astronomy) = \frac{9}{16}$.

(b) Furthermore, there are 8 students who do not study mathematics, one of which does not study astronomy either, hence, $P(\text{not astronomy }|\text{not math}) = \frac{1}{8}$.

(c) We have that $P(\text{math}) = \frac{17}{25}$ and $P(\text{not astronomy}) = \frac{9}{25}$. Therefore, we see that the additional information can substantially alter the probabilities of the events.

Independent Events and Conditional Probability

We can give another definition for the independence of events based on conditional probability. We say that events A and B are independent if each is unaffected by the occurrence of the other. To be more formal, two events A and B are independent, if and only if we have that

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

Interestingly, events do not have to be mutually exclusive to be independent as the following example shows. Consider rolling a fair die and define the events:

A: rolling a multiple of three, i.e., 3 or 6

B: rolling an even number, i.e., 2, 4 or 6

Then the two events are not mutually exclusive, but

$$P(A|B) = \frac{1}{3}\text{and } P(A|B') = \frac{1}{3}$$

Note that when B occurs, the die shows 2, 4, or 6, so there is a chance of one in three to have a 6 which means that A occurs as well. Similarly, if the complement of B occurs, the die shows 1, 3, or 5 and there is a chance again of one in three that we have a 3 which means that A occurs as well. Hence, the two events are independent!

Dependent Events and Conditional Probability

In contrast to independent events, two events are mutually dependent if the occurrence of one event affects the probability of the other and vice versa. A typical example is selecting objects without replacement. In this case the probabilities of the second selection depend on the outcome of the first selection.

We can use the notion of conditional probability to extend the multiplication law for independent events to the multiplication law of probability which applies to dependent as well as independent events. Given two events A and B, we have that

$$\begin{aligned} P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B|A) \\ P(B \text{ and } A) = P(B \cap A) = P(B) \cdot P(A|B) \end{aligned}$$

In particular, we have that

$$P(B) \cdot P(A|B) \equiv P(A) \cdot P(B|A)$$

Importantly, the multiplication law of probability can be used to find conditional probabilities. We just need to rearrange the terms in the above equations, i.e., if we know $P(A)$ and $P(A \cap B)$ we can immediately calculate $P(B|A)$.

Example 11.8: An ordinary die is rolled once. What is the probability that the number obtained is prime, given that it is even?

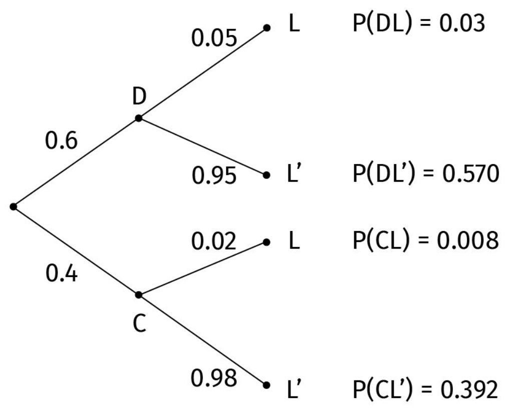
Solution. We need to find $P(\text{prime}|\text{even})$. We have that $P(\text{even}) = \frac{1}{2}$ and $P(\text{prime and even}) = \frac{1}{6}$. Hence,

$$P(\text{prime}|\text{even}) = \frac{P(\text{prime and even})}{P(\text{even})} = \frac{1}{6} \cdot \frac{2}{1} = \frac{1}{3}$$

Example 11.9: A professor drives to university (D) 60% of the time and cycles (C) 40% of the time. She is late to university (L) on 5% of the occasions that she drives, and she is late on 2% of the occasions that she cycles. Given that she is late, find the probability that she cycles, i.e., find $P(C|L)$.

Solution. We can use a tree diagram to visualize the information we are given.

Figure 66: Illustration of Example 11.9



Source: Florian Pausinger, (2022).

First, we calculate the probability that she is late.

$$P(L)=P(D \text{ and } L)+P(C \text{ and } L)=\frac{6}{10}\cdot\frac{5}{100}+\frac{4}{10}\cdot\frac{2}{100}=0.03+0.008=0.038$$

$$P(C|L)=\frac{P(C \text{ and } L)}{P(L)}=\frac{8}{1000}\cdot\frac{1000}{38}=\frac{4}{19}\approx 0.211$$

SUMMARY

In this unit, we saw that picking a particular object at random from a collection of n objects has the probability $\frac{1}{n}$ if we assume all possible outcomes of the event to be equiprobable. We call such an experiment an elementary event. A combination of elementary events is referred to as event and the probability of an event is equal to the ratio of equally likely outcomes that are favorable to the event and all possible outcomes.

Building on this notion, we introduced different types of events such as exhaustive events, mutually exclusive events, independent events, and dependent events. Importantly, the type of event determines how to calculate the probability of the event.

If two events A andW B are mutually exclusive, then the addition law applies, and we have $P(A \text{ or } B)=P(A)+P(B)$. On the other hand, for independent events A and B we have the multiplication law, i.e., $P(A \text{ and } B)=P(A)\cdot P(B)$.

Often, we have additional information about an event which can influence the probability of the outcomes of an experiment. Conditional probability refers to the probability of an event given additional information about the outcome, denoted as $P(A|B)$. Conditional probabilities can be used to extend the multiplication law to also apply to events that might not be independent. We have that

$$P(A \text{ and } B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) = P(B \text{ and } A)$$

UNIT 12

DISCRETE RANDOM VARIABLES

STUDY GOALS

On completion of this unit, you will be able to ...

- explain what a discrete random variable is.
- calculate the expectation and variance of a discrete random variable.
- use the binomial and the geometric distribution in practical situations.

12. DISCRETE RANDOM VARIABLES

Introduction

It is often of great importance to make predictions about the risk of events in the future, about future sales, or chances in gambling based on available information. An insurance company may evaluate the probability that a certain person or group of people will have a car accident before giving a quote for an insurance. A company will assess the most likely outcome of a new marketing campaign before launching it and a casino owner needs to know how likely their customers are to win a jackpot before deciding on how big the main prize of a particular game should be.

In this unit, we investigate the mathematical foundations which are the basis for addressing the above questions. Random experiments and their outcomes are usually modelled as discrete (or continuous) random variables with a certain underlying **probability distribution** that determines the likelihood of certain events. Probability distributions can be determined empirically by collecting the outcomes of many random experiments. However, it is also possible to analyze a random experiment theoretically and match key assumptions with requirements of certain well-known probability distributions such as the binomial or the geometric distribution to obtain a model for a specific situation.

Probability distribution
A probability distribution is a function that returns the probability of the occurrence of a possible outcome of a random experiment.

12.1 Probability Distributions

A discrete random variable is a function from a set of possible outcomes of an experiment to a finite set of values. For example, if we roll two dice, then the set of possible outcomes is a pair (x, y) with $1 \leq x, y \leq 6$ and a possible random variable $X: \{(x, y): 1 \leq x, y \leq 6\} \rightarrow \{0,1,2\}$ is given by the number of 6s obtained in the roll. We often abbreviate the notation and would simply write $X \in \{0,1,2\}$ if the context is clear. In other words, a discrete random variable is a variable that can only take a finite number of values which occur by chance and not according to a **deterministic** rule.

Deterministic
In the context of mathematics, computer science, or physics, deterministic refers to a system where no randomness is involved in the development of future circumstances or states of the system.

