

ELEMENTARY NOTIONS

In this first chapter we will review (quite) *informally* some basic mathematics notions, making especial emphasis on notation. It is meant for those students coming from physics and other sciences, who have never been exposed to rigorous mathematics before.

First, we note the standard symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, which respectively mean the sets of *natural*, *integer*, *rational*, *real* and *complex numbers*. Make sure you remember how each of them is defined (for our purpose here the physics standard of rigor will suffice).^[a]

1.1 Elements of Logic

We proceed to introduce informally a chain of elementary logic and set theoretic notions.

- When we wish to give a name to some object we will use the symbol “ $:=$ ”. For instance,

$$A := 8$$

means that within the scope of the explanation, when we use the symbol A , we mean 8. But note that two paragraphs later the same symbol may be used to denote a new term, and most often it is just a proxy for an arbitrary object of a certain type, say some “positive real number”. Importantly, in mathematics every variable has to be defined explicitly, no matter how obvious its nature or value is. Say, if ε is a proxy variable for a positive real number, no matter how clear it is from its use in the explanation, the mathematician will always write explicitly “let ε be an arbitrary positive real number.” That is because two paragraphs later they could say “let ε be a permutation mapping”, so do not grow fond of proxy variables. In any case, always define them explicitly!

- The curly braces $\{\}$ indicate that we are defining a *set*, which is an unordered grouping of elements. For instance,

$$B := \{1, 2, 3\}$$

means that B is the set composed by the elements 1, 2 and 3. Another example:

$$C := \{\text{odd integers}\},$$

means that C is the set of odd integers.

- The symbol “ $\in$ ” means “*in*”, and it is used to express the belonging of an element to a set. Two examples: (i) by definition, $5 \in C$ (which reads “5 is in C ” or “5 is an element of the set C ”) and (ii) if ε denotes a positive real number, then in particular $\varepsilon \in \mathbb{R}$.

^[a]Note that $\mathbb{N}$ does not contain 0. We use $\mathbb{N}_0$ to denote the set composed by 0 and all elements of $\mathbb{N}$.

- The symbols “:” and “|”, when shown in a mathematical expression mean “*such that*”. We mostly use them to define sets, say,

$$D := \left\{ x \in \mathbb{N} \mid x = 2n \text{ for some } n \in \mathbb{N} \right\} \quad (1.1)$$

means “ D is defined to be the set of elements x in $\mathbb{N}$ such that x equals 2 times n for some natural n ”, i.e., those that can be written as 2 times a natural number, namely, the even naturals. The following more compact notation means exactly the same as the right hand side (r.h.s) of Eq. (1.1):

$$D = \{2n\}_{n \in \mathbb{N}}.$$

We say that $\mathbb{N}$ “*indexes*” the set D because we can use $\mathbb{N}$ ’s elements to label the elements of D . We will see more on this in a moment.

- A *logical statement/sentence/assertion* is a claim which can be either true or false (of course, its truth value might depend on the definition we give to the objects the claim mentions and the axioms assumed in the mathematical theory). For instance

$$S := \text{“every natural number is prime.”}$$

Most likely, the basic logical statements you will see are claims about the existence of elements in a set satisfying a certain property, claims about their number, or claims that certain explicitly defined particular objects satisfy some previously defined properties. Out of them, using the following “assertion-operations”, one builds more elaborated logical statements.

- Given two assertions S_1, S_2 , the assertion claiming that *both* hold is denoted “ S_1 and S_2 ”, or more sophisticatedly as “ $S_1 \wedge S_2$ ” or “ $S_1 \& S_2$ ”. The assertion claiming that *at least one* of the two hold (possibly both), is denoted “ S_1 or S_2 ”, and more sophisticatedly, “ $S_1 \vee S_2$ ”. Lastly, the sentence claiming that *at least one, but not both* hold (an “exclusive or”), is denoted “either S_1 or S_2 ”, sometimes “ $S_1 \vee S_2$ ”.
- One denotes the *negation* (or logical complement) of an assertion S by “ $\neg S$ ”, which is read as “not S ”, and it represents whatever would be true if S were false. It swaps the following keys:

$$\text{“is”} \longleftrightarrow \text{“is not”}, \quad \text{“all/every....is”} \longleftrightarrow \text{“not all/every....is”}, \quad \wedge \longleftrightarrow \vee$$

$$\text{“for all } x/\text{for every } x\text{...”} \longleftrightarrow \text{“at least for one } x \text{ it is not the case that...”}$$

Some examples:

- the logical negation of “ x is divisible by 3 and it is odd” reads “ x is not divisible by 3 or it is not odd”.
- the logical negation of “for all $x \in A$ it holds the property P ” reads “for at least one $x \in A$, P does not hold”.

Note that the double negation of a statement leaves back the original statement. Also, a statement S is true if and only if its negation is false (and vice-versa).

- The symbol $\forall$ means “*for all/each*”. The symbol $\exists$ means “*there exists*”. For example

$$\exists x \in \mathbb{R} : \forall y \in \mathbb{R}, x + y = y$$

reads: “there exists a real number x such that for each real number y , their sum equals y ”. As an example, the negation of this sentence is:

$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R} : x + y \neq y,$$

i.e., “for each real x there exists a real y such that their sum does not equal y ”.

- The symbol $\exists!$ means “*there exists a unique*”, and so,

$$\exists! x \in \mathbb{R} : \forall y \in \mathbb{R}, x + y = y$$

reads: “there exists a unique real number x such that for each real number y , their sum equals y ”.

- The dashed symbols “ $\notin$ ”, and “ $\nexists$ ” mean respectively, “*not in*” and “*there is no*”. For instance, “ $x \notin A$ ” reads “ x is not an element of A ” and “ $\nexists n \in \mathbb{N} : n + 1 = n$ ” reads “there is no natural number such that after adding 1 we get back the same number”.

- The symbol “ $\implies$ ” is the “*logical consequence*” or “*implication*” symbol: it is an operator between two logical statements and it means “if the statement to the left is true, then, the statement to the right is true” or in short “then”. Say, given $x \in \mathbb{R}$,

$$(x > 10) \implies \left(\frac{2}{x} < 7\right).$$

(One reads: “if x is greater than 10, then $2/x$ is smaller than 7”.)

Given S_1, S_2 are any two assertions, “ $S_1 \implies S_2$ ” can be phrased in all the following equivalent ways:

- (a) “If S_1 holds, then S_2 as well”.
- (b) “Only if S_2 holds, may S_1 hold”.
- (c) “ S_1 implies S_2 ”.
- (d) “ S_1 is a sufficient condition for S_2 ”.
- (e) “ S_2 is a necessary condition for S_1 ”.

