Part I

Sets, Relations, Functions

The material in this part is a reasonably complete introduction to basic naive set theory. Unless students can be assumed to have this background, it's probably advisable to start a course with a review of this material, at least the part on sets, functions, and relations. This should ensure that all students have the basic facility with mathematical notation required for any of the other logical sections. NB: This part does not cover induction directly.

The presentation here would benefit from additional examples, especially, "real life" examples of relations of interest to the audience.

It is planned to expand this part to cover naive set theory more extensively.

Chapter 1

Sets

1.1 Basics

sfr:set:bas:sec Sets are the most fundamental building blocks of mathematical objects. In fact, almost every mathematical object can be seen as a set of some kind. In logic, as in other parts of mathematics, sets and set-theoretical talk is ubiquitous. So it will be important to discuss what sets are, and introduce the notations necessary to talk about sets and operations on sets in a standard way. explanation

Definition 1.1 (Set)**.** A *set* is a collection of objects, considered independently of the way it is specified, of the order of the objects in the set, or of their multiplicity. The objects making up the set are called *elements* or *members* of the set. If a is an element of a set X, we write $a \in X$ (otherwise, $a \notin X$). The set which has no elements is called the *empty* set and denoted by the symbol $\emptyset$.

Example 1.2. Whenever you have a bunch of objects, you can collect them together in a set. The set of Richard's siblings, for instance, is a set that contains one person, and we could write it as $S = \{\text{Ruth}\}$. In general, when we have some objects $a_1, \ldots, a_n$, then the set consisting of exactly those objects is written $\{a_1, \ldots, a_n\}$. Frequently we'll specify a set by some property that its elements share—as we just did, for instance, by specifying S as the set of Richard's siblings. We'll use the following shorthand notation for that: $\{x : \ldots x \ldots\}$, where the $\ldots x \ldots$ stands for the property that x has to have in order to be counted among the elements of the set. In our example, we could have specified S also as

$$S = \{x : x \text{ is a sibling of Richard}\}.$$

When we say that sets are independent of the way they are specified, we mean that the elements of a set are all that matters. For instance, it so happens explanation

that

$$\{\text{Nicole}, \text{Jacob}\},$$
$$\{x : \text{is a niece or nephew of Richard}\}, \text{ and}$$
$$\{x : \text{is a child of Ruth}\}$$

are three ways of specifying one and the same set.

Saying that sets are considered independently of the order of their elements and their multiplicity is a fancy way of saying that

$$\{\text{Nicole}, \text{Jacob}\} \text{ and}$$
$$\{\text{Jacob}, \text{Nicole}\}$$

are two ways of specifying the same set; and that

$$\{\text{Nicole}, \text{Jacob}\} \text{ and}$$
$$\{\text{Jacob}, \text{Nicole}, \text{Nicole}\}$$

are also two ways of specifying the same set. In other words, all that matters is which elements a set has. The elements of a set are not ordered and each element occurs only once. When we *specify* or *describe* a set, elements may occur multiple times and in different orders, but any descriptions that only differ in the order of elements or in how many times elements are listed describes the same set.

Definition 1.3 (Extensionality)**.** If X and Y are sets, then X and Y are *identical*, $X = Y$, iff every element of X is also an element of Y, and vice versa.

explanation

Extensionality gives us a way for showing that sets are identical: to show that $X = Y$, show that whenever $x \in X$ then also $x \in Y$, and whenever $y \in Y$ then also $y \in X$.

Problem 1.1. Show that there is only one empty set, i.e., show that if X and Y are sets without members, then $X = Y$.

1.2 Some Important Sets

sfr:set:set:sec

Example 1.4. Mostly we'll be dealing with sets that have mathematical objects as members. You will remember the various sets of numbers: $\mathbb{N}$ is the set of *natural* numbers $\{0, 1, 2, 3, \dots\}$; $\mathbb{Z}$ the set of *integers*,

$$\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\};$$

$\mathbb{Q}$ the set of *rational* numbers ($\mathbb{Q} = \{z/n : z \in \mathbb{Z}, n \in \mathbb{N}, n \neq 0\}$); and $\mathbb{R}$ the set of *real* numbers. These are all *infinite* sets, that is, they each have

infinitely many elements. As it turns out, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ have the same number of elements, while $\mathbb{R}$ has a whole bunch more—$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ are "enumerable and infinite" whereas $\mathbb{R}$ is "non-enumerable".