Note that also selections without replacement can be modelled as a discrete random variable. For example, if we randomly select three balls from a box with four blue and two red balls, then the number of selected blue balls, B, and the number of selected red balls, R, are discrete random variables with $B \in \{1, 2, 3\}$ and $R \in \{0, 1, 2\}$.

Given a discrete random variable, its probability distribution assigns a probability to all the possible values of the random variable. Importantly, since a probability distribution always shows all the possible values of a variable, the sum of the probabilities is always 1.

To return to the example of the two dice, we could simply count how many pairs contain zero, one, or two 6s to obtain the corresponding probability distribution, i.e.,

$$P(X = 0) = \frac{25}{36}, \; P(X = 1) = \frac{10}{36}, \; P(X = 2) = \frac{1}{36}.$$

Thus, the probabilities for the possible values of X are equal to the relative frequencies of the values.

Example 12.1: There are three more tickets for a concert available, but eight teenagers, one man and one woman want to attend the concert. It is decided to select three of the interested people at random. Draw a probability distribution for the random variable T, which represents the number of teenagers selected.

Solution. The selections are made without replacement, so the probabilities $P(T = t)$ can be obtained via combinations. The possible values of T are 1, 2 and 3. In total there are 10 people of which three are chosen. So the set of possible outcomes is the set of all triples that can be chosen. There are ${}^{10}C_3 = \binom{10}{3} = 120$ such triples. From this we get

$$P(T = 1) = \frac{{}^8C_1{}^2C_2}{{}^{10}C_3} = \frac{8 \cdot 1}{120} = \frac{1}{15}$$
$$P(T = 2) = \frac{{}^8C_2{}^2C_1}{{}^{10}C_3} = \frac{28 \cdot 2}{120} = \frac{7}{15}$$
$$P(T = 3) = \frac{{}^8C_3{}^2C_0}{{}^{10}C_3} = \frac{56 \cdot 1}{120} = \frac{7}{15}$$

Table 12: Probability Distribution Table for Example 12.1

y	1	2	3
$P(Y = y)$	$\frac{1}{15}$	$\frac{7}{15}$	$\frac{7}{15}$

Source: Florian Pausinger, (2022).

Example 12.2: Let X be a discrete random variable with $X \in \{3,4,5,6\}$. Assuming that $P(X = x) = c \cdot x^2$, for a constant c, find the value of the constant and the probability $P(X > 4)$.

Solution. We have that

$$1 = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$$

Therefore,

$$1 = 9c + 16c + 25c + 36c \ \Rightarrow c = \frac{1}{86}$$

Furthermore, $P(X > 4) = P(X = 5) + P(X = 6)$. Consequently,

$$P(X > 4) = \frac{25}{86} + \frac{36}{86} = \frac{61}{86}$$

12.2 Expectation and Variance of a Discrete Random Variable

Expected Value
The origins of the expected value can be traced back to the 17th century when Blaise Pascal (1623-1662) considered the problem of points, also known as the problem of division of stakes, which is an even older, classical problem in probability theory.

Random variables correspond to random experiments. When a number of trials is carried out, we can record the values of the random variable and produce a frequency distribution. This distribution has a mean or **expected value**.

The mean of a discrete random variable X is referred to as its expectation and is usually denoted as $E(X)$. Given a frequency table or a probability distribution, we can calculate the mean of the random variable as

$$Mean = E(X) = \frac{\sum x_i \cdot f(x_i)}{\sum f(x_i)} = \sum x_i \cdot P(X = x_i)$$

in which $f(x)$ denotes the frequency of a value x.

The spread around the mean of a discrete random variable X can be measured with the variance and standard deviation of X. We can define both quantities again in terms of frequencies and in terms of probabilities, i.e.,

$$Var(X) = \sum P(X = x_i) \cdot (x_i - E(X))^2$$

Similarly, we can define the variance also in terms of frequencies.

Example 12.3: Given the probability distribution of the random variable X as shown in the table, find its expectation, variance, and standard deviation.

Table 13: Probability Distribution Table for Example 12.3

x	0	5	15	20
$P(X = x)$	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{5}{12}$	$\frac{3}{12}$

Source: Florian Pausinger, (2022).

Solution. We have that

$$\begin{aligned} E(X) &= 0 \cdot \frac{1}{12} + 5 \cdot \frac{3}{12} + 15 \cdot \frac{5}{12} + 20 \cdot \frac{3}{12} = \frac{15 + 75 + 60}{12} = 12.5 \\ Var(X) &= \frac{1}{12}12.5^2 + \frac{3}{12}(5 - 12.5)^2 + \frac{5}{12}(15 - 12.5)^2 + \frac{3}{12}(20 - 12.5)^2 \\ &= 43.75 \end{aligned}$$

Finally,

$$SD(X) = \sqrt{V(X)} = \sqrt{43.75} \approx 6.61$$

12.3 The Binomial Distribution

A discrete random variable X has a **binomial distribution** with the parameters n and p if the following criteria are met:

Binomial Distribution
The "bi" in binomial distribution indicates that there are only two possible outcomes, i.e., failure or success. There exists an important generalization called the multinomial distribution, in which each basic experiment can have more than two outcomes with fixed probabilities.

- There are n repeated and independent trials of the same experiment.
- The number n is finite.
- There are just two possible outcomes for each trial, i.e., 0 (failure) or 1 (success).
- The success probability in each trial is constant and equal to p.

Then, the random variable X is the number of trials whose result is a success. Such a random variable is usually denoted as $X\sim B(n,p)$. Note that the $\sim$ is used to indicate that a random variable is "distributed according to." We have that if X has a binomial distribution $B(n,p)$ then the probability of k successes is

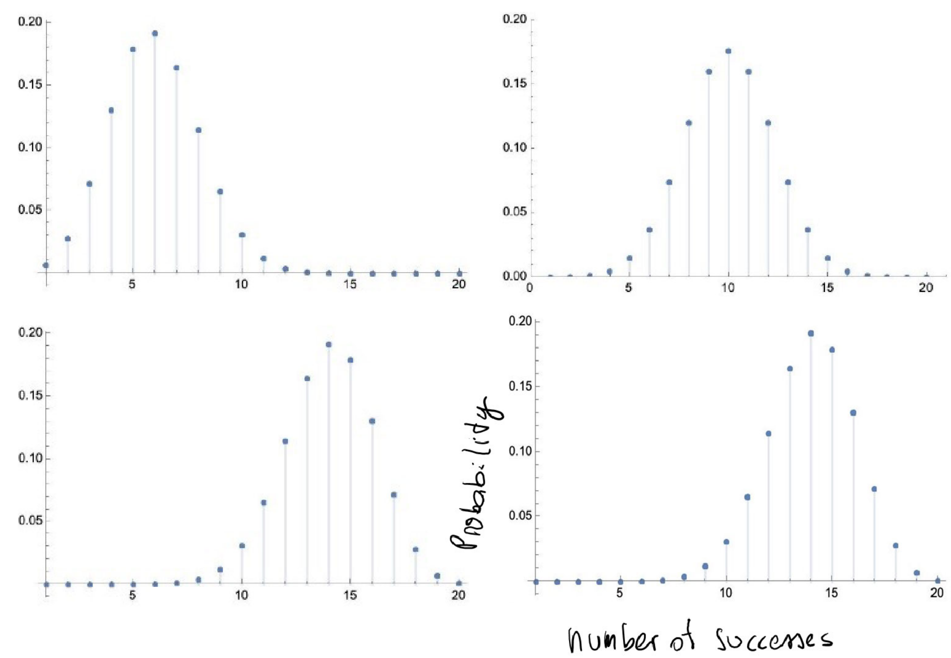
$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

As an example, if $X\sim B(2,p)$ then $X \in \{0,1,2\}$ and we have that

$$P(X=0) = p_0 = (1-p)^2,\ P(X=1) = 2\ p(1-p),\ P(X=2) = p^2$$

According to the definition, the coefficients are symmetric and can be read off Pascal's triangle; see the below figure for an illustration.

