



**Probabilistische Datenwissenschaft
für die Psychologie**

Dirk Ostwald

Table of contents

Welcome	3
Citation	3
License	3
I Mathematical Foundations	4
1 Language and Logic	6
1.1 Mathematics Is a Language	6
1.2 Basic Building Blocks	7
1.3 Propositional Logic	8
1.4 Equivalence Transformations	13
1.5 Proof Techniques	15
2 Sets	16
2.1 Basic Definitions	16
2.2 Operations	19
2.3 Special Sets	20
3 Sums, Products, Powers	25
3.1 Sums	25
3.2 Products	28
3.3 Powers	30
4 Functions	32
4.1 Definition and Properties	32
4.2 Types of Functions	37
4.3 Elementary Functions	40
References	44

Welcome

Welcome to the working version of *Probabilistic Data Science for Psychology (PDSP)*, a textbook on data-scientific methodology at the Institute of Psychology at Otto von Guericke University Magdeburg.



Figure 1 Probabilistic Data Science for Psychology

Citation

Ostwald, D. (2024) Probabilistic Data Science for Psychology. [10.5281/zenodo.10730199](https://doi.org/10.5281/zenodo.10730199)

License

This work is licensed under a Creative Commons Attribution 4.0 International License.

Part I

Mathematical Foundations

As the language of natural-scientific modeling, mathematics is also the language of probabilistic data science in psychology. In this role, mathematics mediates between intuitive scientific language and the variety of programming languages used to implement data-scientific analyses. Mathematical concept formation makes it possible to grasp concepts very precisely and in a generally understandable way, and to communicate them independently of the imprecision of everyday scientific language or the idiosyncrasies of different programming languages. Of course, mathematical language cannot assume this role detached from scientific-language intuition or practical application. In probabilistic data analysis, mathematics is therefore never an end in itself, but must always be understood as applied mathematics.

In this part, after some thoughts on the fundamental concepts of mathematics (Chapter 1), we present sets and functions (Chapter 2 and Chapter 4), the two pillars of modern mathematics, in compressed form. We supplement this presentation with some notational aspects in Chapter 3. Closely interwoven with the concept of a function are differential calculus and integral calculus. Here, too, the aim is not an exhaustive presentation, but in particular an explanation of some foundations that are central to the data-scientific concepts of optimization and to working with probability functions. We supplement these sections with some introductory thoughts on the analytical foundations of differential and integral calculus. In dealing with the large data sets of contemporary science, matrix notation has proved useful. We first consider this subfield of linear algebra for single-column matrices and then for matrices in detail, again with the aim of providing notational and computational foundations for later sections.

Overall, this part therefore primarily serves as a brief introduction to foundations that will be taken up again in other parts, and in particular also as the introduction of a unified notation. Depending on need and interest, more in-depth self-study of the various contents is of course advisable. In the following literature notes, we provide an overview of sources and reading recommendations for the contents considered here.

Literature Notes

Bärwolff (2017) and Arens et al. (2018) form the primary basis for many of the topics presented here and offer an excellent and comprehensive entry point into the material covered. At the international level, Spivak (2008) and Strang (2009) provide very readable introductions to differential calculus and linear algebra, respectively. A deeper understanding, especially of analytical concepts, is provided, for example, by Abbott (2015) and Chiossi (2021). Searle (1982) offers a compact presentation of those aspects of matrix calculus and linear algebra that are used mainly in probabilistic data analysis. Deisenroth et al. (2020) provides an overview of many of the topics treated here under the moniker of machine learning.

1 Language and Logic

1.1 Mathematics Is a Language

Mathematics is the language of natural-scientific model building. For example, the expression

$$F = ma \tag{1.1}$$

corresponds, in the sense of Newton's second axiom, to a theory of the motion of objects under the action of forces (Newton (1687)). Likewise, the expression

$$\max_{q(z)} \int q(z) \ln \left(\frac{p(y, z)}{q(z)} \right) dz \tag{1.2}$$

corresponds, in the sense of variational inference, to contemporary theory about how the brain works (Friston (2005)). Mathematical symbolism serves in particular to communicate scientific insights precisely and aims to describe complex matters exactly and efficiently. As in reflective engagement with any form of language, the question "What is this supposed to mean?" therefore always remains the guiding question when working with mathematical content and symbolism.

As a language system, mathematics has a number of special features. First, its contents are often abstract. This is because mathematics strives for the broadest possible general comprehensibility and applicability. Mathematical approaches to the phenomena of the world are interested in making insights as easy as possible to transfer to other contexts. To make this possible, mathematics tries to work as precisely and understandably as possible, that is, in terms of precise concepts. It proceeds in a strictly hierarchical way, so that concepts introduced later often presuppose a good understanding of the underlying concepts introduced earlier.

The precision of mathematical language implies a high density of information. It is therefore rather sober and leaves out what is superfluous, so that, in the best case, *everything* in a mathematical text is relevant for communicating an idea. As recipients of mathematical texts, we perceive their information density through the high cognitive energy required when reading such a text. This high energy expenditure calls in particular for calm and slowness when reading with the aim of genuine understanding. The following quotation may serve as a guiding principle for working with mathematical texts: "A mathematical text cannot be read like a novel; one has to work through it" (Unger (2000)). After reading a short mathematical text, one should always ask critically whether one has really understood what has been read or whether further sources should be consulted to clarify the matter. It is also helpful, in the spirit of Richard Feynman's famous quotation "What I cannot create, I do not understand", to make one's own notes and to reconstruct mathematical language systems oneself.

If one wants to gain access to the world of natural-scientific model building, then it is helpful, when dealing with its mathematical expressions and symbolism, to use the same

strategies as when learning a foreign language. In addition to immersing oneself in the corresponding language space, that is, constant exposure to mathematical modes of expression, this certainly includes first memorizing concepts and translating texts into everyday language. A deep and secure understanding of mathematical model building then arises in particular through the application of mathematical ways of thinking in written and spoken form.

1.2 Basic Building Blocks

In the following, we introduce three basic building blocks of mathematical communication that will accompany us throughout: the concepts of a *definition*, a *theorem*, and a *proof*.

Definition

A *definition* is a cornerstone of a mathematical system that is neither justified nor deductively derived within that system. Definitions can only be evaluated according to their usefulness within a mathematical system. A definition is best memorized first and questioned only once one has understood its use in application or is not convinced by it. Some relaxation and calm are generally helpful when dealing with definitions that appear complex at first sight. To indicate that we define a symbol as something, we use the notation “:=”. For example, the expression “ $a := 2$ ” defines the symbol a as the number two. In this text, definitions always end with the symbol \bullet .

Theorem

A *theorem* is a mathematical statement that can be recognized as correct (true) by means of a proof. That is, a theorem is always derived from definitions and/or other theorems. In this sense, theorems are the empirical results of mathematics. In applied, data-analytic mathematics, theorems are often helpful for calculations. It is therefore worth memorizing them, because they usually form the basis for data analysis and data interpretation. Equations often appear in theorems. These equations arise from the assumptions of the theorem. To mark equations, we use the equals sign “=”. For example, in a given context, the expression “ $a = 2$ ” says that, because of certain assumptions, the variable a has the value two. In this text, theorems always end with the symbol \circ .

Proof

A *proof* is a chain of argumentation that draws on known definitions and theorems to establish the correctness (truth) of a theorem. Short proofs often contribute to understanding a theorem; long proofs tend to do so less. Proofs are therefore, in particular, answers to the question why a mathematical statement holds (“Why is this so?”). Proofs are not memorized. When proofs are short, it is sensible to work through them, because they usually repeat content that is assumed to be known. When they are long, it is more sensible to skip them at first so as not to get lost in details and lose the main path through the mathematical structure. In this text, proofs always end with the symbol \square .

In addition to definitions, theorems, and proofs, mathematical texts contain other typical building blocks, such as *axioms*, *lemmas*, *corollaries*, and *conjectures*. We will not use these terms and therefore give only a brief overview of them.

- *Axioms* are unproved fundamental assumptions in the sense that they serve as starting points for building mathematical systems. The transition between definitions and axioms is often fluid. Because we will not work at a particularly deep mathematical level, we prefer the term definition in almost all cases.
- A *lemma* is an “auxiliary theorem”, that is, a mathematical statement that is proved but is not as important as a theorem. Because, on the one hand, we focus on important content and, on the other hand, we do not want to discriminate among mathematical statements, we avoid this term and instead use the term theorem throughout.
- A *corollary* is a mathematical statement that follows from a theorem by a simple proof. Because the “simplicity” of mathematical proofs is a relative property, we avoid this term and instead also use the term theorem throughout.
- *Conjectures* are mathematical statements for which it is unknown whether they are provable or refutable. Because we work in the area of applied mathematics, we will not encounter conjectures.

1.3 Propositional Logic

Having now become acquainted with some basic building blocks of mathematical language systems, we turn to *propositional logic*, a simple system that allows relationships between mathematical statements to be established and formalized. Propositional logic plays a central role, for example, in the definition of set operations, in optimization conditions for functions, and in many proofs. In data-analytic applications, propositional logic is the basis of Boolean logic in programming. In mathematical psychology, propositional logic is the basis of the representational theory of measurement.

We begin with the definition of the concept of a mathematical *statement*.

Definition 1.1 (Statement). A *statement* is a sentence to which the property *true* or *false* can be assigned unambiguously.

•

The adjective *true* can also be understood as *correct*. We abbreviate true by “t” and false by “f”. In the field of real numbers, for example, the statement $1 + 1 = 2$ is true and the statement $1 + 1 = 3$ is false. Note that the binarity of the truth value of statements is a basic assumption of propositional logic and therefore to be understood formally and scientifically, not empirically. Truth values do not refer to definitions; definitions fix meanings.

A first way of working with statements is to negate them. This leads to the following definition.

Definition 1.2 (Negation). Let A be a statement. Then the *negation of A* is the statement that is false when A is true and true when A is false. The negation of A is denoted by $\neg A$, read as “not A ”.

•

For example, the negation of the statement “The sun is shining” is the statement “The sun is not shining”. The negation of the statement $1 + 1 = 2$ is the statement $1 + 1 \neq 2$, and the negation of the statement $x > 1$ is the statement $x \leq 1$. The definition of the negation of a statement A is represented in tabular form as follows:

Table 1.1 Truth table of negation

A	$\neg A$
t	f
f	t

Tables of this form are called *truth tables*. They are a popular tool in propositional logic, and we will use them often in the following.

If one wants to connect two statements logically, the concepts of *conjunction* and *disjunction* are available.

Definition 1.3 (Conjunction). Let A and B be statements. Then the *conjunction of A and B* is the statement that is true if and only if both A and B are true. The conjunction of A and B is denoted by $A \wedge B$, read as “ A and B ”.

•

The definition of conjunction implies the following truth table:

Table 1.2 Truth table of conjunction

A	B	$A \wedge B$
t	t	t
t	f	f
f	t	f
f	f	f

As an example, let A be the statement $2 \geq 1$ and let B be the statement $2 > 1$. Since both A and B are true, the statement $(2 \geq 1) \wedge (2 > 1)$ is also true. As another example, let A be the statement $1 \geq 1$ and let B be the statement $1 > 1$. Here, only A is true and B is false. Thus the statement $(1 \geq 1) \wedge (1 > 1)$ is false.

Definition 1.4 (Disjunction). Let A and B be statements. Then the *disjunction of A and B* is the statement that is true if and only if at least one of the two statements A and B is true. The disjunction of A and B is denoted by $A \vee B$, read as “ A or B ”.

