Contents

1 The language of mathematics 2
2 Implications 3
3 Proofs 5
4 Proof by contradiction 6
5 The induction principle 8
6 The language of set theory 12
7 Quantifiers 16
8 Functions 18
9 Properties of functions 20
10 Counting 24
11 Properties of finite sets 27
12 Counting functions and subsets 28
1 The language of mathematics

1.1 Mathematical statements

Mathematical statements is a broad concepts. It includes proposition and predicates.

A proposition is a sentence which is either true or false. A predicate is a sentence which depends on some free variables and it can be both true or false depending on the values of the free variables.

Example of two propositions:

(i) \(1 + 1 = 2\).

(ii) \(\pi = 2\).

Here (i) is true while (ii) is false.

Example 1.1.1. "Every even integer greater than 2 may be written as the sum of two prime numbers" is a proposition. But we don’t know it is true or false. We often call a proposition like this as an open problem.

This is the Goldbach’s Conjecture. This conjecture is raised by Christian Goldbach in 1742. The Chinese famous mathematician Jingrun Chen proved a "close" proposition to the Goldbach’s Conjecture in 1966: every even integer greater than 2 can be written as the sum of a prime and a semiprime (the product of two primes).

Then let us look at two examples which are not propositions.

Example 1.1.2.

(i) \(m < n\),  
(ii) \(x\) is a special number,  
(iii) \(n - m\).

Here (i) is a sentence but it is not a proposition. Because we don’t know whether it is true or not until the values of \(n, m\) are given. Sentences like this are called predicates. The symbols like \(n, m\) are called free variables.

(ii) is a sentence but it is meaningless in mathematics because we don’t know what "special" means. (ii) can be a predicates if we define, for example, “special” to be prime numbers’ square.

(iii) \(n - m\) is not a sentence.

Definition 1.1. The word statement will be used to denote a sentence which is either a proposition or a predicate.

1.2 Logical connectives

In mathematics, we often need to consider a large number of statements. We need logical connectives to connect them or to tell the logical relations of them.

We introduce three of the logical connectives: “or”, “and”, “not”.

Suppose we have two statements: \(P, Q\). If one of \(P\) and \(Q\) is true, then \(P\) or \(Q\) is true.
Example 1.2.1. (1) If \( P = n < m \) and \( Q = n \geq m \), then \( P \) or \( Q \) is true.

(2) A true proposition: \( ab = 0 \) if \( a = 0 \) or \( b = 0 \).

The connective “or” is sometimes hidden in other notations:

\[
\begin{align*}
    &a \leq b : \quad \text{means: } a < b \quad \text{or} \quad a = b; \\
    &a = \pm b : \quad \text{means: } a = b \quad \text{or} \quad a = -b;
\end{align*}
\]

Now we consider “and”. We use “and” when we want to assert that two things are both true.

Example 1.2.2. (1) If \( P = n < m \) and \( Q = n \geq m \), then \( P \) and \( Q \) has to be false.

(2) \( 3 < \pi < 4 \) means \( \pi > 3 \) and \( \pi < 4 \).

Now we discuss “not”. A statement \( P \) is true is and only if not \( P \) is false.

Example 1.2.3. Let \( f(x) \) be a polynomial and \( a \) be a real number. Write the negation of the following statement: if \( f(a) = 0 \) then \( a \) is positive.

Solution. Consider \( P \) to be a statement that \( f(a) = 0 \) and \( Q \) to be a statement that \( a \) is positive. Then the assertion (or statement) here is “\( P \) implies \( Q \)”. The negation of this is “not (\( P \) implies \( Q \))”. The negation is the same as “\( P \) does not implies \( Q \)”. And so it is “\( (P) \) and (not \( Q \)) can happen at the same time”.

Thus the correct negation is: it is possible that \( f(a) = 0 \) and \( a \) is non-positive i.e. there exists \( a \leq 0 \) such that \( f(a) = 0 \).

We will discuss this more in Theorem 2.2.

Example 1.2.4. Read the tables 1.2.1, 1.2.2 in the textbook.

2 Implications

Mathematics is primarily concerned with establishing the truth of statements. This is achieved by giving a proof of the statement. Implication is one of the key in proofs.

2.1 Implications

The definition of implications is the following: given two statements \( P \) and \( Q \), we say \( P \) implies \( Q \) if whenever \( P \) is true then we must have that \( Q \) is true. We use the notation \( P \implies Q \).

Example 2.1.1. Suppose that \( P \) is the statement that \( "x^2 - x - 2 > 0" \), \( Q_1 \) is the statement that \( "x > 2 \) or \( x < -1" \) and \( Q_2 \) is the statement that \( "x \neq 0" \). Then we have

\[
P \implies Q_1 \implies Q_2.
\]
Actually here we also have $Q_1 \implies P$.

But $Q_2$ does not implies $Q_1$ written as $Q_2 \nRightarrow Q_1$.

**Definition 2.1.** In an implication $P \implies Q$, $P$ is called the *hypothesis* or *antecedent* and $Q$ is called the *conclusion* or *consequent*.

**Example 2.1.2.** Consider the following statements:

(i) $(\pi < 4) \implies (1 + 1 = 2)$;
(ii) $(\pi < 4) \implies (1 + 1 = 3)$;
(iii) $(\pi < 3) \implies (1 + 1 = 2)$;
(iv) $(\pi < 3) \implies (1 + 1 = 3)$.

(i)(iii)(iv) is OK. For (i), both the assumption and the conclusion are true. For (iii)(iv), the assumption can never be true, so we do not care what the conclusion is in this case. For (ii), even under the assumption, the conclusion is false.

**Theorem 2.2.** The followings are equivalent (have the same meaning)

- $P \nRightarrow Q$;
- “$P \implies Q$” is false;
- “$P$ and (not $Q$)” is true.

**Reading implications.** $P \implies Q$ is the same as

- $P$ implies $Q$;
- $Q$ if $P$;
- $P$ only if $Q$;
- $P$ is sufficient for $Q$;
- $Q$ is necessary for $P$.

**Definition 2.3.** We define *equivalence* of two statements $P,Q$ (written as $P \iff Q$) as

$$(P \implies Q) \quad \text{and} \quad (Q \implies P).$$

We read $P \iff Q$ as
• $P$ is equivalent to $Q$.
• $P$ is necessary and sufficient for $Q$.
• $P$ if and only if $Q$ (sometimes written $P$ iff $Q$).
• $P$ precisely when $Q$.

**Definition 2.4.** By *universal* statement we mean the statement is always true for all values of variables in the statement. For example “$a^2 + b^2 \geq 0$ for all real numbers $a, b$”.

### 3 Proofs

A proof of a mathematical statement is a logical argument which establishes the truth of the statement. The steps of the logical argument are usually provided by implications.

#### 3.1 Direct proofs

**Example 3.1.1.** *For positive real numbers $a$ and $b$, prove $a < b \implies a^2 < b^2$.***

Before the proof, we need to list the assumptions that people take for granted. The following basic properties of real numbers are often referred to as *Axioms*. Let $a, b, c$ be any real numbers.

(i) Trichotomy law. For each pair of real numbers $a, b$, one of the following holds: $a < b$, $a = b$, $a > b$.

(ii) Addition law.

\[
\begin{align*}
\text{ } a < b &\iff a + c < b + c. \\
\text{ (ii) Multiplication law.} \\
\text{ } a < b &\iff ac < bc \quad \text{if } c > 0, \\
\text{ } a < b &\iff ac > bc \quad \text{if } c < 0.
\end{align*}
\]

(ii) Transitive law.

\[
\begin{align*}
\text{ } \text{ } a < b \text{ and } b < c &\implies a < c.
\end{align*}
\]

Now we are able to write a formal proof of example 3.1.1.

**Proof.** Given positive real numbers $a, b$ and suppose $a < b$. Then by the multiplication law $a^2 < ab$ (multiplying through by $a > 0$) and $ab < b^2$ (multiplying through by $b > 0$). By the transitive law $a^2 < ab < b^2$. Hence $a < b \implies a^2 < b^2$. $\square$
3.2 Constructing proofs backwards

Sometimes the problem would be much easier if we start with the conclusions instead of the assumptions.

Example 3.2.1. For real numbers $a$ and $b$ such that $a < b$, show $4ab < (a + b)^2$.

Actually the conclusion holds for all real $a, b$ that are not equal.

Let us construct a proof backwards:

$4ab < (a + b)^2 \iff 4ab < a^2 + 2ab + b^2$

$\iff 0 < a^2 - 2ab + b^2$

$\iff 0 < (a - b)^2$

$\iff a \neq b$ (think about why?)

$\iff a < b.$

Hence $a < b \implies 4ab < (a + b)^2$. With this backwards proof, if we write it backward, we get a direct proof:

Proof. $a < b \implies a \neq b \implies (a - b)^2 > 0 \implies 0 < a^2 - 2ab + b^2 \implies 4ab < a^2 + 2ab + b^2 \implies 4ab < (a + b)^2.$

4 Proof by contradiction

In the previous section we construct some simple direct proofs. Here we introduce a different idea of proof: proof by contradiction.

4.1 Proving negative statements by contradiction

Let us explain the idea through the following example:

Example 4.1.1. There do not exist integers $m, n$ such that

$$14m + 20n = 101.$$  \hspace{1cm} (1)

Idea: The problem is: known that $m, n$ are integers, and the goal is to show that $14m + 20n \neq 101$.