Figure 67: Illustration of the Binomial Distribution (n=20, p=0.3, 0.5, 0.7)



Source: Florian Pausinger, (2022).

Example 12.4: Consider an experiment in which we roll three ordinary fair dice. In each trial of this experiment, the random variable X returns the number of 3s obtained. Find the distribution of the random variable X.

Solution. We have that $X \in \{0, 1, 2, 3\}$. To find the probability distribution of X, we have to calculate the probabilities $P(X = x)$ for $x = 0, 1, 2, 3$. Since the three dice are independent, we can view the experiment as rolling a single die three times. In each roll, we have a success probability of $\frac{1}{6}$, i.e., the probability to get a 3. Hence, overall we see that $X \sim B\left(3, \frac{1}{6}\right)$and, therefore,

$$P(X = 0) = \left(\frac{5}{6}\right)^3,\ P(X = 1) = 3\frac{1}{6}\left(\frac{5}{6}\right)^2,\ P(X = 2)$$
$$= 3\left(\frac{1}{6}\right)^2\frac{5}{6},\ P(X = 3) = \left(\frac{1}{6}\right)^3$$

Example 12.5: Find $P(X > 6)$, assuming that $X \sim B(9{,}0.7)$.

Solution. $P(X > 6) = P(X = 7) + P(X = 8) + P(X = 9)$. We apply the above formula to get

$$
\begin{aligned}
P(X=7) &= \binom{9}{7} 0.7^7 0.3^2 = 0.267 \\
P(X=8) &= \binom{9}{8} 0.7^8 0.3^1 = 0.156 \\
P(X=9) &= \binom{9}{9} 0.7^9 = 0.040
\end{aligned}
$$

Hence, in total we have $P(X > 6) = 0.463$.

Expectation and Variance of the Binomial Distribution

The expectation, variance, and standard deviation of the binomial distribution can be calculated from the parameters n and p. Using the definition of the expectation together with the definition of the probabilities $P(X = k)$ of a binomially distributed random variable and some identities about binomial coefficients it can be shown that

$$E(X) = \sum k \cdot P(X = k) = \sum k \cdot \binom{n}{k} p^k (1 - p)^{n - k} = np$$

Moreover, the variance is given as

$$\sigma^2 = np(1 - p)$$

Example 12.6: Find the mean, variance, and standard deviation of X~$B(10{,}0.4)$.

Solution. We can apply the above formulas and obtain

$$
\begin{aligned}
E(X) &= 10 \cdot 0.4 = 4 \\
Var(X) &= 10 \cdot 0.4 \cdot 0.6 = 2.4 \\
SD(X) &= \sqrt{Var(X)} = 1.55
\end{aligned}
$$

Alternatively, we can write $\mu = 4,\ \sigma^2 = 2.4,\ \sigma = 1.55$.

Example 12.7: Given a random variable X~$B(n, p)$. Assume that $E(X) = 12$ and $Var(X) = 7.5$. Find the value of n and p as well as $P(X = 12)$.

Solution. Setting $q = 1 - p$, we observe that

$$q = \frac{n\ p\ q}{n\ p} = \frac{Var(X)}{E(X)}$$

Therefore, $q = \frac{7.5}{12} = \frac{15}{24} = \frac{5}{8}$ and $p = \frac{3}{8}$. Since, $E(X) = np$, we get that $n = \frac{12 \cdot 8}{3} = 32$.

From this we get that

$$P(X = 12) = \binom{32}{12} \cdot \left(\frac{3}{8}\right)^{12} \left(\frac{5}{8}\right)^{20} \approx 0.144$$

12.4 The Geometric Distribution

Instead of asking for the probability of having k successes in n trials, we ask now about the probability to have the first success in the k^{th} trial. Note that this is a fundamentally different question even if it sounds very similar at first. The main difference is that in the case of the binomial distribution, the number of trials was always fixed and finite. However, asking for the first success, includes the possibility of having an arbitrarily large number of failures first – this number of potential failures is unbounded. In particular, recall the formula for the expectation of a random variable. Now, we have to potentially sum infinitely more numbers to obtain the expectation and it is a priori not clear whether that even makes sense.

Let's look at an example first. Imagine, we are attempting to roll a 1 with an ordinary fair die. How likely are we to get it on the first, second, third roll?

We have a constant success probability of $\frac{1}{6}$ in each roll and a failure probability of $\frac{5}{6}$. Hence,

$$\begin{aligned} P(\text{first 1 on first roll}) &= \frac{1}{6} \\ P(\text{first 1on second roll}) &= \frac{5}{6} \cdot \frac{1}{6} \\ P(\text{first 1 on third roll}) &= \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} \end{aligned}$$

We see that the probabilities are the terms of a geometric progression with first term p and common ratio $1 - p$. Defining a random variable X as the number of trials needed for the first success, we first observe that the sum of the probabilities is equal to the sum to infinity of a geometric progression:

$$\sum P(X = k) = S_\infty = \frac{\text{first term}}{1 - \text{common ratio}} = \frac{p}{1 - (1 - p)} = 1$$

This is reassuring as it confirms that we are dealing with a proper probability distribution. We say that a discrete random variable, X, has a geometric distribution with parameter p, if it meets the following criteria:

- The repeated trials are all independent.
- The trials can be repeated infinitely often.
- There are just two possible outcomes for each trial, i.e., 0 (failure) or 1 (success).
- The success probability in each trial is constant and equal to p.

A random variable X with a geometric distribution is denoted as $X \sim Geo(p)$ and the probability that the first success occurs on the k^{th} trial is

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

Note that there is only one way to obtain the first success on the k^{th} trial, i.e., when we first have $k-1$ failures followed by one success.

We can apply the following trick when we wish to find probabilities that involve inequalities:

$$\begin{aligned} P(X \leq k) &= P(\text{success on one of the first k trials}) \\ &= 1 - P(\text{failure on the first k trials}) \\ P(X > k) &= P(\text{first success after k trials}) \\ &= P(\text{failure on the first k trials}) \end{aligned}$$

To summarize, if $X \sim Geo(p)$, then$P(X \leq k) = 1 - (1-p)^k$ and $P(X > k) = (1-p)^k$.

Example 12.8: In a random experiment, repeated and independent trials are carried out. The probability of success is $\frac{1}{3}$ in each trial. Find the probability that the first success occurs (a) on the third trial, (b) before the third trial, (c) after the third trial.

Solution. We can define a random variable X representing the number of trials until (and including) the first success. This random variable has a geometric distribution, i.e., we have that $X \sim Geo\left(\frac{1}{3}\right)$. Therefore, we have that

$$\begin{aligned} P(X=3) &= \frac{1}{3}\left(\frac{2}{3}\right)^2 = \frac{12}{81} = \frac{4}{27} \approx 0.148 \\ P(X<3) &= P(X \leq 2) = 1 - \left(\frac{2}{3}\right)^2 = \frac{5}{9} \approx 0.556 \\ P(X>3) &= 1 - P(X=3) - P(X<3) = 1 - \frac{4}{27} - \frac{15}{27} = \frac{8}{27} \approx 0.296 \end{aligned}$$

Alternatively, we can calculate the last probability directly, i.e.,

$$P(X > 3) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}$$

Example 12.9: Find the probability that a fair coin has to be tossed fewer than 8 times until the first head is obtained. Now assume that the coin is biased such that the probability of obtaining heads is $\frac{5}{11}$. How often do we need to toss the biased coin to achieve the same (or a larger) probability of success as in the fair case?