The symbol “ $\Leftarrow$ ” is defined in the obvious other way around.

- The symbol “ $\iff$ ” is the “*logical equivalence*” symbol and it means “if and only if”. More precisely, we denote “ $S_1 \iff S_2$ ” when both, “ $S_1 \implies S_2$ ” and “ $S_1 \Leftarrow S_2$ ”, i.e.,

$$(S_1 \iff S_2) = (S_1 \implies S_2) \wedge (S_1 \Leftarrow S_2).$$

One says any of the following equivalent statements:

- (a) “ S_1 holds if and only if S_2 holds”.
- (b) “ S_1 is a necessary and sufficient condition for S_2 (and vice-versa).”

- Note that “ $S_1 \implies S_2$ ”, “ $S_1 \Leftarrow S_2$ ” and “ $S_1 \iff S_2$ ” are also logical statements, so they can be negated, composed using $\wedge, \vee$ etc. Importantly, the following equivalence holds:

$$(S_1 \implies S_2) = \left((\neg S_1) \Leftarrow (\neg S_2) \right).$$

That is, it is the same to know that “if S_1 holds, necessarily S_2 holds”, than knowing that “whenever S_2 does not hold, then S_1 does not either”. This is called the *contrapositive* statement. Similarly,

$$(S_1 \Leftarrow S_2) = \left((\neg S_1) \implies (\neg S_2) \right),$$

$$(S_1 \iff S_2) = \left((\neg S_1) \iff (\neg S_2) \right).$$

1.2 The Way Modern Mathematics is Written

Full logical statements in mathematics, which we will call “*propositions*”, usually have the following structure: first, in a preamble, a sequence of assertions defining symbols and assuming properties for them. Second, a non-trivial claim, often given as a couple of assertions linked by “ $\Leftarrow$ ”, “ $\Rightarrow$ ” or “ $\iff$ ”. Let us refer to it as “*the claim*”. (Certainly, a proposition in itself is a logical sentence, but one can see the preamble as the construction of the “arena” and the presentation of the “cast” and their “backstories”, while the true “juice” lies in the the central sentence that we called the “claim”.) We will see tons of propositions in other lectures, but here is an example:

The
Claim

 {

 {

 Preamble

 }

 “Let p be an arbitrary prime number (where, by ‘prime’ we mean Definition X), and let m be an arbitrary natural number. Assume there exists an $n \in \mathbb{N}$ such that $m = n^2$. Then,

if p divides m , then, p divides n .”

The core task of a mathematician is to propose such statements and to prove whether they are true or not. This involves checking whether the conditions of the preamble are compatible with each other, making sure the objects presented there are not ill-defined —otherwise, as nice as the main claim could be, it may be talking about impossible stuff. This check is usually easy. For instance, in the example one would need to check whether such n and m exist (and they do: e.g., for $m = 9$ and $n = 3$). The core of the proof comes after that: to check the main claim’s truth under the given conditions.

To prove whether the claim is true, there exist several archetypic techniques, but no genuine algorithmic recipe. It is a purely creative task based on having “happy ideas” that end up working (or leading to new ideas), and it is usually entirely guided by intuitions and inspiration brought from proofs of similar statements. Also, many different proves (or disproves —if false) exist for a same claim, and they are all equally valid *qua proof*, no matter how long or succinct. Nonetheless, the simpler the proof/disproof is, the easier it is going to be for a colleague to understand the truth or falsehood of the claim. With all, in the literal sense, proposing mathematical statements and proving them is an *art*. One of an utmost delicate fineness.

- Four example archetypic techniques you should know about are the following ones:
- (a) Proof by *direct computation*: By a direct, step by step manipulation of an expression known to hold true under the premises of the preamble, we reach the main claim. By a chain of equalities, each of which has a simple justification, the result is proven.
 - (b) Proof by *contradiction* or *reductio ad absurdum*: We assume hypothetically that the main claim is false. If this were true, we then find a chain of arguments that would consequently hold true, but which leads to a logically contradictory statement, also called an absurd statement. That means that the assumption that lead to this conclusion (namely, the negated main claim) must be false, and thus, its logical negation, namely, the original claim must be true.
 - (c) Proof by *contraposition*: Instead of proving the direct sentence, we prove the contrapositive. Say, instead of proving “ $S_1 \implies S_2$ ”, we prove “ $(\neg S_1) \iff (\neg S_2)$ ”. As an example of how this could be simpler, consider the proposition: “Given $n \in \mathbb{N}$, if $3n+2$ is odd, then n is odd”, whose contrapositive is “Given $n \in \mathbb{N}$, if n is not odd, then $3n+2$ is not odd”. The second sentence is trivially proven by direct computation in a single step.
 - (d) Proof by *induction*. This style of proof applies when there is some natural variable in the preamble, say, n and the main claim is that a certain sentence holds for all possible values of n . The technique then is to first prove the $n = 1$ case by direct computation, which is usually easy. Then prove that if we know the sentence holds for an arbitrary n , then it also holds for the case $n + 1$. The gears for the main claim then “turn on their own.”
- Now, it is often the case that we wish to prove a very complex proposition, the proof of which takes several pages. The reason is that in the way we need to prove other smaller statements. This explains why a mathematical text is rarely given as a single proposition and a monolithic proof. Instead, it comes as a plethora of small propositions, each relatively simple to prove after chaining them adequately and all of which together end up proving the complex statement announced. The aforism at work here is “divide and conquer”. Yet, there is a risk with the chain of small propositions: the relevant results may get lost in the thin grain. That is why, mathematicians have developed names for the different types of propositions.
- (a) *Lemma*: it is the name for the smallest type of proposition, small in the sense that its proofs is simple and does not relay on other big results. They are usually very technical and their relevance is only seen in later results, or for the expert reader.

- (b) *Proposition* (with a capital “P”): it is the next tier of proposition regarding its simplicity and relevance to the reader. Alternatively, it is used when we do not want to categorize the result as big nor small and such a judgment is left to the reader.
- (c) *Theorem*: it is one of the big results of the work and the reason why the previous Lemmas were given. If a person wants to present their results to others this tier of result is the one usually shown as central.
- (d) *Corollary*: it is a proposition that follows almost trivially from the previous Lemmas and theorems. Since it usually comes after a theorem, it is often the main result the author wanted to show (although that is often given as a theorem).

The three first types of propositions are expected to be followed by their proofs, while a corollary does not need to when it is obvious how to use the other results to obtain it.