The proof by contradiction method is, instead of starting from what is known, we start with assuming that the goal is incorrect. So in this example, we assume that $14m + 20n = 101$ and $m, n$ are integers. And the goal is to show that this is impossible i.e. contradiction.

This would be easier to argue. Because, when $m, n$ are integers, $14m + 20n$ is an integer and furthermore it equals $2(7m + 10n)$ which is an even number. While 101 is odd, we obtain a contraction. This contraction indicates that the original statement (1) is true.

Now we consider a harder problem:
Example 4.1.2. Let $n$ be a positive integer. Then $4n + 3$ can not be one positive integer’s square.

Proof. Suppose for contradiction that there exist a positive integer $n$ such that $4n + 3 = m^2$ for some positive integer $m$.

Since $m$ is a positive integer, it is either odd or even. If it is even, we can suppose $m = 2k$ for some positive integer $k$. Then $m^2 = 4k^2$ is even. However $4n + 3$ is odd, $m^2 = 4n + 3$ is not possible.

Now if $m$ is odd, suppose $m = 2k + 1$. Then $m^2 = (2k + 1)^2 = 4k^2 + 4k + 1$ and thus the remainder of dividing $m^2$ by 4 equals 1. However the remainder of dividing $4n + 3$ by 4 equals 3. Therefore $m^2$ can not equal $4n + 3$. We get a contradiction.

In all $4n + 3$ can not be one positive integer’s square.

A template for proofs by contradiction. To show $P \implies Q$.

Proof. Suppose, for contradiction, that the statement $Q$ is false. Then [present argument which leads to a contradiction]. Hence our assumption that $Q$ is false must be false. Thus $Q$ is true as required.

4.2 Proving implications by contradiction

Example 4.2.1. If $a, b, c$ are integers such that $a > b$, then

$$ac \leq bc \implies c \leq 0.$$  

We are going to use the Multiplication law:

$$a < b \iff ac < bc \quad \text{if } c > 0,$$

However it is difficult to make use of the multiplicative law of inequalities in a direct proof. The easiest thing to do in this case is again a proof by contradiction.

Proof. For integers $a, b, c$ with $a > b$, suppose that $ac \leq bc$ but, for contradiction, that $c > 0$. Then the given statement $a > b$ implies that $ac > bc$ by the multiplicative law. However this contradicts with the statement that $ac \leq bc$. Hence the assumption that $c > 0$ must be false. Thus $ac \leq bc \implies c \leq 0$.

4.3 Proof by contrapositive

Theorem 4.1. Let $P, Q$ be two statements. Then $P \implies Q$ is equivalent to $(\text{not } Q) \implies (\text{not } P)$.

Can you prove the theorem?

A proof by contrapositive is to use the theorem and so that we show $(\text{not } Q) \implies (\text{not } P)$ to conclude with $P \implies Q$.

Let us again consider Example (4.2.1).
Proof. (of Example (4.2.1)) The contrapositive of the statement
\[ ac \leq bc \implies c \leq 0 \]
is the statement
\[ c > 0 \implies ac > bc \]
which is exactly the multiplicative law. Thus the proposition of Example (4.2.1) is true. □

4.4 Proving “or” statements

In this subsection we constructing a proof for composite statements involving “or”.

Example 4.4.1. If \( a \) and \( b \) are real numbers, then
\[ ab = 0 \iff a = 0 \text{ or } b = 0. \]

Before the proof, let us analyze the statement. We are asked to prove a statement of the form \( P \iff Q \). Then we are required to give two parts of proofs: \( P \implies Q \) and \( P \iff Q \).

Proof. “\( \implies \)” : Known \( ab = 0 \) and the goal is to show that \( a \) or \( b \) equals 0. If \( a = 0 \), the conclusion holds. Otherwise suppose \( a \neq 0 \). Then dividing \( ab = 0 \) through by \( a \) gives \( b = 0 \), as required. Hence \( ab = 0 \implies a = 0 \text{ or } b = 0 \).

“\( \iff \)” : If one of \( a, b \) equals 0, then it is a basic property of 0 that \( ab = 0 \). The conclusion holds. □

Actually we can discuss by cases for \( a, b \). there are three possibility for \( a \) (\( b \)): \( a < 0 \) or \( a = 0 \) or \( a > 0 \) (\( b < 0 \) or \( b = 0 \) or \( b > 0 \)). Discussing by cases leads to nine possibilities in all

<table>
<thead>
<tr>
<th></th>
<th>a&lt;0</th>
<th>a=0</th>
<th>a&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>b&lt;0</td>
<td>ab&gt;0</td>
<td>ab=0</td>
<td>ab&lt;0</td>
</tr>
<tr>
<td>b=0</td>
<td>ab=0</td>
<td>ab=0</td>
<td>ab=0</td>
</tr>
<tr>
<td>b&gt;0</td>
<td>ab&lt;0</td>
<td>ab=0</td>
<td>ab&gt;0</td>
</tr>
</tbody>
</table>

5 The induction principle

The induction principle is a special proof technique which is particularly useful when proving statements about the positive integers.

Suppose we wish to prove some property for all the positive integers: 1,2,3,4... That is we want to prove infinitely many statements labeled by \( n \). Suppose the statement involving positive integer \( n \) is \( P(n) \).

Axiom 5.1. (The induction principle) \( P(n) \) is true for all \( n = 1, 2, 3, \ldots \) if

(i) \( P(1) \) is true, and
(ii) $P(k) \implies P(k+1)$ for all positive integers $k$.

Remark 5.1. (i) is often referred to as the base case. In (ii), we assume $P(k)$ and the goal is to prove $P(k+1)$. The assumption $P(k)$ is called the inductive hypothesis.

Example 5.0.1. For all positive integers $n$ we have $n \leq 2^n$.

Proof. We use induction on $n$ using the Axiom 5.1.

Base case: For $n = 1$, since $n = 1 \leq 2 = 2^n$, then $n \leq 2^n$.

Inductive step: Suppose now as inductive hypothesis that $k \leq 2^k$ for a positive integer $k$. Then

$$2^{k+1} = 2 \times 2^k \geq 2k \geq k + 1.$$ 

So $k + 1 \leq 2^{k+1}$ as required.

Conclusion: Hence, by induction, $n \leq 2^n$ holds for all positive integer $n$. \qed

Example 5.0.2. For all positive integers $n$, the number $n^2 + n$ is even.

Proof. We use induction on $n$. (Base case) For $n = 1$, $n^2 + n = 1 + 1 = 2$ which is even as required. (Inductive step) Now suppose $n^2 + n$ is even for $n = k$ for some positive integer $k$. Then we have

$$k^2 + k = 2q \quad \text{for some positive integer } q.$$ 

For $n = k + 1$,

$$(k + 1)^2 + (k + 1) = k^2 + 2k + 1 + k + 1 = (k^2 + k) + 2(k + 1) = 2q + 2(k + 1)$$

which is again even. (Conclusion) Hence, by induction, $n^2 + n$ is even for all positive integers $n$. \qed

5.1 Changing the base case

Example 5.1.1. For all integers $n$ such that $n \geq 4$, we have the inequality $n^2 \leq 2^n$.

Solution. We use induction on $n$ starting with $n = 4$. (Base case) For $n = 4$, $n^2 = 16 = 2^4$ and so $n^2 \leq 2^n$.

(Inductive step) Suppose now as inductive hypothesis that $k^2 \leq 2^k$ for some $k \geq 4$. Then

$$2^{k+1} \geq 2k^2.$$ 

So we will have proved that $2^{k+1} \geq (k + 1)^2$ if we can show that $2k^2 \geq (k + 1)^2$. Notice

$$2k^2 - (k + 1)^2 = 2k^2 - k^2 - 2k - 1 = k^2 - 2k - 1 = k(k - 2) - 1 \geq 0$$

when $k \geq 4$. Therefore we proved $2^{k+1} \geq (k + 1)^2$ to complete the inductive step.

(Conclusion) Hence, by induction, $n^2 \leq 2^n$ for all $n \geq 4$. \qed

9
5.2 Definition by induction

Suppose given a sequence of numbers \( a_1, a_2, \ldots \), recall the summation notation: \( \sum_{i=1}^{n} a_i \) for positive integers \( n \). We can interpret the summation notation as a definition by induction:

(i) For \( n = 1 \), \( \sum_{i=1}^{1} a_i \) is defined as \( a_1 \);

(ii) (Suppose \( \sum_{i=1}^{k} a_i \) is defined.) For \( n = k + 1 \), \( \sum_{i=1}^{k+1} a_i = (\sum_{i=1}^{k} a_i) + a_{k+1} \).

Example 5.2.1. For positive integers \( n \), show

\[
\sum_{i=1}^{n} i = \frac{1}{2} n(n + 1),
\]

\[
\sum_{i=1}^{n} i^2 = \frac{1}{6} n(n + 1)(2n + 1).
\]

Proof. Let me only prove the first equality here.

We use induction on \( n \). (Base case) For \( n = 1 \), \( \sum_{i=1}^{1} i = 1 \) and \( \frac{1}{2} n(n + 1) = \frac{1}{2} \times 1 \times 2 = 1 \) and therefore the equality holds for \( n = 1 \).