•

The definition of disjunction implies the following truth table:

Table 1.3 Truth table of disjunction

A	B	$A \vee B$
t	t	t
t	f	t
f	t	t
f	f	f

$A \vee B$ is therefore also true when both A and B are true. The “or” considered here is thus more precisely an “and/or”. For this reason, disjunction is also called a “non-exclusive or”. As an example, let A be the statement $2 \geq 1$ and let B be the statement $2 > 1$. A is true and B is true. Thus the statement $(2 \geq 1) \vee (2 > 1)$ is true. Now again let A be the statement $1 \geq 1$ and let B be the statement $1 > 1$. Then A is true and B is false. Thus the statement $(1 \geq 1) \vee (1 > 1)$ is true.

One way of placing statements in a mechanical logical relationship is *implication*. It is defined as follows.

Definition 1.5 (Implication). Let A and B be statements. Then the *implication*, denoted by $A \Rightarrow B$, is the statement that is false if and only if A is true and B is false. A is called the *assumption (premise)* and B the *conclusion* of the implication. $A \Rightarrow B$ is read as “from A follows B ”, “ A implies B ”, or “if A , then B ”.

•

The definition of implication can be represented by the following truth table:

Table 1.4 Truth table of implication

A	B	$A \Rightarrow B$
t	t	t
t	f	f
f	t	t
f	f	t

An intuitive understanding of the definition of implication in the sense of the truth table above is obtained most readily by reading it as an attempt to capture and formalize the intuitive idea of a consequence in the context of propositional logic. To understand this, it is best to read the truth states in Table 1.4 in the order truth state of A , truth state of $A \Rightarrow B$, and finally truth state of B . Read in this way, the truth table shows that if A is true and $A \Rightarrow B$ is true, then B is true. Thus, if one constructs a true implication based on a true statement, for example by transforming equations, it follows that B is also true. If this is not possible, that is, if A is true and $A \Rightarrow B$ is always false, then B is also false. This is how statements may be refuted. Finally, one sees that if A is false and $A \Rightarrow B$ is true, then B can be either true or false. Only from a true premise does a true implication always yield a true conclusion. In particular, the definition of implication therefore satisfies the principle “anything follows from falsehood (ex falso sequitur quodlibet)”. Thus one cannot infer anything meaningful from false statements by means of implication.

In the context of implication, the concepts of *sufficient* and *necessary* conditions arise: If $A \Rightarrow B$ is true, one says that “ A is *sufficient* for B ” and that “ B is *necessary* for A ”. This terminology is explained in the context of implication as follows: If $A \Rightarrow B$ is true, then if A is true, B is also true. The truth of A is therefore enough for the truth of B . Thus A is sufficient for B . Furthermore, if $A \Rightarrow B$ is true, then if B is false, A is also false. The truth of B is therefore necessary for the truth of A .

Finally, to illustrate Table 1.4, we give a concrete everyday example. Let A be the statement “The bridge is damaged”, let B be the statement “The bridge is closed”, and let $A \Rightarrow B$ be the implication “If the bridge is damaged, then the bridge is closed”. We discuss the four cases of Table 1.4.

- A true, B true, $A \Rightarrow B$ true. If A is true, then the bridge is damaged, and if B is also true, then the bridge is also closed. In this case, the implication “If the bridge is damaged, then the bridge is closed” is therefore obviously true.
- A true, B false, $A \Rightarrow B$ false. If A is true, then the bridge is damaged, and if B is false, then the bridge is not closed. In this case, the implication “If the bridge is damaged, then the bridge is closed” is therefore obviously false.
- A false, B true, $A \Rightarrow B$ true. If A is false, then the bridge is not damaged, and if B is true, then the bridge is closed. Since the bridge is not damaged, however, no conclusion can be drawn regarding the implication “If the bridge is damaged, then the bridge is closed”; in particular, one cannot show that it would be false. Thus one assumes that the statement “If the bridge is damaged, then the bridge is closed” is true.
- A false, B false, $A \Rightarrow B$ true. If A is false, then the bridge is not damaged, and if B is false, then the bridge is not closed. Since the bridge is not damaged, however, no conclusion can be drawn regarding the implication “If the bridge is damaged, then the bridge is closed”; in particular, one cannot show that it would be false. Thus one again assumes that the statement “If the bridge is damaged, then the bridge is closed” is true.

A very frequently occurring relation between two statements is their *equivalence*.

Definition 1.6 (Equivalence). Let A and B be statements. The *equivalence of A and B* is the statement that is true if and only if A and B are both true or if A and B are both false. The equivalence of A and B is denoted by $A \Leftrightarrow B$ and read as “ A if and only if B ” or “ A is equivalent to B ”.

•

The definition of equivalence implies the following truth table:

Table 1.5 Truth table of equivalence

A	B	$A \Leftrightarrow B$
t	t	t
t	f	f
f	t	f
f	f	t

The definition of the concept of *logical equivalence* allows, among other things, the equivalence of two statements to be established by means of implications.

Definition 1.7 (Logical Equivalence). Two statements are called *logically equivalent* if their truth tables are the same. •

As examples of logical equivalences that are frequently used in proof arguments, we show the statements of the following theorem.

Theorem 1.1 (Logical Equivalences). *Let A and B be two statements. Then the following statements are logically equivalent:*

- (1) $A \Leftrightarrow B$ and $(A \Rightarrow B) \wedge (B \Rightarrow A)$
- (2) $A \Rightarrow B$ and $(\neg B) \Rightarrow (\neg A)$

◦

Proof. By the definition of logical equivalence, we have to show that the truth tables of the statements under consideration are the same. We show first (1), then (2).

(1) We recall the truth table of $A \Leftrightarrow B$:

A	B	$A \Leftrightarrow B$
t	t	t
t	f	f
f	t	f
f	f	t

We further consider the truth table of $(A \Rightarrow B) \wedge (B \Rightarrow A)$:

A	B	$A \Rightarrow B$	$B \Rightarrow A$	$(A \Rightarrow B) \wedge (B \Rightarrow A)$
t	t	t	t	t
t	f	f	t	f
f	t	t	f	f
f	f	t	t	t

Comparing the truth table of $A \Leftrightarrow B$ with the first two and the last column of the truth table of $(A \Rightarrow B) \wedge (B \Rightarrow A)$ shows their equality.

(2) We recall the truth table of $A \Rightarrow B$:

A	B	$A \Rightarrow B$
t	t	t
t	f	f
f	t	t
f	f	t

We further consider the truth table of $(\neg B) \Rightarrow (\neg A)$:

A	B	$\neg B$	$\neg A$	$(\neg B) \Rightarrow (\neg A)$
t	t	f	f	t
t	f	t	f	f
f	t	f	t	t
f	f	t	t	t

Comparing the truth table of $A \Rightarrow B$ with the first two and the last column of the truth table of $(\neg B) \Rightarrow (\neg A)$ shows their equality. \square

The first statement of Theorem 1.1 says that the statement “ A and B are equivalent” is logically equivalent to the statement “from A follows B and from B follows A ”. This is the basis for many so-called *direct proofs* using equivalence transformations. The second statement of Theorem 1.1 says that the statement “from A follows B ” is logically equivalent to the statement “from not B follows not A ”. This is the basis for the technique of *indirect proof*. We consider these proof techniques more closely in the following section. First, we summarize the meanings of the symbols introduced in this section once more in the table below.

Table 1.10 Symbol overview. A and B are statements, that is, A and B are either true or false.

Symbol	Meaning	Note
$\neg A$	Not A	True when A is false, and vice versa
$A \wedge B$	A and B	True only when both A and B are true
$A \vee B$	A and/or B	True when at least one of the statements is true
$A \Rightarrow B$	From A follows B	B is necessary for A , A is sufficient for B
$A \Leftrightarrow B$	A is equivalent to B	Both $A \Rightarrow B$ and $B \Rightarrow A$ hold

1.4 Equivalence Transformations

Mathematical problems often lead to the case in which information about an unknown variable is represented implicitly by means of an equation or an inequality. To represent the information about the corresponding variable explicitly, that is, to determine its value in the case of equations or to determine the range of numbers in which the value of the variable lies in the case of inequalities, one uses *equivalence transformations*. Here, we consider only the equivalence transformations of equations and inequalities with respect to real variables that are familiar from school mathematics. Equivalence transformations of equations and inequalities have the property that the truth value of the statement formulated by an equation or inequality does not change when the corresponding transformation is applied. Both sides of the equation or inequality are always transformed. For the application of a mathematical operation that transforms an equation or inequality statement A into an equation or inequality statement B to be an equivalence transformation of the form $A \Leftrightarrow B$, both $A \Rightarrow B$ and $B \Rightarrow A$ must hold, as is well known. This implies that equivalence transformations are reversible (invertible) operations.

For equations, admissible equivalence transformations include in particular

- adding a number to both sides of the equation,
- subtracting a number from both sides of the equation,
- multiplying both sides of the equation by a nonzero number,
- dividing both sides of the equation by a nonzero number, and

- applying an invertible function to both sides of the equation.

Example

Anticipating Section 4.3, we consider the statement

$$2 \exp(x) - 2 = 0. \quad (1.3)$$

Then

$$\begin{aligned} 2 \exp(x) - 2 &= 0 \\ \Leftrightarrow 2 \exp(x) - 2 + 2 &= +2 \\ \Leftrightarrow 2 \exp(x) &= 2 \\ \Leftrightarrow \frac{1}{2} \cdot 2 \exp(x) &= \frac{1}{2} \cdot 2 \\ \Leftrightarrow \exp(x) &= 1 \\ \Leftrightarrow \ln(\exp(x)) &= \ln(1) \\ \Leftrightarrow x &= 0. \end{aligned} \quad (1.4)$$

In summary, therefore,

$$2 \exp(x) - 2 = 0 \Leftrightarrow x = 0. \quad (1.5)$$

For inequalities, admissible equivalence transformations include in particular

- adding a number to both sides of the inequality,
- subtracting a number from both sides of the inequality,
- multiplying both sides of the inequality by a nonzero number, with multiplication by a negative number reversing the inequality sign,
- dividing both sides of the inequality by a nonzero number, with division by a negative number reversing the inequality sign, and
- applying an invertible monotonically increasing function to both sides of the inequality; for monotonically decreasing functions, the inequality sign is reversed.

Example

We consider the statement

$$-5x - 2 \geq 8. \quad (1.6)$$

Then

$$\begin{aligned} -5x - 2 &\geq 8 \\ \Leftrightarrow -5x - 2 + 2 &\geq 8 + 2 \\ \Leftrightarrow -5x &\geq 10 \\ \Leftrightarrow -\frac{1}{5} \cdot 5x &\geq \frac{1}{5} \cdot 10 \\ \Leftrightarrow -x &\geq 2 \\ \Leftrightarrow x &\leq -2. \end{aligned} \quad (1.7)$$

In summary, therefore,

$$-5x - 2 \geq 8 \Leftrightarrow x \leq -2. \quad (1.8)$$

1.5 Proof Techniques

In this section, we outline three fundamental proof techniques: *direct* and *indirect proofs*, and *proof by contradiction*. In what follows, especially the first technique will be used repeatedly to justify theorems.

We have:

- *Direct proofs* use equivalence transformations to show $A \Rightarrow B$.
- *Indirect proofs* use the logical equivalence of $A \Rightarrow B$ and $(\neg B) \Rightarrow (\neg A)$.
- *Proofs by contradiction* show that $(\neg B) \wedge A$ is false.

Equipped with this, we now prove the following theorem by means of a direct proof, an indirect proof, and a proof by contradiction (cf. Arens et al. (2018)).