- Now, there are three other types of parts that constitute mathematical treaties (and lectures):
- (e) *Definition*: it is a declaration of what a certain word, concept or symbol means and how it relates to other objects. It should not contain any non-trivial claim needing a proof. Definitions are given before they are used in the propositions, so the beginning of a mathematical text usually starts with Definitions.
- (f) *Example*: it is a particular instance of a concept defined in a Definition. If checking it satisfies indeed the defining conditions is not trivial or a matter of simple computation, it is often given instead as a Lemma or a Proposition. It is allowed not to prove that an example satisfies the defining conditions.
- (g) *Remark* or *Observation*: It is an idea, a moral, a thought or a question that is remarkable, worthy of distinction from the main body of text surrounding the propositions, examples and definitions. It can range from an informal comment to a quasi-proposition (a statement that is almost a proposition but is a mere suggestion of it). It is not to be considered as a rigorous result of the work, nor to be taken as a binding claim.
- Organizing mathematical reports and textbooks in this fashion has many advantages. Not only it makes easier to read and follow the arguments (and to skip the parts we are not interested), but it allows something key in mathematics: to use results proven for one purpose also for other purposes! The most productive proof technique is the “*proof by literature*”, which consists on citing a proposition proven elsewhere in order to partially or fully check your proposition.

1.3 Functions

Given two sets X, Y the expression

$$f : X \rightarrow Y$$

means that the symbol f assigns to each element $x \in X$ an element $f(x) \in Y$. It is called a *function*, a *map* or a *mapping* (from X to Y). We say that X is the *domain* (of definition) of f , while Y is called the *co-domain* of f . Informally (but very commonly) one says that f “*eats*” elements of X (or that we “*plug in*” elements of X in f) and that it “*spits out*” elements of Y . For each $x \in X$, $f(x)$ is called the *image of x by f* and x is called a *pre-image of $f(x)$* . Note that multiple $x \in X$ may have the same $y \in Y$ as image, so a single $y \in Y$ may have multiple pre-images. The *range of f* (also called the *image of f*) is the subset of the co-domain Y that is actually ranged by f , i.e.,

$$\text{range}(f) := \left\{ y \in Y : f(x) = y \text{ for some } x \in X \right\} = \left\{ y \in Y \mid \exists x \in X : f(x) = y \right\},$$

often also denoted $\text{Im}(f)$.

- In order to define the assignments of a function f (also called, its *action* on X), one often uses the following extended notation:

$$\begin{array}{ccc} f : & X & \longrightarrow & Y \\ & x & \longmapsto & f(x) \end{array},$$

where $\mapsto$ is read “*maps to*” and tells where an arbitrary $x \in X$ is mapped to. For example,

$$\begin{array}{ccc} h : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^2 \end{array}$$

means that h is defined to be a mapping taking a real number to its square. Alternatively, one may define it in a single line as $h : \mathbb{R} \rightarrow \mathbb{R}$, $h(x) := x^2$ (where it is implied that $x \in \mathbb{R}$). Even more succinctly, one may denote

$$h(x) := x^2 \quad \forall x \in \mathbb{R}.$$

- Given three sets X, Y, Z and two maps $f : X \rightarrow Y$, $g : Y \rightarrow Z$, one defines the *composition map* $g \circ f : X \rightarrow Z$ as

$$(g \circ f)(x) := g(f(x)), \quad \forall x \in X.$$

- Consider two arbitrary sets X, Y and a function $f : X \rightarrow Y$.

- f is said to be *injective* (or “1-to-1”) when each element of X is mapped to a different element of Y , i.e., when

$$f(x) \neq f(y) \quad \forall x, y \in X : x \neq y$$

(read: $f(x)$ is different from $f(y)$ for all elements x, y in X such that x is different from y).

- f is said to be *surjective* (or “onto”) when $\text{range}(f) = Y$, i.e., when every element of the codomain Y has a pre-image. In symbols, when

$$\forall y \in Y \quad \exists x \in X : f(x) = y.$$

- f is said to be *bijective* (or a bijection) when it is both, injective and surjective. In symbols, when

$$\forall y \in Y \quad \exists! x \in X : f(x) = y.$$

- f is said to be *invertible* when there exists a function $g : Y \rightarrow X$ such that

$$g(f(x)) = x \quad \forall x \in X \quad \text{and} \quad f(g(y)) = y \quad \forall y \in Y.$$

If it exists, such a function (called *inverse function*) can be proven to be unique and one denotes $f^{-1} := g$.

Lemma 1. Given sets X, Y and $f : X \rightarrow Y$,

$$f \text{ is invertible} \iff f \text{ is bijective.} \quad \blacklozenge$$

Proof: Exercise. $\square$

Lemma 2. Given sets X, Y, Z and $f : X \rightarrow Y$, $g : Y \rightarrow Z$,

- If f, g are both injective, then $g \circ f$ is injective.
- If f, g are both surjective, then $g \circ f$ is surjective.
- If f, g are both invertible, then $g \circ f$ is invertible. Its inverse $(g \circ f)^{-1}$ equals $f^{-1} \circ g^{-1}$. $\blacklozenge$

Proof: Exercise. $\square$

- There is a very important trivial fact you should always remember. Given sets X, Y and $f : X \rightarrow Y$, consider the same function but now as a mapping to the set $\text{range}(f)$, let us call it $\tilde{f} : X \rightarrow \text{range}(f)$, $\tilde{f}(x) := f(x)$. It is trivially surjective. Thus, if f is injective, $\tilde{f}$ is bijective. In that sense, an injective map is essentially an invertible map—we need to consider it as a mapping to the image set.

1.4 Operations Between Sets

- The synonymous symbols $\subset$ and $\subseteq$ are used to denote that a set is a *subset* of another set (which includes the case that the two sets are actually the same set). More precisely, given two sets X, Y , we denote $X \subseteq Y$ when it holds that

$$(x \in X) \implies (x \in Y).$$

The symbol $\subsetneq$ on the other hand, is used to denote that a set is a subset of another set but the two sets are not equal. In precise terms, $X \subsetneq Y$ when

$$(X \subseteq Y) \wedge (X \neq Y).$$

One says that X is a *proper subset* of Y .

- The symbols “ $\cup$ ” and “ $\cap$ ” represent respectively, the *union* and *intersection* operations for sets. They are defined as follows. Given arbitrary sets X, Y ,

$$X \cup Y := \{z \mid z \in X \text{ or } z \in Y\}$$

$$X \cap Y := \{z \mid z \in X \text{ and } z \in Y\}.$$

- The symbol “ $\emptyset$ ” is the empty set $\{\}$, namely, the set with no elements.
- Given two sets X, Y , we saw they are *disjoint* when they have no element in common. This is exactly the same as saying that

$$X \cap Y = \emptyset.$$

It is conventional to denote the union of two disjoint sets with either a “ $\sqcup$ ” (or a “ $\uplus$ ”) instead of a “ $\cup$ ”, but this is not a must.