(Inductive step) Suppose as inductive hypothesis that \( \sum_{i=1}^{k} i = \frac{1}{2} k(k + 1) \) for some positive integer \( k \). Then

\[
\sum_{i=1}^{k+1} i = \sum_{i=1}^{k} i + (k + 1) \quad \text{(by definition of summation notation)}
\]

\[
= \frac{1}{2} k(k + 1) \quad \text{(by inductive hypothesis)}
\]

\[
= \frac{1}{2} (k + 1)(k + 2) \quad \text{(by direct computations)}
\]

and so

\[
\sum_{i=1}^{k+1} i = \frac{1}{2} (k + 1)(k + 2)
\]

as required.

(Conclusion) Hence, by induction, \( \sum_{i=1}^{n} i = \frac{1}{2} n(n + 1) \) holds for all positive integer \( n \). \( \square \)

Another example is the factorial \( n \).

Definition 5.2. For non-negative integers \( n \), the number factorial \( n \) written \( n! \), are defined inductively by

(i) \( 0! = 1 \); and

(ii) \( (k + 1)! = k! \times (k + 1) \) for all non-negative integer \( k \).
5.3 The strong induction principle

**Axiom 5.2. (The strong induction principle)** Suppose that \( P(n) \) is a statement involving a general positive integer \( n \). Then \( P(n) \) is true for all positive integers \( n \) if

(i) \( P(1) \) is true, and

(ii) \( \left[ P(n) \text{ holds for all positive integers } n \leq k \right] \implies P(k + 1), \text{ for all positive integers } k. \)

In this case, the basic template is as follows.

**Proof.** We use (strong) induction on \( n \).

(Base case) \( \left[ \text{Prove the statement } P(1) \right] \).

(Inductive step) Suppose now as inductive hypothesis that \( \left[ P(n) \text{ is true for all positive integers } n \leq k \right] \) for some positive integer \( k \). Then \( \left[ \text{deduce that } P(k + 1) \text{ is true} \right] \). This proves the inductive step.

(Conclusion) Hence, by induction, \( \left[ P(n) \text{ is true for all positive integers } n \right] \). \( \Box \)

As an illustration of the use of this form of induction we introduce the Fibonacci numbers.

**Definition 5.3.** For each positive integer \( n \) define the number \( F_n \) inductively as follows.

\[
F_1 = 1, \quad F_2 = 1, \\
F_{k+1} = F_k + F_{k-1} \quad \text{for } k \geq 2.
\]

In this way, we defined a sequence of numbers. The beginning of this sequence is

\[ 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots, \]

and the sequence is known as the Fibonacci sequence.

**Proposition 5.4.** The Fibonacci numbers are given by the following formula:

\[
F_n = \left( \frac{\alpha^n - \beta^n}{\sqrt{5}} \right),
\]

where \( \alpha, \beta = \left(1 \pm \sqrt{5}\right)/2. \)

**Remark 5.5.** Here \( \alpha, \beta \) are the roots of

\[ x^2 - x - 1 = 0. \]

It is surprising here that these integers \( F_n \) involve \( \sqrt{5} \) which is not an integer (which is not even a rational number). Actually we will see later that there is a general procedure for finding a general formula for sequences defined inductively in this way.
Proof. (Base case) For $n = 1$, the formula gives $(\alpha - \beta)/\sqrt{5} = 1 = F_1$. Since the inductive formula does not apply until $F_3$, we have to calculate $F_3$. For $n = 2$, the formula gives $(\alpha^2 - \beta^2)/\sqrt{5}$. Using that $\alpha^2 = \alpha + 1$, $\beta^2 = \beta + 1$, we get

$$(\alpha^2 - \beta^2)/\sqrt{5} = 1 = F_2.$$  

(Inductive step) Now suppose as inductive hypothesis that the formula holds for all positive integers $n$ such that $n \leq k$ for some positive integer $k \geq 2$. Then

$$F_{k+1} = F_k + F_{k-1} \quad \text{(by definition of the sequence)}$$

$$= ((\alpha^k - \beta^k) + (\alpha^{k-1} - \beta^{k-1}))/\sqrt{5} \quad \text{(by inductive hypothesis)}$$

$$= ((\alpha^{k-1}(\alpha + 1) - \beta^{k-1}(\beta + 1))/\sqrt{5}$$

$$= (\alpha^{k+1} - \beta^{k+1})/\sqrt{5} \quad \text{(using } \alpha^2 = \alpha + 1, \beta^2 = \beta + 1)$$

as required to prove the formula for $n = k + 1$.

(Conclusion) Hence, by induction, the formula holds for all positive integers.  

6 The language of set theory

The language of set theory is used throughout mathematics. For example, real numbers, integers etc. are sets of numbers.

We usually use the following symbols:

- \( \mathbb{Z} \) : the set of all integers;
- \( \mathbb{Z}^+ \) : the set of all positive integers;
- \( \mathbb{Z}_{\geq 0} \) : the set of all non-negative integers (same as \( \mathbb{N} \), the natural numbers);
- \( \mathbb{Q} \) : the set of all rational numbers;
- \( \mathbb{R} \) : the set of all real numbers;
- \( \mathbb{C} \) : the set of all complex numbers.

6.1 Sets

More generally, a set is defined as a collection of objects (not necessarily numbers). The objects in a set are called the elements, members or points of the set. We write

$$x \in A$$

to denote that $x$ is an element of the set $A$. Here $x$ can be a number, a ball, a space or an apple etc. The negation of the statement $x \in A$ is written $x \notin A$. For example $\sqrt{2} \notin \mathbb{Q}$.

If only talking about a set of elements, we DO NOT consider any order of elements in a set; we DO NOT repeat the same element for more than one times in one set.

There are three ways to specify a set:
• **List the elements.** E.g. \( A = \{1, 3, \pi, -14, \text{“banana”}\} \).

• **Conditional definition.** E.g. \( B = \{n \in \mathbb{Z} \mid 0 < n < 6\} \).

• **Constructive definition** by giving a formula of elements. E.g. \( \{n^2 \mid n \in \mathbb{Z}\}, \{a/b \mid a, b \in \mathbb{Z}, b \neq 0\} \) (this is the same as \( \mathbb{Q} \)). The two set examples here contains infinitely many elements.

**Definition 6.1. Equality of sets.** Two set \( A, B \) are equal, written \( A = B \), if they have precisely the same elements, i.e. \( A = B \) means \( x \in A \iff x \in B \).

**Example 6.1.1.**

\[ \{x \in \mathbb{R} \mid x^2 - x - 2 = 0\} = \{-1, 2\} = \{2, -1\} = \{-1, -1, 2\}. \]

**Definition 6.2.** The empty set is the unique set which has no element at all, denoted as \( \emptyset \).

**Definition 6.3.** Given sets \( A, B \), we say that \( A \) is a subset of \( B \), written \( A \subseteq B \), when every element of \( A \) is an element of \( B \), i.e. \( x \in A \implies x \in B \). If in addition \( A \neq B \), then we say that \( A \) is a proper subset of \( B \) and write \( A \subset B \).

**Remark 6.4.** It can be shown that

\[ A = B \iff A \subseteq B \text{ and } B \subseteq A. \]

Next \( \in \) and \( \subseteq \) have very different meaning. For example \( a \in A \) means that \( a \) is an element of \( A \). While for \( B \subseteq A \), \( B \) is a set, but just smaller. We can write

\[ a \in A \iff \{a\} \subseteq A, \]

here \( \{a\} \) is viewed as a set which contains a single element.

### 6.2 Operations on sets

Given two sets \( A, B \).

**Definition 6.5.** The intersection of \( A \) and \( B \) is denoted by \( A \cap B \). We have

\[ A \cap B = \{x \mid x \in A \text{ and } x \in B\}. \]

\( A \) and \( B \) are said to be disjoint if \( A \cap B = \emptyset \).

The union of \( A \) and \( B \) is denoted by \( A \cup B \). We have

\[ A \cup B = \{x \mid x \in A \text{ or } B\}. \]

We can also form a set of elements which lie in \( A \) but not \( B \). This is called the difference of \( A \) and \( B \), denoted as \( A - B \). Thus

\[ A - B = \{x \mid x \in A \text{ and } x \notin B\}. \]

We call use Venn diagram to illustrate the above operations. See page 68 of the textbook.

**Example 6.2.1.** Given any two sets \( A \) and \( B \), the three sets \( A \cap B \), \( A - B \) and \( B - A \) are pairwise disjoint (i.e. each pair of these sets is disjoint) and

\[ A \cup B = (A \cap B) \cup (A - B) \cup (B - A). \]
6.3 The power set

The power set is a more complicated way of constructing sets from one set.

Definition 6.6. The power set of a set $X$ is defined as the set of all subsets of the set $X$, denoted as $\mathcal{P}(X)$.

Thus $A \in \mathcal{P}(X)$ is an alternative way of writing $A \subseteq X$.

We have $\emptyset \in \mathcal{P}(X)$, and $X \in \mathcal{P}(X)$.

Example 6.3.1. If $X = \{a, b\}$, then

$$\mathcal{P}(X) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$  

It is often the case that all the sets we are considering are subsets of a fixed set. We refer to this largest set as the universal set.

Definition 6.7. Suppose $U$ is a universal set and $A \subseteq U$. Then the complement of $A$ in $U$ is defined as

$$A^c := U - A.$$  

Here we can also write $A, A^c \in \mathcal{P}(U)$.