Solution. Let X represent the number of times the fair coin is tossed up to and including the first head. Let Y represent the number of times the biased coin is tossed up to and including the first head. Then$X \sim Geo\left(\frac{1}{2}\right)$ and $Y \sim \left(\frac{5}{11}\right)$.

$$P(X < 8) = P(X \leq 7) = 1 - \left(\frac{1}{2}\right)^7 = \frac{127}{128}$$

Now we need to solve the inequality:

$$P(Y < k) \geq \frac{127}{128}$$

We have that $P(Y < k) = P(Y \leq k - 1) = 1 - \left(\frac{7}{11}\right)^{k-1}$ such that we look for k with $1 - \left(\frac{7}{11}\right)^{k-1} \geq \frac{127}{128}$. We get that $\frac{1}{128} \geq \left(\frac{7}{11}\right)^{k-1}$ and collect different values for the right-hand side in the following table:

Table 14: Illustration of Example 12.9

k	10	11	12	13
$\left(\frac{7}{11}\right)^{k-1}$	0.0171	0.0109	0.0069	0.0044

Source: Florian Pausinger, (2022).

Since$1/128 \approx 0.0078$, we conclude from the table that$k = 12$ is the smallest integer such that the inequality$\frac{1}{128} \geq \left(\frac{7}{11}\right)^{k-1}$ is satisfied.

Mode and Expectation of the Geometric Distribution

An important observation about geometric distributions is that $P(X = 1)$ is always greatest among all individual probabilities. This is easy to see from the fact that p and $1 - p$ are both numbers smaller than 1 and multiplying a number with a number smaller than 1 can only decrease this number. Therefore, the most likely value of X is always when $X = 1$. However, do not confuse this with the fact that $P(X > 1)$ can still be much larger than $P(X = 1)$. With the same logic, we always have that $P(X = k) > P(X = m)$, whenever $k < m$. In conclusion, the mode of all geometric distributions is 1.

Moreover, the expectation can be computed, and it turns out that if $X \sim Geo(p)$, then

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

Intuitively, this means that a larger success probability implies a shorter expected waiting time.

Example 12.10: One in five chocolate eggs contains a free toy. Let the random variable X be the number of eggs that a child opens, up to and including the one in which they find their first toy. Find the mode and expectation of X as well as the probability to find a toy in the first five eggs that were opened. Interpret your results.

Solution. We have that $X \sim Geo(0.2)$. The mode is 1 and

$$E(X) = \frac{1}{p} = 5$$

Furthermore, $P(X \leq 5) = 1 - 0.8^5 = 0.672$.

Consequently, the child is most likely to find their first toy in the first chocolate egg they open. However, on average, a child will find their first toy in the fifth egg that they open. The probability to find a toy within the first five eggs is approximately 2/3.

SUMMARY

A discrete random variable is a function that maps a finite set of possible outcomes of a random experiment to a finite set of values. However, this assignment of values is random and not deterministic. A probability distribution governs this assignment and contains a probability for every possible value of the discrete random variable such that the sum of all probabilities is always 1.

The binomial distribution is an example of a probability distribution of a discrete random variable. It is used to model the number of successes in a series of n repeated and independent trials such that each trial has the same success probability.

If X has a binomial distribution $B(n, p)$, then the probability of k successes is

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

The geometric distribution is a second example of a discrete probability distribution. It is used to model the number of trials up to and including the first success in a series of repeated and independent trials, again with a constant success probability in each trial.

If X has a geometric distribution $Geo(p)$, then the probability of having the first success in the k^{th} trial is:

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

UNIT 13

THE NORMAL DISTRIBUTION

STUDY GOALS

On completion of this unit, you will be able to ...

- use normal curves to describe distributions and probabilities.
- apply the normal distribution and normal distribution tables when modelling a continuous random variable.
- approximate the binomial distribution with an appropriate normal distribution, use a continuity correction, and explain why such an approximation is not possible in certain situations.

13. THE NORMAL DISTRIBUTION

Introduction

Whenever we take measurements of continuous quantities such as time, mass, or distance as part of an experiment in the sciences, we have to be aware that such measurements are subject to error. It lies in the very nature of continuous quantities that they cannot be measured precisely. Inaccuracies often stem from imprecise measurement tools or from imprecise measurements.

It was a great discovery of Carl Friedrich Gauss in the 18th century to realize that small errors are more likely to happen than large errors and that in our measurements underestimates are usually as likely as overestimates. In particular, when repeated measurements are taken, errors are likely to cancel each other out, so that the average error is close to zero. This also means that the average of the measurements is basically error-free.

Gauss introduced what is known today as normal distribution to model measurement errors in astronomical observations. Through the years, the normal distribution has become one of the most important tools in statistics. This unit introduces the basics on continuous random variables, presents properties of the normal distribution, and gives an idea of how to use the normal distribution to approximate the binomial distribution.

13.1 Continuous Random Variables

A continuous random variable is a random variable with a continuous cumulative distribution function. In other words, and in contrast to a discrete random variable, there is an infinite number of possible outcomes of a given random experiment. Therefore, the probability that a continuous random variable takes a certain value is always 0, since the number of possible values is infinite. In this case, we rather ask for the probability that a value lies in a certain interval of values. For example, we can ask for the height of people. Within the range of possible heights, say 40 cm to 250 cm for humans, the random variable H can take any value such as 82.33449… or 171.3 or 200 or 188.12341… and so on. Instead of asking for the probability of $H = 175$, we may ask for the probability that $170 \leq H < 180$ or $174 \leq H < 176$. The probability distribution of a continuous random variable shows its range of values and the probabilities for intervals within this range. To summarize, when X is a discrete random variable, we have only finitely many possible outcomes and, thus, we ask for $P(X = x)$. However, when X is a continuous random variable, there are infinitely many outcomes, and we ask for $P(a \leq X \leq b)$.