- Given a non-empty set I and a fixed set X_j per $j \in I$, one defines the *union of all X_j with $j \in I$* as

$$\bigcup_{j \in I} X_j := \{z \mid z \in X_j \text{ for some } j \in I\} = \{z \mid \exists j \in I : z \in X_j\}.$$

Similarly, the *intersection of all X_j with $j \in I$* is

$$\bigcap_{j \in I} X_j := \{z \mid z \in X_j \forall j \in I\}.$$

Then $X \cup Y$ and $X \cap Y$ from above are their particular cases when $I = \{1, 2\}$, $X_1 := X$ and $X_2 := Y$.

- The symbol “ $\setminus$ ” represents *set subtraction*. Namely, $Y \setminus X$ denotes the set of elements of Y that are not elements of X ,

$$Y \setminus X := \{y \in Y : y \notin X\}.$$

If $X \subseteq Y$, one calls $Y \setminus X$ the *complement of X (in Y)* and (if it is clear from the context which “ambient set” Y we mean, one denotes it also by X^c).

1.5 The Power Set Map

- A set of sets is usually called a “*family*” of sets. Given a set X , we denote by $\mathcal{P}(X)$ the family of all its subsets, i.e.,

$$\mathcal{P}(X) := \{A \subseteq X\}.$$

It is called the *power set of X* . Note that by definition, $\emptyset \in \mathcal{P}(X)$ for any set X .

- An “*abuse*” of notation, is the act of using the same symbol to denote two different objects. It is obviously not an advisable practice, but sometimes it is done to avoid overloading the notation. Let us see an example.

- Given arbitrary sets X, Y each $f : X \rightarrow Y$ comes with an associated function

$$\begin{array}{ccc} f_{set} : \mathcal{P}(X) & \longrightarrow & \mathcal{P}(Y) \\ A & \longmapsto & \{f(x) : x \in A\} \end{array} ,$$

called *power set map*. In words, f_{set} eats $A \subseteq X$ and applies f to each of its elements. For example, $f_{set}(X) = \text{range}(f)$. Now, if one wrongly wrote $f(A)$ for an $A \subseteq X$ (“wrongly” because f eats elements of X one by one instead of not many at once), the obvious interpretation would be that the writer meant $f_{set}(A)$ (to apply f to every element of A). This is why, it is a common abuse of notation to write f when we mean f_{set} . The argument we put will tell us whether we really mean f or f_{set} .

- Now, recall that for $f : X \rightarrow Y$, in general the inverse map $f^{-1} : Y \rightarrow X$ is not definable, because a same $y \in Y$ may have multiple pre-images (or even none!). However, one can perfectly define the associated set mapping

$$\begin{array}{ccc} f^{rev} : \mathcal{P}(Y) & \longrightarrow & \mathcal{P}(X) \\ B & \longmapsto & \{x \in X : f(x) \in B\} \end{array} ,$$

which eats $B \subseteq Y$, takes all the pre-images of each $y \in B$ and spits out the set they form in X . We call it the *reverse power set map*. If f was invertible, with inverse f^{-1} , one would have that $f^{rev}(B) = f^{-1}(B)$ for all $B \subseteq Y$. This is why, in a savage abuse of notation, one writes f^{-1} in place of f^{rev} , even in the cases when f has no inverse. The ambiguity disappears when we analyze the type of object fed to f^{-1} .

Note that f^{rev} is not either the inverse of f_{set} ! It is the “reverse” function heuristically, but not the inverse: whenever there is a $y_0 \in Y$ with no pre-image, then f_{set} does not reach every $B \subseteq Y$. For instance, $\{y_0\}$ is not in the image of f_{set} , so it is not surjective and thus f_{set} is not invertible. Likewise, $f^{rev}(\{y_0\}) = \emptyset$, so, f^{rev} is not injective because it is also the case that $f^{rev}(\emptyset) = \emptyset$. Thus, f^{rev} cannot be inverted either. They will be inverses of each other exactly when $f : X \rightarrow Y$ is bijective.

- f^{prev} is a very privileged map, because it commutes with all the relevant set theoretic operators (although f_{set} does not in general):

Proposition 1. Consider non-empty sets X, Y, I , a subset $A_j \subseteq X$, $B_j \subseteq Y$ per $j \in I$ and $A, C \subseteq X$, $B, D \subseteq Y$. For every function $f : X \rightarrow Y$, the following hold true:

$$\begin{aligned} (i) \ f\left(\bigcup_{j \in I} A_j\right) &= \bigcup_{j \in I} f(A_j), & (ii) \ f\left(\bigcap_{j \in I} A_j\right) &\subseteq \bigcap_{j \in I} f(A_j), & (iii) \ f(A) \setminus f(C) &\subseteq f(A \setminus C), \\ (iv) \ f^{-1}\left(\bigcup_{j \in I} B_j\right) &= \bigcup_{j \in I} f^{-1}(B_j), & (v) \ f^{-1}\left(\bigcap_{j \in I} B_j\right) &= \bigcap_{j \in I} f^{-1}(B_j), & (vi) \ f^{-1}(B) \setminus f^{-1}(D) &= f^{-1}(B \setminus D), \\ (vii) \ f(A) \cap B &= f\left(A \cap f^{-1}(B)\right). \end{aligned}$$

Items (ii) and (iii) hold with an equality if f is injective. ♦

1.6 The Cartesian Product

We know that the sets $\{1, 2, 3\}$ and $\{3, 1, 2\}$ are the same set because the position in which we write the elements has no significance. There is no “first, second and third slot of the set”. In the same way, the sets $\{1, 1, 2, 3\}$ and $\{1, 2, 3\}$ are identical because there is no structure to distinguish the “first copy of 1” from the “second copy of 1”: since the position where we write the element is immaterial, we cannot say something like “this is the 1 in the first slot of the set and this is the 1 in the second slot of the set”. Instead of “slots”, one can also think in terms of “labels”. In $\{1, 1, 2, 3\}$, we could label each element by their position, such that we could put the label “first” 1 and “second” 1. Such a labeling however, would be an extra structure for a set. This “label”/“slot” structure is exactly what a *Cartesian product* embodies and one uses the word “*index*” to mean “label” or “slot” in the sense we suggested.

