Chapter 6 The mathematics of the natural numbers

Section 6.1 The system of the natural numbers

The most important concept of mathematics is that of *natural number*. These are the numbers used for *counting*, among others; we talk about a set having 0, or 1, or 2, 3, etc. elements. Before we discuss counting in detail, we need to give an overview of certain basic properties of the natural numbers.

The natural numbers are 0, 1, 2, 3, ...; all numbers obtained by repeatedly adding 1 to the previous number, starting with 0. This is a rather poor "definition"; and in fact, rather than defining, one has to postulate the existence of the (set of the) natural numbers. Here we do not attempt to build up the theory of the natural number system in an axiomatic way. Rather, we only summarize the main points, some of which should be very familiar.

The set of all the natural numbers is denoted by $\mathbb{N}$. The most fundamental operation on the natural numbers is the *successor operation*:

$$\begin{aligned} S : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto n+1 \ , \end{aligned}$$

that is, the operation of adding 1 to a number. This operation satisfies the following properties:

$$0 \neq S(n) \quad \text{for all } n \in \mathbb{N} ;$$

$$S(n) = S(m) \implies n = m \quad \text{for all } n, m \in \mathbb{N} .$$

The first of these says that 0 is not the successor of any natural number, the second says that $S:\mathbb{N}\longrightarrow\mathbb{N}$ is injective. Of course, every child knows these; the reason for pointing them out is that, together with the principle of mathematical induction (see below), they are enough for establishing all the needed properties of natural numbers (which is certainly a surprising fact).

One considers many other operations on natural numbers; foremost among them are those of addition and multiplication:

$$\begin{aligned} + : \mathbb{N} \times \mathbb{N} &\longrightarrow \mathbb{N} \\ (k, n) &\longmapsto k + n \end{aligned}$$

$$\begin{aligned} \cdot : \mathbb{N} \times \mathbb{N} &\longrightarrow \mathbb{N} \\ (k, n) &\longmapsto k \cdot n \ . \end{aligned}$$

(In the notation of multiplication, the dot may be omitted if no confusion arises as a result: kn may be written for $k \cdot n$. On the other hand, especially with numerical factors, sometimes we use × in place of $\cdot$.)

We may define these by the method of *recursion*, or *recurrence*, a very important concept not only in mathematics, but also in computer science. The recursive definition of addition, via the successor function, is as follows:

$$k + 0 = k , \qquad \text{(ADD 1)}$$

$$k + S(n) = S(k + n) \ . \qquad \text{(ADD 2)}$$

Certainly, these equations are *true* as we know addition (the second says that $k + (n + 1) = (k + n) + 1$), but in what sense do they give a definition of addition? The answer is that, although these equations do not explicitly specify what $k + n$ is in general, by repeated application of them, we can calculate any value of $k + n$. To see this, first recall that any natural number is obtained, in fact in a unique way, by applying the S operation to 0 . If $m = SS...S0$ (we omitted parentheses in the function-value notation), then, according to the second equation, (ADD 2),

$$k + m = k + SS...S0 = S(k + S...S0) \ ,$$

and on the right-hand-side, inside the parentheses, we now have an instance of addition with *one less* S *operators in the second argument*. That is, we have *reduced* the calculation of

$k + m$ to another calculation, namely to that of $k + n$ where n is the *predecessor* of m, $m = Sn$. We may continue reducing in this way until we hit 0 in the second argument, in which case we apply the first recursion equation (ADD 1).

For instance,

$$\begin{aligned} 4 + 3 = 4 + SSS0 &= S(4 + SS0) \quad (\text{ADD 2 with } n = SS0) \\ &= SS(4 + S0) \\ &= SSS(4 + 0) \\ &= SSS4 \qquad (\text{ADD 1}) \\ &\quad = SS5 = S6 = 7\;. \end{aligned}$$

As another example for recursion, here is a recursive definition for multiplication among the natural numbers:

$$k \cdot 0 \;=\; 0 \qquad\qquad (\text{MULT 1})$$

$$k \cdot Sn = \; k \cdot n + k \qquad\qquad (\text{MULT 2})$$

This recursion uses the addition as already given, but otherwise it operates in the same way as the previous definition: the second equation *reduces* the calculation of an instance of multiplication, $k \cdot Sn$, to one, namely $k \cdot n$, in which the second argument is *one less* than in the first instance. This circumstance, together with the first equation, allows us to calculate any instance of multiplication of natural numbers in a series of steps each of which is an application of the recursion equations (MULT 1), (MULT 2); when doing so, we have to be able to calculate addition for any pair of arguments. E.g.,

$$\begin{aligned} 4\times 3 = 4\times SSS0 = 4\times SS0 + 4 = (4\times S0 + 4) + 4 &= \\ = ((4\times 0 + 4) + 4) + 4 & \\ = ((0 + 4) + 4) + 4 & \\ = 12\;. & \end{aligned}$$

In the case of addition and multiplication, their definition via recursion has a theoretical significance only; however, recursion is used to specif) many functions on the natural numbers which would be harder to define without recursion. An important example is the factorial function:

$$! : \mathbb{N} \longrightarrow \mathbb{N} ::$$

$$0! = 1 \tag{!1}$$

$$(Sn)! = n! \cdot Sn \tag{!2}$$

E.g., $4! = 3! \cdot 4 = 2! \cdot 3 \cdot 4 = 1! \cdot 2 \cdot 3 \cdot 4 = 0! \cdot 1 \cdot 2 \cdot 3 \cdot 4 = 1 \cdot 1 \cdot 2 \cdot 3 \cdot 4 = 24$.

A related but more general recursion defines *sums* and *products*. Suppose x_i is a quantity, one for each natural number i (in other words, $i \longmapsto x_i$ is a given function defined on $\mathbb{N}$). Then, for any $n \in \mathbb{N}$, we may define the quantities

$$\sum_{i<n} x_i ,$$

$$\prod_{i<n} x_i$$

by recursion on the variable n :

$$\sum_{i<0} x_i = 0 \tag{SUM 1}$$

(the sum of zero-many terms is 0)

$$\sum_{i<Sn} x_i = \sum_{i<n} x_i + x_n \tag{SUM 2}$$

$$\prod_{i<0} x_i = 1 \tag{PROD 1}$$

(the product of zero-many terms is 1)

$$\prod_{i<Sn} x_i = (\prod_{i<n} x_i) \cdot x_n . \tag{PROD 2}$$

$\sum_{i<n} x_i$ is the sum of all terms x_i where the subscript runs over the given range, in this case the set of natural numbers less than n. We may write

$$\sum_{i<n} x_i = x_0 + x_1 + \ldots + x_{n-1} ;$$

the recursive definition of the expression makes the three dots precise. Similarly,

$$\prod_{i<n} x_i = x_0 \cdot x_1 \cdot \ldots \cdot x_{n-1} .$$

We may write sum and product expressions with ranges of subscripts different from " $i < n$ ". E.g.,

$$\sum_{i=2}^{n+1} x_i \text{ is the sum } \quad x_2 + x_3 + \ldots + x_n + x_{n+1} ;$$

of course, it is really just another case of the original kind of expression:

$$\sum_{i=2}^{n+1} x_i = \sum_{j<n} x_{j+2}$$

(here, $j = i-2$, i.e., $i = j+2$; while i ranges from 2 to $n+1$, j ranges from 0 to $n-1$).

The factorial function is a special case of the product-expression;

$$n! = \prod_{i=1}^{n} i = 1 \cdot 2 \cdot \ldots \cdot n .$$

The famous *Fibonacci numbers* are defined by recursion:

$$f_0 = 0 ,$$

$$f_1 = 1 \,,$$

$$f_{n+2} = f_n + f_{n+1} \qquad\qquad (\, n \in \mathbb{N} \,)$$

In other words, a sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ is defined, by saying that the first two terms of the sequence (the ones indexed by 0 and 1) be equal to 1 , and any other term equal the sum of the two previous terms. The first few Fibonacci numbers are:

$$f_0 = 0 \,,\;\; f_1 = 1 \,,\;\; f_2 = f_0 + f_1 = 1 \,,\;\; f_3 = f_1 + f_2 = 2 \,,$$

$$f_4 = f_3 + f_2 = 3 \,,\;\; f_5 = f_3 + f_4 = 5 \,,\;\; \ldots \,.$$

The fundamental method of proving facts about the natural numbers is *mathematical induction*. In the axiomatic introduction of the natural number system, the principle of mathematical induction is taken as a basic axiom. The principle says that in order to prove that all natural numbers have a certain property , it suffices to convince oneself of two things: **one**, that 0 has the property, and **two**, that the property is inherited from any natural number to the next. One sees that this is correct, by the following intuitive argument. 0 has the property, as assumed. But then, since it is inherited from 0 to 1 , 1 has it . Since it is inherited from 1 to 2 , 2 has it. Etc. Since by starting with 0 , applying the successor operation repeatedly, every natural number will be eventually reached, we obtain that the every natural number has the property.