Representation of a Continuous Probability Distribution

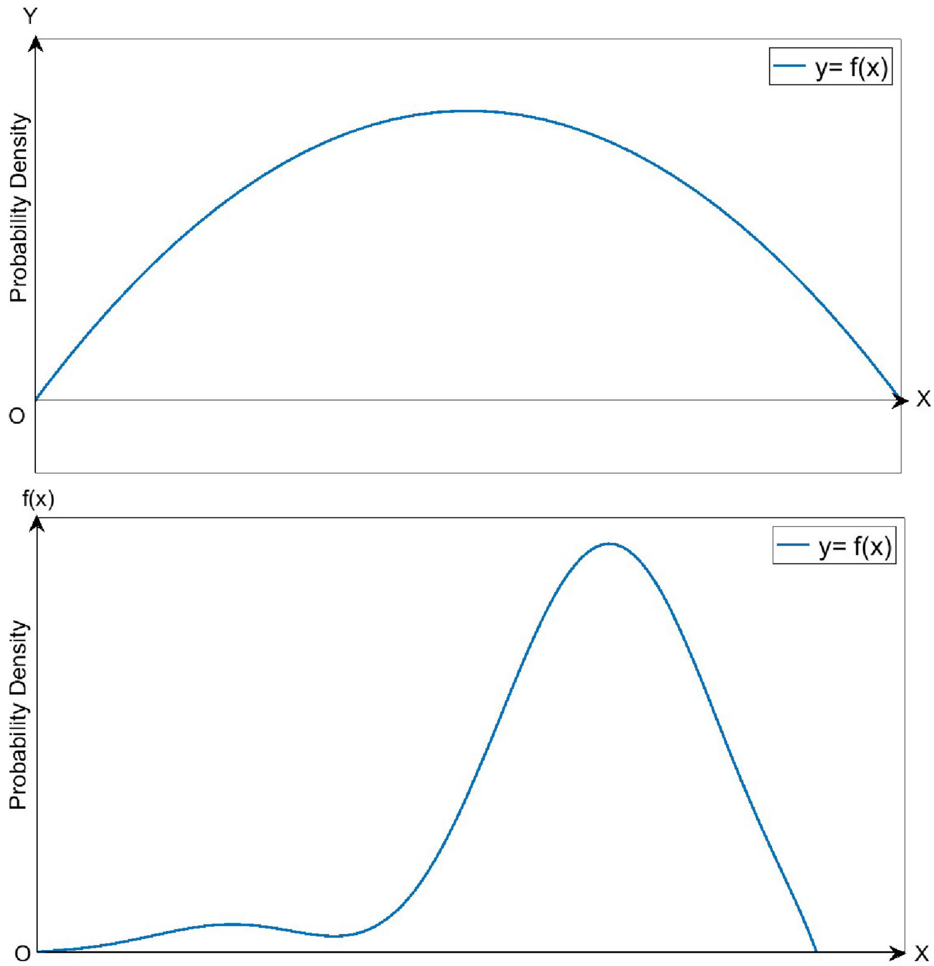
Continuous data is usually illustrated in a histogram, in which column areas are proportional to frequencies. The corresponding continuous probability distribution function of the set of data can be approximated by drawing a graph that is based on the shape of the histogram.

To obtain such an approximation, we first have to change the frequency density values on the vertical axis to relative frequency density values. The key observation is that

$$\text{relative frequency density} = \text{relative density divided by class width}$$

Hence, column areas will now represent relative frequencies, which in turn are estimates of probabilities and the vertical axes can now be labelled $\text{probability density}$. If all class intervals are of equal width, then this process ensures that the total area of all columns changes from the total sum of all frequencies to 1. The probability distribution of a large set of continuous data can now be approximated by drawing a continuous curve over the columns of the equal-width interval histogram. The continuous function is called the probability density function of the continuous random variable and is usually abbreviated to PDF or pdf. The area under the graph of a PDF is always 1.

Figure 68: Examples of Probability Density Functions

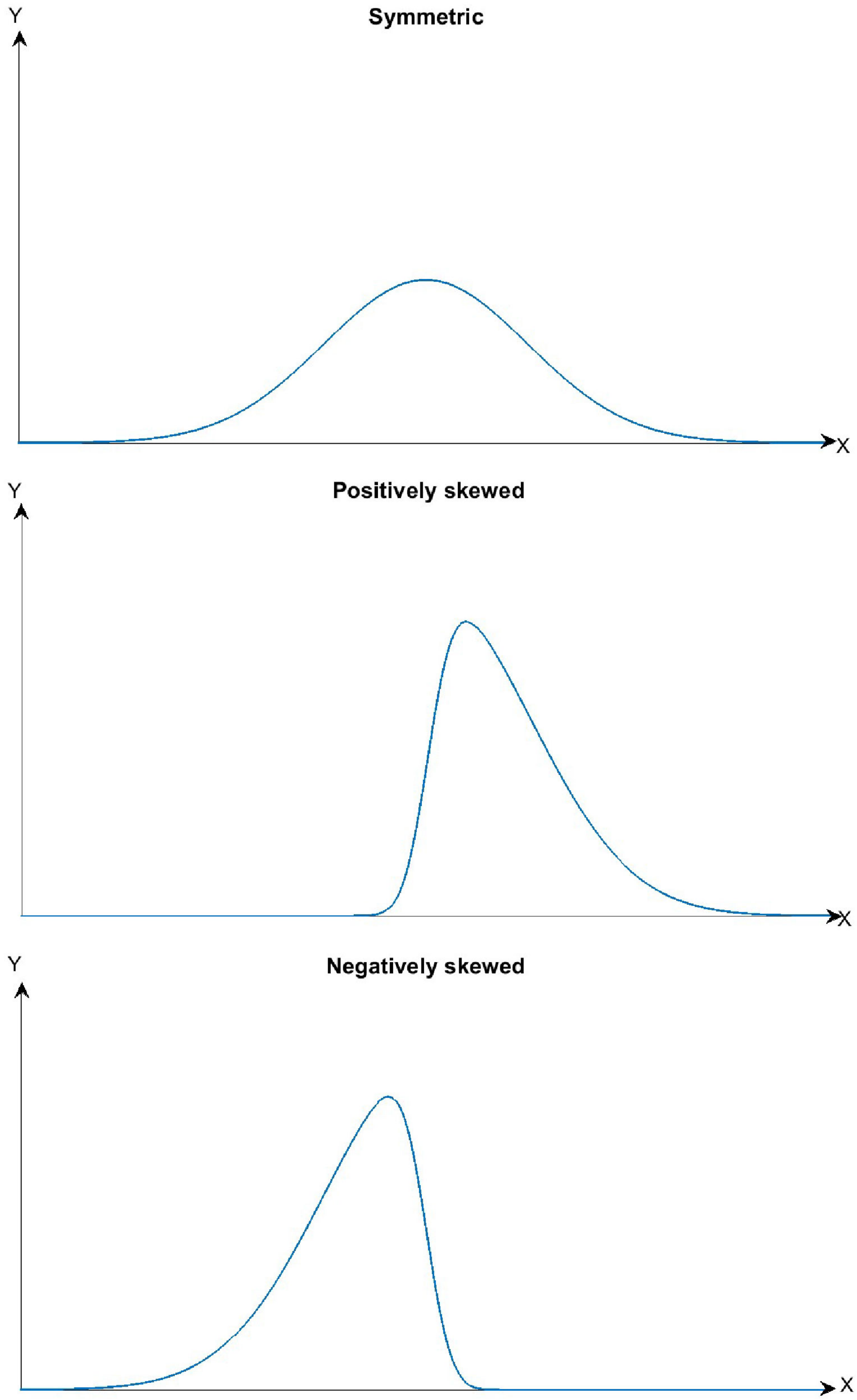


Source: Florian Pausinger, (2022).

While any shape of the curve is possible (in particular, there can of course be more than one local maximum), it is often the case that the curve can be roughly classified as having one of the following shapes.

Three commonly occurring types of graphs are negatively skewed graphs, symmetric graphs, and positively skewed graphs. A negatively skewed graph has one peak and a longer tail to the left, while a positively skewed graph has a longer tail to the right of the peak. A symmetric graph has its axis of symmetry at the location of the peak. Note however, that any such graph is always positive and the area below the curve is always equal to 1.

Figure 69: Common Types of Graphs



Source: Florian Pausinger, (2022).