Definition 1. Given a set I and a set X_j per $j \in I$, their *Cartesian product* is the set:

$$\prod_{j \in I} X_j := \left\{ x : I \rightarrow \bigcup_{j \in I} X_j \mid x(j) \in X_j \ \forall j \in I \right\}. \quad (1.2)$$

That is, it is the set of mappings x assigning to each $j \in I$ an element $x(j) \in X_j$. As an alternative notation for the mapping x we use

$$x_j := x(j) \quad \forall j \in I \quad \text{and} \quad (x_j)_{j \in I} := x.$$

We call $(x_j)_{j \in I}$ a *tuple*, I an *index set* and for each fixed $j \in I$, x_j is called the *j -th element or the element of index j* (in x). We adopt the following conventions:

- (i) If all X_j equal the same set X , we denote “ $(x_j)_{j \in I} \subseteq X$ ” to mean “ $(x_j)_{j \in I} \in \prod_{j \in I} X$ ”.
- (ii) If $I = \{1, \dots, n\}$ for some $n \in \mathbb{N}$, we use the equivalent symbols:

$$X_1 \times X_2 \times \dots \times X_n := \prod_{j=1}^n X_j := \prod_{j \in \{1, \dots, n\}} X_j,$$

$$(x_1, \dots, x_n) := (x_j)_{j=1}^n := (x_j)_{j \in \{1, \dots, n\}}.$$

If all X_j equal the same set X , one writes $X^n := \prod_{j=1}^n X$.

- (iii) If $I = \mathbb{N}$, we use the equivalent symbols:

$$X_1 \times X_2 \times \dots := \prod_{j=1}^{\infty} X_j := \prod_{j \in \mathbb{N}} X_j,$$

$$(x_1, \dots) := (x_j)_{j=1}^{\infty} := (x_j)_{j \in \{1, 2, \dots\}}. \quad \blacklozenge$$

Note that a tuple is the precise notion of “a set where the slot matters”. Despite I itself need not have any order whatsoever, when we query x for its j -th entry/slot, we get unambiguously x_j . This is why, the tuples $(1, 1, 2, 3)$, $(1, 2, 3)$, $(3, 1, 2)$ are all different from each other.

- As an example of the use of tuples, given sets X, Y and a function $f : X \rightarrow Y$ one defines the *graph of f* as the set:

$$\text{graph}(f) := \left\{ (x, f(x)) \in X \times Y \mid x \in X \right\}.$$

In fact, note that if we were given $\text{graph}(f)$ alone, we would still know everything about f . For each $x \in X$, we would only need to look for the element $(x, y) \in \text{graph}(f)$, then we would know that y is $f(x)$. In fact this is another way of giving functions, as subsets of $X \times Y$ where there is one and only one element per $x \in X$.

1.7 Relation

- This takes us to the notion of *relation*. Given an index set I and a set X_j per $j \in I$, each subset $R \subseteq \prod_{j \in I} X_j$ is called a *relation*. Why? Because it provides a particular relation between elements of each X_j : for each choice $x_j \in X_j$, $j \in I$, we can tell if they are related or not by checking if $(x_j)_{j \in I} \in R$. When $I = \{1, 2\}$, this is called a *binary relation* and one denotes $x_1 R x_2$ when $(x_1, x_2) \in R$, saying that *according to the relation R , x_1 and x_2 are related*. We denote $x_1 \bar{R} x_2$ when $(x_1, x_2) \notin R$ and say that *according to the relation R , x_1 and x_2 are not related*. The graph of a function $f : X \rightarrow Y$ (and hence, quite literally, the function itself) is a special type of binary relation, one in which every element of $x \in X$ is related to one and only one element of Y —but not vice-versa unless f is a bijection.

- To define a binary relation it is standard to use the following notation: for $x_1 \in X_1$, $x_2 \in X_2$,

$$x_1 R x_2 \quad :\Longleftrightarrow \quad (\text{some condition for } x_1, x_2),$$

and it means

$$R := \left\{ (x_1, x_2) \in X_1 \times X_2 \mid \text{some condition for } x_1, x_2 \right\}.$$

- Given a single set X , the binary relations of elements in X with each other, namely, the subsets of $X \times X$ are called *homogeneous relations*. There is a very important type there that we denote with symbols like $\approx$ and $\sim$ instead of a letter like R , because it is a “same-ness” or “identification” relation:

Definition 2. Given a set X , an *equivalence relation* is an homogeneous relation $\approx \subseteq X \times X$ satisfying all the following:

- (i) It is *reflexive*, i.e., every element of X is related to itself: $x \approx x \quad \forall x \in X$.
- (ii) It is *symmetric*, i.e., if an element $x \in X$ is related to $y \in Y$, then, y is also related to x :

$$(x \approx y) \implies (y \approx x) \quad \forall x, y \in X.$$

- (iii) It is *transitive*, i.e., when $x \in X$ is related to $y \in X$, who is related to $z \in X$, then x is also related to z :

$$(x \approx y) \wedge (y \approx z) \implies (x \approx z) \quad \forall x, y, z \in X.$$

We say that $\approx$ *identifies* elements in X with some other elements in X . For each $x \in X$, we define the *equivalence class of x* as

$$[x] := \left\{ y \in X \mid y \approx x \right\}.$$

We call any $y \in [x]$ a *representative of the class $[x]$* . ♦

For this to be a notion of identification or same-ness, it should happen that the class of elements related to x and those related to y are disjoint if $x \not\approx y$. That is exactly what the three conditions give us.

Definition 3. Given a set X , a *partition of X* is a family Γ of non-empty subsets of X (i.e., $\Gamma \subseteq \mathcal{P}(X) \setminus \{\emptyset\}$) that

- (i) *cover X* , i.e., $\bigcup_{C \in \Gamma} C = X$,
- (ii) *are pairwise disjoint*, i.e., $A \cap B = \emptyset \quad \forall A, B \in \Gamma : A \neq B$. ♦

Then, indeed:

Lemma 3. Given a set X and an equivalence relation $\approx$ on X , the family of equivalence classes

$$\left\{ [x] : x \in X \right\}$$

is a partition of X . ♦

Proof: Exercise. $\square$

What if some other conditions on a relation also gave us that the classes give a partition? Then, that relation would be an equivalence relation because of the following:

Lemma 4. Given a set X and a partition $\Gamma \subseteq \mathcal{P}(X)$, consider the relation R given by the membership to the same $A \in \Gamma$, i.e., $x, y \in X$,

$$x R y \iff \left(\exists A \in \Gamma : x, y \in A \right),$$

or equivalently,

$$R := \bigcup_{A \in \Gamma} \left\{ (a, b) \in X \times X \mid a, b \in A \right\}.$$

Then, R is an equivalence relation. ♦

Proof: Exercise. $\square$

1.8 Order

Consider a set X . There is no order in it, meaning there is no given way to say whether an $x \in X$ is smaller or bigger than y . But certainly, to give such a structure is to give a relation. Such relations are denoted with the symbol $\leq$.