Theorem 6.8. Let $A, B, C$ be subsets of a set $U$. Then the following identities hold:

(i) Associativity:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C.$$  

(ii) Commutativity:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$  

(iii) Distributivity:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$  

(iv) De Morgan laws:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$  

(v) Complementation:

$$A \cup A^c = U, \quad A \cap A^c = \emptyset.$$  

(vi) Double complement:

$$(A^c)^c = A.$$
Proof. (of \((A \cup B)^c = A^c \cap B^c\). In the textbook one of the distributive law is proved. Please read it.) (This is like showing “equivalence \(\iff\); we prove “=” from two directions.)

Proof of “\(\subseteq\)”: Suppose \(x \in (A \cup B)^c\). Then \(x \notin (A \cup B)\) and so \(x \notin A\) and \(x \notin B\). \(x \notin A\) implies that \(x \in A^c\) and \(x \notin B\) implies that \(x \in B^c\). Therefore we get \(x \in A^c \cap B^c\).

Proof of “\(\supseteq\)”: Suppose \(x \in A^c \cap B^c\). Then \(x \in A^c\) and \(x \in B^c\). We get \(x \notin A\) and \(x \notin B\), and so \(x \notin A \cup B\) which implies that \(x \in (A \cup B)^c\). \(\square\)

**Proposition 6.9.** Show

\((A \cup B) \cap (C \cup D) = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D)\).

Proof. Again here we are asked to prove an equality of sets. In the following, we apply Proposition 6.9 and show the two directions at the same time.

First let us view \(A \cup B\) as one single set and apply the distributivity law:

\[(A \cup B) \cap (C \cup D) = ((A \cup B) \cap C) \cup ((A \cup B) \cap D).\]

By distributivity again

\[((A \cup B) \cap C) = (A \cap C) \cup (B \cap C),\]

\[((A \cup B) \cap D) = (A \cap D) \cup (B \cap D).\]

The associativity of the union operation implies that

\[(A \cup B) \cap (C \cup D) = (A \cap C) \cup (B \cap C) \cup (A \cap D) \cup (B \cap D) \]

\[= (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D).\]

\(\square\)

### 6.4 Cartesian product of two sets

Cartesian is one way to construct a set from two set.

**Definition 6.10.** Given sets \(X, Y\), the Cartesian product of \(X, Y\), denoted by \(X \times Y\), is the set of all ordered pairs \((x, y)\) where \(x \in X\) and \(y \in Y\). Thus

\[X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}.\]

We say that the ordered pair \((x, y)\) has coordinates \(x\) and \(y\).

**Example 6.4.1.** \(\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}\) is the familiar 2-dimensional Euclidean space. \(\mathbb{R}^d = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}\) is the \(d\)-dimensional Euclidean plane.

**Proposition 6.11.** For all sets \(A, B, C\) and \(D\) the following hold:

(i) \(A \times (B \cup C) = (A \times B) \cup (A \times C)\);
(ii) \( A \times (B \cap C) = (A \times C) \cap (A \times C); \)

(iii) \( (A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D); \)

(iv) \( (A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D). \)

**Proof.** (of part (iii), the textbook proved (ii).) A proof that two sets are equal requires us to prove two inclusions. In this case we can do them together as follows.

\[
(x, y) \in (A \times B) \cap (C \times D) \iff (x, y) \in (A \times B) \text{ and } (x, y) \in (C \times D)
\]

\[
\iff x \in A, y \in B \text{ and } x \in C, y \in D
\]

\[
\iff x \in A \cap C \text{ and } y \in B \cap D
\]

\[
\iff (x, y) \in (A \cap C) \times (B \cap D).
\]

\[
\square \]

7 Quantifiers

In this section we consider a set of statements \( P(a) \) with a single free variable \( a \) taking values in a set \( A \). Recall that \( P(a) \) is called a predicate. The statement that all predicates \( P(a) \) for \( a \in A \) are true is called a *universal statement*.

Next we introduce two notations:

**Definition 7.1.** The notation \( \forall a \in A; P(a) \) means: for each (or for any, for all) \( a \in A \), \( P(a) \) is true. (\( \forall \) is called the *universal quantifier symbol*.)

The notation \( \exists a \in A; P(a) \) means there exists (at least) one \( a \in A \) such that \( P(a) \) is true. (\( \exists \) is called the *existential quantifier symbol*. It reads: “there exists”, “for some” etc.)

7.1 Proving statements involving quantifiers

**Example 7.1.1.** Prove \( \exists n \in \mathbb{Z}^+, 2n + 1 \text{ is a prime number}. \)

*Proof.* Let \( n = 3 \) and then \( n \in \mathbb{Z}^+ \). We find that \( 2n + 1 = 7 \) is a prime number. \( \square \)

**Example 7.1.2.** For integers \( n \), if \( n \text{ is even then } n^2 \text{ is even.} \)

This is a universal implication: \( \forall n \in \mathbb{Z} \text{ etc.} \) The hypothesis that \( n \) is even is an existence statement, because \( n \text{ is even is the same as } \exists q \in \mathbb{Z} \text{ such that } n = 2q. \)

*Proof.* Since \( n \text{ is even, } \exists q \in \mathbb{Z} \text{ such that } n = 2q. \) Then \( n^2 = (2q)^2 = 2(2q^2) \) which is again even. \( \square \)

**Example 7.1.3.** To *disprove the statement* \( \forall x \in \mathbb{R}, x \leq 10^{10}. \)

16
Proof. Let $x = 10^{10} + 1$. Then $x \in \mathbb{R}$ and $x > 10^{10}$.

Here in the example, we are asked to prove:

$\exists a \in A$, not $P(a)$.

We gave a counterexample as a proof. This is often called disproof by counterexample to $P(a)$.

### 7.2 Predicates involving more than one variable

Let $P(a, b)$ be a predicate involving two free variables $a \in A, b \in B$. Then there are several ways to form a proposition.

**Example 7.2.1.** $\forall a, b \in \mathbb{R}^+, a < b \implies a^2 < b^2$.

Here in the examples, “$\forall a, b \in A$” is a shorthand for “$\forall a \in A, \forall b \in A$” and similarly for “$\exists$”.

**Example 7.2.2.** It is not true that $\exists m, n \in \mathbb{Z}, 14m + 20n = 9$.

Statements might also involve both quantifiers:

**Example 7.2.3.** $\forall \epsilon \in \mathbb{R}^+, \exists N \in \mathbb{R}^+$ such that $\frac{1}{N} < \epsilon$.

**Proof.** Given any $\epsilon > 0$, we can pick $N = \frac{1}{\epsilon} + 1$. Then

$$\frac{1}{N} = \frac{1}{\frac{1}{\epsilon} + 1} = \frac{\epsilon}{1 + \epsilon} < \epsilon.$$  

**Example 7.2.4.** $\exists N > 0, \forall \epsilon \in \mathbb{R}^+, N > \frac{2 + \epsilon}{1 + \epsilon}$.

**Proof.** Let us pick $N = 2$. Then

$$2 > \frac{2 + \epsilon}{1 + \epsilon}$$

is true because it is equivalent to

$$2(1 + \epsilon) > 2 + \epsilon \iff \epsilon > 0.$$  

**Use quantifiers to define limit.**

**Definition 7.2.** Let $\{a_n, n \in \mathbb{Z}^+\}$ be a sequence of real numbers. Then $\lim_{n \to \infty} a_n = c$ if and only if

$$\forall \epsilon > 0, \exists N \in \mathbb{Z}^+ \text{ such that } \forall n \geq N, \quad |a_n - c| \leq \epsilon.$$
8 Functions

The notion of function is one of the most fundamental in mathematics.

Definition 8.1. Suppose that $X$ and $Y$ are sets. A function or map from $X$ to $Y$ is the assignment of a unique element of $Y$ to each element of $X$. We write $f : X \to Y$. We denote the element $y \in Y$ assigned to $x \in X$ as $y = f(x)$. Here $y$ or $f(x)$ is called the value or the image of $x$. The set $X$ is called the domain and the set $Y$ is called the codomain.

Example 8.0.1. Suppose $X = \{a, b\}$ and $Y = \{0, 1\}$. Then a function $f_1$ from $X$ to $Y$ can be defined though

$$f_1(a) = 0, \quad f_1(b) = 0.$$  

There are precisely 4 functions from $X$ to $Y$ given as

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_1(x)$</th>
<th>$f_2(x)$</th>
<th>$f_3(x)$</th>
<th>$f_4(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$b$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Here $f_1, f_4$ are called constant functions, since the image of $X$ is one fixed element of $Y$.

Of course a function can have the same set as its domain and codomain i.e. $X = Y$. For example the maps $f_i : X \to X$ for $i = 1, 2, 3, 4$ with 0, 1 by $a, b$. Then $f_3$ becomes a function such that $f_3(a) = a$ and $f_3(b) = b$. This function is called the identity function. The definition of identity function is: $f : X \to X$ is given by $f(x) = x$ for all $x \in X$. We often write such functions as $I_X$.

We can use explicit formulae to define a function. But we need to make sure that we assign exactly one value to each element in the domain.