13.2 Normal Curves and the Normal Distribution

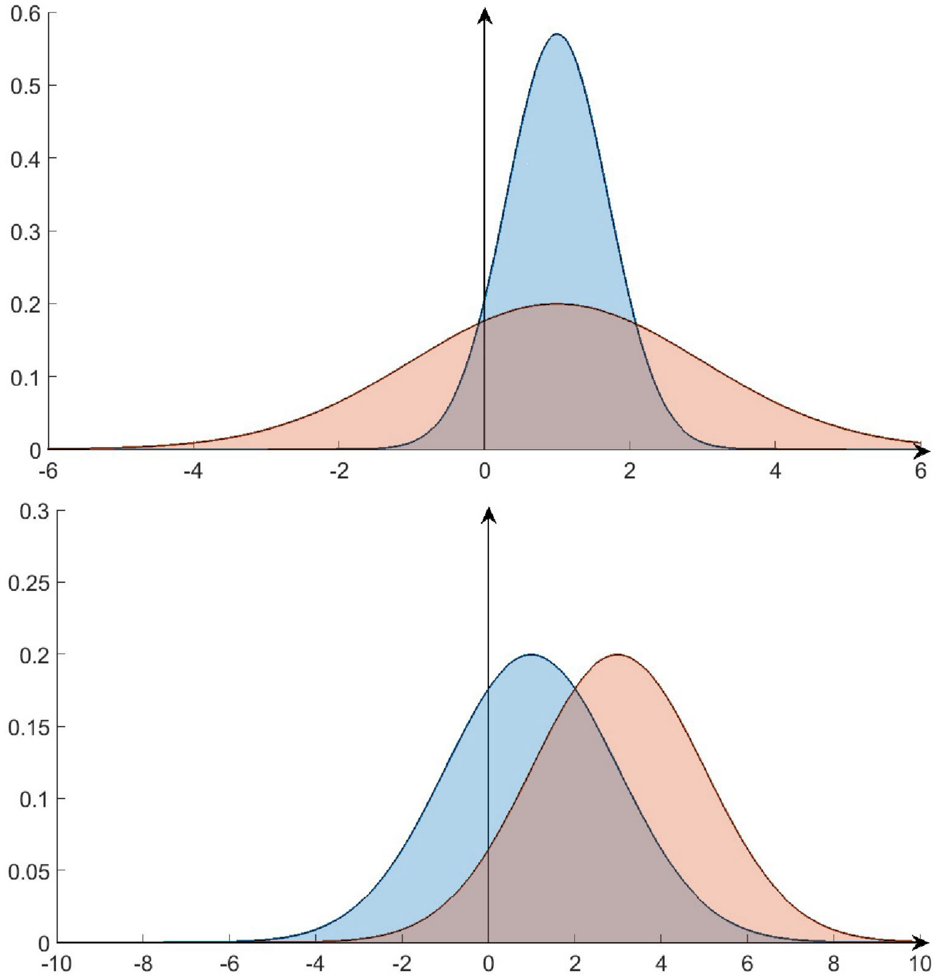
The prime example of a probability density function belongs to the so-called normal distribution. The shape of its graph, the normal curve, is symmetric and bell-shaped.

The Normal Curve

If a probability density function is represented by a normal curve, then:

1. Mean = median = mode
2. The line of symmetry of the graph is at the local maximum (the peak) which is at the mean.
3. The probability density monotonically decreases as we move away from the mean on both sides of the peak, i.e., values that are further away from the mean occur with smaller probability.
4. The standard deviation controls the spread of the values, i.e., since the area under the graph is always 1, a larger standard deviation decreases the height of the peak and increases the width of the curve.

Figure 70: Examples of Normal Curves



Source: Florian Pausinger, (2022).

Note that in both figures, the area under the curves is 1. On the left, we see the PDF of two random variables A and B with the same mean, but a different standard deviation. On the right, we see that the PDFs of the random variables C and D have identical shapes, because they have the same standard deviation, but their positions on the real line are different since they have different means.

The Normal Distribution

Abstractly speaking, the probability distribution of a continuous variable is a mathematical function that provides a method of determining probabilities for the occurrence of different outcomes of observations. An important thing to keep in mind is that extremely rare events are still assigned a positive, albeit small, probability. This may lead to differences between histograms and probability density functions. We can, for example, collect data from 1000 random experiments with range $[0{,}100]$. However, if $P(X \leq 2) = 0.00002$ then it is very likely that we will not have a data point with value ≤ 2 in our sample and, thus, the probability derived from the histogram will be zero for such an event.

The following formula is displayed for completeness, and you are not expected to know it in the exam. If a random variable X is normally distributed with mean μ and variance σ^2, then its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

for all real values of x. Here, $exp(\ldots)$ denotes the exponential function, obtained by raising the number $e = 2.71828\ \ldots$ to the power in the bracket. Note that $exp(\ldots) > 0$ for any power.

Looking at the formula, and back to what you need to memorize, we see that the parameters defining a normally distributed random variable are its mean μ and its variance σ^2. In short, we write $X \sim N(\mu, \sigma^2)$. Importantly, the probability that X takes a value between a and b is equal to the area under the curve between the x-axis and the boundary lines $x = a$ and $x = b$. The area under the graph can be found through integration, i.e.,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Note that the area under the curve is the same whether or not the boundary values are included; hence,

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X < b)$$

This also corresponds to the fact that the probability is zero at a single point of the range. As you might guess looking at the definition of $f(x)$ it is not possible to find a closed formula for this integral. Fortunately, there are many numerical schemes to approximate the definite integrals corresponding to probabilities. Before looking into these approximations, we will discuss some further properties of the normal distribution.

Table 15: Properties of Normal Distribution

Properties	Probabilities
Half of the values are less than the mean and half are greater than the mean.	$P(X < \mu) = P(X \leq \mu) = 0.5$
Approximately 68% of the values lie within one standard deviation from the mean.	$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826$
Approximately 95% of the values lie within two standard deviations from the mean.	$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545$
Approximately 99.72% of the values lie within three standard deviations from the mean.	$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$

Source: Florian Pausinger, (2022).

The Standard Normal Distribution

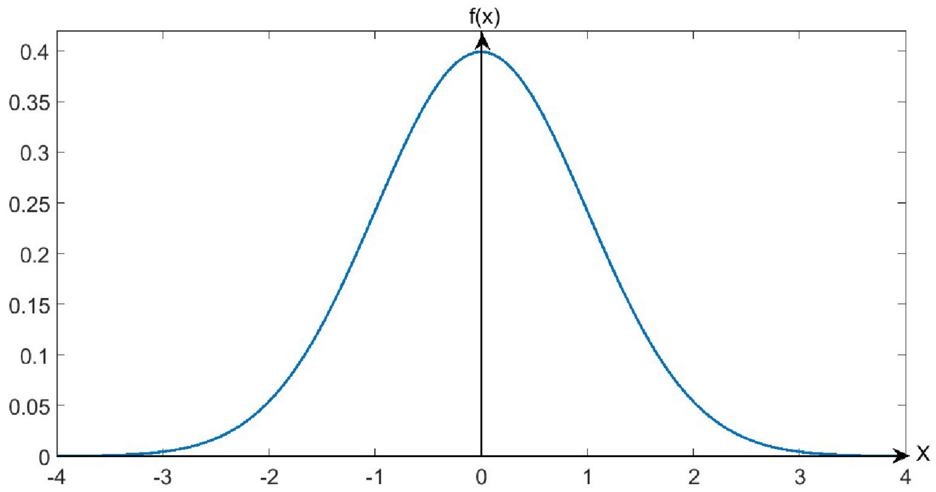
Among all the infinitely many different values for the parameters of the normal distribution, there is one particular pair of values. The standard normal variable Z has mean 0 and variance 1, i.e., $Z \sim N(0\ ,1\ )$. Importantly, most problems involving a normally distributed random variable can be solved by transforming the random variable into a standard normal variable. For this reason, we study the standard normal distribution.