Definition 4. Given a set X , a *paritla order* is an homogeneous relation $\leq \subseteq X \times X$ satisfying all the following:

- (i) It is *reflexive*: $x \leq x \quad \forall x \in X$.
- (ii) It is *anti-symmetric*, i.e., if an element is related by $\leq$ to than another and vice-versa, then they are the same element:

$$(x \leq y) \wedge (y \leq x) \implies (x = y) \quad \forall x, y \in X.$$

- (iii) It is *transitive*: $(x \leq y) \wedge (y \leq z) \implies (x \leq z) \quad \forall x, y, z \in X$.

The pair $(X, \leq)$ is called a *partially ordered set*. If $x \leq y$ we say that x is *smaller than* y . When $x \leq y$ and $x \neq y$, we denote $x < y$. ♦

This is the weakest notion that fits with our intuitions of what having an “order”, or being able to compare two elements means. But note that none that in that definition nothing ensures that we can compare any two pair of elements. Thus the name “partial” in “partial order”.

Definition 5. Given a set X , a partial order that “allows to compare every pair of elements”, i.e., such that $\forall x, y \in X$

$$x \leq y \quad \text{or} \quad y \leq x,$$

is called a *total order* or simply an *order*. ♦

Example 1. All, $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ are totally ordered sets when equipped with the usual symbol $\leq$ as a relation. ♦

This now allows to have a generalized notion of maximum and minium:

Definition 6. Let $(X, \leq)$ be a partially ordered set and let $A \subseteq X$. An element $M \in A$ is called a *maximum* of A when

$$a \leq M \quad \forall a \in A.$$

An element $m \in A$ is called a *minimum* of A when

$$m \leq a \quad \forall a \in A. \quad \blacklozenge$$

Lemma 5. Given a partially ordered set $(X, \leq)$, each subset $A \subseteq X$ has at most one maximum and at most one minimum. ♦

Proof: Suppose that $M_1 \in A$ and $M_2 \in A$ are both maxima of A . By definition of maximum,

$$M_1 \leq M_2 \quad \text{and} \quad M_2 \leq M_1.$$

But then, by anti-symmetry of $\leq$, $M_1 = M_2$. The proof for the minimum is analogous. o.e.δ.

Thus, we refer to *the* maximum and *the* minimum of a set A and if they exist, we can safely use the symbols $\max A$ and $\min A$ to refer to them.

Note that the minimum and maximum of a set may not exist. For (non-empty) subsets of $\mathbb{N}$, you will prove in the exercises there is always a minimum. But for example, the interval $(0, 1) \subseteq \mathbb{R}$ has no minimum nor maximum. And yet, you might have a feeling that 0 and 1 are somewhat related. That is because you are thinking on *bounds*.

Definition 7. Given a partially ordered set $(X, \leq)$ and $A \subseteq X$, an element $u \in X$ (note, u is not necessarily in A) is called an *upper bound* of A when

$$a \leq u \quad \forall a \in A.$$

Let us denote the set of all upper bounds of A as

$$\mathfrak{B}_{up}(A) := \left\{ u \in X \mid a \leq u \quad \forall a \in A \right\}.$$

We say that A is *bounded above* when $\mathfrak{B}_{up}(A) \neq \emptyset$. Similarly, an element $\ell \in X$ (note, ℓ is not necessarily in A) is called a *lower bound* of A when

$$\ell \leq a \quad \forall a \in A.$$

Let us denote the set of all lower bounds of A as

$$\mathfrak{B}_{low}(A) := \left\{ \ell \in X \mid \ell \leq a \quad \forall a \in A \right\}.$$

We say that A is *bounded below* when $\mathfrak{B}_{low}(A) \neq \emptyset$. ♦

This allows us to formulate the substitute for maximum and minimum we were thinking about for $(0, 1) \subseteq \mathbb{R}$:

Definition 8. Given a partially ordered set $(X, \leq)$ and $A \subseteq X$, an element $s \in X$ is called the *supremum* of A when it is the *minimum upper bound*, $s = \min \mathfrak{B}_{up}(A)$, i.e.,

$$s \in \mathfrak{B}_{up}(A) \quad \text{and} \quad \left(s \leq u \quad \forall u \in \mathfrak{B}_{up}(A) \right).$$

If so, we write $\sup A := s$. Similarly, an element $t \in X$ is called the *infimum* of A when it is the *maximum lower bound*, $t = \max \mathfrak{B}_{low}(A)$, i.e.,

$$t \in \mathfrak{B}_{low}(A) \quad \text{and} \quad \left(\ell \leq t \quad \forall \ell \in \mathfrak{B}_{low}(A) \right).$$

If so, we write $\inf A := t$. ♦

Note that for a general partial order, the supremum and infimum of a set might not exist. Not even for a total order. For instance, $\mathbb{R}$ itself has no supremum nor infimum —after all, $\mathfrak{B}_{up}(A) = \emptyset = \mathfrak{B}_{low}(A)$. In fact, even if a set is bounded, such that the set of upper or lower bounds is non-empty, it may happen there is no supremum or infimum.

Example 2. Consider the totally ordered set $(\mathbb{Q}, \leq)$ and the subset

$$A := \left\{ q \in \mathbb{Q} \mid 0 < q \text{ and } q^2 < 2 \right\}.$$

A has no minimum nor maximum, but it is both a lower and upper bounded set (say, lower bounded by 0 and upper bounded by 2). Moreover, 0 is clearly its infimum. However, it has no supremum: it should be $\sqrt{2}$, but this is not in $\mathbb{Q}$. ♦

One thing is clear though: supremum and infimum are a generalization of the notion of maximum and minimum of a set. Trivially, if A has a maximum, it is also its supremum and if it has a minimum it is also its infimum. We say that the supremum and infimum of A are *realized* in such a case.

Now, you do not need to worry really, because in most practical cases you will use the supremum or infimum for subsets of $\mathbb{R}$ (say, not considered as subsets of $\mathbb{Q}$). And for them, there is the following nice property (so-called *completeness condition* of $\mathbb{R}$):

Theorem 1. For each non-empty $A \subseteq \mathbb{R}$,

- (i) if A has an upper bound, then it has a supremum.