Theorem 1.2 (Squares of Positive Numbers). *Let a and b be two positive numbers. Then $a^2 < b^2 \Rightarrow a < b$.*

◦

Proof. We first give a *direct proof*. Let $a^2 < b^2$ be the statement A and let $a < b$ be the statement B . Then

$$a^2 < b^2 \Leftrightarrow 0 < b^2 - a^2 \Leftrightarrow 0 < (b + a)(b - a) \Leftrightarrow 0 < (b - a) \Leftrightarrow a < b. \quad (1.9)$$

We now give an *indirect proof*. Let $a^2 \geq b^2$ be the statement $\neg A$. Furthermore, let $a \geq b$ be the statement $\neg B$. Then

$$a \geq b \Leftrightarrow a^2 \geq ab \wedge ab \geq b^2 \Leftrightarrow a^2 \geq b^2. \quad (1.10)$$

Finally, we give a *proof by contradiction*. To do so, we show that the assumption $(\neg B) \wedge A$ leads to a false statement. We have

$$a \geq b \wedge a^2 < b^2 \Leftrightarrow a^2 \geq ab \wedge a^2 < b^2 \Leftrightarrow ab \leq a^2 < b^2. \quad (1.11)$$

Furthermore,

$$a \geq b \wedge a^2 < b^2 \Leftrightarrow ab \geq b^2 \wedge a^2 < b^2 \Leftrightarrow a^2 < b^2 \leq ab. \quad (1.12)$$

Overall, this yields the false statement

$$ab \leq a^2 < b^2 \leq ab \Leftrightarrow ab < ab. \quad (1.13)$$

□

2 Sets

2.1 Basic Definitions

Sets collect mathematical objects and form the foundation of modern mathematics. We begin with the following definition.

Definition 2.1 (Sets). Following Cantor (1895), a *set* M is defined as “a collection into a whole of definite, distinct objects m of our intuition or of our thought (which are called the elements of the set)”. We write

$$m \in M \text{ or } m \notin M \tag{2.1}$$

to express that m is an element or not an element of M , respectively.

•

There are at least the following ways of defining sets:

- Listing the elements in curly braces, for example $M := \{1, 2, 3\}$.
- Specifying the properties of the elements, for example $M := \{x \in \mathbb{N} \mid x < 4\}$.
- Equating the set with another uniquely defined set, for example $M := \mathbb{N}_3$.

The notation $\{x \in \mathbb{N} \mid x < 4\}$ is read as “ $x \in \mathbb{N}$ such that $x < 4$ ”, where the meaning of \mathbb{N} will be explained below. It is important to recognize that sets are *unordered* mathematical objects, that is, the order in which the elements of a set are listed does not matter. For example, $\{1, 2, 3\}$, $\{1, 3, 2\}$, and $\{2, 3, 1\}$ denote the same set, namely the set of the first three natural numbers.

Basic relations between several sets are fixed in the next definition.

Definition 2.2 (Subsets and Set Equality). Let M and N be two sets.

- A set M is called a *subset* of a set N if, for every element $m \in M$, we also have $m \in N$. If M is a subset of N , we write

$$M \subseteq N \tag{2.2}$$

and call M a *subset* of N and N a *superset* of M .

- A set M is called a *proper subset* of a set N if, for every element $m \in M$, we also have $m \in N$, but there is at least one element $n \in N$ for which $n \notin M$. If M is a proper subset of N , we write

$$M \subset N. \tag{2.3}$$

- Two sets M and N are called *equal* if, for every element $m \in M$, we also have $m \in N$, and if, for every element $n \in N$, we also have $n \in M$. If the sets M and N are equal, we write

$$M = N. \quad (2.4)$$

•

Example

For example, consider the sets $M := \{1\}$, $N := \{1, 2\}$, and $O := \{1, 2\}$. Then, by the definitions above, $M \subset N$, because $1 \in M$ and $1 \in N$, but $2 \in N$ and $2 \notin M$. Furthermore, $N \subseteq O$, because $1 \in N$ and $1 \in O$, as well as $2 \in N$ and $2 \in O$, and there is no element of O that is not in N . Likewise, $O \subseteq N$, because $1 \in O$ and $1 \in N$, as well as $2 \in O$ and $2 \in N$, and there is no element of N that is not in O . Finally, we even have $N = O$, because, for every element $n \in N$, we also have $n \in O$, and at the same time, for every element $o \in O$, we also have $o \in N$. We represent these relations schematically by means of *Venn-Euler diagrams* in Figure 2.1.

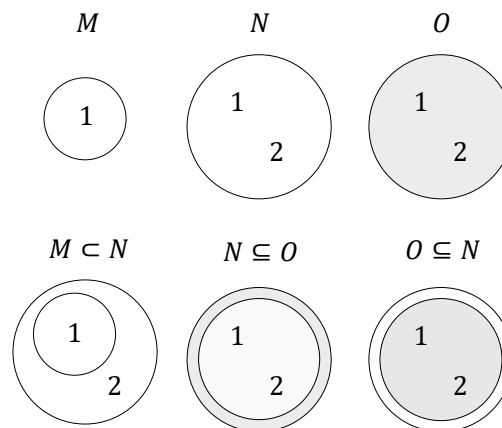


Figure 2.1 Venn-Euler diagrams of the subset properties of the sets $M := \{1\}$, $N := \{1, 2\}$, and $O := \{1, 2\}$.

An important property of a set is the number of elements it contains. This is called the *cardinality* of the set.

Definition 2.3 (Cardinality). The number of elements of a set M is called its *cardinality* and is denoted by $|M|$.

•

A special set is the set without elements.

Definition 2.4. A set with cardinality zero is called the *empty set* and is denoted by \emptyset .

•

As examples, let $M := \{1, 2, 3\}$, $N = \{a, b, c, d\}$, and $O := \emptyset$. Then $|M| = 3$, $|N| = 4$, and $|O| = 0$.

For every set, one can consider the set of all subsets of this set. This leads to the important concept of the *power set*.

Definition 2.5 (Power Set). The set of all subsets of a set M is called the *power set* of M and is denoted by $\mathcal{P}(M)$.

•

Note that the empty subset of M and M itself are also always elements of $\mathcal{P}(M)$. We first consider four examples of the concept of a power set.

- Let $M_0 := \emptyset$ be the empty set. Then

$$\mathcal{P}(M_0) = \{\emptyset\}. \quad (2.5)$$

- Let M_1 be the one-element set $M_1 := \{a\}$. Then

$$\mathcal{P}(M_1) = \{\emptyset, \{a\}\}. \quad (2.6)$$

- Let $M_2 := \{a, b\}$. Then M_2 has both one-element and two-element subsets, and

$$\mathcal{P}(M_2) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}. \quad (2.7)$$

- Finally, let $M_3 := \{a, b, c\}$. Then M_3 has, among others, one-element, two-element, and three-element subsets, and

$$\mathcal{P}(M_3) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}. \quad (2.8)$$

Theorem 2.1 (Cardinality of the Power Set). *Given a set M with cardinality $|M| = n$, and let $\mathcal{P}(M)$ be its power set. Then $|\mathcal{P}(M)| = 2^n$.*

◦

Proof. To prove the statement of the theorem, we associate each element P of the power set of M uniquely with a binary sequence of length n , where the entry at the i -th position represents whether the i -th element of M is an element of P or not. For example, let $M := \{m_1, m_2, m_3\}$ and $P := \{m_2, m_3\}$. Then P corresponds to the binary sequence 011. The empty set $P := \emptyset$ corresponds to the binary sequence 000, and the original set $P = M$ corresponds to the binary sequence 111. Thus the question is how many unique binary sequences of length n there are. Since each element of the sequence has two possible states, there are n factors $2 \cdot 2 \cdots 2$, that is, 2^n . \square

In the examples above, we have the cases

- $|M_0| = 0 \Rightarrow |\mathcal{P}(M_0)| = 2^0 = 1,$
- $|M_1| = 1 \Rightarrow |\mathcal{P}(M_1)| = 2^1 = 2,$
- $|M_2| = 2 \Rightarrow |\mathcal{P}(M_2)| = 2^2 = 4,$
- $|M_3| = 3 \Rightarrow |\mathcal{P}(M_3)| = 2^3 = 8,$

as can be verified by counting the elements of the corresponding power sets.

2.2 Operations

Two sets can be combined with one another in different ways. The result of such a combination is another set. We call the combination of two sets a *set operation* and give the following definitions.

Definition 2.6 (Set Operations). Let M and N be two sets.

- The *union of M and N* is defined as the set

$$M \cup N := \{x | x \in M \vee x \in N\}, \quad (2.9)$$

where \vee is to be understood according to Definition 1.4 as a *non-exclusive or*, that is, as and/or.

- The *intersection of M and N* is defined as the set

$$M \cap N := \{x | x \in M \wedge x \in N\}. \quad (2.10)$$

If M and N satisfy $M \cap N = \emptyset$, then M and N are called *disjoint*.

- The *difference of M and N* is defined as the set

$$M \setminus N := \{x | x \in M \wedge x \notin N\}. \quad (2.11)$$

The difference of M and N is also called, especially when $N \subseteq M$, the *complement of N with respect to M* and is denoted by N^c . In this context, M is also called the *basic set* or the *universe*.

- The *symmetric difference of M and N* is defined as the set

$$M \Delta N := \{x | (x \in M \vee x \in N) \wedge x \notin M \cap N\}. \quad (2.12)$$

The symmetric difference can therefore be understood as an *exclusive or*.

•

Example

As an example, consider the sets $M := \{1, 2, 3\}$ and $N := \{2, 3, 4, 5\}$. Then

- $M \cup N = \{1, 2, 3, 4, 5\}$, because $1 \in M$, $2 \in M$, $3 \in M$, $4 \in N$, and $5 \in N$.
- $M \cap N = \{2, 3\}$, because only for 2 and 3 do we have $2 \in M, 3 \in M$ and also $2 \in N, 3 \in N$. For 1, we only have $1 \in M$, and for 4 and 5, we only have $4 \in N$ and $5 \in N$.
- $M \setminus N = \{1\}$, because $1 \in M$, but $1 \notin N$, and $2 \in M$, but also $2 \in N$.
- $N \setminus M = \{4, 5\}$, because $4 \in N$ and $5 \in N$, but $4 \notin M$ and $5 \notin M$. This shows in particular that the difference of M and N is *not* symmetric, that is, it is *not* necessarily the case that $M \setminus N$ equals $N \setminus M$.
- $M \Delta N = \{1, 4, 5\}$, because $1 \in M$, but $1 \notin \{2, 3\}$, $2 \in M$, but $2 \in \{2, 3\}$, $3 \in M$, but $3 \in \{2, 3\}$, $4 \in N$, but $4 \notin \{2, 3\}$, and $5 \in N$, but $5 \notin \{2, 3\}$.

Figure 2.2 visualizes the set operations considered in this example.

Finally, we introduce the concept of a partition of a set.

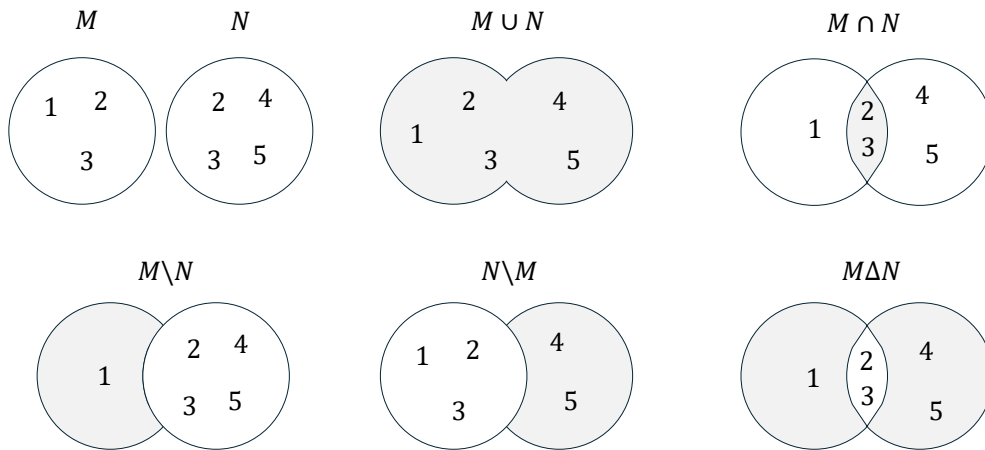


Figure 2.2 Venn-Euler diagrams of the set operations considered in the example for $M := \{1, 2, 3\}$ and $N := \{2, 3, 4, 5\}$. The gray shaded regions correspond to the resulting sets.