Let us analyze the principle of mathematical induction in terms of sets. With a property P of natural numbers, let us take the set of all those natural numbers that have property P , and let us call this set X . Thus, " $n \in X$ " is now synonymous with " n has property P ". To say that 0 has property P is the same as to say that

$$0 \in X \,. \tag{1}$$

To say that the property is inherited from any natural to its successor is expressed more mathematically in this way:

for all $n \in \mathbb{N}$, if $n \in X$, then $Sn \in X$, or, with logical abbreviations,

$$\forall n \in \mathbb{N} \ (n \in X \Longrightarrow Sn \in X) \ . \qquad (2)$$

(Read " $\forall n \in \mathbb{N}$ " as "for all n in $\mathbb{N}$ ".) The principle says that from the two assumptions (1), (2) it follows that every natural number has property P ; this latter is simply the statement that every natural number is in X , or even more simply, since X is already a subset of $\mathbb{N}$, that

$$X = \mathbb{N} \ . \qquad (3)$$

Thus, finally, the principle of mathematical induction is as follows.

Principle of Mathematical Induction (PMI) Let X be any subset of $\mathbb{N}$. Assume that

$$0 \in X \qquad (1)$$

and that

$$\forall n \in \mathbb{N} \ (n \in X \Longrightarrow Sn \in X) \ . \qquad (2)$$

Then

$$X = \mathbb{N} \ . \qquad (3)$$

Here is a formulation, directly in terms of properties rather than sets. Let us write $P(n)$ for: " n has property P ". Then the PMI may be stated as follows:

PMI (second form). Let P be any property of natural numbers. Assume

$$P(0) \qquad (1')$$

and

$$\forall n \in \mathbb{N} \ (P(n) \implies P(Sn)) \ . \qquad (2')$$

Then

$$\forall n \in \mathbb{N} \ P(n) \ . \qquad (3')$$

Induction may be used, on a very basic level, to establish the fundamental laws of arithmetic for addition, multiplication and exponentiation on the natural numbers (the same laws concerning more comprehensive number systems are established as later steps in the process of building up mathematics axiomatically). This happens very naturally, because those operations are defined by recursion, and recursion and induction go hand in hand.

To see the Principle of Mathematical Induction at work, that is, for an example for a proof by induction, let us consider a simple example concerning sums. The sum of the first n odd numbers may be written as $\sum_{i=1}^{n} (2i-1)$. Experimenting with the first few values, we find

$$\sum_{i=1}^{1} (2i-1) = 1 \ ,$$

$$\sum_{i=1}^{2} (2i-1) = 1 + 3 = 4 \ ,$$

$$\sum_{i=1}^{3} (2i-1) = 1 + 3 + 5 = 9 \ ,$$

$$\sum_{i=1}^{4} (2i-1) = 1 + 3 + 5 + 7 = 16 \ ,$$

from which we may conjecture that

$$\sum_{i=1}^{n} (2i-1) = n^2 \qquad (4)$$

for all values of $n \in \mathbb{N}$. We propose to show that (4) holds for all natural numbers n , by *induction on* n. This phrase means that we consider the property P of an arbitrary natural number n that (4) holds true, or equivalently, we consider the set X of all those natural numbers n such that (4) is true, and we apply the PMI to this property/set. First, we have to show that $0 \in X$ (see (1)); in other words, that (4) holds for $n = 0$; this is called the ***basis step*** of the induction.

Basis step: $n = 0$ in (4).

The sum $\sum_{i=1}^{0} (2i-1)$ is empty, hence, by definition, it is 0. Also, $0^2 = 0$. The basis step is complete.

Secondly, we have to show that the property in question is inherited from *any* n to $Sn = n+1$. In other words, we want to show the statement under (2). To this end, let n be an *arbitrary* natural number, and *assume* $n \in X$; using this assumption, we will show that $Sn = n+1 \in X$. This is called the ***induction step***.

Induction step: For all $n \in \mathbb{N}$, (4) for n implies (4) for $n+1$.

We fix an arbitrary natural number n, and we assume (4) for n, that is,

$$!: \quad \sum_{i=1}^{n} (2i-1) = n^2 . \qquad (5)$$

We call (5) the ***induction hypothesis***; in general, when establishing (2), or (2'), we *assume* $n \in X$, respectively $P(n)$, and we call this assumption the induction hypothesis.

We will show that (4) holds for $n+1$ in place of n, that is

$$?: \quad \sum_{i=1}^{n+1} (2i-1) = (n+1)^2 \qquad (6)$$

But, according to the recursive definition of summation,

$$\sum_{i=1}^{n+1} (2i-1) = \sum_{i=1}^{n} (2i-1) + \underset{\underset{2(n+1)-1}{\uparrow}}{(2n+1)} \quad .$$

Using the induction hypothesis (5), we find that the latter equals

$$= n^2 + (2n+1)$$

and this is indeed the same as $(n+1)^2$, as needed for (6). The induction step is complete.

According to the PMI, we may now conclude that (4) holds generally, for all $n \in \mathbb{N}$.

There are other forms of the PMI. For one thing, we may start with any fixed number, rather than 0 , and seek to prove that a property holds from that number on. E.g., if we had not wanted to consider the empty sum in the previously proved identity, we might have asserted it for integers $n = 1$ and greater; in this case, the Basis Step would have been the case $n = 1$, and in the Induction Step, we would have argued for an arbitrary positive natural number n , rather than an arbitrary natural number.

More importantly, we have the version of the PMI in which we infer the truth of the induction statement at n from the truth of the statement at *all* numbers less than n , rather than at the immediately preceding value.

Wellordering Principle (WOP). Let X be a subset of $\mathbb{N}$. Assume that for any $n \in \mathbb{N}$,

$$\begin{aligned} &\text{if } k \in X \text{ holds for all } k \in \mathbb{N} \text{ such that } k < n , \\ &\qquad \text{then } n \in X \end{aligned} \tag{7}$$

(in logical abbreviation:

$$((\forall k \in \mathbb{N})\,(k < n \Longrightarrow k \in X) \Longrightarrow n \in X \qquad) .$$

Then

$X = \mathbb{N}$.

WOP, second form. Let P be any property of natural numbers. Assume that for any $n \in \mathbb{N}$,

if $P(k)$ holds for all $k < n$, then $P(n)$

(in logical abbreviation:

$$((\forall k \in \mathbb{N})(k < n \Longrightarrow P(k)) \Longrightarrow P(n) \qquad).$$

Then

for all $n \in \mathbb{N}$, $P(n)$ holds.

In comparison with the PMI, we find only one assumption (7) in place of the two in the PMI. First of all, let us note that $0 \in X$, the "basis step" occurring as an assumption in the PMI, follows from the present "global induction step" (7). This is because the proposition

" $k \in X$ holds for all $k \in \mathbb{N}$ such that $k < 0$ " (8)

is always true, no matter what the set X is, since there is no natural number less than 0 . (The proposition asserts that something holds for all members of the empty set; and any such proposition is automatically true; we also say it is *vacuously true*.) Since the global induction step (7) says that from (8), $0 \in X$ follows, we have that $0 \in X$.

To give an application, let us prove that every natural number can be written as the sum of distinct integral powers of 2 :

for any $n \in \mathbb{N}$, *there are natural numbers* m *and* $i_0 < i_1 < \ldots < i_{m-1}$ *such that*

$$n = \sum_{j<m} 2^{i_j} = 2^{i_0} + 2^{i_1} + \ldots + 2^{i_{m-1}} .$$

(Note that when $n=0$, then also $m=0$, and the sum is an empty one, with value equal to 0 as it should be.)

Now, the property $P(n)$, the one that we want to prove to hold for all $n\in\mathbb{N}$, is this:

$P(n)$: *there are natural numbers* m *and* $i_0 < i_1 < \dots < i_{m-1}$ *such that*

$$n = \sum_{j<m} 2^{i_j} = 2^{i_0} + 2^{i_1} + \dots + 2^{i_{m-1}} .$$

Global Induction Step. Let $n \in \mathbb{N}$ be arbitrary. Assume

$$\forall k\in\mathbb{N}\ (k < n \Longrightarrow P(k)) , \tag{9}$$

to show

$$P(n) . \tag{10}$$

(9) is called the ***induction hypothesis.***

In the proof of the implication (9) $\Longrightarrow$ (10), we make a *case distinction*; we distinguish the cases $n=0$ and $n>0$; this has nothing to do with the "basis step" and the "induction step" of the earlier form of induction; more remarks on this later.

Case 1. $n = 0$. As we noted before, in this case, we can take $m = 0$; 0 is the empty sum of powers of 2.

Case 2. $n > 0$. Let 2^i be the largest (integral) power of 2 for which $2^i \le n$ (we will justify this intuitively obvious step below by what we'll call the **Greatest Number Principle**; for the time being, it should be enough to note that, certainly, there is at least one integral power 2^h of 2 for which $2^h \le n$, namely the one with the exponent $h = 0$; $2^0 = 1 \le n$; it is here that we use the case assumption $n > 0$). Let $k = n - 2^h$. Since $2^h \ge 1$, we have that $k < n$. Now, let us apply the induction hypothesis (9) to this k;

$P(k)$ holds, that is, $k = \sum_{j<\ell} 2^{i_j}$ with strictly increasing $i_j \in \mathbb{N}$. Then,

$$n = k + 2^h = \sum_{j<\ell} 2^{i_j} + 2^h ;$$

the only thing left to show is that h is strictly greater than all the indices i_j, $i<\ell$. But, otherwise, we must have that $\ell \geq 1$, $h \leq i_{\ell-1}$, and

$$n \geq 2^{i_{\ell-1}} + 2^h \geq 2^h + 2^h = 2 \cdot 2^h = 2^{h+1} ,$$

and this is in contradiction with the choice of 2^h as the largest power of 2 still $\leq n$; this shows what we wanted.