Example 8.0.2. The modulus function $x \to |x|$ is defined from $\mathbb{R}$ to $\mathbb{R}^+$ as

$$|x| = \begin{cases} 
  x & \text{if } x \geq 0, \\
  -x & \text{if } x \leq 0.
\end{cases}$$

The value of the function at 0 is given twice. Think about why this function is still well-defined?

The following function is not well defined on $\mathbb{R}$ to $\mathbb{R}$, but it is well-defined on $\mathbb{R}^+$ to $\mathbb{R}$ (or $\mathbb{R}^+$):

$$f(x) = 1 + \frac{1}{x}.$$  

Definition 8.2. Two functions $f, g : X \to Y$ are equal, written $f = g$, if $f(x) = g(x)$ for all $x \in X$.

Definition 8.3. Suppose that $f : X \to Y$ is a function and $A$ is a subset of $X$, i.e. $A \subseteq X$. Then we can define a function $g : A \to Y$ by setting $g(a) = f(a)$ for all $a \in A$. This function is called the restriction of $f$ to $A$ and is denoted as $g = f|A$. 

18
8.1 Composition of functions

**Definition 8.4.** Given two functions \( f : X \to Y \) and \( g : Y \to Z \), the **composite** of \( f \) and \( g \), denoted by \( g \circ f \) is a map from \( X \to Z \) defined by

\[
    g \circ f (x) = g(f(x)) \quad \text{for all } x \in X.
\]

The order of \( f, g \) is important. In general when \( g \circ f \) is well-defined, \( f \circ g \) might not be well-defined (e.g. \( Z \not\subseteq X \)). Even if they are both well-defined (e.g. \( X = Y = Z \)), they are different functions in general. We can write the following to indicate the order:

\[
    g \circ f : X \xrightarrow{f} Y \xrightarrow{g} Z.
\]

**Example 8.1.1.** Let \( f, g : \mathbb{R} \to \mathbb{R} \) be defined by \( f(x) = x + 1 \) and \( g(x) = x^2 \). Then

\[
    g \circ f (x) = (x + 1)^2, \quad f \circ g (x) = x^2 + 1.
\]

8.2 Sequences

By “sequence” (alone), we usually mean a sequence of real numbers. Below we use functions to define a sequence of elements in one set \( A \).

**Definition 8.5.** A function \( f : \mathbb{Z}^+ \to A \) is called a **sequence in** \( A \).

In the following use quantifiers to give a rigorous definition of a sequence converging to 0.

**Definition 8.6.** Given a sequence \( f : \mathbb{Z}^+ \to \mathbb{R} \) of real numbers, we say that the sequence is **null**, written \( \lim_{n \to \infty} f(n) = 0 \), when

\[
    \forall \epsilon \in \mathbb{R}^+, \exists N \in \mathbb{Z}^+ \text{ such that } \forall n \geq N, \text{ we have } |f(n)| < \epsilon.
\]

**Example 8.2.1.** The sequence \( n \to \frac{1}{\sqrt{n}} \) is null.

**Constructing a proof.** By the definition, we are required to show the that sequence converges to 0 i.e.

\[
    \forall \epsilon \in \mathbb{R}^+, \exists N \in \mathbb{Z}^+ \text{ such that } (n \geq N \implies |f(n)| < \epsilon).
\]

The key point is, after fixing \( \epsilon > 0 \), to find out the value of \( N \).

We study for what \( n \), \( |f(n)| < \epsilon \). Notice

\[
    \frac{1}{\sqrt{n}} < \epsilon \iff \frac{1}{n} < \epsilon^2 \iff n > \frac{1}{\epsilon^2}.
\]

Hence we only need to select \( N \) to be one integer which is greater than \( \frac{1}{\epsilon^2} \).

**Proof.** For any positive \( \epsilon \), let us take \( N \) be the smallest positive integer which \( > \frac{1}{\epsilon^2} \). Then \( n \geq N \) implies that \( n \geq N > \frac{1}{\epsilon^2} \). So

\[
    \frac{1}{\sqrt{n}} < \epsilon
\]

as required. \( \square \)
8.3 The image and the graph of a function

Definition 8.7. Given a function \( f : X \to Y \), the image of \( f \), denoted as \( \text{Im} f \) is defined as

\[
\text{Im} f = \{ f(x) \mid x \in X \}.
\]

For any \( y \in Y \), a pre-image of \( y \) is an element \( x \in X \) such that \( y = f(x) \). We also often call the following set the pre-image of \( y \)

\[
\{ x \in X \mid f(x) = y \}.
\]

Hence the image of \( f \) is a subset of the codomain \( Y \). The pre-image of \( y \) of \( f \) is a subset of \( X \).

Definition 8.8. Suppose that \( f : X \to Y \) is a function. The graph of \( f \) is defined as

\[
G_f = \{(x, f(x)) \in X \times Y \}.
\]

Hence the graph of \( f \) is a subset of the Cartesian product of \( X \times Y \).

In the case that \( X = Y = \mathbb{R} \), this definition of the graph of a function is the same as we have seen before.

Example 8.3.1. Find the graph of \( f : \mathbb{R} \to \mathbb{R} \) where \( f(x) = x^2 \).

Find the graph of \( f_1 \) in Example 8.0.1.

9 Properties of functions

9.1 Injections, surjections and bijections

Definition 9.1. Let \( X, Y \) be two sets. Suppose \( f : X \to Y \) is a function.

(i) If \( \forall x_1, x_2 \in X \) and \( x_1 \neq x_2 \), we have \( f(x_1) \neq f(x_2) \), then we say that the function \( f \) is injective or one-to-one.

(ii) If \( \forall y \in Y \), we can find \( x \in X \) such that \( f(x) = y \), then \( f \) is surjective or onto.

(iii) If \( f \) is both an injection (\( f \) is injective) and a surjection (\( f \) is surjective), then we say that it is bijection or that it is bijective or one-to-one and onto.

Example 9.1.1. Show \( f_1 : \mathbb{R} \to \mathbb{R} \) is bijective, where \( f_1(x) = x + 1 \). Show \( f_2 : \mathbb{R} \to \mathbb{R}^+ \) is bijective, where \( f_2(x) = e^x \).

Proof. Let us only prove for \( f_2 \). First we show that \( f_2 \) is injective. Take any two different real numbers \( x_1, x_2 \). Then \( e^{x_1} \neq e^{x_2} \) and so it is injective. Next we show that \( f_2 \) is surjective. Fix any \( y \in \mathbb{R}^+ \). Let \( x = \ln y \) and then \( e^x = e^{\ln y} = y \). So \( f_2 \) is surjective. In all, it is a bijection. \( \Box \)
For a function \( f : X \to Y \), given an element \( y \in Y \), recall that a pre-image of \( y \) is an element of \( x \in X \) such that \( y = f(x) \).

So \( f \) is an injection if and only if every \( y \in Y \) has at most one pre-image; \( f \) is a surjection if and only if every \( y \in Y \) has at least one pre-image.

**Remark 9.2.** Recall that the image of \( f : X \to Y \) is defined as

\[
\text{Im } f = \{ f(x) \mid x \in X \}.
\]

Hence we can convert \( f \) into a surjection by changing its codomain from \( Y \) to \( \text{Im } f \subseteq Y \).

**Example 9.1.2.** To determine whether the function \( f : \mathbb{R}^+ \to \{ x \geq 1 \} \) given by \( f(x) = 4x^2 - 4x + 2 \) is injective, surjective or bijective.

**Solution.** Notice that \( y = f(x) \) is equivalent to

\[
y = 4x^2 - 4x + 2 \iff y = (2x - 1)^2 + 1 \iff x = \frac{1 \pm \sqrt{y - 1}}{2}.
\]

For any \( y \) in the codomain, we have \( y \geq 1 \) and then \( x = \frac{1 \pm \sqrt{y - 1}}{2} \) are its pre-image. Thus \( f \) is surjective. However it is not injective since for all \( y > 1 \), there are two pre-images \( x \). And the function is not bijective.

### 9.2 Bijections and inverses

**Definition 9.3.** A function \( f : X \to Y \) is called invertible if there exists a function \( g : Y \to X \) such that

\[
y = f(x) \iff x = g(y)
\]

for all \( x \in X \) and \( y \in Y \). In this case, we call \( g \) an inverse (function) of \( f \) and write \( g = f^{-1} \).

**Remark 9.4.** If \( g = f^{-1} \), then \( f = g^{-1} \).

**Theorem 9.5.** Let \( f : X \to Y \). Then \( f \) is invertible if and only if it is a bijection. Furthermore, if \( f \) is invertible, its inverse function is unique.

Intuitively the result is straightforward. Let me present it below demonstrating that how a rigorous proof often looks like.

**Proof.** The first statement is an equivalence. We need to prove two directions.

(a) “\( f \) invertible \( \implies \) \( f \) bijective”: Suppose \( f \) is invertible, and then let us write its inverse as \( g \).

Below we show that \( f \) is injective. Take any \( x_1, x_2 \in X \). We only need to show that

\[
f(x_1) = f(x_2) \implies x_1 = x_2.
\]
(Compare this with the definition. Here we used that \( A \implies B \) is the same as \( \neg B \implies \neg A \).) Since \( f(x_1) = f(x_2) \), then \( g(f(x_1)) = g(f(x_2)) \). Since \( g = f^{-1} \), this implies that \( x_1 = x_2 \).