Definition 2.7 (Partition). Let M be a set, and let $P := \{N_i\}$ be a set of sets N_i with $i = 1, \dots, n$, such that

$$(M = \cup_{i=1}^n N_i) \wedge (N_i \cap N_j = \emptyset \text{ for } i, j = 1, \dots, n \text{ and } i \neq j). \tag{2.13}$$

Then P is called a *partition of M* .

•

The partition of a set therefore corresponds to splitting the set into disjoint subsets. Partitions are generally not unique, that is, there are usually several ways to partition a given set.

As an example, consider the set $M := \{1, 2, 3, 4, 5, 6\}$. Then $P_1 := \{\{1\}, \{2, 3, 4, 5, 6\}\}$, $P_2 := \{\{1, 2, 3\}, \{4, 5, 6\}\}$, and $P_3 := \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ are three possible partitions of M . Figure 2.3 visualizes the partitions considered in this example.

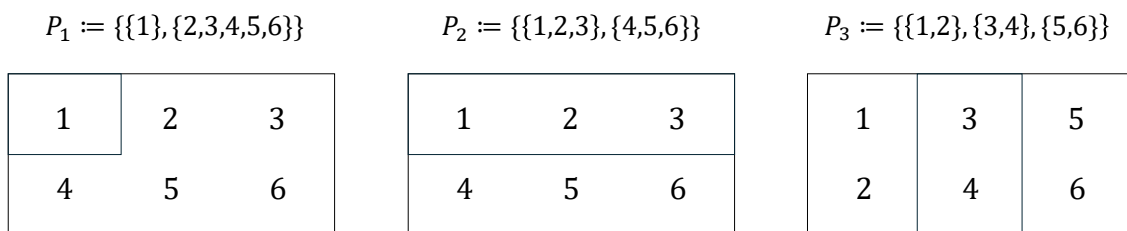


Figure 2.3 Diagrams of the partitions considered in the example for $M := \{1, 2, 3, 4, 5, 6\}$.

2.3 Special Sets

Number Sets

In natural science, one attempts to describe phenomena of the world that are intuitively identified as discrete or continuous by means of numbers. Depending on the type of

phenomenon, different number sets are suitable for this purpose. Mathematics provides, among others, the number sets given in the following definition.

Definition 2.8 (Number Sets). The following denote:

- $\mathbb{N} := \{1, 2, 3, \dots\}$ the *natural numbers*,
- $\mathbb{N}_n := \{1, 2, 3, \dots, n\}$ the *natural numbers of order n* ,
- $\mathbb{N}^0 := \mathbb{N} \cup \{0\}$ the *natural numbers and zero*,
- $\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ the *integers*,
- $\mathbb{Q} := \{\frac{p}{q} | p \in \mathbb{Z}, q \in \mathbb{N}\}$ the *rational numbers*,
- \mathbb{R} the *real numbers*, and
- $\mathbb{C} := \{a + ib | a, b \in \mathbb{R}, i := \sqrt{-1}\}$ the *complex numbers*.

•

The natural numbers and the integers are suitable for quantifying discrete phenomena. The rational numbers and especially the real numbers are suitable for quantifying continuous phenomena. \mathbb{R} includes the rational numbers and the so-called *irrational numbers* $\mathbb{R} \setminus \mathbb{Q}$. Rational numbers are numbers that can be expressed as fractions of integers and natural numbers. These include all integers as well as negative and positive decimal numbers such as $-\frac{9}{10} = -0.9$, $\frac{1}{3} = 0.\bar{3}$, and $\frac{196}{100} = 1.96$. Irrational numbers are numbers that cannot be expressed as rational numbers. Examples of irrational numbers are *Euler's number* $e \approx 2.71$, the *circle constant* $\pi \approx 3.14$, and the square root of 2, $\sqrt{2} \approx 1.41$.

The real numbers contain the natural numbers, the integers, and the rational numbers as subsets. Thus there are very many real numbers. Indeed, Cantor (1874) proved that there are more real numbers than natural numbers, even though there are infinitely many real numbers and infinitely many natural numbers. This property of the real numbers is called their *uncountability*. We will deepen this concept in Chapter 4. In particular,

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}. \quad (2.14)$$

Between any two real numbers there are infinitely many further real numbers. Positive infinity ∞ and negative infinity $-\infty$ are not numbers with which one could calculate in standard mathematics. They also do not belong to the number sets given in the definition above; thus $\infty \notin \mathbb{R}$ and $-\infty \notin \mathbb{R}$. Complex numbers are suitable for describing two-dimensional continuous phenomena. The values of the first dimension are represented in the real part a , and the values of the second dimension in the complex part b of a complex number. Complex numbers are used, for example, in the modeling of physical phenomena and in Fourier analysis.

Important subsets of the real numbers are the so-called *intervals*. We give the following definitions.

Intervals

Definition 2.9. Connected subsets of the real numbers are called *intervals*. For $a, b \in \mathbb{R}$, one distinguishes

- the *closed interval*

$$[a, b] := \{x \in \mathbb{R} | a \leq x \leq b\}, \quad (2.15)$$

- the *open interval*

$$]a, b[:= \{x \in \mathbb{R} | a < x < b\}, \tag{2.16}$$

- and the *half-open intervals*

$$]a, b] := \{x \in \mathbb{R} | a < x \leq b\} \text{ and } [a, b[:= \{x \in \mathbb{R} | a \leq x < b\}. \tag{2.17}$$

•

As examples, Figure 2.4 graphically represents the intervals $[1, 2]$, $]1, 2]$, $[1, 2[$, and $]1, 2[$. Here one imagines the real numbers as a continuous number line and must in each case note whether the left and right endpoints are part of the interval or not. As mentioned above, positive infinity (∞) and negative infinity $-\infty$ are not elements of \mathbb{R} . Thus one always writes $] - \infty, b]$ or $] - \infty, b[$ and $]a, \infty[$ or $[a, \infty[$, as well as $\mathbb{R} =] - \infty, \infty[$.

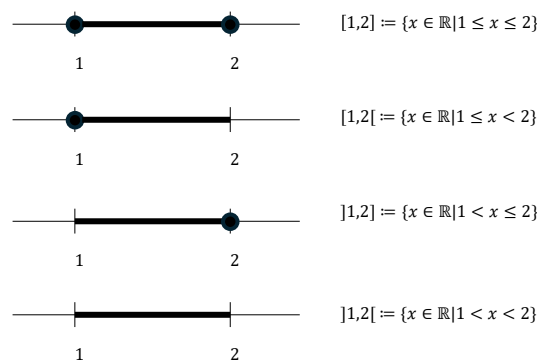


Figure 2.4 Representation of intervals on the number line. The black dot indicates that the corresponding number is part of the interval.

Cartesian Products

Often one wants to quantitatively describe several independent properties of a phenomenon at the same time. For this purpose, the one-dimensional number sets defined above can be extended to multidimensional number sets by forming *Cartesian products*. The elements of Cartesian products are called *ordered tuples* or *vectors*.

Definition 2.10 (Cartesian Products). Let M and N be two sets. Then the *Cartesian product of the sets M and N* is the set of all ordered tuples (m, n) with $m \in M$ and $n \in N$, formally

$$M \times N := \{(m, n) | m \in M, n \in N\}. \tag{2.18}$$

The Cartesian product of a set M with itself is denoted by

$$M^2 := M \times M. \tag{2.19}$$

Furthermore, let M_1, M_2, \dots, M_n be sets. Then the *Cartesian product of the sets M_1, \dots, M_n* is the set of all ordered n -tuples (m_1, \dots, m_n) with $m_i \in M_i$ for $i = 1, \dots, n$, formally

$$\prod_{i=1}^n M_i := M_1 \times \dots \times M_n := \{(m_1, \dots, m_n) | m_i \in M_i \text{ for } i = 1, \dots, n\}. \tag{2.20}$$

The n -fold Cartesian product of a set M with itself is denoted by

$$M^n := \prod_{i=1}^n M := \{(m_1, \dots, m_n) \mid m_i \in M \text{ for } i = 1, \dots, n\}. \quad (2.21)$$

•

In contrast to sets, the tuples introduced in Definition 2.10 are *ordered*. This means, for example, that for sets we have $\{1, 2\} = \{2, 1\}$, but for tuples we have $(1, 2) \neq (2, 1)$.

Example

Let $M := \{1, 2\}$ and $N := \{1, 2, 3\}$. Then the Cartesian product $M \times N$ is given by

$$M \times N := \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\} \quad (2.22)$$

and the Cartesian product $N \times M$ is given by

$$N \times M := \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)\}. \quad (2.23)$$

The Cartesian product is therefore generally not commutative, that is, it is not necessarily the case that $M \times N = N \times M$. One may construct the sets $M \times N \neq N \times M$ in this example from Table 2.1 and Table 2.2.

Table 2.1 Cartesian product $M \times N$

(m, n)	$n = 1$	$n = 2$	$n = 3$
$m = 1$	(1, 1)	(1, 2)	(1, 3)
$m = 2$	(2, 1)	(2, 2)	(2, 3)

Table 2.2 Cartesian product $N \times M$

(n, m)	$m = 1$	$m = 2$
$n = 1$	(1, 1)	(1, 2)
$n = 2$	(2, 1)	(2, 2)
$n = 3$	(3, 1)	(3, 2)

\mathbb{R} to the Power of n

As described above, the real numbers are particularly suitable for describing continuous phenomena. For the simultaneous description of several aspects of a continuous phenomenon, the *set of real tuples of n -th order*, or \mathbb{R} to the power of n for short, is correspondingly useful.

Definition 2.11 (Set of Real Tuples of n -th Order). The n -fold Cartesian product of the real numbers with themselves is denoted by

$$\mathbb{R}^n := \prod_{i=1}^n \mathbb{R} := \{x := (x_1, \dots, x_n) \mid x_i \in \mathbb{R}\} \quad (2.24)$$

and is read as “ \mathbb{R} to the power of n ”. We write the elements of \mathbb{R}^n as columns

$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (2.25)$$

and call them *n-dimensional vectors*. For distinction, we also call the elements of $\mathbb{R}^1 = \mathbb{R}$ *scalars*.

•

Examples

Familiar examples of \mathbb{R}^n are \mathbb{R}^1 as the set of real numbers, \mathbb{R}^2 as the set of real tuples in the model of the two-dimensional plane, and \mathbb{R}^3 as the set of real triples in the model of three-dimensional space, as visualized in Figure 2.5.