The Global Induction Step is completed; by the Wellordering Principle, $P(n)$ holds for all n, which is what we wanted.

Note that within the proof of the Global Induction Step, there was a case distinction that resembled the distinction between the "Basis Step" and "Induction Step" in the PMI. However, this is merely coincidental. The important point is that the proof uses the Wellordering Principle in an essential way. Note that in the "induction step", we inferred the truth of $P(n)$ *not* from the truth of $P(n-1)$ as in a proof by the ordinary PMI, but from the truth of $P(k)$ for *some* $k<n$, namely $k=n-2^h$, where we only know that $k<n$, and k may well be different from $n-1$.

Uses of the WOP are also called proofs by induction; the distinction in the names serves the clarity of the discussion here, and later the principles mentioned in this section will all be referred to as "induction".

The WOP can be *proved* on the basis of the PMI. The proof consists in showing that, under the assumption of the WOP, the property

$$Q(n) \equiv \text{for all } k \in \mathbb{N} \text{, if } k < n \text{, } P(k) \text{ holds}$$

satisfies the assumptions of the PMI; we will not give the details here (the reader may try to complete the proof; it is not hard!). By the PMI, then $Q(n)$ holds for all n, which, applied to $n+1$ in place of n, one gets that $P(n)$ holds for all n.

There is an equivalent version of the WOP, the

Least Number Principle (LNP). Let X be any subset of $\mathbb{N}$. If X is non-empty, then there is a least element of X: there is $n \in X$ such that $x \leq y$ for all $y \in X$.

x n

The connection between the WOP and the LNP can be summarized by saying that applying the LNP to X is the same as applying the WOP to its complement, the set $\mathbb{N} - X$.

[Let us prove the LNP for X by applying the WOP to $\mathbb{N} - X$. Assume X is non-empty, but, contrary to the assertion, there is no least element in it; we will derive a contradiction. We claim that the Global Induction Step (7) holds for $\mathbb{N} - X$ in place of X. Assume that for all $k < n$, we have $k \in \mathbb{N} - X$. Then $n \in \mathbb{N} - X$: otherwise $n \in X$ and since for all $k < n$, we have $k \in \mathbb{N} - X$, this is exactly to say that x is the least element of X, which is not supposed to exist. This shows (7) for the set $\mathbb{N} - X$. By the WOP, $\mathbb{N} - X = \mathbb{N}$, which is to say that $X = \emptyset$ (since $X \subset \mathbb{N}$). We have arrived at the desired contradiction.]

Perhaps it is not superfluous to mention that the LNP is a special property of the system of the natural numbers. When you replace $\mathbb{N}$ by $\mathbb{Z}$, the principle becomes incorrect: do you see that?

As a first application of the LNP, let us state and prove the

Greatest Number Principle (GNP). Let N be any natural number, and let X be a subset of $\{i \in \mathbb{N} : i < N\}$ (in other words, $X \subset \mathbb{N}$ and it is *strictly bounded* by N). If X is non-empty, then there is a greatest element of X: there is $n \in X$ such that $x \leq n$ for all $x \in X$.

Note that, unlike in the LNP, now the side-condition of X being bounded is essential; do you see that?

Here is the proof of the GNP. Let X be as in the statement. Consider the set of all *strict upper* *bounds* of X, and call it Y:

$$Y = \{y\in\mathbb{N} : x<y \text{ for all } x\in X\} .$$

Y is non-empty: $N\in Y$ (why?). Therefore, by the LNP, Y has a least element; call it m. m cannot be 0 since m is a strict upper bound of X, and X is non-empty. Now, consider $m-1\in\mathbb{N}$. I claim that $m-1\in X$, and $n=m-1$ is the maximal element of X. We have that $x<m$ for all $x\in X$; that is,

$x\le m-1$ for all $x\in X$;

and in yet different words,

either $x<m-1$ (Case 1) or $x=m-1$ (Case 2) for all $x\in X$.

It cannot be that we have Case 1 for all $x\in X$, since then $m-1$ would be also a strict upper bound of X, which is impossible since $m-1<m$, and m was the *least* strict upper bound of X. Therefore, there must be an $x\in X$ for which Case 2 holds, that is, $m-1=x$; but this means that $m-1\in X$. But we already know that $x\le m-1$ for all $x\in X$; thus, x is the largest element of X.

Note that we are in fact using the GNP in our proof above of the "binary decomposition of numbers"; in Case 2 there, we said: "Let 2^h be the largest integral power of 2 for which $2^h\le n$." Can you justify this step formally by the GNP?

Let us give another application of the LNP.

We will prove in the next section that, for every $n\in\mathbb{N}$, there are prime numbers greater than n. Using this, we may define the sequence $\langle p_n\rangle_{n\in\mathbb{N}}$ by recursion as follows:

$$p_0 \underset{\text{def}}{=} 2 ,$$

$$p_n \underset{\text{def}}{=} \text{the least prime number greater than } p_{n-1} .$$

The point is that the set

$$\{p\in\mathbb{N} \mid p \text{ is prime and } p>p_{n-1}\}$$

is non-empty, by what we just said; hence, by the LNP, p_n is well-defined. The sequence

$$p_0 = 2 , \quad p_1 = 3 , \quad p_2 = 5 , \quad \dots$$

is the list of all primes in increasing order; p_n is the $n+1^{st}$ prime number.

Finally in this section, let us mention still another equivalent form of the "wellorderedness" of the natural numbers, the method of "infinite descent". This is the form the principle of mathematical induction takes explicitly in the work of Pierre de Fermat (1601-1665), one of the greatest of all mathematicians. Actually, the principle should be called "*the impossibility of infinite descent*". What it says is that if we have a sequence

$$n_1 > n_2 > n_3 > n_4 > \dots$$

of strictly decreasing natural numbers, then the sequence cannot be infinite: there must be a stage k such that n_{k+1} is not defined any more. The truth of this fact is seen by applying the Least Number Principle: consider the *set*

$$\{n_1 , n_2 , n_3 , n_4 , \dots\}$$

of all the numbers in the sequence; this non-empty set has to have a least element; but if that is n_k, then n_{k+1} cannot be defined, since if it were, it would be smaller than n_k, and thus the latter would not be the least element of the said set.

In the next section, we will see an application of the "impossibility of infinite descent".

Section 6.2 Divisibility among the integers

An integer $a \varepsilon \mathbb{Z}$ is *divisible* by $b \varepsilon \mathbb{Z}$ if there is an integer $c \varepsilon \mathbb{Z}$ such that $a = bc$. Note that 0 is divisible by any integer b, since $0 = b \cdot 0$. On the other hand, a is divisible by 0 only if $a = 0$: from $a = 0 \cdot c$ it follows that $a = 0$. The symbolic way of writing " a is divisible by b " is: $b|a$. Instead of " a is divisible by b " we also say that " b divides a ", or that " b is a divisor of a " , or " a is a multiple of b " .

Note the obvious

Transitivity law for divisibility:

$$a|b \text{ and } b|c \implies a|c.$$

In case $b \neq 0$, a being divisible by b is the same as to say that $\frac{a}{b}$ is an integer; we cannot say this, however, if $b = 0$, since $\frac{a}{0}$ is meaningless. In particular, if $b|a$, then either $a=0$, or else $|b| \leq |a|$; in other words, for positive integers a and b such that $a<b$, $b|a$ is impossible (since then $0<\frac{a}{b}<1$, and $\frac{a}{b}$ cannot be an integer).

As far as divisibility is concerned, any integer a and its negative $-a$ behave in the same way: $b|a$ iff $b|-a$ iff $-b|a$. Therefore, e.g., when we want to account for all the divisors of an integer, we may restrict our search to the non-negative numbers. Always, $a|a$ and $a|-a$. Moreover, if both $a|a'$ and $a'|a$ hold, then either $a' = a$ or $a' = -a$.

In what follows, variables a, b, ... range over $\mathbb{Z}$, the set of all integers, unless otherwise stated.

Given any a and b such that $b > 0$, we may divide a by b with a remainder: we can find q and r such that

$$a = qb + r, \quad 0 \leq r < b. \tag{1}$$

E.g., with $a = 17$, $b = 5$, we have $q = 3$ and $r = 2$:

$$17 = 3 \cdot 5 + 2, \quad 0 \le 2 < 5 .$$

In (1), q is the *quotient*, r is *remainder* when a is divided by b. The remainder being equal to 0 signifies, of course, that b divides a, $b|a$.

To prove the existence of the quotient/remainder representation, first let us assume that $a \ge 0$. The set X of all non-negative multiples of b that are less than or equal to a is nonempty ($0 \in X$), and bounded by a; thus, by the Greatest Number Principle (see the last section), it has a maximal element, say qb. Thus we have that $qb \in X$ but $(q+1)b \notin X$ (since $b \ne 0$, $qb < (q+1)b$). This means that $qb \le a < (q+1)b$. It follows that for $r = a - qb$, we have the relations in (1).

For the case when $a < 0$, we write $-a = qb + r$ by what we already know; from this, $a = (-q-1)b + (b-r)$ is the desired decomposition.

The *common divisors* of a and b are those integers that divide both a and b.