For surjectivity suppose \( y \in Y \). Take \( x = g(y) \in X \). Then

\[
f(x) = f(g(y)) = y
\]

which implies that \( f \) is also surjective. In all \( f \) is a bijection.

(b) “\( f \) bijective \implies \( f \) invertible”: Suppose that \( f \) is bijective. To show that it is invertible we need to construct a function \( g : Y \to X \) such that \( g = f^{-1} \). For any \( y \in Y \), since \( f \) is surjective, there is a unique \( x \in X \) such that \( f(x) = y \). Then we define \( g(y) = x \). By this definition we have for any \( x \in X \), \( y \in Y \)

\[
y = f(x) \iff x = g(y)
\]

and so this function \( g \) is the inverse of \( f \).

(c) “\( f \) invertible \implies \) unique inverse”: Suppose \( f \) has two inverse functions \( g_1, g_2 \). We need to show that \( g_1 = g_2 \). To do this we pick any \( y \in Y \) and it suffices to show that \( g_1(y) = g_2(y) \). Let us write \( x_1 = g_1(y) \) and \( x_2 = g_2(y) \). Then

\[
f(x_1) = f(g_1(y)) = y = f(g_2(y)) = f(x_2).
\]

Since \( f \) is bijective, this shows that we must have \( x_1 = x_2 \) which concludes part (c).

Example 9.2.1. The function \( \sin : \mathbb{R} \to \mathbb{R} \) is neither injective nor surjective. However notice the image of \( \sin \) function is \( [-1, 1] \), and so \( \sin : \mathbb{R} \to [-1, 1] \) is surjective. But it is not injective.

The function \( \sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1] \) is a bijection.

Remark 9.6. Once we change the domain or the codomain of a function, we view this function as a different one. However with a slight abuse of notation, people sometimes use the same symbol, for example \( \sin \) in the above example.

Remark 9.7. By this example, we know that \( \sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1] \) has an inverse

\[
\sin^{-1} \text{ (or arcsin)} : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}].
\]

In making this definition we had to choose a subset of \( \mathbb{R} \) on which \( \sin \) is a bijection. Notice that the choice \( [-\frac{\pi}{2}, \frac{\pi}{2}] \) is not unique. Here, with this choice, \( \sin^{-1} \) is called the principal value of the inverse.

In the same way, we can obtain

\[
\cos^{-1} : [-1, 1] \to [0, \pi],
\]

and

\[
\tan^{-1} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2}).
\]
Remark 9.8. So far, the bijections we gave satisfy that: if the domain is an open (close) set, then the codomain is also an open (close) set. Can you construct a function $f : (0, 1) \to [0, 1]$ such that $f$ has an inverse?

To end this subsection, let me present one proposition. I refer the readers to Proposition 9.2.5 in the book for the proof.

**Proposition 9.9.** The functions $f : X \to Y$ and $g : Y \to X$ are inverses of each other if and only if

$$g \circ f = I_X \quad \text{and} \quad f \circ g = I_Y.$$  

Here $\circ$ is the composite operator, $I_X, I_Y$ are identity functions on $X, Y$.

### 9.3 Functions and subsets

We can extend the definition of a function $f : X \to Y$ to $f : \mathcal{P}(X) \to \mathcal{P}(Y)$. And then $f^{-1}$ is well-defined without requiring $f$ to be bijective.

**Definition 9.10.** Suppose that $f : X \to Y$ is a function.

(i) The function $\mathcal{f} : \mathcal{P}(X) \to \mathcal{P}(Y)$ is defined by

$$\mathcal{f}(A) = \{ f(x) \mid x \in A \}$$

for $A \in \mathcal{P}(X)$.

(ii) The function $\mathcal{f} : \mathcal{P}(Y) \to \mathcal{P}(X)$ is defined by

$$\mathcal{f}(B) = \{ x \in X \mid f(x) \in B \}$$

for $B \in \mathcal{P}(Y)$.

**Remark 9.11.** It can be seen that $\mathcal{f}, \mathcal{f}$ are really extensions of $f, f^{-1}$. Because for example, $f(x) = y$ is the same as $\mathcal{f}(\{x\}) = \{y\}$ and if $f$ is invertible, $f^{-1}(y) = x$ is the same as $\mathcal{f}(\{y\}) = \{x\}$. While if $f$ is not a injection, $\mathcal{f}(\{y\})$ will not be a singleton subset of $Y$. Think about what happens if $f$ is not a surjection?

The notations $\mathcal{f}, \mathcal{f}$ are non-standard. Most mathematicians denote $\mathcal{f}$ simply by $f$ and denote $\mathcal{f}$ by $f^{-1}$. In the words of the book, this is an example of how ambiguity can be better than pedantry. However we will still use $\mathcal{f}, \mathcal{f}$, since it is potentially confusing to denote different functions by one symbol.

**Example 9.3.1.** Let $f : X \to Y$ be a function and $B_1, B_2 \in \mathcal{P}(Y)$. Prove that

$$B_1 \subseteq B_2 \implies \mathcal{f}(B_1) \subseteq \mathcal{f}(B_2).$$
Proof. By definition of \( \overleftarrow{f} \), for \( i = 1, 2 \)

\[
x \in \overleftarrow{f}(B_i) \iff f(x) \in B_i.
\]

Assume \( B_1 \subseteq B_2 \). Now for any \( x \in \overleftarrow{f}(B_1) \), \( f(x) \in B_1 \). Since \( B_1 \subseteq B_2 \), we have \( f(x) \in B_2 \) and then \( x \in \overleftarrow{f}(B_2) \). We obtain

\[
x \in \overleftarrow{f}(B_1) \implies x \in \overleftarrow{f}(B_2),
\]

which implies the statement. \qed

**Axiom 9.1.** (Peano) The set of positive integers \( \mathbb{Z}^+ \) is a set with a function \( s : \mathbb{Z}^+ \to \mathbb{Z}^+ \) and an element \( 1 \in \mathbb{Z}^+ \) such that

(i) \( s \) is an injection,

(ii) \( 1 \) is not in the image of \( s \),

(iii) for \( A \subseteq \mathbb{Z}^+ \), if \( 1 \in A \) and \( (n \in A) \implies s(n) \in A \), then \( A = \mathbb{Z}^+ \).

**Remark 9.12.** Once known the set is \( \mathbb{Z}^+ \), the the function \( s \) described in the above is the successor function and it is given by \( s(n) = n + 1 \).

With the choice of \( s(n) = n + 1 \), we can check that (i) holds i.e. \( s \) is an injection; (ii) holds i.e. \( s(n) \geq 2 \) for \( n \in \mathbb{Z}^+ \); (iii) if \( 1 \in A \), by the assumption \( s(1) = 2 \in A \). By induction it can be shown that \( A = \mathbb{Z}^+ \). So for the normal \( \mathbb{Z}^+ \), the criteria (i)-(iii) hold.

However what is important here is that the Axiom is announcing that any set satisfying (i)-(iii) can be seen as a copy of \( \mathbb{Z}^+ \).

## 10 Counting

Counting is a very common acting in real life. Here we provide another way to think of counting in mathematics (we use bijection).

We start with a sequence of standard set

\[
\mathbb{N}_n := \{1, 2, ..., n\}.
\]

**Definition 10.1.** Given a set \( X \), if there is a bijection \( f : \mathbb{N}_n \to X \) then we say that the cardinality of \( X \), or the number of elements in \( X \), is \( n \) and write \( |X| = n \). In particular \( |\emptyset| = 0 \).

For example the cardinality of

\[
A = \{k \in \mathbb{Z} \mid 25 \leq k < 30\}
\]

is 5. The cardinality of

\[
A = \{k \in \mathbb{R} \mid 25 \leq k < 30\}
\]

is infinity. However let me mention that, there are infinitely many different “infinite” that can be defined through bijection.
**Proposition 10.2.** Suppose that \( f : \mathbb{N}^n \to X \) and \( g : \mathbb{N}^n \to X \) are bijections with the same codomain \( X \). Then \( m = n \). (But in general we don’t have \( f = g \).)

The proposition is very important to our Definition 10.1. Imagine, if the proposition fails, then according to Definition 10.1, the cardinality of \( X \) can then be both \( n \) and \( m \), which certainly does not make sense.

The proof of the proposition can be found in the textbook. The key idea is to use that 1. bijections have inverse functions. 2. the following lemma.

**Lemma 10.3.** If there exists an injection \( \mathbb{N}^m \to \mathbb{N}^n \), then \( m \leq n \).

**Proof.** Let us give a proof by induction on \( n \).

Base case: \( n = 1 \). Suppose for contradiction that if \( m > 1 \) and we have \( f : \mathbb{N}^m \to \mathbb{N}^1 \) is an injection. Then \( 1, 2 \in \mathbb{N}_m \) and \( f(1), f(2) \in \mathbb{N}_1 = \{1\} \). Thus we have \( f(1) = f(2) \) which contradicts with the assumption that \( f \) is injective, and so \( m = 1 \leq n \).

Inductive step: suppose for some \( k \geq 1 \), the result holds for \( n = k \). We need to show the result for \( n = k + 1 \). We consider two cases:

1. Suppose \( f(i) < k + 1 \) for all \( i \in \mathbb{N}^m \). Then we can restrict the codomain of the function \( f \) from \( N_{k+1} \) to \( N_k \). Then by inductive hypothesis, \( m \leq k < k + 1 \).