We'll sometimes also use the set of positive integers $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$ and the set containing just the first two natural numbers $\mathbb{B} = \{0, 1\}$.

Example 1.5 (Strings)**.** Another interesting example is the set A^* of *finite strings* over an alphabet A: any finite sequence of elements of A is a string over A. We include the *empty string* Λ among the strings over A, for every alphabet A. For instance,

$$\mathbb{B}^* = \{\Lambda, 0, 1, 00, 01, 10, 11, \\ 000, 001, 010, 011, 100, 101, 110, 111, 0000, \dots\}.$$

If $x = x_1 \dots x_n \in A^*$is a string consisting of n "letters" from A, then we say *length* of the string is n and write $\operatorname{len}(x) = n$.

Example 1.6 (Infinite sequences)**.** For any set A we may also consider the set A^ω of infinite sequences of elements of A. An infinite sequence $a_1 a_2 a_3 a_4 \dots$ consists of a one-way infinite list of objects, each one of which is an element of A.

1.3 Subsets

sfr:set:sub:sec Sets are made up of their elements, and every element of a set is a part of that set. But there is also a sense that some of the elements of a set *taken together* are a "part of" that set. For instance, the number 2 is part of the set of integers, but the set of even numbers is also a part of the set of integers. It's important to keep those two senses of being part of a set separate. explanation

Definition 1.7 (Subset)**.** If every element of a set X is also an element of Y, then we say that X is a *subset* of Y, and write $X \subseteq Y$.

Example 1.8. First of all, every set is a subset of itself, and $\emptyset$ is a subset of every set. The set of even numbers is a subset of the set of natural numbers. Also, $\{a, b\} \subseteq \{a, b, c\}$.

But $\{a, b, e\}$ is not a subset of $\{a, b, c\}$.

Note that a set may contain other sets, not just as subsets but as elements! In particular, a set may happen to *both* be an element and a subset of another, e.g., $\{0\} \in \{0, \{0\}\}$ and also $\{0\} \subseteq \{0, \{0\}\}$. explanation

Extensionality gives a criterion of identity for sets: $X = Y$ iff every element of X is also an element of Y and vice versa. The definition of "subset" defines $X \subseteq Y$ precisely as the first half of this criterion: every element of X is also an element of Y. Of course the definition also applies if we switch X and Y: $Y \subseteq X$ iff every element of Y is also an element of X. And that, in turn, is exactly the "vice versa" part of extensionality. In other words, extensionality amounts to: $X = Y$ iff $X \subseteq Y$ and $Y \subseteq X$. explanation

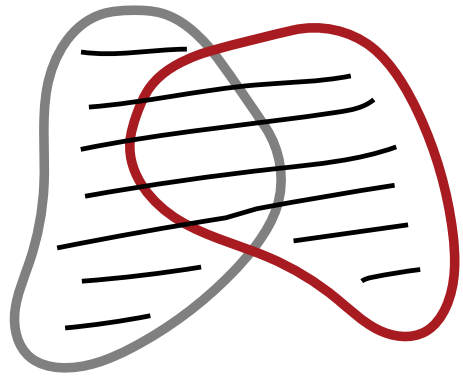


Figure 1.1: The union $X \cup Y$ of two sets is set of elements of X together with those of Y.

Definition 1.9 (Power Set)**.** The set consisting of all subsets of a set X is called the *power set of* X, written $\wp(X)$.

$$\wp(X) = \{Y : Y \subseteq X\}$$

Example 1.10. What are all the possible subsets of $\{a, b, c\}$? They are: $\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{a, b, c\}$. The set of all these subsets is $\wp(\{a, b, c\})$:

$$\wp(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

Problem 1.2. List all subsets of $\{a, b, c, d\}$.

Problem 1.3. Show that if X has n elements, then $\wp(X)$ has 2^n elements.