The following formula is again displayed for completeness, and you are not expected to know it in the exam. Setting $\mu = 0$ and $\sigma^2 = 1$, we can find the probability density function for $Z \sim N(0,1)$, which is usually denoted as $\phi(z)$ (lower case Greek 'phi'), i.e.,

$$\phi\left(z\right), = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right)$$

for all real values of z. The graph of $\phi(z)$ is shown below.

Figure 71: PDF of Standard Normal Distribution



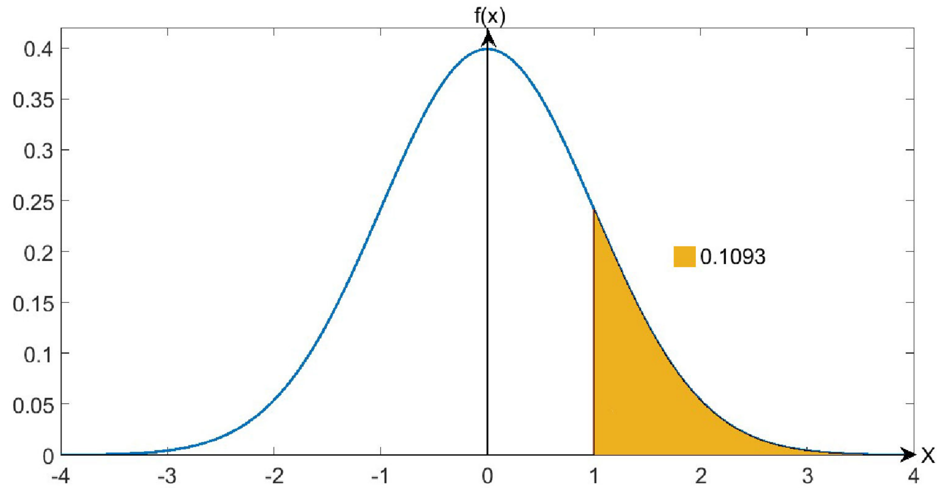
Source: Florian Pausinger, (2022).

Note that the axis of symmetry is the y-axis. Therefore, any $z < 0$ represents a value smaller than the mean, while any $z > 0$ represents a value that is greater than the mean. Moreover, notice that for $z < \ -3$ or $z > 3$ we have $\phi(z) \approx 0$. As usual, the area under the graph is equal to 1 and any vertical line drawn at a value of Z divides the area under the graph into two parts corresponding to $P(Z \leq z)$ and $P(Z > z)$. Importantly, the value $P(Z \leq z)$ is usually denoted as $\Phi(z)$ (upper case Greek 'Phi'). Since we cannot find such values by integration, there are large tables showing the value of $\Phi(z)$ for different values of z. Modern calculators as well as computer algebra systems are also able to give the values of $\Phi(z)$ as well as of its inverse $\Phi(z)^{-1}$.

Example 13.1: Given that $Z \sim N(0,1)$, find $P(Z \leq 1.23)$ and $P(Z > 1.23)$.

Solution. Looking t the table, we find that $P(Z \leq 1.23) = 0.8907$ and $P(Z > 1.23) = 0.1093$.

Figure 72: Illustration of Example 13.1



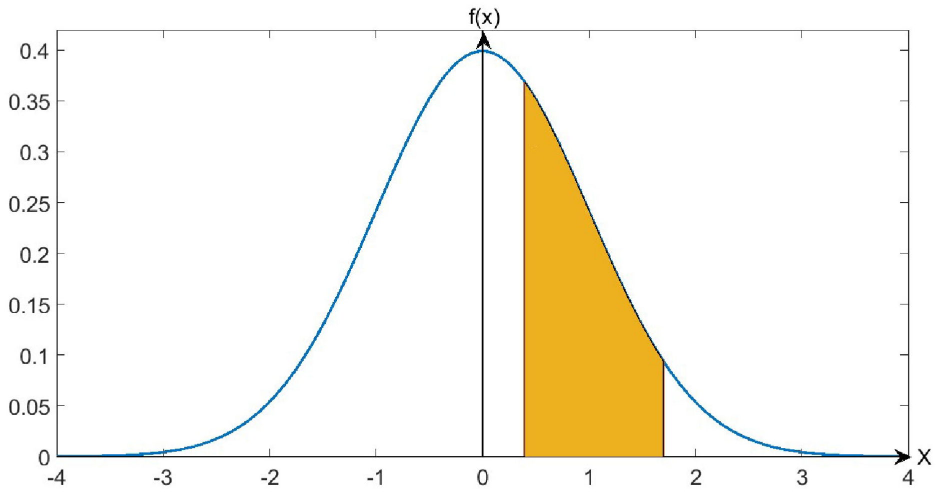
Source: Florian Pausinger, (2022).

Example 13.2: Given that $Z \sim N(0,1)$, find $P(0.4 \leq Z < 1.7)$ correct to three decimal places.

Solution. We have $\Phi(0.4) = 0.6554$ and $\Phi(1.7) = 0.9554$. Hence,

$$P(0.4 \leq Z < 1.7) = P(Z < 1.7) - P(Z < 0.4) = \Phi(1.7) - \Phi(0.4)$$
$$= 0.300$$

Figure 73: Illustration of Example 13.2



Source: Florian Pausinger, (2022).

Finally, we can use the symmetry properties of the normal curve to relate values of $\Phi(z)$ and $\Phi(-z)$. If $z > 0$ then

$$\Phi(z) = P(Z \leq z) \text{ and } 1 - \Phi(z) = P(Z \geq z)$$

On the other hand, by symmetry,

$$\Phi(z) = P(Z \geq -z) \text{ and } 1 - \Phi(z) = P(Z \leq -z)$$

Hence, we have that $P(Z \leq z) = P(Z \geq -z)$ and $P(Z \geq z) = P(Z \leq -z)$.

Standardizing a Normal Distribution

Given that we cannot evaluate the integrals associated to a normal distribution, it would be horrendous if we needed a table of values for each parameter pair μ and σ^2. Fortunately, it is possible to transform any normal curve into a standard normal. Let $X \sim N(\mu, \sigma^2)$. If we subtract μ from X, then the PDF is translated horizontally by $-\mu$ and is thus centred at 0. Now the new random variable $X - \mu$ has mean 0 and standard deviation σ. In a next step, we can multiply $X - \mu$ with $1/\sigma$. This gives a new random variable with standard deviation (and variance) equal to 1. In summary, let $X \sim N(\mu, \sigma^2)$, then

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

Probabilities in terms of X are now equal to probabilities involving the corresponding values of Z, which can be found from normal distribution value tables.

Example 13.3: Let $X \sim N(20{,}4)$. Find $P(X < 23)$.

Solution. $P(X < 23) = P\Big(Z < \frac{23-20}{2}\Big) = P(Z < 1.5) = 0.9332$

13.3 Modelling Discrete Situations

Gauss already showed that measurement errors can be modelled by a normal distribution. Since then, the normal distribution turned out to be an immensely useful and versatile tool in statistics. In the following, we discuss some examples of real-life applications in which the normal distribution turns out to be useful.