2. On the other hand, suppose that \( k + 1 \) is a value of \( f \), say \( f(i_0) = k + 1 \). By re-arranging \( \mathbb{N}^n \), without loss of generality, we can assume that \( i_0 = m \). By injectivity of \( f \), \( f(i) \leq k \) for \( i \leq m - 1 \). Hence if we define \( f_1 : \mathbb{N}^m \to \mathbb{N}^k \) by

\[
f_1(i) := f(i),
\]

the function \( f_1 \) is an injection. By inductive hypothesis, \( m - 1 \leq k \) and therefore we get \( m \leq k + 1 \) as required.

Conclusion: by induction on \( n \), if \( f : \mathbb{N}^m \to \mathbb{N}^n \) is an injection then \( m \leq n \). \( \square \)

**Definition 10.4.** Given a set \( X \), if \( |X| = n \) for some \( n \), then we say the set is *finite*. Otherwise, we say that the set is *infinite*.

For example, natural numbers \( \mathbb{N} \) is infinite, and real numbers \( \mathbb{R} \) is also infinite.

**10.1 Properties**

**Theorem 10.5.** Suppose that \( X \) and \( Y \) are disjoint finite sets. Then \( X \cup Y \) is finite and

\[
|X \cup Y| = |X| + |Y|.
\]

**Proof.** Let \( |X| = n \) and \( |Y| = m \). If one of \( n, m \) is 0, then one of \( X, Y \) is an empty set. And the statement holds automatically. So let us consider the case that \( n, m \geq 1 \).

We know there are bijections \( f : \mathbb{N}^n \to X \) and \( g : \mathbb{N}^m \to Y \). The goal is define a bijection \( h : \mathbb{N}^{n+m} \to X \cup Y \). Let

\[
h(i) = \begin{cases} 
  f(i) & \text{if } 1 \leq i \leq n, \\
  g(i - n) & \text{if } n + 1 \leq i \leq n + m.
\end{cases}
\]

25
This this is indeed the bijection required. Because, injectivity follows from the injectivity of \( f, g \) and \( X \cap Y = \); surjectivity follows from that \( f, g \) are surjective.

**Theorem 10.6.** Suppose that \( X, Y \) are finite sets such that \( |X| = n \) and \( |Y| = m. \) Then the Cartesian product of \( X \times Y \) is a finite set and \( |X \times Y| = mn. \)

We omit the proof.

**Example 10.1.1.** The cardinality of \( \mathbb{N}_9 \times \mathbb{N}_9 \) is \( 9 \times 9 = 81. \)

The diagonal set in \( \mathbb{N}_9 \times \mathbb{N}_9 \) is \([ (x, x) \mid x \in \mathbb{N}_9 ]. \)

We denote the set as \( \Delta(\mathbb{N}_9) \) and then \( |\Delta(\mathbb{N}_9)| = 9. \)

What is the number of ordered pair of distinct positive integers no greater than 9? The answer is 36. One reason can be:

\[
\frac{|\mathbb{N}_9 \times \mathbb{N}_9| - |\Delta(\mathbb{N}_9)|}{2} = \frac{81 - 9}{2} = 36.
\]

Think about why.

The following proposition is often referred to as the inclusion-exclusion principle.

**Proposition 10.7.** Suppose that \( X, Y \) are finite set. Then

\[
|X \cup Y| = |X| + |Y| - |X \cap Y|.
\]

More generally, Suppose \( X, Y, Z \) are finite sets. Then

\[
|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|.
\]

**Example 10.1.2.** At an international conference of 100 people, 80 speak English, 60 speak Spanish and 40 speak Chinese (any everyone present speaks at least one of these language). What is the maximum possible number of the people who can speak the three languages.

**Solution.** Let \( X \) be the set of people who speak English, \( Y \) Spanish and \( Z \) Chinese. Then from the assumption

\[
|X \cup Y \cup Z| = 100, \quad |X| = 80, \quad |Y| = 60, \quad |Z| = 40.
\]

From the inclusion-exclusion principle,

\[
100 = 80 + 60 + 40 - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|,
\]

which simplifies to

\[
|X \cap Y| + |X \cap Z| + |Y \cap Z| = 80 + |X \cap Y \cap Z|.
\]

The people who can speak the three languages are \( X \cap Y \cap Z. \) Since \( |X \cap Y|, |X \cap Z|, |Y \cap Z| \geq |X \cap Y \cap Z|, \) we have

\[
80 + |X \cap Y \cap Z| \geq 3|X \cap Y \cap Z|.
\]

So

\[
|X \cap Y \cap Z| \leq 40.
\]

The maximum number of people who can speak the three languages are 40. This can be obtained (why?).
11 Properties of finite sets

11.1 The pigeonhole principle

The principle says: you want to put objects into holes. If you have more objects than you have holes, at least one hole must have multiple objects in it.

**Theorem 11.1.** Suppose that \( f : X \to Y \) is a function between non-empty finite sets such that \( |X| > |Y| \). Then \( f \) is not an injection, i.e. there exists distinct elements \( x_1, x_2 \in X \) such that \( f(x_1) = f(x_2) \).

The proof can be constructed using Lemma 10.3.

**Example 11.1.1.** In a party there are more than 365 people. Then there must be two people with birthdays in the same month and same date (probably different years).

**Theorem 11.2.** Suppose that \( f : X \to Y \) is a function between non-empty finite sets such that \( |X| < |Y| \). Then \( f \) is not a surjection.

**Theorem 11.3.** Suppose that \( X, Y \) are non-empty finite sets of the same cardinality. Then a function \( X \to Y \) is an injection if and only if it is a surjection.

This theorem is not true if \( X, Y \) are not finite sets. For example \( X = Y = \mathbb{Z}^+ \), and \( f : X \to Y \) is defined by

\[
f(n) = n + 1.
\]

Then \( f \) is injection but it is not a surjection.

So it is not hard to believe that finite sets have many nice properties that infinite sets do not have.

The following shows that finite non-empty set of real numbers has a greatest and a smallest element.

**Definition 11.4.** Let \( A \) be a set of real numbers. Then \( b \) is a minimum element of \( A \) or \( b = \min A \) when (i) \( b \in A \), (ii) \( a \in A \implies b \leq a \).

Similarly \( c \) is a maximum element of \( A \) or \( c = \max A \) when (i) \( c \in A \), (ii) \( a \in A \implies c \geq a \).

**Proposition 11.5.** If \( A \) is a finite non-empty set of real numbers, \( A \) has a minimum element and a maximum element.

The textbook provides us a proof by induction on \( n \). Think about, how you would prove the proposition.

Now we discuss the common divisor.

Recall that we say a non-zero integer \( d \) divides \( a \) when there is an integer \( q \) such that \( a = dq \). In this case, we say that \( d \) is a divisor or a factor of \( a \). Alternatively we say that \( a \) is a multiple of \( d \).

**Proposition 11.6.** Every non-zero integer divides 0.
Let us write the set of divisors of $a$ as,

$$D(a) = \{ n \in \mathbb{Z} \mid n \text{ divides } a \}.$$  

For example,

$$D(-12) = \{ \pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12 \}.$$

For any $a \in \mathbb{Z}$, $D(a)$ is a finite set and

$$\min D(a) = -|a|, \quad \max D(a) = |a|.$$

Now consider two integers $a, b$, not both zero, then the set of common divisors of $a, b$ is given by $D(a) \cap D(b)$. Notice that

$$\pm 1 \in D(a) \cap D(b).$$

We define the largest number of $D(a) \cap D(b)$ as the greatest common divisor or highest common factor of $a$ and $b$. And we denote it as $(a, b)$ or gcd$(a, b)$.

**Example 11.1.2.** Find gcd$(−12, 28)$.

**Solution.** Since

$$D(-12) = \{ \pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12 \}$$

and

$$D(28) = \{ \pm 1, \pm 2, \pm 4, \pm 7, \pm 14, \pm 28 \},$$

then

$$\gcd(-12, 28) = \max\{D(-12) \cap D(28)\} = \max\{\pm 1, \pm 2, \pm 4\} = 4.$$

**Definition 11.7.** Two integers $a$ and $b$, not both zero, are called coprime or relatively prime when $(a, b) = 1$.

**Definition 11.8.** A positive number $n$ is said to be prime when $n > 1$ and the only positive divisor of $n$ are 1 and $n$.

## 12 Counting functions and subsets

### 12.1 Counting functions

Let us count the number of functions from $X$ to $Y$ where $X, Y$ are finite sets.

**Example 12.1.1.** Let $X = \{ a, b, c \}$ and $Y = \{ d, e \}$. In order to define a function $f : X \to Y$, we must list the elements $f(a), f(b), f(c)$. Since each of them has two choices: either $d$ or $e$, the there are $2 \times 2 \times 2 = 8$ ways. So altogether we have 8 functions from $X$ to $Y$.

In general, we have
**Proposition 12.1.** Suppose that \( X, Y \) are non-empty finite sets with \(|X| = m\) and \(|Y| = n\). Then the number of functions \( X \to Y \) is given by \( n^m \).

**Definition 12.2.** Given sets \( X, Y \), we denote the set of functions from \( X \) to \( Y \) by \( \text{Fun}(X,Y) \).

**Proposition 12.3.** Let \( X, Y \) as in Proposition 12.1. Then the number of injections \( X \to Y \) is given by \( n(n-1)\cdots(n-m+1) \). Let us denote the injections as \( \text{Inj}(X,Y) \).