1.4 Unions and Intersections

explanation We can define new sets by abstraction, and the property used to define the new set can mention sets we've already defined. So for instance, if X and Y are sets, the set $\{x : x \in X \vee x \in Y\}$ defines a set which consists of all those objects which are elements of either X or Y, i.e., it's the set that combines the elements of X and Y. This operation on sets—combining them—is very useful and common, and so we give it a name and a symbol. sfr:set:uni:sec

Definition 1.11 (Union)**.** The *union* of two sets X and Y, written $X \cup Y$, is the set of all things which are elements of X, Y, or both.

$$X \cup Y = \{x : x \in X \vee x \in Y\}$$

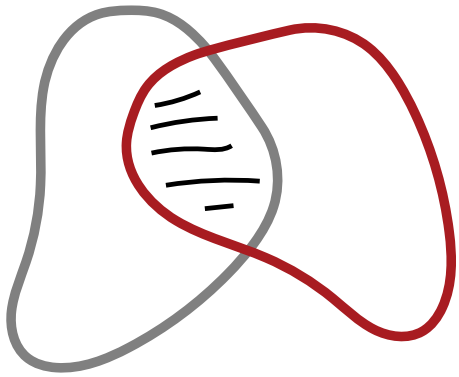


Figure 1.2: The intersection $X \cap Y$ of two sets is the set of elements they have in common.

Example 1.12. Since the multiplicity of elements doesn't matter, the union of two sets which have an element in common contains that element only once, e.g., $\{a, b, c\} \cup \{a, 0, 1\} = \{a, b, c, 0, 1\}$.

The union of a set and one of its subsets is just the bigger set: $\{a, b, c\} \cup \{a\} = \{a, b, c\}$.

The union of a set with the empty set is identical to the set: $\{a, b, c\} \cup \emptyset = \{a, b, c\}$.

Problem 1.4. Prove rigorously that if $X \subseteq Y$, then $X \cup Y = Y$.

The operation that forms the set of all elements that X and Y have in common is called their *intersection*. explanation

Definition 1.13 (Intersection)**.** The *intersection* of two sets X and Y, written $X \cap Y$, is the set of all things which are elements of both X and Y.

$$X \cap Y = \{x : x \in X \wedge x \in Y\}$$

Two sets are called *disjoint* if their intersection is empty. This means they have no elements in common.

Example 1.14. If two sets have no elements in common, their intersection is empty: $\{a, b, c\} \cap \{0, 1\} = \emptyset$.

If two sets do have elements in common, their intersection is the set of all those: $\{a, b, c\} \cap \{a, b, d\} = \{a, b\}$.

The intersection of a set with one of its subsets is just the smaller set: $\{a, b, c\} \cap \{a, b\} = \{a, b\}$.

The intersection of any set with the empty set is empty: $\{a, b, c\} \cap \emptyset = \emptyset$.

Problem 1.5. Prove rigorously that if $X \subseteq Y$, then $X \cap Y = X$.

explanation We can also form the union or intersection of more than two sets. An elegant way of dealing with this in general is the following: suppose you collect all the sets you want to form the union (or intersection) of into a single set. Then we can define the union of all our original sets as the set of all objects which belong to at least one element of the set, and the intersection as the set of all objects which belong to every element of the set.

Definition 1.15. If Z is a set of sets, then $\bigcup Z$ is the set of elements of elements of Z:

$$\bigcup Z = \{x : x \text{ belongs to an element of } Z\}, \text{ i.e.,}$$
$$\bigcup Z = \{x : \text{there is a } Y \in Z \text{ so that } x \in Y\}$$

Definition 1.16. If Z is a set of sets, then $\bigcap Z$ is the set of objects which all elements of Z have in common:

$$\bigcap Z = \{x : x \text{ belongs to every element of } Z\}, \text{ i.e.,}$$
$$\bigcap Z = \{x : \text{for all } Y \in Z, x \in Y\}$$

Example 1.17. Suppose $Z = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$. Then $\bigcup Z = \{a, b, d, e\}$ and $\bigcap Z = \{a\}$.