(ii) if A has a lower bound, then it has an infimum. $\diamond$

Lastly, consider not $\mathbb{R}$, but the extended real line

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\},$$

equipped with the obvious extension of the total order of $\mathbb{R}$ by which

$$-\infty \leq +\infty \quad \text{and} \quad -\infty \leq x \leq +\infty \quad \forall x \in \mathbb{R}.$$

Then, all subset $A \subseteq \mathbb{R}$ have an upper bound and a lower bound in $\overline{\mathbb{R}}$ (possibly $\pm\infty$). As such:

Corollary 1. *Every non-empty $A \subseteq \overline{\mathbb{R}}$ has both a supremum and an infimum.* $\diamond$

1.9 Cardinality

Finally, all this allows us to talk about types of sets with different sizes, allowing us to make rigorous sense of what we mean by *different types of infinities*.

Definition 9. Given two sets X, Y , they are said to have the same *cardinality* (informally, “the same number of elements”) when there exists a bijection $f : X \rightarrow Y$ (i.e., when we can put the elements of one set in one to one correspondence with all the elements of the other set). If so, we write $|X| = |Y|$, otherwise, we write $|X| \neq |Y|$.

- (i) We say that X is a *finite set* when it is empty or $\exists n \in \mathbb{N}$ such that $|X| = |\{1, 2, \dots, n\}|$ (i.e., when there exists a bijection $f : \{1, 2, \dots, n\} \rightarrow X$ that labels the elements of X with numbers from 1 to n , informally, f counts the elements of X as the 1st, ..., n -th). We write $|X| = n$ and say that X has n elements, or *finitely many* elements (in the empty set’s case use $|\emptyset| := 0$).
- (ii) We say that X is an *infinite set* or simply *infinite*, when it is not a finite set (i.e., you cannot count them with finitely many indices). For instance, $\mathbb{N}$ is clearly an infinite set. In particular:
 - (a) If $|X| = |\mathbb{N}|$ we say that X is a *countably infinite set* or simply *countable*. (After all, this means there exists a way to enumerate all the elements of X as the 1st, 2nd, 3rd... and still cover all of X .)
 - (b) If X is infinite and is not countable ($|X| \neq |\mathbb{N}|$), we say that X is *uncountable*. $\diamond$

- The following lemma sounds trivial, but its proof requires the order theory we formulated above.

Lemma 6. Every infinite subset of $\mathbb{N}$ is countable. $\diamond$

Proof: Let $A \subseteq \mathbb{N}$ be infinite. By Lemma ??, every non-empty subset of $\mathbb{N}$ has a minimum and so, we can define the map $f : \mathbb{N} \rightarrow A$,

$$f(1) := \min A, \quad f(n+1) := \min \left(A \setminus \{f(1), \dots, f(n)\} \right) \quad \forall n \in \mathbb{N}.$$

We can do this for all $n \in \mathbb{N}$ (say, A always has more elements) because otherwise A would be a finite set. Essentially by definition, f is injective. Moreover, it is surjective because given $a \in A$, there are at most a many elements in A that are smaller than a . Thus, for some $n \in \{1, \dots, a\}$, $f(n) = a$. But then, f is a bijection and hence, $|A| = |\mathbb{N}|$. **o.e.d.**

- Now, $\mathbb{Q}$ is incredibly big as an infinite set, after all, between every two points in the real line there are always infinitely many rational numbers, no matter how small the interval. And still, as incredible as it may sound, they can be counted, namely, put into a one to one correspondence with the natural numbers.

Lemma 7. The set of rational numbers $\mathbb{Q}$ is countable. $\diamond$

Proof: $\mathbb{N} \subseteq \mathbb{Q}$, so $\mathbb{Q}$ is infinite. Now, consider the infinite matrix

$$\left(\frac{p}{q}\right)_{(p,q) \in \mathbb{N}^2} = \begin{pmatrix} \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \cdots \\ \frac{1}{2} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \cdots \\ \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Enumerate its entries along successive diagonals:

$$r := \left(\frac{1}{1}, \quad \frac{1}{2}, \frac{2}{1}, \quad \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \quad \cdots \right).$$

Note that the fractions of the n -th diagonal have that the sum of the numerator and denominator is $n+1$. In particular, given an arbitrary positive rational $\frac{p}{q}$, it occurs on the diagonal $p+q-1$. Therefore, the enumeration gives a sequence r containing every positive rational number. There are repetitions like $\frac{1}{2} = \frac{2}{4}$, but we can omit them keeping only the first occurrence of each rational number. The result is a tuple $\tilde{r} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \dots)$. Now, defining the tuple

$$R := \left(0, \tilde{r}_1, -\tilde{r}_1, \tilde{r}_2, -\tilde{r}_2, \tilde{r}_3, -\tilde{r}_3, \dots \right),$$

we have also included the negative rationals (only once each). Thus, the tuple $R : \mathbb{N} \rightarrow \mathbb{Q}$ is injective and surjective by construction, proving that $|\mathbb{Q}| = |\mathbb{N}|$. **o.e.d.**

This a priori “incredible” fact (now credible after seeing the proof) suggests that maybe there is no uncountable set at all. However, there are. Here is an example:

Lemma 8. The set $\prod_{n \in \mathbb{N}} \{0, 1\}$ (i.e., the tuples $(s_1, s_2, s_3, \dots)$ with $s_j \in \{0, 1\}$), is uncountable. $\blacklozenge$

Proof: First, $\prod_{n \in \mathbb{N}} \{0, 1\}$ is infinite. For any fixed $n \in \mathbb{N}$, there are at least $n+1$ tuples:

$$(0, 0, \dots), (1, 0, 0, \dots), (1, 1, 0, \dots), \dots, \underbrace{(1, 1, \dots, 1, 0, 0, \dots)}_n.$$

Hence, $\prod_{n \in \mathbb{N}} \{0, 1\}$ cannot have n elements.