**Proof.** The proof is by induction on \( m \).

Base case: For \( m = 1 \) suppose that \( X = \{x_1\} \) and \( Y \) is any finite set. Then any function \( X \to Y \) is a injection and a function is determined by the value of \( f(x_1) \). Thus \( |\text{Inj}(X,Y)| = |Y| = n \) as required.

Inductive step: Suppose that, for some \( k \geq 1 \), the result holds for \( m = k \) and all positive \( n \) and the goal is to deduce the result for \( m = k + 1 \). Let \( X \) be a set of \( k+1 \) elements:

\[ X = \{x_1, \ldots, x_{k+1}\}, \]

and \( Y \) be a set of \( n \) elements:

\[ Y = \{y_1, \ldots, y_n\}. \]

Then we can rewrite the set of injections as

\[ \text{Inj}(X,Y) = \bigcup_{i=1}^{n} \{f \in \text{Inj}(X,Y) \mid f(x_{k+1}) = y_i\}, \]

a disjoint union. For each set, we can view the elements (functions) as those from \( X - \{x_{k+1}\} \) to \( Y - \{y_i\} \). In other words, we have the bijection

\[ \{f \in \text{Inj}(X,Y) \mid f(x_{k+1}) = y_i\} \leftrightarrow \text{Inj}(X - \{x_{k+1}\}, Y - \{y_i\}). \]

The latter are injections from a set of \( k \) elements to a set of \( n-1 \) elements. For then, by inductive hypothesis

\[ |\{f \in \text{Inj}(X,Y) \mid f(x_{k+1}) = y_i\}| = |\text{Inj}(X - \{x_{k+1}\}, Y - \{y_i\})| \]

\[ = (n-1)(n-2)\cdots(n-k). \]

Hence, by the addition principle,

\[ |\text{Inj}(X,Y)| = \sum_{i=1}^{n} |\{f \in \text{Inj}(X,Y) \mid f(x_{k+1}) = y_i\}| \]

\[ = n \times (n-1)(n-2)\cdots(n-k) \]

as required and so completing the inductive step.

Conclusion: Hence, by induction on \( m \), the results holds for all positive \( n, m \). \( \square \)

**Definition 12.4.** Given a set \( X \), a bijection \( X \to X \) is called a *permutation* of the set \( X \).

And then from the above proposition, we know that for a set \( X \) of cardinality \( n \), the number of permutations of \( X \) is given by \( n! \).
12.2 Counting sets of subsets

Example 12.2.1. Consider $X = \{1, 2, 3\}$. Recall that $\mathcal{P}(X)$ denotes the power set of $X$. Then

$$|\mathcal{P}(X)| = |\{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}| = 6.$$ 

Proposition 12.5. Suppose that $X$ is a set of cardinality $n$. Then

$$|\mathcal{P}(X)| = 2^{|X|}.$$ 

Definition 12.6. For any set $X$ and $A \subseteq X$, the characteristic function $\chi_A : X \to \{0, 1\}$ is defined by

$$\chi_A(x) = \begin{cases} 
0 & \text{if } x \notin A, \\
1 & \text{if } x \in A.
\end{cases}$$

Lemma 12.7. The function $\mathcal{P}(X) \to \text{Fun}(X, \{0, 1\})$ given by $A \to \chi_A$ is a bijection.

Proof. The inverse function is given by

$$\chi \to \{x \in X \mid \chi(x) = 1\}.$$ 

Note that the above also equals $\chi \to \chi^{-1}(1)$. \hfill \square

Proof. (of Proposition 12.5) From the previous Lemma

$$|\mathcal{P}(X)| = |\text{Fun}(X, \{0, 1\})|.$$ 

It follows from Proposition 12.1 that $|\mathcal{P}(X)| = 2^{|X|}$. \hfill \square

Definition 12.8. Given a set $X$ and a non-negative integer $r$, an $r$-subset of $X$ is a subset $A \subseteq X$ of cardinality $r$. We denote the set of $r$-subset by $\mathcal{P}_r(X)$ i.e.

$$\mathcal{P}_r(X) = \{A \subseteq X \mid |A| = r\}.$$ 

We define the binomial coefficient or binomial number $\binom{n}{r}$ (read “$n$ choose $r$”) to be the cardinality of the set $\mathcal{P}_r(X)$ when $|X| = n$.

$\binom{n}{r}$ describes the numbers of ways that we can choose $r$ objects from $n$ projects without considering the order.

Example 12.2.2. Compute $\binom{4}{3}$.

Solution. Suppose $X = \{1, 2, 3, 4\}$, then

$$\mathcal{P}_3(X) = \{\{2, 3, 4\}, \{1, 3, 4\}, \{1, 2, 4\}, \{1, 2, 3\}\}.$$ 

Thus $\binom{4}{3} = 4$. \hfill \square
Proposition 12.9. For $n, r$ non-negative integers

(i) \( \binom{n}{r} = 0 \) if \( r > n \),

(ii) \( \binom{n}{0} = 1, \binom{n}{1} = n, \binom{n}{n} = 1 \),

(iii) \( \binom{n}{r} = \binom{n}{n-r} = 1 \) for \( 0 \leq r \leq n \).

Proof. Suppose \( X \) is a set with \( n \) elements.

(i) \( X \) has no subsets of cardinality greater than \( n \).

(ii) Since \( P_0(X) = \{\emptyset\} \), \( P_n(X) = \{X\} \), \( \binom{n}{0} = 1, \binom{n}{n} = 1 \). Considering subsets with 1 element, \( P_1(X) \) has \( n \) elements.

(iii) There is a bijection \( P_r(X) \rightarrow P_{n-r}(X) \). For any \( A \in P_r(X) \), \( A \) is a subset of cardinality \( r \). We map it to \( A^c \) which is then a subset of cardinality \( n - r \). Hence \( \binom{n}{r} = \binom{n}{n-r} \). \( \square \)

Proposition 12.10. (Pascal's rule) For positive integers \( n, r \) such that \( 1 \leq r \leq n \)

\[
\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.
\]

Proof. Here let us give a combinatorial argument.

Recall that \( \binom{n}{r} \) equals the number of subsets with \( r \) elements from a set \( X \) with \( n \) elements. Suppose one particular element is uniquely labeled \( x \) in \( X \).

Every subset of \( X \) either contains \( x \) or not. To construct a subset of \( r \) elements containing \( x \), we need to choose \( r - 1 \) elements from the remaining \( n - 1 \) elements in \( X \). Thus there are \( \binom{n-1}{r-1} \) such subsets. For subsets not containing \( x \), we choose \( r \) elements from the remaining \( n - 1 \) elements. The total number of subsets is the sum of the above two. We have

\[
\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.
\]

\( \square \)

Remark 12.11. The proposition has been first discovered in the eleventh century in China. It is referred to as Yanghui triangle:

\[
\begin{array}{cccccccc}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
\end{array}
\]

Theorem 12.12. For non-negative integers \( n, r \) such that \( r \leq n \),

\[
\binom{n}{r} = \frac{n(n-1)...(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.
\]

31
Proof. The textbook uses a proof by induction. Let us sketch a combinatorial argument.

Recall that \( \binom{n}{r} \) equals the number of subsets with \( r \) elements from a set \( X \) with \( n \) elements. If we randomly pick elements one by one from \( X \) for \( r \) times, we get a subset of cardinality \( r \). The first time we pick an element, we have \( n \) choices. The second time we pick an element, we have \( n - 1 \) choices... Thus to pick \( r \) times, altogether the number of choices we have is

\[
n(n - 1)(n - 2)...(n - r + 1).
\]

Now consider a \( r \)-subset of \( X: \{x_1, x_2, ..., x_r\} \). Each permutation of it provides a way of picking \( r \) elements one by one from \( X \). We have proved that the a set with cardinality \( r \), there are \( r! \) permutations. Therefore

\[
\binom{n}{r} \times r! = n(n - 1)(n - 2)...(n - r + 1)
\]

and we proved the theorem. \( \square \)

**Theorem 12.13.** (Binomial Theorem) For any real numbers \( a, b \) and a non-negative integers \( n \)

\[
(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.
\] (2)

Special Cases of binomial theorem were known since at least the 4th century BC mentioned by Greek mathematician Euclid for exponent 2. The earliest known reference to this problem is by the Indian lyricist Pingala in 200 BC.

**Example 12.2.3.**

\[
(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,
\]

\[
(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 3ab^3 + b^4...
\]

Proof. The textbook uses a proof by induction. Here we present a combinatorial argument.

Expanding \((a + b)^n\) yields the sum of the \( 2^n \) products of the form \( e_1 e_2 ... e_n \) where each \( e_i \) is a or b. Rearranging factors shows that each product equals \( a^{n-k} b^k \) for some \( 0 \leq k \leq n \). For each \( k \), the number of copies \( a^{n-k} b^k \) in the expansion is the same as the number of \( n \)-character \( a, b \) strings having \( b \) exactly \( k \) positions. This number is then the same as picking \( k \) positions from \( n \) positions which equals by definition \( \binom{n}{k} \). \( \square \)

**Corollary 12.14.** If picking \( a = b = 1 \) in (2), we get

\[
2^n = \sum_{k=0}^{n} \binom{n}{k}.
\]