Example 13.4: Let the mass of a newborn baby in a certain city be denoted by M. Assume that $M \sim N(3.35, 0.0858\ )$, in which the numbers are in kg and kg². Estimate how many of the 1356 born in the previous year had masses of less than 3.5 kg.

Solution. First, we transform M to get a random variable with standard normal distribution, i.e., $Z = \frac{M - 3.35}{\sqrt{0.0858}}$. Now, we calculate

$$\Phi\Big(\frac{3.5 - 3.35}{\sqrt{0.0858}}\Big) = \Phi(0.512) = 0.6957$$

Therefore, $P(mass < 3.5\text{kg}) = 0.6957$. In other words, about 69.57% of all babies had a mass below 3.5 kg, i.e., about 943 of the 1356 babies.

Example 13.5: Let W denote the waiting time in minutes for patients in the emergency unit of a hospital. Assume that $W \sim N(13, 16)$.

(a) Calculate the probability that a randomly selected patient has to wait for more than 16.5 minutes.

(b) In a given month 468 patients attended the unit. Calculate an estimate of the number of people who would wait for less than nine minutes.

Solution. (a) We calculate

$$\Phi\Big(\frac{16.5 - 13}{\sqrt{16}}\Big) = \Phi(0.875) = 0.8092$$

Hence, $P(W > 16.5) = 1 - 0.8092 = 0.1908$.

(b) We calculate

$$\Phi\Big(\frac{9 - 13}{\sqrt{16}}\Big) = \Phi(-1) = 1 - \Phi(1) = 0.1587$$

Therefore, since $468 \cdot 0.1587 = 74.272$ about 74 people waited less than 9 minutes.

13.4 Using the Normal Distribution to Approximate the Binomial Distribution

Recall that the binomial distribution can be used to solve problems such as: Find the probability of obtaining exactly 40 heads with 70 tosses of a fair coin. The probability of this event is

$$\binom{70}{40} 0.5^{40} 0.5^{30}$$

And now imagine, you have to calculate this number without a computer or calculator. Even worse, imagine you need the probability of obtaining 40 or more heads. In this case you would need to calculate the probability of obtaining 40, 41, and so on and add them all together. Naturally, people thought of ways to approximate the binomial distribution.

It turns out that in certain situations, we can approximate a probability like the one above, with fewer calculations using the normal distribution. If $n \cdot p$ and $n \cdot q$ are both greater than 5, then the binomial distribution has approximately a normal shape. In this case $X{\sim}B(n, p)$ can be approximated by $N(np, np(1-p))$. Remember that np is the mean of the binomial distribution and $npq = np(1-p)$ is the variance.

For example, the distribution $X{\sim}B(40, 0.95)$ cannot be well approximated by a normal distribution because $n \cdot q = 40 \cdot 0.05 = 2 < 5$. On the other hand, the distribution $X{\sim}B(250, 0.8)$ can be well approximated by a normal distribution because $n \cdot p = 250 \cdot 0.8 = 200$ and $n \cdot q = 250 \cdot 0.2 = 50$.

When approximating a discrete distribution by a continuous distribution, we need to make use of a trick. Remember that $P(X = x) = 0$ for all x if X is a continuous random variable. Therefore, a discrete value such as $Y = 13$ for a discrete random variable Y is represented by the interval of values $12.5 \leq X < 13.5$ when approximated by the continuous random variable X. For this reason, $Y = 13$ must be replaced by either $X = 12.5$ or $X = 13.5$ depending on whether we wish to include the value 13 in the probability we are calculating or not. To be more concrete, $P(Y < 13)$ would be approximated by $P(X < 12.5)$, while $P(Y \leq 13)$ would be approximated by $P(X < 13.5)$. This trick is known as *"making a continuity correction."*

Example 13.6: Given a random variable $X{\sim}B(100, 0.4)$. Use suitable normal approximation and continuity correction to find $P(X < 43)$ and $P(X > 43)$.

Solution. We have that $\mu = np = 40$ and $\sigma^2 = np(1-p) = 24$. Since, both np and nq are greater than 5, we can approximate the binomial distribution by $N(40, 24)$.

For $P(X < 43)$ we use the continuity correction 42.5. Hence,

$$P(X < 43) \approx P\Big(Z < \frac{42.5 - 40}{\sqrt{24}}\Big) = \Phi(0.5103) = 0.695$$

For $P(X > 43)$ we use the continuity correction 43.5. Hence,

$$P(X > 43) \approx 1 - P(Z < (43.5 - 40)/\sqrt{24}) = 1 - \Phi(0.7144) = 0.2375$$

Example 13.7: Groups of 8000 people are randomly selected. It is found that 0.2% of the population have a favorable gene mutation. Use a suitable approximation to find the probability that (a) a sample group contains 20 or more people with this gene mutation; (b) exactly three out of four sample groups contain 20 or more people with this gene mutation.

Solution. We have that $X \sim B(8000, 0.002)$ which can be approximated with $N(16, 15.968)$ since both np and nq are greater than 5.

(a) To calculate $P(X \geq 20)$, we use the continuity correction 19.5. Hence,

$$P(X \geq 20) \approx 1 - P\left(Z < \frac{19.5 - 16}{\sqrt{15.968}}\right) = 1 - \Phi\left(0.876\right) = 0.1906$$

The probability that a group contains 20 or more people with this mutation is approximately 19%.

(b) Let W be the number of sample groups containing 20 or more people with this gene mutation. Then,

$$P(W = 3) = \binom{4}{3} \cdot 0.1906^3 \cdot 0.8094^1 = 0.0224$$

The probability that three out of four sample groups contain more than 20 people with this mutation is approximately 2.2%.

SUMMARY

A continuous random variable can take any value within a continuous range. The probability distribution of a continuous random variable is represented by a function known as the probability density function (PDF). An important example of a continuous random variable is the normal distribution, which is defined by two parameters, the mean and the variance. We write $X \sim N(\mu, \sigma^2)$ and say that X is normally distributed with mean μ and variance σ^2.

If $X \sim N(\mu, \sigma^2)$ we can always transform it into a so-called standard normal random variable $Z \sim N(0, 1)$ via the transformation $Z = \frac{X - \mu}{\sigma}$. This is important since the PDF of a normal distribution cannot be explicitly evaluated and, hence, we normally use large tables to find probabilities. Standardization of normally distributed random variables ensures we only need to keep one table with the value for the standard normal distribution.

The normal distribution can be used to approximate the binomial distribution. If $X \sim B(n, p)$ and if $n \cdot p$ and $n \cdot q$ are both greater than 5, then the binomial distribution can be approximated by $N(np, np(1-p))$.

Importantly, continuity corrections must be made whenever a discrete distribution is approximated by a continuous distribution.

BACKMATTER

LIST OF REFERENCES

Books

Abbott, S. (2016). *Understanding analysis* (2nd ed.). Springer.

Adrignola. (2011). *Radian.* Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Radian_-_en.svg

Pemberton, S. (2018). *Cambridge international AS & A Level mathematics: Pure mathematics 1 coursebook*. Cambridge University Press.

Rohieb. (2016). *Pascal triangle small.* Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Pascal_triangle_small.svg

LIST OF TABLES AND FIGURES

IU Internationale Hochschule GmbH
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt

Mailing Address
Albert-Proeller-Straße 15-19
D-86675 Buchdorf

media@iu.org
www.iu.org

Help & Contacts (FAQ)
On myCampus you can always find answers
to questions concerning your studies.