Now, we prove that $|\prod_{n \in \mathbb{N}} \{0, 1\}| \neq |\mathbb{N}|$ by *reductio ad absurdum*. (It is the famous *Cantor's diagonal argument*.) Assume that $|\prod_{n \in \mathbb{N}} \{0, 1\}| = |\mathbb{N}|$. Then, there is a bijection $F : \mathbb{N} \rightarrow \prod_{n \in \mathbb{N}} \{0, 1\}$ (so, for each $n \in \mathbb{N}$, $F(n)$ is a different string $(F(n)_1, F(n)_2, \dots)$ and we claim that these are all possible tuples of 1's and 0's.) Now, consider the tuple $S = (S_1, S_2, \dots)$ defined as

$$S_n := \begin{cases} 0 & \text{if } F(n)_n = 1 \\ 1 & \text{if } F(n)_n = 0 \end{cases}, \quad \forall n \in \mathbb{N}.$$

In words, we take $F(n)$ (the n -th tuple according to F 's enumeration), and within that tuple the 1 or 0 in the n -th position. Then, we define the n -th entry of S as the opposite number. If so, S is a tuple that is different from $F(n)$ because we made sure that their n -th entry is different. But then, S is different to all $F(n)$, $n \in \mathbb{N}$. Therefore, S is not in the range of the map F . But S is by definition, an element of $\prod_{n \in \mathbb{N}} \{0, 1\}$ and so, it should be in the image of F because it is surjective. Absurd! Therefore, it must be that $|\prod_{n \in \mathbb{N}} \{0, 1\}| \neq |\mathbb{N}|$. **o.e.d.**

In fact, $\mathbb{R}$ is uncountable. To prove it, we notice the following first:

Lemma 9. Given a set Ω , the relation between its subsets defined by possessing the same cardinality, namely, $X, Y \in \mathcal{P}(\Omega)$,

$$X \sim Y \quad :\Longleftrightarrow \quad |X| = |Y|,$$

is an *equivalence relation*. We denote an equivalence class $[X]$ a *cardinal*, and abusing notation, we still

denote it by $|X|$. ♦

Lemma 10. $\mathbb{R}$ is an uncountable set. ♦

Proof: $\mathbb{N} \subseteq \mathbb{R}$, so $\mathbb{R}$ is infinite. We prove that $|\mathbb{R}| \neq |\mathbb{N}|$ by contradiction. Assume that $|\mathbb{R}| = |\mathbb{N}|$. Then, there would exist a bijection $G : \mathbb{R} \rightarrow \mathbb{N}$. Now, define the function $H : \prod_{n \in \mathbb{N}} \{0, 1\} \rightarrow \mathbb{R}$, such that $H((s_1, s_2, \dots)) := 0.s_1s_2s_3\dots$, which sends the tuple of 0's and 1's to the real in the interval $[0, 1]$ with those decimal digits. It is clearly injective. Thus, $F := G \circ H$ is an injective map and considered as a mapping to its image, $\tilde{F} : \prod_{n \in \mathbb{N}} \{0, 1\} \rightarrow \text{range}(F) \subseteq \mathbb{N}$, it is a bijection. Hence, $|\prod_{n \in \mathbb{N}} \{0, 1\}| = |\text{range}(F)|$. Since $\text{range}(F) \subseteq \mathbb{N}$ is infinite (because $\prod_{n \in \mathbb{N}} \{0, 1\}$ is so), by Lemma 6, it is countable, i.e., $|\text{range}(F)| = |\mathbb{N}|$. But then, by transitivity proven in Lemma 9, $|\prod_{n \in \mathbb{N}} \{0, 1\}| = |\mathbb{N}|$, which is absurd by Lemma 8! o.e.δ.

This suggests there is a way to say when an infinity is bigger than another one. Consider two sets X, Y . If there exists an injective map $f : X \rightarrow Y$, that means we can place every element of X “inside” Y and thus, the number of elements of X is smaller or equal to those of Y . With this intuition in mind we define:

Definition 10. Given a set Ω , we define the *cardinality comparison* relation $\leq$ on the equivalence classes of cardinals on $\mathcal{P}(\Omega)$: given $X, Y \in \mathcal{P}(\Omega)$ and their cardinals $|X|, |Y|$ (recall, we mean $[[X]], [[Y]]$),

$$|X| \leq |Y| \quad :\Longleftrightarrow \quad \left(\exists \text{ injective } f : X \rightarrow Y \right).$$

We write $|X| < |Y|$ when $|X| \leq |Y|$ and $|X| \neq |Y|$. ♦

It has the following properties:

Lemma 11. For all sets $X, Y, Z \subseteq \Omega$:

1. $\leq$ is reflexive: $|X| \leq |X|$,
2. $\leq$ is transitive: if $|X| \leq |Y|$ and $|Y| \leq |Z|$, then $|X| \leq |Z|$,
3. if $X \subseteq Y$, then $|X| \leq |Y|$. ♦

Proof: Exercise. $\square$

Consequently, since $\mathbb{N} \subseteq \mathbb{R}$, $|\mathbb{N}| \leq |\mathbb{R}|$. But we just proved that $|\mathbb{N}| \neq |\mathbb{R}|$, so, $|\mathbb{N}| < |\mathbb{R}|$. $\mathbb{R}$ is a strictly bigger infinite set than $\mathbb{N}$.

• This Lemma tells us that $\leq$ is almost a partial order for cardinals (i.e., for classes of cardinalities). And it is indeed, but the anti-symmetry condition is non-trivial. It follows from the Cantor-Schröder-Bernstein theorem:

Theorem 2 (Cantor–Schröder–Bernstein). Given sets X, Y , if there exist injective maps $f : X \rightarrow Y$ and $g : Y \rightarrow X$, then, there exists a bijection $h : X \rightarrow Y$. That is, cardinal $\leq$ is anti-symmetric. ♦

Corollary 2. For any set Ω , the set of its cardinals equipped with $\leq$ is a partially ordered set. ♦

Furthermore, one can prove (using the so-called axiom of choice) that it is a total order (so, we can compare the cardinals of every pair of sets!):

Theorem 3. For every pair of sets X, Y , it holds that $|X| \leq |Y|$ or $|Y| \leq |X|$. Thus, for any set Ω , its cardinals equipped with $\leq$ are a totally ordered set. ♦

Finally, one may think that all uncountable sets have the same cardinality, so that there are only two possible infinite cardinals: the countable and uncountable one. This is far from true. Cantor proved that:

Theorem 4. For every set X , $|X| < |\mathcal{P}(X)|$. ♦

Therefore, there are actually infinitely many infinite cardinals: $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < \dots$.

Exercise Sheet 1

1. Prove Lemmas 1 and 2.
2. Prove Lemmas 3 and 4.
3. Prove the following: (*Hint: Use a proof that uses both reductio ad absurdum and induction.*)

Lemma 12. If $A \subseteq \mathbb{N}$ is not empty, it has a minimum. ♦

4. Prove Lemmas 9 and 11.
5. Prove the following:

Lemma 13. Given a set X , $|\mathcal{P}(X)| = |\prod_{x \in X} \{0, 1\}|$. ♦

This tells that we can identify each subset of X with a choice of 0 or 1 per $x \in X$. Note that it is common to denote $2^X := \prod_{x \in X} \{0, 1\}$.