With any integers a and b, a(n *integer*) *linear combination* of a and b is any integer of the form $xa + yb$, with x and y also integers (although we usually say "linear combination" without the qualification "integer", we insist that the coefficients should also be integers!). Note that

any linear combination of linear combinations of a *and* b *is a linear combination of* a *and* b:

if $c = xa + yb$, $d = ua + vb$ and $e = sc + td$, then

$$e = s(xa + yb) + t(ua + vb) = (sx + tu)a + (sy + tv)b .$$

Also note that

any common divisor of a *and* b *is a divisor of any linear combination of* a *and* b:

if $c|a$, $c|b$, that is $a = uc$, $b = vc$, then

$$xa + yb = xuc + yvc = (xu + yv)c .$$

Now, if

$$a = qb + r , \tag{2}$$

then a is a linear combination of b and r (since $a = qb + 1 \cdot r$) and also, since $r = a - qb = 1 \cdot a + (-q)b$, r is a linear combination of a and b. We may conclude that

under (2), *the common divisors of* a and b, *and the common divisors of* b and r *are the same.*

c is a *greatest common divisor* (gcd) of a and b if it is a common divisor of a and b, and a multiple of every common divisor of a and b at the same time; in other words,

$$c|a \text{ and } c|b$$

and

$$\text{for all } d \text{ such that } d|a \text{ and } d|b, \text{ we have } d|c .$$

Another way of putting the defining property of c is to say the common divisors of a and b are the same as the divisors of (the single) c: for any d,

$$d|a \text{ and } d|b \iff d|c .$$

Note that it is not clear, at this point, that any pair of numbers a and b *has* a gcd; we will prove this soon. However, one thing is pretty clear, namely that the gcd, if it exists, is essentially unique: if both c and c' are gcd's of a and b, then $c = c'$ or $c = -c'$; the reason is that, from the definition it follows that both $c \mid c'$ and $c' \mid c$ hold. To make the gcd completely unique, we agree that $\mathrm{gcd}(a, b)$ should denote the non-negative one of the two possible values.

A remark on the name "greatest common divisor". Assume that both a and b are positive (the only "interesting" case for $\mathrm{gcd}(a, b)$). Then $c{=}\mathrm{gcd}(a, b)$ is certainly the *greatest one* among the common divisors of a and b, since it is positive, and it divides all of them. One might then say that it is *obvious* that there is a *greatest one* among these common divisors, as there is always a *greatest one* among finitely many integers. However, if we denote this greatest of the common divisors by c, it is not clear that for every common divisor d of a and b we have $d \mid c$ as required in the definition of "gcd"; we only have that $d{\le}c$, which, of course, is not enough for $d \mid c$. It is important to realize that the definition of "greatest common divisor" imposes a *stronger* condition than it appears from the wording of the concept.

These remarks explain why, to prove the existence of the gcd, we have to go through the considerably more sophisticated argument than just saying "take the largest of the common divisors". The argument that follows is not only one of the most important ones in all of mathematics, but it is also one of earliest ones: it appears in Euclid's "Elements", the classic ancient Greek treatise on mathematics.

Note that if $b \mid a$, then $\mathrm{gcd}(a, b) = |b|$; hence, $\mathrm{gcd}(0, b) = |b|$.

For the proof of the existence of the gcd, the first remark is that

$$\textit{if } a = qb + r \textit{, then } \mathrm{gcd}(a, b) = \mathrm{gcd}(b, r) \,, \qquad (3)$$

meaning that if one gcd exists, so does the other, and they are equal. The reason is that, in this case, the common divisors of the pair (a, b) and those of (b, r) are the same, as we noted above.

Let a and b be arbitrarily given integers; we want to compute $\mathrm{gcd}(a, b)$. We may assume that $b > 0$; if $b = 0$, then $\mathrm{gcd}(a, 0) = |a|$ as said above, and if $b < 0$, we may pass to $-b$: $\mathrm{gcd}(a, b) = \mathrm{gcd}(a, -b)$. Now, assuming $b>0$, we can define, by recursion, the sequence

$$a_0 , \ a_1 , \ a_2 , \quad , \ a_n , \ a_{n+1} \tag{3'}$$

by

$$a_0 \underset{\text{def}}{=} a$$

$$a_1 \underset{\text{def}}{=} b$$

and for any $i \geq 0$, if we have already defined a_i and a_{i+1} ,

$$\textit{and if } a_{i+1} \textit{ is greater than } 0 , \tag{3''}$$

a_{i+2} is defined as the remainder of a_i divided by a_{i+1} . In other words, the relations

$$a_i = q_i \cdot a_{i+1} + a_{i+2} , \quad 0 \leq a_{i+2} < a_{i+1} \tag{4}$$

hold with suitable q_i . When $a_{i+1} = 0$, we stop, that is, we do not define a_{i+2} , and we put $n = i$; thus, the sequence (3') will have been defined. Since the a_i's are strictly decreasing after $i = 1$ (see the second relation in (4)), by the "principle of the impossibility of infinite descent" (see the last section), we must reach a stage $i+1$ when a_{i+2} is no longer defined, that is, the condition (3") fails, that is, $a_{i+1} = 0$. Denote this i by n . Therefore, since $a_{n+1}=0$, we have by (4), for $i=n-1$, that

$$a_{n-1} = q_{n-1} \cdot a_n \ . \tag{5}$$

Now, since a_n is a divisor of a_{n-1} , $\mathrm{gcd}(a_{n-1}, a_n) = a_n$. The first relation in (4) tells us that

$$\mathrm{gcd}(a_i, a_{i+1}) = \mathrm{gcd}(a_{i+1}, a_{i+2}) \qquad (i+2 \leq n)$$

(see (3)). Thus, we have that

$$\gcd(a, b) = \gcd(a_0, a_1) = \gcd(a_1, a_2) = \ldots = \gcd(a_{n-1}, a_n) = a_n .$$

We have shown that $\gcd(a, b)$ exists, and in fact, have shown how to compute it. We can summarize the procedure this way: we construct the sequence the first two terms of which are the given numbers, and in which every term is the remainder when the previous term divides the term preceding it. The construction terminates when 0 is reached; the term previous to the zero term is the desired gcd.

E.g., let $a = 3293$, $b = 4107$. Then

$$3293 = 0\times4107 + 3293 ,$$

$$4107 = 1\times3293 + 814 ,$$

$$3293 = 4\times814 + 37 ,$$

$$814 = 22\times37 .$$

That is, in this case, $n = 3$, $a_2 = 814$ and $a_3 = 37$, and $\gcd(3293, 4107) = 37$.

The procedure described is called the *Euclidean algorithm*. It was known to the ancient Greeks; it appears in Euclid's "Elements". An important fact about it is that it is an *efficient* algorithm; for relatively large numbers, it terminates quite fast.

Besides a way of computing the gcd, the Euclidean algorithm also gives us an important theoretical conclusion:

the gcd *of any two numbers* a *and* b *is a linear combination of* a *and* b .

To see this, we prove by induction on $i \leq n$ that a_i is a linear combination of a and b . For $i = 0$ and $i = 1$, this is certainly true: $a = 1\cdot a + 0\cdot b$ and $b = 0\cdot a + 1\cdot b$.

Assuming the result for all indices less than $i+2$, we have that

$$a_{i+2} = a_i - q_i a_{i+1} \tag{6}$$

that is, a_{i+2} is a linear combination of a_i and a_{i+1}. Since, by the induction hypothesis, a_i and a_{i+1} are linear combinations of a and b, it follows that a_{i+2} is a linear combination of a and b as desired.

In the example,

$$37 = 3293 - 4\times 814 ,$$

$$814 = 4107 - 1\times 3293 ,$$

hence,

$$\gcd(4107, 3293) = 37 = 3293 - 4\times(4107 - 1\times 3293) = (-4)\times 4107 + 5\times 3293 .$$

A *prime number* is any integer p which is not a unit, that is, not 1 or -1, but which is not divisible by any number other than 1, -1, p and $-p$. Clearly, p is prime iff $-p$ is prime; therefore, it is customary to restrict attention to positive primes; in what follows, by "prime number" we always mean a positive prime. Restated, p is prime if $p > 1$, and the only positive divisors of p are 1 and p.

A fundamental property of primes is this:

if p *is a prime, and* $p \mid ab$, *then either* $p \mid a$, *or* $p \mid b$ (*or both*).

Indeed, assume also that p does not divide a, to show that $p \mid b$. Then $\gcd(p, a) = 1$, since $\gcd(p, a)$ is a divisor of p, therefore it cannot be anything else but 1 or p, and it cannot be p, since then p would divide a. Since $\gcd(p, a)$ is a linear combination of p and a,

$$1 = xp + ya$$

for suitable integers x and y. Multiplying this equality with b, we get

$$b = xbp + yab\ .$$

Since, by assumption, ab is divisible by p, $ab = zp$ for a suitable z, we have

$$b = (xb + yz)p\ ,$$

that is, b is divisible by p, which is what we wanted to show.

An obvious generalization of the last fact is this:

if p *is a prime, and* $p \mid \prod_{i<k} a_i$, *then* $p \mid a_i$ *for at least one* $i < k$.

We **claim**:

Every non-zero, non-unit integer has at least one prime divisor.