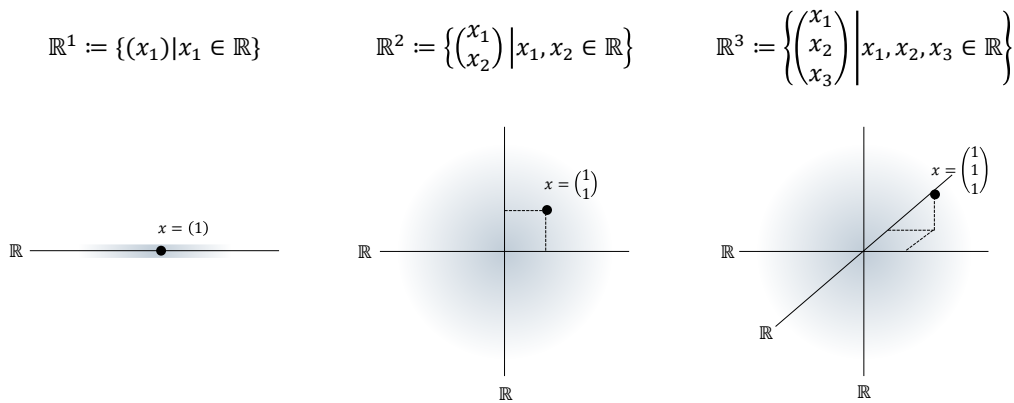


Figure 2.5 \mathbb{R}^n for $n = 1$, $n = 2$, and $n = 3$. The black dots each represent an element of the corresponding set.

An example of an $x \in \mathbb{R}^4$ is

$$x = \begin{pmatrix} 0.16 \\ 1.76 \\ 0.23 \\ 7.11 \end{pmatrix}. \quad (2.26)$$

3 Sums, Products, Powers

This unit introduces notation for the elementary arithmetic operations.

3.1 Sums

Definition 3.1 (Summation Symbol). It denotes

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n. \quad (3.1)$$

Here,

- Σ denotes the Greek letter *Sigma*, mnemonically standing for *sum*,
- the subscript $i = 1$ denotes the *running index* and the *starting index*,
- the superscript n denotes the *terminal index*, and
- x_1, x_2, \dots, x_n denote the *summands*.

For meaningful use of the summation symbol, it is essential to specify the beginning and the end of the summation by means of the subscript and the superscript. The precise name of the running index, by contrast, is irrelevant for the value of the sum; thus,

$$\sum_{i=1}^n x_i = \sum_{j=1}^n x_j. \quad (3.2)$$

Sometimes the running index is also given as an element of an *index set*. If, for example, the index set $I := \{1, 5, 7\}$ is defined, then

$$\sum_{i \in I} x_i := x_1 + x_5 + x_7. \quad (3.3)$$

In the following, we briefly consider a few examples of how the summation symbol is used.

- *Summation of predefined summands*. Let $x_1 := 2$, $x_2 := 10$, and $x_3 := -4$. Then

$$\sum_{i=1}^3 x_i = x_1 + x_2 + x_3 = 2 + 10 - 4 = 8. \quad (3.4)$$

- *Summation of weighted predefined summands.* Again let $x_1 := 2$, $x_2 := 10$, and $x_3 := -4$. In addition, let the *weighting coefficients* $a_1 := \frac{1}{2}$, $a_2 := \frac{1}{5}$, and $a_3 := 2$ be defined. Then

$$\sum_{i=1}^3 a_i x_i = a_1 x_1 + a_2 x_2 + a_3 x_3 = \frac{1}{2} \cdot 2 + \frac{1}{5} \cdot 10 + 2 \cdot (-4) = 1 + 2 - 8 = -5. \quad (3.5)$$

Expressions of the form $\sum_{i=1}^n a_i x_i$ are also called *linear combinations* of x_1, \dots, x_n with the *coefficients* or *weighting parameters* a_1, \dots, a_n .

- *Summation of natural numbers.* We have

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15. \quad (3.6)$$

- *Summation of even natural numbers.* We have

$$\sum_{i=1}^5 2i = 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 = 2 + 4 + 6 + 8 + 10 = 30. \quad (3.7)$$

- *Summation of odd natural numbers.* We have

$$\sum_{i=1}^5 (2i-1) = 2 \cdot 1 - 1 + 2 \cdot 2 - 1 + 2 \cdot 3 - 1 + 2 \cdot 4 - 1 + 2 \cdot 5 - 1 = 1 + 3 + 5 + 7 + 9 = 25. \quad (3.8)$$

Working with the summation symbol can often be simplified by applying the following calculation rules.

Theorem 3.1 (Calculation Rules for Sums).

- (1) *Sums of identical summands*

$$\sum_{i=1}^n x = nx \quad (3.9)$$

- (2) *Associativity for sums of equal length*

$$\sum_{i=1}^n x_i + \sum_{i=1}^n y_i = \sum_{i=1}^n (x_i + y_i) \quad (3.10)$$

- (3) *Distributivity under multiplication by a constant*

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad (3.11)$$

- (4) *Splitting sums for $1 < m < n$*

$$\sum_{i=1}^n x_i = \sum_{i=1}^m x_i + \sum_{i=m+1}^n x_i \quad (3.12)$$

- (5) *Reindexing*

$$\sum_{i=0}^n x_i = \sum_{j=m}^{n+m} x_{j-m} \quad (3.13)$$

◦

Proof. These calculation rules can be verified by writing out the sums and applying the rules of addition and multiplication. Here, we show associativity for sums of equal length and distributivity under multiplication by a constant as examples. For the former, we have

$$\begin{aligned} \sum_{i=1}^n x_i + \sum_{i=1}^n y_i &= x_1 + x_2 + \cdots + x_n + y_1 + y_2 + \cdots + y_n \\ &= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\ &= \sum_{i=1}^n (x_i + y_i). \end{aligned} \quad (3.14)$$

For the latter, we have

$$\begin{aligned} \sum_{i=1}^n ax_i &= ax_1 + ax_2 + \cdots + ax_n \\ &= a(x_1 + x_2 + \cdots + x_n) \\ &= a \sum_{i=1}^n x_i. \end{aligned} \quad (3.15)$$

□

Examples

As a first example of applying the calculation rules recorded in Theorem 3.1, we consider the evaluation of a *mean* (sometimes also called an *average*). Let x_1, x_2, \dots, x_n be real numbers. The mean of these numbers is the sum of x_1, x_2, \dots, x_n divided by the number of numbers n . By statement (3) of Theorem 3.1, it is irrelevant whether the numbers are first added up and the resulting sum is then divided by n , or whether the individual numbers are each divided by n and the corresponding results are then added up. More precisely, by applying Theorem 3.1 (3) with $a = 1/n$, we obtain

$$\frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{x_i}{n}. \quad (3.16)$$

For example, the mean of $x_1 := 1, x_2 := 4, x_3 := 2$, and $x_4 := 1$ is given by

$$\frac{1}{4} \sum_{i=1}^4 x_i = \frac{1}{4}(1 + 4 + 2 + 1) = \frac{8}{4} = 2 = \frac{8}{4} = \frac{1}{4} + \frac{4}{4} + \frac{2}{4} + \frac{1}{4} = \sum_{i=1}^4 \frac{x_i}{4}. \quad (3.17)$$

As a second example, we consider the reindexing rule recorded in Theorem 3.1 (5). Let $n := 3$ and $m := 2$, and let $x_0 := 2, x_1 := 3, x_2 := 5$, and $x_3 := 10$. Then, evidently,

$$\begin{aligned} \sum_{i=0}^3 x_i &= x_0 + x_1 + x_2 + x_3 \\ &= 2 + 3 + 5 + 10 \\ &= 20. \end{aligned} \quad (3.18)$$

But we also have

$$\begin{aligned}
 \sum_{j=m}^{n+m} x_{j-m} &= \sum_{j=2}^{3+2} x_{j-2} \\
 &= \sum_{j=2}^5 x_{j-2} \\
 &= x_{2-2} + x_{3-2} + x_{4-2} + x_{5-2} \\
 &= x_0 + x_1 + x_2 + x_3 \\
 &= 2 + 3 + 5 + 10 \\
 &= 20.
 \end{aligned} \tag{3.19}$$

Double Sums

In applications, one often encounters expressions that carry out several summations one after another. For each of these sums, the definitions and calculation rules listed above apply. The meaning of double sums is best made clear by writing them out from the inside to the outside, while noting that the running index of the outer sum remains constant during each iteration of the inner sum.

The following examples may illustrate this.

(1)

$$\begin{aligned}
 \sum_{i=1}^2 \sum_{j=1}^3 (i+j) &= \sum_{i=1}^2 (i+1 + i+2 + i+3) \\
 &= (1+1 + 1+2 + 1+3) + (2+1 + 2+2 + 2+3) \\
 &= 9 + 12 \\
 &= 21
 \end{aligned} \tag{3.20}$$

(2)

$$\begin{aligned}
 \sum_{i=1}^2 \sum_{j=1}^3 (x_i + y_j) &= \sum_{i=1}^2 (x_i + y_1 + x_i + y_2 + x_i + y_3) \\
 &= (x_1 + y_1 + x_1 + y_2 + x_1 + y_3) + (x_2 + y_1 + x_2 + y_2 + x_2 + y_3)
 \end{aligned} \tag{3.21}$$

(3)

$$\begin{aligned}
 \sum_{i=1}^3 \sum_{j=1}^2 (x_i y_j) &= \sum_{i=1}^3 (x_i y_1 + x_i y_2) \\
 &= (x_1 y_1 + x_1 y_2) + (x_2 y_1 + x_2 y_2) + (x_3 y_1 + x_3 y_2)
 \end{aligned} \tag{3.22}$$

3.2 Products

The product symbol provides notation for products that is analogous to the summation symbol.

Definition 3.2 (Product Symbol). It denotes

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \cdots \cdot x_n. \quad (3.23)$$

Here,

- \prod denotes the Greek letter *Pi*, mnemonically standing for *product*,
- the subscript $i = 1$ denotes the *running index* and the *starting index*,
- the superscript n denotes the *terminal index*, and
- x_1, x_2, \dots, x_n denote the *product terms*.

•

Analogously to the summation symbol, the product symbol is meaningful only with a subscript and a superscript specifying the running and terminal index. The precise name of the running index is again irrelevant; thus,

$$\prod_{i=1}^n x_i = \prod_{j=1}^n x_j. \quad (3.24)$$

Here, too, the running index is in rare cases given as an element of an index set. If, for example, the index set $J := \mathbb{N}_2^0$ is defined, then

$$\prod_{j \in J} x_j := x_0 \cdot x_1 \cdot x_2. \quad (3.25)$$

One example of the use of the product symbol is the definition of the *factorial* of a natural number n by

$$n! := \prod_{i=1}^n i. \quad (3.26)$$

For example,

$$3! := \prod_{i=1}^3 i = 1 \cdot 2 \cdot 3 = 6. \quad (3.27)$$

There are also a number of calculation rules for products that often simplify working with them. We list some in the following theorem. In doing so, we make anticipatory use of the notation of *powers* defined in Section 3.3.

Theorem 3.2 (Calculation Rules for Products).

(1) *Products of identical factors*

$$\prod_{i=1}^n x = x^n \quad (3.28)$$

(2) *Raising constants to powers*

$$\prod_{i=1}^n ax_i = a^n \prod_{i=1}^n x_i \quad (3.29)$$

(3) *Splitting products for $1 < m < n$*

$$\prod_{i=1}^n x_i = \prod_{i=1}^m x_i \prod_{j=m+1}^n x_j \quad (3.30)$$

(4) *Product of products*

$$\prod_{i=1}^n x_i y_i = \prod_{i=1}^n x_i \prod_{i=1}^n y_i \quad (3.31)$$

◦

3.3 Powers

Products of numbers with themselves can be abbreviated using power notation.

Definition 3.3 (Power). For $a \in \mathbb{R}$ and $n \in \mathbb{N}^0$, the n -th power of a is defined by

$$a^0 := 1 \text{ and } a^{n+1} := a^n \cdot a. \quad (3.32)$$

Furthermore, for $a \in \mathbb{R} \setminus \{0\}$ and $n \in \mathbb{N}^0$, the *negative n -th power of a* is defined by

$$a^{-n} := (a^n)^{-1} := \frac{1}{a^n}. \quad (3.33)$$

a is called the *base*, and n is called the *exponent*.