We could also do the same for a sequence of sets X_1, X_2, ...

$$\bigcup_i X_i = \{x : x \text{ belongs to one of the } X_i\}$$
$$\bigcap_i X_i = \{x : x \text{ belongs to every } X_i\}.$$

Definition 1.18 (Difference)**.** The *difference* $X \setminus Y$ is the set of all elements of X which are not also elements of Y, i.e.,

$$X \setminus Y = \{x : x \in X \text{ and } x \notin Y\}.$$

1.5 Pairs, Tuples, Cartesian Products

explanation Sets have no order to their elements. We just think of them as an unordered collection. So if we want to represent order, we use *ordered pairs* $\langle x, y\rangle$. In an unordered pair $\{x, y\}$, the order does not matter: $\{x, y\} = \{y, x\}$. In an ordered pair, it does: if $x \neq y$, then $\langle x, y\rangle \neq \langle y, x\rangle$. sfr:set:pai:sec

Sometimes we also want ordered sequences of more than two objects, e.g., *triples* $\langle x, y, z\rangle$, *quadruples* $\langle x, y, z, u\rangle$, and so on. In fact, we can think of triples as special ordered pairs, where the first element is itself an ordered pair: $\langle x, y, z\rangle$ is short for $\langle\langle x, y\rangle, z\rangle$. The same is true for quadruples: $\langle x, y, z, u\rangle$ is short for $\langle\langle\langle x, y\rangle, z\rangle, u\rangle$, and so on. In general, we talk of *ordered n-tuples* $\langle x_1, \ldots, x_n\rangle$.

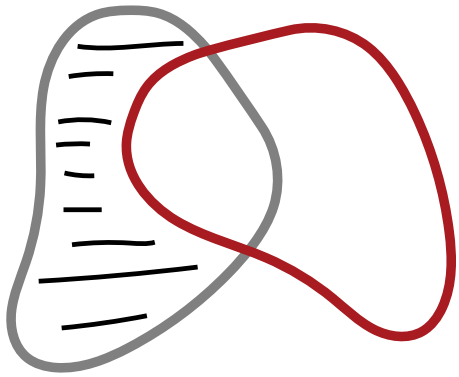


Figure 1.3: The difference $X \setminus Y$ of two sets is the set of those elements of X which are not also elements of Y.

Definition 1.19 (Cartesian product)**.** Given sets X and Y, their *Cartesian product* $X \times Y$ is $\{\langle x, y\rangle : x \in X \text{ and } y \in Y\}$.

Example 1.20. If $X = \{0, 1\}$, and $Y = \{1, a, b\}$, then their product is

$$X \times Y = \{\langle 0, 1\rangle, \langle 0, a\rangle, \langle 0, b\rangle, \langle 1, 1\rangle, \langle 1, a\rangle, \langle 1, b\rangle\}.$$

Example 1.21. If X is a set, the product of X with itself, $X \times X$, is also written X^2. It is the set of *all* pairs $\langle x, y\rangle$ with $x, y \in X$. The set of all triples $\langle x, y, z\rangle$ is X^3, and so on. We can give an inductive definition:

$$\begin{aligned} X^1 &= X \\ X^{k+1} &= X^k \times X \end{aligned}$$

Problem 1.6. List all elements of $\{1, 2, 3\}^3$.

Proposition 1.22. *If X has n elements and Y has m elements, then $X \times Y$ has $n \cdot m$ elements.*

Proof. For every element x in X, there are m elements of the form $\langle x, y\rangle \in X \times Y$. Let $Y_x = \{\langle x, y\rangle : y \in Y\}$. Since whenever $x_1 \neq x_2$, $\langle x_1, y\rangle \neq \langle x_2, y\rangle$, $Y_{x_1} \cap Y_{x_2} = \emptyset$. But if $X = \{x_1, \ldots, x_n\}$, then $X \times Y = Y_{x_1} \cup \cdots \cup Y_{x_n}$, and so has $n \cdot m$ elements.