Let a be any integer, $a \neq 1$, $a \neq -1$. We may assume that $a > 1$. The set X of all divisors of a that are greater than 1 is a non-empty set; a itself is an element of it. By the LNP, let p be the least element of X . p must be prime; otherwise, there would be a divisor x of p which is greater than 1 but less than p ; x would be a non-unit divisor of a smaller than p , contrary to the choice of p . This proves the **claim**.

There are many prime numbers; in fact, there are *infinitely* many:

for any $n \varepsilon \mathbb{N}$, *there is a prime number greater than* n .

Indeed, consider the number $n! + 1$, and let p be a prime divisor of this number. p cannot be $\leq n$, since then p would be a divisor of $n!$, and hence also a divisor of $(n! + 1) - n! = 1$, which is absurd since p is not a unit. p is a prime number greater than n .

Next, we see that

Every non-zero number is the product of prime numbers.

Let n be any positive integer. If $n = 1$, n is the empty product of prime numbers. We treat the general case by induction, more precisely, by the WOP. Let $n > 1$. We know that n has at least one prime divisor; let p one such; let $m = \frac{n}{p}$. Since $m < n$, we may apply the induction hypothesis, and have that m is the product of prime numbers, $m = \prod_{i<k} p_i$. But then, $n = m \cdot p$, and $n = (\prod_{i<k} p_i) \cdot p$, and n is also a product of primes.

Let us use the notation p_i for the $i+1^{\text{st}}$ prime; see the end of the last section. With the fixed meaning of the p_i , we may write every positive n in the form

$$n = \prod_{i<k} p_i^{\alpha_i}$$

with suitable natural exponents α_i . Indeed, we know that n is the product of a certain number of prime factors; by bringing together the equal factors into powers, and using the exponent 0 in case a specific p_i does not occur in the product, we get the form mentioned. E.g.,

$$2420 = 20 \times 121 = 2^2 \times 5 \times 11^2 = 2^2 \times 3^0 \times 5^1 \times 7^0 \times 11^2 ;$$

now, $k = 5$.

We have:

Prime factorization is unique:

$$\textit{if } n = \prod_{i<k} p_i^{\alpha_i} = \prod_{i<k} p_i^{\beta_i} , \tag{7}$$

$$\textit{then } \alpha_i = \beta_i \textit{ for all } i < k .$$

The proof is by induction on n (via the WOP). If $n = 1$, it is clear that $\alpha_i = \beta_i = 0$ for all $i < k$. Otherwise, for some $i < k$, say i_0 , we have that $\alpha_{i_0} \geq 1$; let $p = p_{i_0}$. p divides $n = \prod_{i<k} p_i^{\beta_i}$, and since p is prime, p divides at least one $p_i^{\beta_i}$. But if $i \neq i_0$, p does not divide $p_i^{\beta_0}$ (why?). Thus, p must divide $p_{i_0}^{\beta_{i_0}}$, which implies that $\beta_{i_0} \geq 1$. Now, dividing (7) by the factor p_{i_0} , we get

$$m \underset{\text{def}}{=} \prod_{i<k} p_i^{\alpha'_i} = \prod_{i<k} p_i^{\beta'_i}$$

where $\alpha'_i = \alpha_i$ for $i \neq i_0$, $\alpha'_{i_0} = \alpha_{i_0} - 1$, and similarly for the β'_i . Clearly, $m < n$. By the induction hypothesis, prime factorization for m is unique; hence, $\alpha'_i = \beta'_i$ for all $i < k$. This means that $\alpha_i = \beta_i$ for all $i \neq i_0$, and $\alpha_{i_0} = \alpha'_{i_0} + 1 = \beta'_{i_0} + 1 = \beta_{i_0}$, that is, $\alpha_i = \beta_i$ for all $i < k$, as desired.

In terms of prime factorization, divisibility may be characterized as follows:

$$\textit{if } n = \prod_{i<k} p_i^{\alpha_i} ,\ m = \prod_{i<k} p_i^{\beta_i} , \textit{ then } n \mid m \textit{ iff } \alpha_i \leq \beta_i \textit{ for all } i < k .$$

The reason is simple: if $n|m$, then $m = n\cdot\ell$ for some ℓ; hence, if $\ell = \prod_{i<k} p_i^{\gamma_i}$ (with possibly a greater k; extend the range of the α's and β's by inserting 0's), we have that

$$m = \prod_{i<k} p_i^{\alpha_i} \cdot \prod_{i<k} p_i^{\gamma_i} = \prod_{i<k} p_i^{\alpha_i+\gamma_i} .$$

By the uniqueness of prime factorization, $\beta_i = \alpha_i+\gamma_i$; and since each $\gamma_i \geq 0$, we get that $\beta_i \geq \alpha_i$ as claimed.

Counting means assigning consecutive natural numbers to the elements of a set in a one-to-one fashion. Let us formulate counting in a mathematical style.

With n a natural number, let $[n)$ denote the set

$$[n) \underset{\text{def}}{=} \{k\in\mathbb{N} \mid k<n\} = \{0,\ 1,\ \ldots,\ n-2,\ n-1\}\ ;$$

$[n)$ is the primordial set with exactly n elements. If $n=0$, $[n)=\emptyset$, the empty set; $[1)=\{0\}$, $[2)=\{0, 1\}$, etc.

We say that *the cardinality of* the set X *is* n, or that *the number of elements of* X *is* n if there is a bijection $f\colon [n) \xrightarrow{\cong} X$; the function f provides the counting of X.

The question arises if one could have a bijection $f\colon [n) \xrightarrow{\cong} X$ and another $g\colon [m) \xrightarrow{\cong} X$ with the same set X, but with different n and m; if so, the notion of cardinality would not be well-defined. The answer to the question is "no"; the described situation is impossible. Namely, if we had that situation, $h \underset{\text{def}}{=} g^{-1}\circ f\colon [n) \longrightarrow [m)$ would be a bijection (see Chapter 1, p.25, where it is stated that the composite of two bijections is a bijection), and we would have a bijection between two different sets of the form $[n)$, contrary to the third of the following propositions:

If $h\colon [n) \longrightarrow [m)$ *is an injection,* $n\le m$;
if $h\colon [n) \longrightarrow [m)$ *is a surjection,* $n\ge m$;
if $h\colon [n) \longrightarrow [m)$ *is a bijection,* $n=m$.

The proof of these assertions use induction; of course, the last part is a consequence of the two previous parts.

We call a set A *finite* if there exists $n\in\mathbb{N}$ such that the cardinality of A is n. That is, A

is finite if there exists a bijection of the form $[n) \xrightarrow{\cong} A$, with $n \in \mathbb{N}$. The cardinality of the set A is denoted by $|A|$; $|A|$ is defined just in case A is finite (in set theory, one talks about the cardinality of infinite sets too; we will not do so here). Thus, e.g.,

$$|[n)| = n \qquad (n \in \mathbb{N})$$

(exemplified by the identity function $1_{[n)} : [n) \longrightarrow [n)$).

A frequently applied method of finding out the cardinality of a set X is to find another set Y the cardinality of which is known, and to establish a bijection of X and Y; in this case we know that the cardinality of X is the same as that of Y. The principle is

If $|Y| = n$ *and* $f: Y \xrightarrow{\cong} X$, *then* $|X| = n$.

This is obvious, since, under the assumptions here, we have some $g: [n) \xrightarrow{\cong} Y$, and then $f \circ g: [n) \xrightarrow{\cong} X$.

One simple law concerning finiteness is that

any subset of a finite set is finite; *moreover, a proper subset of a finite set has a strictly smaller cardinality.*

The rigorous proof is by an induction: one proves by induction on the natural number n that any subset of a set of cardinality n is finite, in fact, of cardinality $\leq n$.

The three propositions stated above immediately generalize in the following forms:

If $h: A \longrightarrow B$ *is an injection,* $|A| \leq |B|$;
if $h: A \longrightarrow B$ *is a surjection,* $|A| \geq |B|$;
if $h: A \longrightarrow B$ *is a bijection,* $|A| = |B|$.

The third proposition expresses the fundamental fact of life according to which if we count a pile of pebbles on two occasions, and in the meantime, no pebble was added or taken away, then the numbers arrived at must be the same.

The first proposition is equivalent to the so-called **pigeon-hole principle,** according to which

if we have put n things into less than n holes, then in at least one hole, we have put at least two things.

Namely, let the set of the things be A and the set of holes B; $|A| = n$ and $|B| < n$; let the function $f:A \longrightarrow B$ map every thing in A to the hole it is put into; since $|A| > |B|$, f cannot be an injection, that is, there are $a_1 \neq a_2$ in A for which $f(a_1) = f(a_2)$, i.e., a_1 and a_2 are put into the same hole.

E.g., among thirteen people, there must always be at least two who were born in the same month. Among thirteen integers, there always are two distinct ones whose difference is divisible by 12 : there are two that give the same remainder when divided by 12, and their difference is divisible by 12.

If $f:A \longrightarrow A$ is an injective function of a finite set A into itself, then f is a bijection; if $f:A \longrightarrow A$ is a surjective function of a finite set A into itself, then f is a bijection.

To see the first assertion, assume that $f:A \longrightarrow A$ is injective. Suppose f is not surjective, to derive a contradiction. There is some $a \in A$ such that $a \notin \mathrm{range}(f)$. Then the same function f can be considered a function from A to $A-\{a\}$; that is, $f:A \longrightarrow A-\{a\}$; f so construed is still injective. But then we would have $|A| \leq |A-\{a\}|$, contradicting the fact that $A-\{a\}$ is a proper subset of A. This contradiction proves that f must be surjective.