•

The way a^{n+1} is defined by referring back to the power a^n in the above definition is called *recursive*. The definition $a^0 := 1$ is called the *initial case* of the recursion; it is what makes the recursive definition of a^{n+1} possible in the first place. The definition $a^{n+1} := a^n \cdot a$ is also called the *recursion step*. The following calculation rules simplify computations with powers.

Theorem 3.3 (Calculation Rules for Powers). For $a, b \in \mathbb{R}$ and $n, m \in \mathbb{Z}$ with $a \neq 0$ for negative exponents, the following calculation rules hold:

$$a^n a^m = a^{n+m} \quad (3.34)$$

$$(a^n)^m = a^{nm} \quad (3.35)$$

$$(ab)^n = a^n b^n \quad (3.36)$$

◦

Instead of a proof, we consider the following three examples.

(1)

$$2^2 \cdot 2^3 = (2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) = 2^5 = 2^{2+3} \quad (3.37)$$

(2)

$$(3^2)^3 = (3 \cdot 3)^3 = (3 \cdot 3) \cdot (3 \cdot 3) \cdot (3 \cdot 3) = 3^6 = 3^{2 \cdot 3} \quad (3.38)$$

$$(3) \quad (2 \cdot 4)^2 = (2 \cdot 4) \cdot (2 \cdot 4) = (2 \cdot 2) \cdot (4 \cdot 4) = 2^2 \cdot 4^2 \quad (3.39)$$

Closely related to powers is the definition of the n -th root:

Definition 3.4 (n -th Root). For $a \in \mathbb{R}_{\geq 0}$ and $n \in \mathbb{N}$, the n -th root of a is defined as the nonnegative real number r with

$$r^n = a. \quad (3.40)$$

•

When computing with roots, power notation for roots is often helpful because it allows the calculation rules for powers to be applied directly.

Theorem 3.4 (Power Notation for the n -th Root). Let $a \in \mathbb{R}_{\geq 0}$, let $n \in \mathbb{N}$, and let r be the n -th root of a . Then

$$r = a^{\frac{1}{n}} \quad (3.41)$$

◦

Proof. We have

$$\left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n}} \cdot a^{\frac{1}{n}} \cdot \dots \cdot a^{\frac{1}{n}} = a^{\sum_{i=1}^n \frac{1}{n}} = a^1 = a. \quad (3.42)$$

Thus, by Definition 3.4, $r = a^{\frac{1}{n}}$. □

Computing with square roots is greatly facilitated by the power notation

$$\sqrt{x} = x^{\frac{1}{2}} \quad (3.43)$$

For example,

$$\frac{2\pi}{\sqrt{2\pi}} = \frac{2\pi}{(2\pi)^{\frac{1}{2}}} = (2\pi)^1 \cdot (2\pi)^{-\frac{1}{2}} = (2\pi)^{1-\frac{1}{2}} = (2\pi)^{\frac{1}{2}} = \sqrt{2\pi}. \quad (3.44)$$

The following relation between square, root, and absolute value is used quite often.

Theorem 3.5 (Root and Absolute Value). Let $x \in \mathbb{R}$, and let

$$|x| := \begin{cases} x & \text{for } x \geq 0 \\ -x & \text{for } x < 0 \end{cases} \quad (3.45)$$

denote the absolute value of x . Then

$$\sqrt{x^2} = |x|. \quad (3.46)$$

◦

Proof. We consider the cases $x \geq 0$ and $x < 0$. First, let $x \geq 0$. Then

$$x \geq 0 \Rightarrow x^2 \geq 0 \Rightarrow \sqrt{x^2} = x = |x|. \quad (3.47)$$

Now let $x < 0$. Then

$$x < 0 \Rightarrow x^2 > 0 \Rightarrow \sqrt{x^2} = -x = |x|. \quad (3.48)$$

□

4 Functions

Alongside sets, functions are the second foundational pillar of modern mathematics. In this unit, we define the concept of a function, introduce first properties of functions, and give an overview of several elementary functions.

4.1 Definition and Properties

Definition 4.1 (Function). A *function* or *mapping* f is an assignment rule that assigns exactly one element of a set Z to each element of a set D . D is called the *domain* of f , and Z is called the *codomain* of f . We write

$$f : D \rightarrow Z, x \mapsto f(x), \quad (4.1)$$

where $f : D \rightarrow Z$ is read as “the function f maps all elements of the set D uniquely to elements in Z ” and $x \mapsto f(x)$ is read as “ x , which is an element of D , is mapped by the function f to $f(x)$, where $f(x)$ is an element of Z .” The arrow \rightarrow denotes the mapping between the sets D and Z , while the arrow \mapsto denotes the mapping between an element of D and an element of Z .

•

It is essential to distinguish between the *function* f as an assignment rule and a *value of the function* $f(x)$ as an element of Z . x is the *argument* of the function (the function’s *input*), and $f(x)$ is the value that the function f takes for the argument x (the function’s *output*). Usually, the definition of a function specifies, after $f(x)$, the *functional form of f* , that is, a rule for forming the value $f(x)$ from x . For example, in the following definition of a function

$$f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto f(x) := x^2 \quad (4.2)$$

the definition of a power is used. Note that functions are always *unique* in the sense that, whenever the function is applied, they always assign one and the same $f(x) \in Z$ to each $x \in D$.

Figure 4.1 visualizes several aspects of the definition of a function. Figure 4.1 A first illustrates the central terms of the definition. Figure 4.1 B shows a first example of a function in which the function values assigned to the arguments are determined not by a computational rule but by direct definition. Figure 4.1 C and Figure 4.1 D use the same pictorial language to emphasize two aspects of the definition by means of counterexamples: the drawing in Figure 4.1 C is not the representation of a function, because by Definition 4.1 a function assigns *each element* of a set D exactly one element of a codomain Z . The sketched assignment rule does not assign any element in Z to the element $2 \in D$, and therefore it is not a function. Similarly, according to Definition 4.1, a function assigns exactly one element of a codomain Z to each element of a set D . The assignment rule

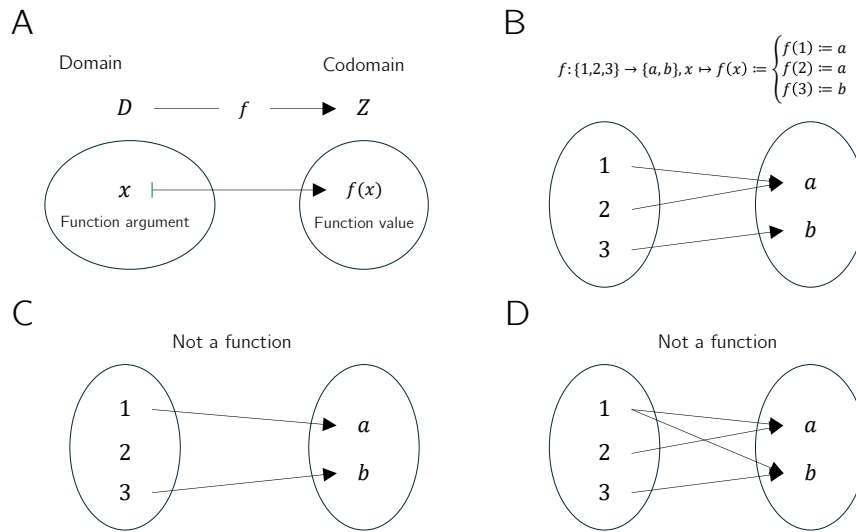


Figure 4.1 Aspects of the definition of a function

sketched in Figure 4.1 D assigns both the element $a \in Z$ and the element $b \in Z$ to $1 \in D$, and therefore it is incompatible with Definition 4.1.

Thus, functions put elements of sets into relation with one another. The sets of these elements receive special names.

Definition 4.2 (Image, Range, Preimage Set, Preimage). Let $f : D \rightarrow Z, x \mapsto f(x)$ be a function and let $D' \subseteq D$ and $Z' \subseteq Z$. The set

$$f(D') := \{z \in Z \mid \text{there exists an } x \in D' \text{ with } z = f(x)\} \tag{4.3}$$

is called the *image of D'* , and $f(D) \subseteq Z$ is called the *range of f* . Furthermore, the set

$$f^{-1}(Z') := \{x \in D \mid f(x) \in Z'\} \tag{4.4}$$

is called the *preimage set of Z'* . An $x \in D$ with $z = f(x) \in Z$ is called a *preimage of z* .

•

Note that the range $f(D)$ of f and the codomain Z of f need not be identical.

Example

To clarify the concepts introduced in Definition 4.2, we consider the function shown in Figure 4.2 A,

$$f : \{1, 2, 3, 4, 5\} \rightarrow \{a, b, c, d\}, x \mapsto f(x) := \begin{cases} f(1) := b \\ f(2) := d \\ f(3) := c \\ f(4) := c \\ f(5) := d \end{cases} . \tag{4.5}$$

According to Definition 4.2, an image is always defined with respect to a subset D' of the domain D . Thus, let $D' := \{2, 3\} \subset D$, as shown in Figure 4.2 B. Then the image

of D' is the set of those $z \in Z$ for which there exists an $x \in D'$ such that $z = f(x)$. Here, the relevant $z \in Z$ for which such an $x \in D'$ exists are precisely $c, d \in Z$, because $f(2) = d$ and $f(3) = c$. Beyond these, there is no $z \in Z$ with $f(x) = z$ and $x \in \{2, 3\}$. By Definition 4.2, the range $f(D)$ is the subset of those $z \in Z$ for which there exists an $x \in D$ such that $z = f(x)$. This is true for all elements of Z except for $a \in Z$, because there is no $x \in D$ with $a = f(x)$. For the function considered here, the range of f is therefore given by $f(D) := \{b, c, d\}$.

Conversely, according to Definition 4.2, a preimage set is always defined with respect to a subset Z' of the codomain Z . Thus, let $Z' := \{c, d\} \subset Z$, as shown in Figure 4.2 C. Then the preimage set of Z' is the set of those $x \in D$ for which $f(x) \in Z'$. The elements of D whose function values under f are elements of Z' are precisely $\{2, 3, 4, 5\}$. By contrast, $1 \in D$ is not an element of this preimage set, because f maps $1 \in D$ to $b \in Z$ and $b \notin Z'$. Nevertheless, $1 \in D$ is of course a preimage of b .

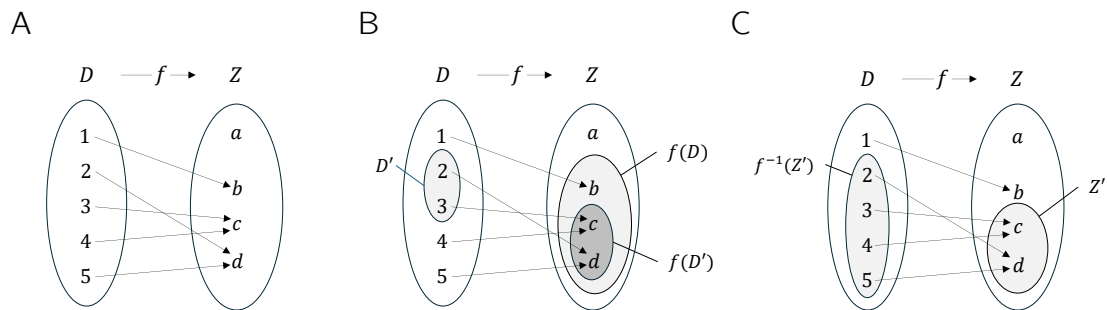


Figure 4.2 (A) Visualization of the function Equation 4.5. (B) Range and an image of f . (C) Preimage set of a set Z' under f

Basic properties of functions are named in the following definition.