The other half of the proposition is proved similarly; now, we take away an element from the domain, rather than from the codomain.

An application of the last-stated principle is the important proposition that

the congruence $ax \equiv b \pmod{n}$ *is solvable for the unknown* x *provided* $\gcd(a, n) = 1$.

(For congruences, see section 2.2 in Chapter 2.)

To see this, first of all, recall that we denoted the set of all equivalence classes of the congruence $\bmod\ n$ by $\mathbb{Z}/n$, and that we proved that $\mathbb{Z}/n$ has exactly n elements; in particular, it is a finite set. Now, consider the mapping $f:\mathbb{N}/n \longrightarrow \mathbb{N}/n$ that takes $[x]$ to $[ax]$. Is f well-defined? For this, we need that if $[x]=[y]$, then $[ax]=[ay]$. But this is true: see Exercise 2 on page 45 of Chapter 2.

Under the assumption that $\gcd(a, n)=1$, f is an injective map: if $[ax]=[ay]$, then $ax \equiv ay \pmod{n}$, that is, $n \mid ax-ay=a(x-y)$; and since $\gcd(a, n)=1$, that is, a and n have no common prime factor, we must have that $n \mid x-y$, which means $x \equiv y \pmod{n}$, and so $[x]=[y]$.

By our last stated principle, f is surjective. This is what we want: the surjectivity of f means that for any $[b] \in \mathbb{Z}/n$ there exists $[x] \in \mathbb{Z}/n$ such that $f([x])=[b]$, that is, $[ax]=[b]$, that is, there is $x \in \mathbb{Z}$ such that $ax \equiv b \pmod{n}$.

The basic laws of counting connect operations on sets with operations on numbers. Here are the most important ones; the sets A , B , etc. are assumed to be finite.

$$|A \cup B| = |A| + |B| \qquad (1)$$

provided $A \cap B = 0$ (A *and* B *are disjoint*),

$$|A \times B| = |A| \cdot |B| , \qquad (2)$$

$$|B^A| = |B|^{|A|} . \qquad (3)$$

These laws contain the information that the operations of union, Cartesian product, and exponentiation, when applied to finite sets, result in finite sets again. The laws can be proved by appropriate inductions; e.g., the last one by induction on $|A|$. Let us see how this proof goes.

Basis Step. $|A| = 0$. In this case, A is empty, and there is exactly one function from $A = 0$ to (any set) B. Therefore, $|B^A| = 1$. Also, by the definition of exponentiation of numbers, $|B|^{|A|} = |B|^0 = 1$. This shows that the desired equality holds in this case.

Induction Step. $|A| = n + 1$, that is, there is a bijection $f\colon [n+1) \longrightarrow A$. Let $a = f(n)$, and let $A' = A - \{a\}$. The function f restricted to the subset $[n)$ of its domain, $g = f\upharpoonright[n)$, is now a bijection from $[n)$ to A'; in particular, $|A'| = n$. Now, we set up a bijection

$$g : B^{A'} \times B \xrightarrow{\cong} B^A$$

as follows: to any pair $(s, b) \in B^{A'} \times B$ where s is a function $s{:}A' \longrightarrow B$ and $b \in B$, g assigns the function $t{:}A \longrightarrow B$ for which

$$t(x) = \begin{matrix} s(x) & \text{if } x \neq a \text{ (hence, } x \varepsilon A'\text{)} \\ b & \text{if } x = a \end{matrix}$$

It is easy to check that g is indeed a bijection. It follows that

$$|B^A| = |B^{A'} \times B| \qquad = |B^{A'}| \cdot |B| \qquad \text{(by (2))}$$

$$= |B|^n \cdot |B| \qquad \text{(by the induction hypothesis,}$$

$|B^{A'}| = |B|^n$, since $|A'| = n$)

$$= |B|^{n+1} \qquad \text{(by the laws } m^1 = m\text{,}$$

$m^{n+\ell} = m^n \cdot m^\ell$ for exponentiation of numbers)

$$= |B|^{|A|}$$

as desired.

An application of the last fact is the proof of the relation

$$|\mathcal{P}(A)| = 2^{|A|},$$

or in words, the number of all subsets of an n-element set is 2^n. The reason for this is that the subsets of A are in a one-to-one correspondence with the functions $A \longrightarrow \{0, 1\}$: if $X \subseteq A$, we consider $\chi_X : A \longrightarrow \{0, 1\}$, the *characteristic function* of X, defined by

$$\chi_X(a) = \begin{array}{l} 1 \text{ if } a \in X \\ 0 \text{ if } a \notin X \end{array} .$$

Any function $\chi : A \longrightarrow \{0, 1\}$ is the characteristic function of a unique subset X of A, namely of $X = \{a \in A \mid \chi(a) = 1\}$. Thus, we have the bijection

$$\begin{tikzcd}
\mathcal{P}(A) \arrow[r, "\cong"] & \{0, 1\}^A \\
X \arrow[r, mapsto] & \chi_X
\end{tikzcd}$$

and therefore, $|\mathcal{P}(A)| = |\{0, 1\}^A| = 2^{|A|}$ as claimed.

The laws (1), (2), (3) are generalized to many-termed unions/sums and Cartesian products/products as follows. In what follows, I and each A_i are assumed to be finite sets.

Sum rule:

$$|\bigcup_{i \in I} A_i| = \sum_{i \in I} |A_i| \qquad \textit{provided the } A_i \textit{ are pairwise disjoint:}$$

$A_i \cap A_j = \emptyset$ *whenever* $i, j \in I$ *and* $i \neq j$.

Product rule:

$$|\prod_{i \in I} A_i| = \prod_{i \in I} |A_i|$$

(thus, the use of the same symbol $\prod$ for the product of numbers and the Cartesian product of sets is justified). The proofs of these identities are by induction on $|I|$. (1) is the special case of the sum rule when $|I| = 2$; (2) is the special case of the sum rule when $B = I$ and $A_b = \{b\} \times A$ (essentially, the case of equal-cardinality terms); (3) is the special case of the product rule when $I = A$, and $A_i = B$ for all $i \in I$.

The sum rule can be expressed in the following informal way. We have a set A which is *partitioned into* certain subsets A_i, for $i \in I$, or A is the *disjoint union* of the A_i's, meaning that $A = \bigcup_{i \varepsilon I} A_i$ and the A_is are pairwise disjoint. Note that this is the same as to say that every $a \in A$ belongs to A_i for exactly one index $i \in I$. To count the elements of A it suffices to count the elements of each A_i, and to add up the numbers obtained. We write $A = \dot{\bigcup_{i \varepsilon I}} A_i$ to indicate that A is the disjoint union of the A_i's.

To consider a kind of situation when the sum rule is useful, let $f: A \longrightarrow B$ an arbitrary function. Then the sets $f^{-1}(\{b\})$ when b runs over B form a partition of A: every $a \varepsilon A$ is in exactly one of the sets $f^{-1}(\{b\})$, namely the one for which $b = f(a)$. The sum rule says that $|A| = \sum_{b \varepsilon B} |f^{-1}(\{b\})|$. If we also assume that the sets $f^{-1}(\{b\})$ are all of equal cardinality, say m, then this says that $|A| = m \cdot |B|$.

The product rule is paraphrased as follows. An element of $\prod_{i \varepsilon I} A_i$ is the result of m independent choices ($m = |I|$), the i^{th} choice constrained to lie in the set A_i. The number of such compound selections consisting of m independent choices is the product of the numbers of the possibilities of the m individual choices.

The product rule has a generalized form which is the really useful version in practice. In this, we have selections in which the individual choices are not independent of each other, but the numbers of them are. We consider a subset A of a Cartesian product $\prod_{i<n} B_i$ determined as follows. The sequence $\langle a_i \rangle_{i<n}$ from $\prod_{i<n} B_i$ belongs to A iff each a_i belongs to a certain *constraint-set* $A(\langle a_j \rangle_{j<i})$, a subset of B_i depending on the segment $\langle a_j \rangle_{j<i}$ of the a_j preceding a_i. The essential assumption is that the cardinality of the

constraint-set $A(\langle a_j\rangle_{j<i})$ does not depend on $\langle a_j\rangle_{j<i}$, just on i ; let us say, this cardinality is n_i :

$$|A(\langle a_j\rangle_{j<i})| = n_i ,$$

at least when $\langle a_j\rangle_{j<i}$ is *properly constrained*: $a_j \in A(\langle a_k\rangle_{k<j})$ for all $j<i$.

In this case, $|A| = \prod_{i<k} n_i$. Let us call this rule the ***product rule for dependent selections.***

The product rule for dependent selections can be proved by induction on k , the length of the selections made.

Let us take a simple case illustrating the last mentioned rule. Let C be an alphabet of size n , and let us compute the number of strings in C^* in which there are no identical letters next to each other. The set of such strings being called A , A is a subset of $\prod_{i<k} C$ (we identify strings with sequences), and $\langle a_i\rangle_{i<k}$ from $\prod_{i<k} C$ belongs to A just in case for each i in the range $1\le i\le k$, we have $a_i \ne a_{i-1}$. In other words, in this case

$$A(\langle a_j\rangle_{j<i}) = \{a\in C \mid a\ne a_{i-1}\}$$

if $1\le i\le k$, and

$$A(\langle a_j\rangle_{j<0}) = C .$$

Thus, the numbers n_i are: $n_0 = n$, $n_i = n-1$ when $1\le i\le k$, and the desired number is $n\cdot(n-1)^{k-1}$.