Definition 4.3 (Injectivity, Surjectivity, Bijectivity). Let $f : D \rightarrow Z, x \mapsto f(x)$ be a function. f is called *injective* if, for every image element $z \in f(D)$, there is exactly one preimage $x \in D$. Equivalently, f is injective if $x_1, x_2 \in D$ and $x_1 \neq x_2$ imply $f(x_1) \neq f(x_2)$. f is called *surjective* if $f(D) = Z$, that is, if every element of the codomain Z has a preimage in the domain D . Finally, f is called *bijective* if it is both injective and surjective. Bijective functions are also called *one-to-one correspondences* or *one-to-one mappings*.

•

Examples

We first illustrate Definition 4.3 using three (counter)examples in Figure 4.3. Figure 4.3 A shows the *non-injective* function

$$f : \{1, 2, 3\} \rightarrow \{a, b\}, x \mapsto f(x) := \begin{cases} f(1) := a \\ f(2) := a \\ f(3) := b \end{cases} . \quad (4.6)$$

The function is non-injective because the element a in the range of f has more than one preimage in the domain of f , namely the elements 1 and 2. Figure 4.3 B shows the

non-surjective function

$$g : \{1, 2, 3\} \rightarrow \{a, b, c, d\}, x \mapsto g(x) := \begin{cases} g(1) := a \\ g(2) := b \\ g(3) := d \end{cases} . \quad (4.7)$$

The function is non-surjective because the element c in the codomain of g has no preimage in the domain of g . Finally, Figure 4.3 C shows the bijective function

$$h : \{1, 2, 3\} \rightarrow \{a, b, c\}, x \mapsto h(x) := \begin{cases} h(1) := a \\ h(2) := b \\ h(3) := c \end{cases} . \quad (4.8)$$

For *every* element in the codomain of h there is *exactly one* preimage, so the function is injective and surjective and therefore bijective.

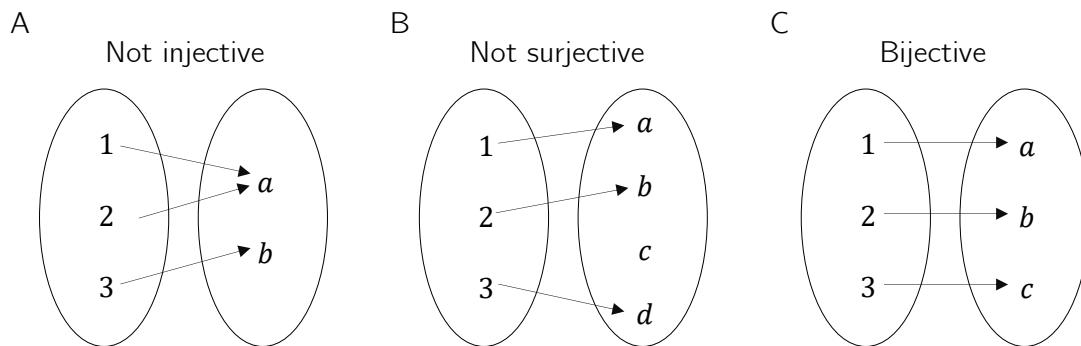


Figure 4.3 Injectivity, surjectivity, bijectivity.

As a further example, consider the function

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2 \quad (4.9)$$

This function is not injective because, for example, with $x_1 = 2 \neq -2 = x_2$ we have $f(x_1) = 2^2 = 4 = (-2)^2 = f(x_2)$. Moreover, f is not surjective because, for example, $-1 \in \mathbb{R}$ has no preimage under f . However, if the domain of f is restricted to the non-negative real numbers, that is, if we define the function

$$\tilde{f} : [0, \infty[\rightarrow [0, \infty[, x \mapsto \tilde{f}(x) := x^2, \quad (4.10)$$

then, in contrast to f , \tilde{f} is injective and surjective, hence bijective.

On the uncountability of the real numbers

At this point, we use the concepts of surjectivity and bijectivity of functions to deepen the idea of the *uncountability of the real numbers* a little. This idea says that there are different kinds of infinities, which is certainly not easy to grasp intuitively at first. Moreover, some aspects of probability theory are formulated in ways that accommodate the uncountability of the real numbers, so even a rough understanding of this property can help motivate some of the subtleties of probability theory. Specifically, we briefly sketch *Cantor’s diagonal argument* (Cantor (1891), Gray (1994)) for the uncountability of the real numbers. We begin by recording what, against the background of bijective mappings, is meant by a *finite* and a *countably infinite* set.

Definition 4.4. A set M is called *finite* if there exists a bijective mapping of the form

$$f : \mathbb{N}_n \rightarrow M \tag{4.11}$$

onto the elements of M . Furthermore, a set M is called *countably infinite* if there exists a bijective mapping of the form

$$f : \mathbb{N} \rightarrow M \tag{4.12}$$

onto the elements of M .

•

In the case of a finite set, each element of the set can thus be assigned a natural number with maximum value $n < \infty$ in the sense of a bijective mapping. In the case of a countably infinite set, each element of the set can at least be assigned a natural number in the sense of a bijective mapping. The elements of these sets can therefore be counted.

Cantor (1891) showed that this is not true in general for the real numbers: no bijective mapping between the real numbers and the natural numbers can be constructed, and hence there must be more real numbers than natural numbers. More specifically, one can argue that even the interval $[0, 1]$ already contains more real numbers than there are natural numbers, that is, no surjective function from \mathbb{N} to $[0, 1]$ can be constructed.

Cantor’s argument has the following form. Suppose that $f : \mathbb{N} \rightarrow [0, 1]$ is an injective function represented in tabular form, for example as in Table 4.1.

Table 4.1 Mapping $f : \mathbb{N} \rightarrow [0, 1]$. The dots \dots and $:$ indicate that the table continues indefinitely to the right and downward.

n	$f(n)$										
1	0.	3	1	4	1	5	9	2	6	5	...
2	0.	8	9	4	5	9	7	8	1	0	...
3	0.	9	6	2	3	2	1	5	8	7	...
4	0.	5	3	6	8	8	8	9	7	3	...
5	0.	7	4	3	8	1	8	0	9	7	...
\vdots	\vdots										

Now construct another number in $[0, 1]$ by adding 1 to each of the bold diagonal entries in Table 4.1 whenever the corresponding value is less than 9, and otherwise setting the value to 0. This is shown in Table 4.2.

Table 4.2 Construction of another real number in $[0, 1]$.

n	$f(n)$										
1	0.	4	1	4	1	5	9	2	6	5	...
2	0.	8	0	4	5	9	7	8	1	0	...
3	0.	9	6	3	3	2	1	5	8	7	...
4	0.	5	3	6	9	8	8	9	7	3	...
5	0.	7	4	3	8	2	8	0	9	7	...
\vdots	\vdots										

The number thus constructed,

$$0.40392... \tag{4.13}$$

cannot occur in Table 4.1, because it differs from $f(1)$ in its first decimal place, from $f(2)$ in its second decimal place, from $f(3)$ in its third decimal place, ..., from $f(n)$ in its n th decimal place, and so on indefinitely. Thus there is at least one number in $[0, 1]$ to which no natural number can be assigned. The injective mapping f is not surjective and therefore not bijective. Consequently, $[0, 1]$ is not countably infinite and is accordingly called *uncountable*.

4.2 Types of Functions

By composition, further functions can be formed from given functions.

Definition 4.5 (Composition of Functions). Let $f : D \rightarrow Z$ and $g : Z \rightarrow S$ be two functions, where the range of f is assumed to agree with the domain of g . Then

$$g \circ f : D \rightarrow S, x \mapsto (g \circ f)(x) := g(f(x)) \quad (4.14)$$

defines a function called the *composition of f and g* .

•

The notation for composed functions takes a little getting used to. It is important to recognize that $g \circ f$ denotes the composed function and $(g \circ f)(x)$ denotes an element in the codomain of the composed function. Intuitively, when evaluating $(g \circ f)(x)$, one first applies the function f to x and then applies the function g to the element $f(x)$ of Z . This is captured by the functional form $g(f(x))$. For simplicity, the composition of two functions is often denoted by a single letter; for example, one writes $h := g \circ f$ with $h(x) = g(f(x))$. A mild source of confusion can arise when elements in the codomain of f are denoted by y , so that the notation $y = f(x)$ and $h(x) = g(y)$ is used. However, this notation is sometimes needed for notational simplicity. We visualize Definition 4.5 in Figure 4.4.

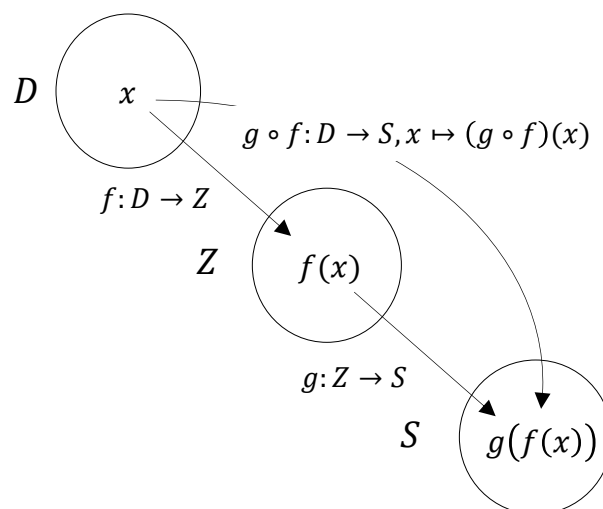


Figure 4.4 Composition of functions.

As an example of the composition of two functions, consider

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := -x^2 \quad (4.15)$$

and

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := \exp(x). \quad (4.16)$$

In this case, the composition of f and g is

$$g \circ f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto (g \circ f)(x) := g(f(x)) = \exp(-x^2). \quad (4.17)$$

A first application of function composition appears in the following definition.

Definition 4.6 (Inverse Function). Let $f : D \rightarrow Z, x \mapsto f(x)$ be a bijective function. Then the function f^{-1} with

$$f^{-1} \circ f : D \rightarrow D, x \mapsto (f^{-1} \circ f)(x) := f^{-1}(f(x)) = x \quad (4.18)$$

is called the *inverse function*, *inverse mapping*, or simply the *inverse of f* . •

Inverse functions are always bijective. This follows because f is bijective and therefore each $x \in D$ is assigned exactly one $f(x) = z \in Z$. Hence, each $z \in Z$ is also assigned exactly one $x \in D$, namely $f^{-1}(f(x)) = x$. Intuitively, the inverse function of f undoes the effect of f on an element x . We visualize Definition 4.6 in Figure 4.5 A.

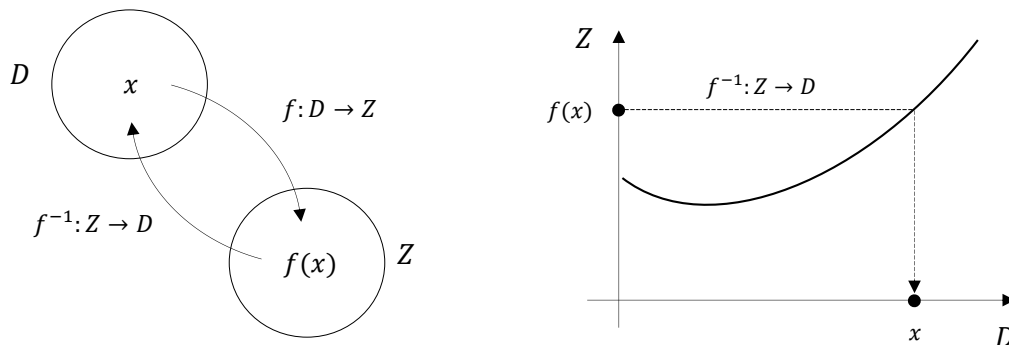


Figure 4.5 Inverse function.