It is customary to express the above argument in the following informal way. To have a string in which there are no two identical letters next to each other, we may take n different letters as the first letter of the string. But for the second letter, we can take only $n-1$, since the first one is now excluded. This says that the number of compound choices for the first two positions is $n(n-1)$. For the third letter we can again choose from $n-1$ letters, the ones that are different from the second letter, whatever that was; thus, there are $n(n-1)(n-1)$ possibilities for the segment in the first three positions. Etc.; the number of such strings of length k is $n(n-1)^{k-1}$.

We can see that the informal argument actually reproves the product rule by induction on k.

Let us determine the cardinality of some important finite sets.

Suppose that $|A| = m$, $|B| = n$ *and* $m \leq n$. *Then the number of injections* $A \xrightarrow{\cong} B$ *between two given sets* A *and* B is

$$\prod_{i<m} (n-i) = n\cdot(n-1)\cdot \ldots \cdot(n-m+2)(n-m+1)$$

This can be easily shown by the product rule for dependent selections. First of all, we may assume without loss of generality that $A = [m)$. A function $A \longrightarrow B$ is a sequence $\langle b_i\rangle_{i<m}$ with each $b_i \in B$. The sequence $\langle b_i\rangle_{i<m}$ is an injection iff for all $i < m$, b_i differs from b_j for each $j < i$. This means that b_i in $\langle b_i\rangle_{i<n}$ is constrained to lie in the set

$$B(\langle b_j\rangle_{j<i}) = \{b\in B \mid b \neq b_j \text{ for all } j < i\}.$$

The latter set has cardinality $n - i$, since the b_j's are all distinct (the selection $\langle b_j\rangle_{j<i}$ being "properly constrained"), and hence, there are exactly i of them. We see that the cardinality of the constraint-set $B(\langle b_j\rangle_{j<i})$ is independent of the segment $\langle b_j\rangle_{j<i}$, it depends on i only. The product rule, for the variant for dependent selections, gives that the desired number is $\prod_{i<m} (n-i)$ as promised.

A special case of the last proposition, for the case $m = n$, is the following.

The number of bijections between two fixed sets of the same cardinality n *is* $n!$ *; in particular, the number of permutations of a set of cardinality* n *is* $n!$.

Indeed, this follows from the previous proposition, since any injection from a set to another of the same cardinality is a bijection, as we stated above.

Note that the injections from $[m)$ into an alphabet A are the same as the strings in A^* of length m in which no letter is repeated; the proposition above gives a formula for the number of such strings.

Let $\binom{n}{k}$ (read: " *n-choose-k* ") denote the number of k-element subsets (more briefly: k-subsets) of an n-element set. Clearly, if $n < k$, then $\binom{n}{k} = 0$. Also, $\binom{n}{0} = \binom{n}{n} = 1$. We claim that

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \quad \textit{whenever } k \leq n .$$

To show this, let us fix k and n, $k \leq n$. Let the set of all permutations of $[n)$ be called P, and let the set of all k-subsets of $[n)$ be S. We partition the permutations of $[n)$ into as many disjoint sets as there are k-subsets of $[n)$. Let $\sigma\colon [n) \xrightarrow{\cong} [n)$ be any permutation; consider the set of values of σ at the first k arguments $0, 1, \ldots,$ $k-1$, that is, the set

$$X_\sigma \overset{=}{\text{def}} \{\sigma(0), \sigma(1), \ldots, \sigma(k-1)\} .$$

Since σ is one-to-one, X_σ is a k-subset of $[n)$. Consider the function

$$\begin{aligned} f\colon P &\longrightarrow S \\ \sigma &\longmapsto X_\sigma . \end{aligned}$$

For any k-subset $X \in S$ *of* $[n)$, $f^{-1}(\{X\})$ consists of those permutations σ for which X_σ is the given set X;

$$f^{-1}(\{X\}) = \{\sigma \mid X_\sigma = X\} .$$

We **claim** that

$$|f^{-1}(\{X\})| = k!(n-k)! \qquad (4)$$

independently of X. This equality is based on the following general fact.

Suppose $A=A_1\dot{\cup}A_2$ and $B=B_1\dot{\cup}B_2$, assume that $|A_1|=|B_1|$, $|A_2|=|B_2|$ (and, as a consequence, $|A|=|B|$), and let us consider the set T of those bijections $\sigma:A\xrightarrow{\cong}B$ for which $\sigma[A_1]=B_1$, that is, σ maps A_1 onto B_1. Then, writing (temporarily) $\mathrm{Bij}(U,V)$ for the set of all bijections $U\xrightarrow{\cong}V$, we have a *bijective* mapping

$$\begin{array}{ccc} T & \xrightarrow{\qquad\cong\qquad} & \mathrm{Bij}(A_1,B_1)\times\mathrm{Bij}(A_2,B_2) \\ \sigma & \longmapsto & (\ \sigma\upharpoonright A_1\ ,\ \ \sigma\upharpoonright A_2\ ) \end{array} \qquad . \qquad (4')$$

The point is that if the bijection $\sigma:A\xrightarrow{\cong}B$ maps (bijectively) A_1 onto B_1, then it necessarily maps the rest of A, A_2, bijectively onto B_2, the rest of B. In other words, if $\sigma\in T$, then $\theta_1=\sigma\upharpoonright A_1\in\mathrm{Bij}(A_1,B_1)$ and $\theta_2=\sigma\upharpoonright A_2\in\mathrm{Bij}(A_2,B_2)$. Conversely, if $\theta_1\in\mathrm{Bij}(A_1,B_1)$, $\theta_2\in\mathrm{Bij}(A_2,B_2)$, then σ defined by

$$\sigma(a) \;=\; \begin{array}{ll} \theta_1(a) & \text{if } \ a\in A_1 \\[1ex] \theta_2(a) & \text{if } \ a\in A_2 \end{array}$$

is a bijection $\sigma:A\xrightarrow{\cong}B$ for which $\sigma\upharpoonright A_1=\theta_1$ and $\sigma\upharpoonright A_2=\theta_2$.

In our application, $A=B=[n)$, $A_1=[k)$, $A_2=[n)-[k)$, $B_1=X$, $B_2=[n)-X$. Then the set T is what we called $f^{-1}(\{X\})$. Since $|A_1|=B_1|=k$, $|A_2|=B_2|=n-k$, we have $|\mathrm{Bij}(A_1,B_1)|=k!$, $|\mathrm{Bij}(A_2,B_2)|=(n-k)!$. The relation (4') therefore tells us that $|T|=k!\cdot(n-k)!$, as desired. This shows (4).

Since for each k-subset X of $[n)$, $f^{-1}(\{X\})$ is of the same cardinality, namely

$k!(n-k)!$, and P is partitioned into $\binom{n}{k}$ sets $f^{-1}(\{X\})$, we have

$$|P| = \binom{n}{k} \cdot k!(n-k)! \ .$$

But we know that $|P| = n!$. The desired expression for $\binom{n}{k}$ follows by dividing by $k!(n-k)!$.

The numbers $\binom{n}{k}$ are called the *binomial coefficients*, because their appearance in the

Binomial theorem:

$$(x+y)^n = \sum_{k\leq n} \binom{n}{k} x^k y^{n-k} \qquad\qquad (n \in \mathbb{N},\ n \geq 1) \ .$$

This equality is immediate when one considers that in the product $(x+y) \ldots (x+y)$ (n factors), when written out via the distributive law as a sum of monomials $x^k y^{n-k}$, the number of terms with exactly k x-factors (and hence exactly $n-k$ y-factors) is the same as the number of ways we can select k factors $(x+y)$ out of the n such; the latter number is, by definition, $\binom{n}{k}$.

The binomial coefficients satisfy many identities. One such is

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \ .$$

The reason for this is the fact that the set S of $k+1$-subsets of $[n+1)$ is partitioned into two disjoint subsets, S_1 and S_2 , according to whether $X \varepsilon S$ does or does not contain the element n . The elements of S_1 are in one-to-one correspondence with the k-subsets of $[n)$: with $X \varepsilon S_1$, take away from X the fixed element n , and get a k-element subset of $[n)$. S_2 is nothing but the set of all $k+1$-subsets of $[n)$. Thus

$$|S| = \binom{n+1}{k+1}, \quad |S_1| = \binom{n}{k}, \quad \text{and} \quad |S_2| = \binom{n}{k+1},$$

and since $S = S_1 \dot{\cup} S_2$, the assertion follows.

The last-proved identity gives a recursive definition of the binomial coefficients. The successive calculation of the binomial coefficients is suggested by the *Pascal triangle*:

$$\begin{array}{ccccccccccc}
 & & & & & \binom{0}{0} & & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

in which every coefficient is the sum of the two immediately above it, and in which all the values on the two sloping sides are equal to 1.