If one considers the graph of a function in a Cartesian coordinate system, then applying a function to a value on the x -axis leads to a value on the y -axis. Applying the inverse function correspondingly leads from a value on the y -axis to a value on the x -axis. We visualize this in Figure 4.5 B. For example, consider the function

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := 2x =: y. \quad (4.19)$$

Then the inverse function of f is given by

$$f^{-1} : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto f^{-1}(y) := \frac{1}{2}y, \quad (4.20)$$

because for every $x \in \mathbb{R}$,

$$(f^{-1} \circ f)(x) := f^{-1}(f(x)) = f^{-1}(2x) = \frac{1}{2} \cdot 2x = x. \quad (4.21)$$

An important class of functions consists of *linear maps*.

Definition 4.7 (Linear Map). A map $f : D \rightarrow Z, x \mapsto f(x)$ is called a *linear map* if, for $x, y \in D$ and a scalar c ,

$$f(x + y) = f(x) + f(y) \quad (\text{additivity})$$

and

$$f(cx) = cf(x) \quad (\text{homogeneity})$$

hold. A map for which the above properties do not hold is called a *nonlinear map*. •

Linear maps are often known as “straight lines.” The general definition of linear maps is not completely congruent with this intuition. In particular, linear maps are only those functions that map zero to zero. We show this in the following theorem.

Theorem 4.1 (Linear Map of Zero). *Let $f : D \rightarrow Z$ be a linear map. Then*

$$f(0) = 0. \quad (4.22)$$

◦

Proof. First note that by additivity of f ,

$$f(0) = f(0 + 0) = f(0) + f(0). \quad (4.23)$$

Adding $-f(0)$ to both sides of the above equation gives

$$f(0) - f(0) = f(0) + f(0) - f(0) \Leftrightarrow f(0) = 0 \quad (4.24)$$

and this proves the claim. □

We illustrate the concept of a linear map with two examples.

- For $a \in \mathbb{R}$, the map

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax \quad (4.25)$$

is a linear map because

$$f(x + y) = a(x + y) = ax + ay = f(x) + f(y) \text{ and } f(cx) = acx = cax = cf(x). \quad (4.26)$$

- For $a, b \in \mathbb{R}$, by contrast, the map

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax + b \quad (4.27)$$

is nonlinear because, for example, for $a := b := 1$,

$$f(x + y) = 1(x + y) + 1 = x + y + 1 \neq x + 1 + y + 1 = f(x) + f(y). \quad (4.28)$$

A map of the form $f(x) := ax + b$ is called a *linear-affine map* or *linear-affine function*. Somewhat imprecisely, functions of the form $f(x) := ax + b$ are sometimes also called *linear functions*.

Besides the types of functions discussed so far, there are many further classes of functions. In the following definition, we classify functions by the dimensionality of their domains and codomains. This type of function classification is often helpful for gaining an initial overview of a mathematical model.

Definition 4.8 (Function Types). We distinguish

- *univariate real-valued functions* of the form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x), \quad (4.29)$$

- *multivariate real-valued functions* of the form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n), \quad (4.30)$$

- and *multivariate vector-valued functions* of the form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto f(x) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}, \quad (4.31)$$

where $f_i, i = 1, \dots, m$ are called the *components* or *component functions* of f .

•

In physics, multivariate real-valued functions are called *scalar fields*, and multivariate vector-valued functions are called *vector fields*. In some applications, *matrix-variate matrix-valued functions* also occur.

4.3 Elementary Functions

By *elementary functions* we mean a small set of univariate real-valued functions that frequently appear as building blocks of more complex functions. These are the *polynomial functions*, the *exponential function*, the *logarithm function*, and the *gamma function*. In the following, we give the definitions of these functions and state their main properties as theorems, and then present their graphs. For proofs of the properties introduced here, we refer to the advanced literature.

Definition 4.9 (Polynomial Functions). A function of the form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{i=0}^k a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \dots + a_k x^k \quad (4.32)$$

is called a *polynomial function* of degree k with coefficients $a_0, a_1, \dots, a_k \in \mathbb{R}$. Typical polynomial functions are listed in Table 4.3.

Table 4.3 Polynomial functions.

Name	Form	Coefficients
Constant function	$f(x) = a$	$a_0 := a, a_i := 0, i > 0$
Identity function	$f(x) = x$	$a_0 := 0, a_1 := 1, a_i := 0, i > 1$
Linear-affine function	$f(x) = ax + b$	$a_0 := b, a_1 := a, a_i := 0, i > 1$
Square function	$f(x) = x^2$	$a_0 := 0, a_1 := 0, a_2 := 1, a_i := 0, i > 2$

•

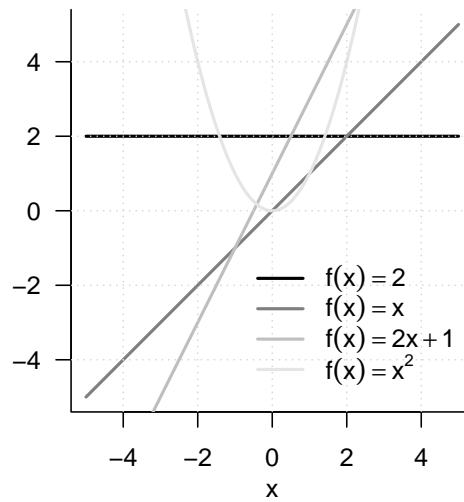


Figure 4.6 Selected polynomial functions

Figure 4.6 shows the graphs of the polynomial functions listed in Definition 4.9.

An important pair of functions consists of the exponential function and the logarithm function.

Theorem 4.2 (Exponential Function and Its Properties). *The exponential function is defined as*

$$\exp : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \exp(x) := e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \quad (4.33)$$

Among others, the exponential function has the properties listed in Table 4.4.

Table 4.4 Properties of the exponential function.

Property	Meaning
Range	$x \in]-\infty, 0[\Rightarrow \exp(x) \in]0, 1[$ $x \in]0, \infty[\Rightarrow \exp(x) \in]1, \infty[$
Monotonicity	$x < y \Rightarrow \exp(x) < \exp(y)$
Special values	$\exp(0) = 1$ and $\exp(1) = e$
Summation property	$\exp(x + y) = \exp(x) \exp(y)$
Subtraction property	$\exp(x - y) = \frac{\exp(x)}{\exp(y)}$

◦

In particular, the exponential function only takes positive values and intersects the y -axis at $x = 0$. The number $\exp(1) := e \approx 2.71\dots$ is called *Euler's number*. Finally, the special values of the exponential function imply

$$\exp(x) \exp(-x) = \exp(x - x) = \exp(0) = 1. \quad (4.34)$$

Theorem 4.3 (Logarithm Function and Its Properties). *The logarithm function is defined as the inverse function of the exponential function,*

$$\ln :]0, \infty[\rightarrow \mathbb{R}, x \mapsto \ln(x) \text{ with } \ln(\exp(x)) = x \text{ for all } x \in \mathbb{R}. \quad (4.35)$$

Among others, the logarithm function has the properties listed in Table 4.5.

Table 4.5 Properties of the logarithm function.

Property	Meaning
Range	$x \in]0, 1[\Rightarrow \ln(x) \in]-\infty, 0[$ $x \in]1, \infty[\Rightarrow \ln(x) \in]0, \infty[$
Monotonicity	$x < y \Rightarrow \ln(x) < \ln(y)$
Special values	$\ln(1) = 0$ and $\ln(e) = 1$
Product property	$\ln(xy) = \ln(x) + \ln(y)$
Power property	$\ln(x^c) = c \ln(x)$
Division property	$\ln\left(\frac{1}{x}\right) = -\ln(x)$

◦

In contrast to the exponential function, the logarithm function takes both negative and positive values. The logarithm function intersects the x -axis at $x = 1$. The product property and the power property are central when calculating with the logarithm function. Intuitively, they can be remembered as: “The logarithm function turns products into sums and powers into products.” The graphs of the exponential and logarithm functions are shown in Figure 4.7.

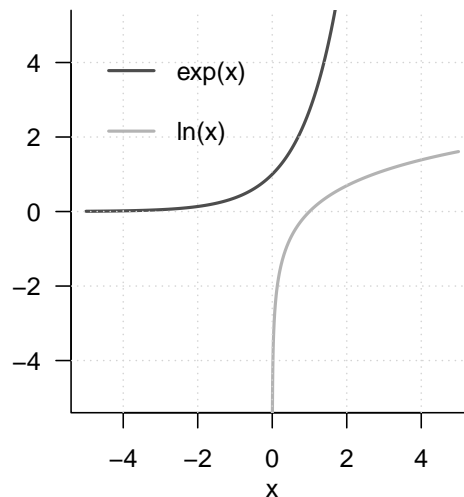


Figure 4.7 Exponential function and logarithm function

A frequent companion in probability theory is the *gamma function*.

Definition 4.10 (Gamma Function). The *gamma function* is defined by

$$\Gamma : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \Gamma(x) := \int_0^{\infty} \xi^{x-1} \exp(-\xi) d\xi \quad (4.36)$$

Among others, the gamma function has the properties listed in Table 4.6.

Table 4.6 Properties of the gamma function.

Property	Meaning
Special values	$\Gamma(1) = 1$ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$
Recursion property	For $x > 0$, $\Gamma(x+1) = x\Gamma(x)$

•

A section of the graph of the gamma function is shown in Figure 4.8.

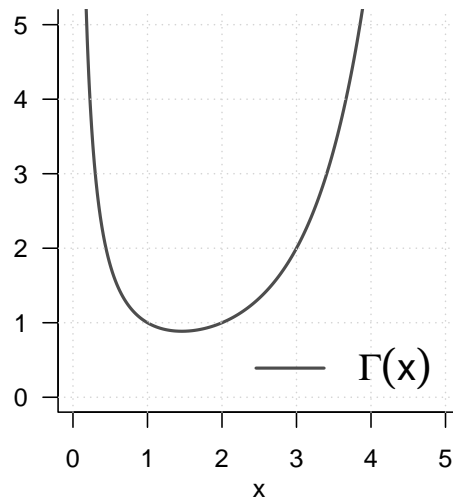


Figure 4.8 Gamma function

References

- Abbott, S. (2015). *Understanding Analysis*. Springer New York. <https://doi.org/10.1007/978-1-4939-2712-8>
- Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K., & Stachel, H. (2018). *Mathematik*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-56741-8>
- Bärwolff, G. (2017). *Höhere Mathematik für Naturwissenschaftler und Ingenieure*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-55022-9>
- Cantor, G. (1874). Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Jahresbericht Der Deutschen Mathematiker-Vereinigung*, 1.
- Cantor, G. (1891). Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht Der Deutschen Mathematiker-Vereinigung*, 1, 75–78.
- Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46(4), 481–512. <https://doi.org/10.1007/BF02124929>
- Chiossi, S. G. (2021). *Essential Mathematics for Undergraduates: A Guided Approach to Algebra, Geometry, Topology and Analysis*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-87174-1>
- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for Machine Learning* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108679930>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Gray, R. (1994). Georg Cantor and Transcendental Numbers. *The American Mathematical Monthly*, 101(9), 819–832. <https://doi.org/10.1080/00029890.1994.11997035>
- Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*. Royal Society.
- Searle, S. (1982). *Matrix Algebra Useful for Statistics*. Wiley-Interscience.
- Spivak, M. (2008). *Calculus* (4th ed.). Publish or Perish, Inc.
- Strang, G. (2009). *Introduction to Linear Algebra*. Cambridge University Press.
- Unger, L. (2000). *Grundkurs Mathematik*.