Substituting particular values for x and y in the binomial theorem, we get various identities involving the binomial coefficients. E.g., if we put $x = -1$, $y = 1$, we obtain

$$(-1+1)^n = 0 = \sum_{k \leq n} (-1)^k \binom{n}{k} \qquad (n \geq 1),$$

that is,

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^{n-1}\binom{n}{n-1} + (-1)^n\binom{n}{n} = 0 .$$

Since $\binom{n}{0} = 1$, we may rewrite this as

$$\binom{n}{1} - \binom{n}{2} + \ldots + (-1)^{k+1}\binom{n}{k} + \ldots + (-1)^{n+1}\binom{n}{n} = 1 ,$$

that is,

$$\sum_{k=1}^{n} (-1)^{k+1}\binom{n}{k} = 1 \qquad (n \geq 1). \tag{5}$$

We will make use of the last identity in establishing the so-called *sieve principle*, or *inclusion/exclusion principle*.

The principle mentioned concerns the way one can compute the cardinality of a union of sets. The sum rule gives the answer when the sets involved are pairwise disjoint. In the general case, the answer involves the cardinalities of the various intersections of the given sets (which are all equal to 0 in the disjoint case).

Consider the special case of the union of two sets. We have

$$| A_1 \cup A_2 | = | A_1 | + | A_2 | - | A_1 \cap A_2 | ;$$

the reason is that "when we add up the cardinalities of A_1 and A_2 , we count the elements in the intersection $A_1 \cap A_2$ twice; subtracting the cardinality of the intersection corrects this".

The case of three sets is like this:

$$| A_1 \cup A_2 \cup A_3 | =$$
$$| A_1 | + | A_2 | + | A_3 | - | A_1 \cap A_2 | - | A_1 \cap A_3 | - | A_2 \cap A_3 | + |A_1 \cap A_2 \cap A_3 | .$$

An argument justifying this would say that the corrections afforded by the three subtractions over-correct precisely for the elements that are in at least two of the double intersections; but these are exactly the elements which are in the triple intersection; hence, we have to compensate by adding the cardinality of that triple intersection.

We have to admit that these arguments, although intuitive, fall somewhat short of the ideal of a clear mathematical proof. Considering that the general case of an arbitrary number of sets is likely to be more involved, we are drawn to a more serious mathematical approach.

First of all, let us state the general result:

$$\boxed{\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}|}$$

or in a more detailed form

$$|A_1 \cup A_2 \cup \ldots \cup A_n| =$$
$$\sum_{i=1}^{n} |A_i| - \sum_{1\le i_1<i_2\le n} |A_{i_1} \cap A_{i_2}| + \sum_{1\le i_1<i_2<i_3\le n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \ldots$$
$$+ (-1)^{n+1} |A_1 \cap A_2 \cap \ldots \cap A_n| .$$

Here, e.g. the sum $\sum_{1\le i_1<i_2\le n} |A_{i_1} \cap A_{i_2}|$ is taken over all pairs (i_1, i_2) of integers between 1 and n inclusive such that $i_1 < i_2$.

To prove this, we introduce the concept of *multiset*. Let X be a large set so that every set we may want to consider is a subset of X . A *multiset* is a function assigning a positive, negative or zero integer to every element of X ; briefly, a function M from X to $\mathbb{Z}$, $M: X \longrightarrow \mathbb{Z}$. Intuitively, M is a "set" for which the things in X may be in M with various "multiplicities"; $M(x)$ is the *multiplicity of* x *in* M . E.g., with $X = \mathbb{N}$, we may consider the multiset M for which $M(n) = 0$ for all $n \ge 5$, and $M(0)=1$, $M(1)=-4$, $M(2)=0$, $M(3)=1$, $M(4)=2$. We consider only *finite* multisets, that is, ones in which only finitely many elements have a multiplicity different from 0 .

A simple notation for concrete multisets follows the notation for functions; the multiset in the example may be denoted by

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & -4 & 0 & 1 & 2 \end{pmatrix} . \qquad (6)$$

It is understood that for any $x \varepsilon X$ not in the upper row of the notation, the multiplicity is 0 .

Any ordinary set A (a subset of X) is considered as a multiset $\dot{A}$ for which

$$\dot{A}(x) \quad = \quad \begin{matrix} 1 & \text{if } x \varepsilon A \\ 0 & \text{if } x \notin A \ . \end{matrix}$$

In other words, $\dot{A}$ is the characteristic function of A as a subset of X . E.g., if $A = \{0, 2, 5\}$, then $\dot{A}$ is

$$\dot{A} = \binom{0\ 1\ 2\ 3\ 4\ 5}{1\ 0\ 1\ 0\ 0\ 1} \ .$$

The *cardinality of* a multiset M, $|M|$, is, by definition, the sum of the multiplicities of the elements: $|M| = \sum_{x \varepsilon X} M(x)$; since we assume that only finitely many $M(x)$ are different from 0 , the sum is a well-defined integer. E.g., in the example, $|M| = 0$, although M is far from being the same as the empty set.

Note that for a finite set A , the usual cardinality of A and the cardinality of it as a multiset are the same: $|A| = |\dot{A}|$.

We define *addition* of multisets by simply adding multiplicities: the multiset $M + N$ is defined by the equality

$$(M + N)(x) \underset{\text{def}}{=} M(x) + N(x)$$

In other words, the multiplicity of an element x in the sum-multiset $M+N$ is, by definition, the sum of the multiplicities of x in M and N .

E.g., for M as above, and for $A = \{0\ ,\ 2\ ,\ 5\}$, $M + \dot{A}$ is the multiset
$\binom{0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5}{2\ -4\ \ 1\ \ 1\ \ 2\ \ 1}$.

If a is an integer, $a \cdot M$ or more simply aM , (*scalar multiplication*) is the multiset for which

$$(aM)(x) \underset{\text{def}}{=} aM(x) .$$

E.g., the multiset $(-1)\dot{A}$ for A as above is $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ -1 & 0 & -1 & 0 & 0 & -1 \end{pmatrix}$.

$-M$ means $(-1)M$, and $M - N$ means $M + (-1)N$. E.g., with M and A as before, $M - \dot{A}$ is $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & -4 & -1 & 1 & 2 & -1 \end{pmatrix}$.

The usual rules concerning addition and scalar multiplication (commutativity, associativity, etc) familiar from linear algebra are valid for addition and scalar multiplication of multisets, since they are inherited from those operations on numbers.

We have the following rules connecting cardinality and the operations just introduced:

$$| M + N | = | M | + | N | ,$$

$$| aM | = a| M | .$$

These are immediate from the definitions. As a consequence, the cardinality of a linear combination of multisets is the corresponding linear combination of the cardinalities of the terms.

The main point is the following equality of multisets: for any (ordinary) sets $A_1 , A_2 , \ldots , A_n ,$ we have

$$\Big(\bigcup_{i=1}^{n} A_i\Big)^{\cdot} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} (A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})^{\cdot} . \quad (7)$$

To prove this, we take an arbitrary $x \varepsilon X$, and show that the multiplicity of x in the left-hand side equals the multiplicity of x in the right-hand side. If x does not belong to any of the A_i , that is, the multiplicity of x in the left side is 0 , then it does not belong to any of the sets involved in the right side either, and thus its multiplicity on the right, being a sum of 0's , is also 0 . Let us then assume that x does belong to at least one A_i ; thus, the multiplicity of x on the left is 1 . Let those indices $\ell = 1, \dots, n$ for which $x \varepsilon A_\ell$ be $\ell_1 < \ell_2 < \dots < \ell_m$; in particular the number of these ℓs is m ; $m \geq 1$. Let us also write

$$L \underset{\text{def}}{=} \{\ell_1 , \ell_2 , \dots \ell_m\} .$$

Take an arbitrary selection $i_1 < i_2 < \dots < i_k$ of indices between 1 and n (inclusive) , and ask what the multiplicity

$$(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})^{\cdot}(x) \tag{8}$$

is. Clearly, this is 1 or 0 depending on whether x does or does not belong to the set $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$. On the other hand, x belongs to the latter set if and only if x belongs to each one of the sets A_{i_1} , A_{i_2} , $\dots A_{i_k}$, that is, if each of i_1 , i_2 , $\dots$, i_k is the same as one of ℓ_1 , ℓ_2 , $\dots \ell_m$, that is, if

$$\{i_1 , i_2 , \dots i_k\} \subseteq L . \tag{9}$$

We have shown that (8) is equal to 1 if (9) holds; otherwise (8) is 0 . Therefore, with a fixed k between 1 and n , the sum

$$\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} (A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})^{\cdot}(x)$$

equals the number of selections $i_1 < i_2 < \dots < i_k$ for which (9) holds. But this number is nothing but the number of k-subsets of L , and this is $\binom{m}{k}$. It follows that the right-hand side of (7), when evaluated at x , equals $\sum_{k=1}^{n} (-1)^{k+1} \binom{m}{k}$, which is the same as $\sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k}$, since $\binom{m}{k} = 0$ for $k > m$. By (5) and $m \geq 1$, the last sum is equal to 1 . We have shown that the multiplicity of x on the right in (7) is 1 ; since the multiplicity on the left is also 1 , we have proved (7).

Having proved (7), we may take the cardinality of the two multisets in (7). The cardinality of the left side is the same as the cardinality of the ordinary union-set. The cardinality of the right side may be taken term by term, as we pointed out above. The cardinalities of the intersection-multisets are just the cardinalities of the intersections as sets. We get the right-hand side expression in the framed equality; that equality is thus proved.

Note that the cases of two and of three sets stated earlier are the special cases of the general formula for $n = 2$ and $n = 3$.