To visualize this, arrange the elements of $X \times Y$ in a grid:

$$\begin{array}{rlllll} Y_{x_1} = & \{\langle x_1, y_1\rangle & \langle x_1, y_2\rangle & \ldots & \langle x_1, y_m\rangle\} \\ Y_{x_2} = & \{\langle x_2, y_1\rangle & \langle x_2, y_2\rangle & \ldots & \langle x_2, y_m\rangle\} \\ \vdots & & \vdots & & \\ Y_{x_n} = & \{\langle x_n, y_1\rangle & \langle x_n, y_2\rangle & \ldots & \langle x_n, y_m\rangle\} \end{array}$$

Since the x_i are all different, and the y_j are all different, no two of the pairs in this grid are the same, and there are $n \cdot m$ of them. □

Problem 1.7. Show, by induction on k, that for all $k \geq 1$, if X has n elements, then X^k has n^k elements.

Example 1.23. If X is a set, a *word* over X is any sequence of elements of X. A sequence can be thought of as an n-tuple of elements of X. For instance, if $X = \{a, b, c\}$, then the sequence "*bac*" can be thought of as the triple $\langle b, a, c\rangle$. Words, i.e., sequences of symbols, are of crucial importance in computer science, of course. By convention, we count elements of X as sequences of length 1, and $\emptyset$ as the sequence of length 0. The set of *all* words over X then is

$$X^* = \{\emptyset\} \cup X \cup X^2 \cup X^3 \cup \ldots$$

1.6 Russell's Paradox

We said that one can define sets by specifying a property that its elements share, e.g., defining the set of Richard's siblings as sfr:set:rus:sec

$$S = \{x : x \text{ is a sibling of Richard}\}.$$

In the very general context of mathematics one must be careful, however: not every property lends itself to *comprehension*. Some properties do not define sets. If they did, we would run into outright contradictions. One example of such a case is Russell's Paradox.

Sets may be elements of other sets—for instance, the power set of a set X is made up of sets. And so it makes sense, of course, to ask or investigate whether a set is an element of another set. Can a set be a member of itself? Nothing about the idea of a set seems to rule this out. For instance, surely *all* sets form a collection of objects, so we should be able to collect them into a single set—the set of all sets. And it, being a set, would be an element of the set of all sets.

Russell's Paradox arises when we consider the property of not having itself as an element. The set of all sets does not have this property, but all sets we have encountered so far have it. $\mathbb{N}$ is not an element of $\mathbb{N}$, since it is a set, not a natural number. $\wp(X)$ is generally not an element of $\wp(X)$; e.g., $\wp(\mathbb{R}) \notin \wp(\mathbb{R})$ since it is a set of sets of real numbers, not a set of real numbers. What if we suppose that there is a set of all sets that do not have themselves as an element? Does

$$R = \{x : x \notin x\}$$

exist?

If R exists, it makes sense to ask if $R \in R$ or not—it must be either $\in R$ or $\notin R$. Suppose the former is true, i.e., $R \in R$. R was defined as the set of all sets that are not elements of themselves, and so if $R \in R$, then R does not have this defining property of R. But only sets that have this property are in R, hence, R cannot be an element of R, i.e., $R \notin R$. But R can't both be and not be an element of R, so we have a contradiction.

Since the assumption that $R \in R$ leads to a contradiction, we have $R \notin R$. But this also leads to a contradiction! For if $R \notin R$, it does have the defining property of R, and so would be an element of R just like all the other non-self-containing sets. And again, it can't both not be and be an element of R.

Chapter 2

Relations

2.1 Relations as Sets

explanation You will no doubt remember some interesting relations between objects of some of the sets we've mentioned. For instance, numbers come with an *order relation* $<$ and from the theory of whole numbers the relation of *divisibility without remainder* (usually written $n \mid m$) may be familar. There is also the relation *is identical with* that every object bears to itself and to no other thing. But there are many more interesting relations that we'll encounter, and even more possible relations. Before we review them, we'll just point out that we can look at relations as a special sort of set. For this, first recall what a *pair* is: if a and b are two objects, we can combine them into the *ordered pair* $\langle a, b\rangle$. Note that for ordered pairs the order *does* matter, e.g, $\langle a, b\rangle \neq \langle b, a\rangle$, in contrast to unordered pairs, i.e., 2-element sets, where $\{a, b\} = \{b, a\}$. sfr:rel:set:sec

If X and Y are sets, then the *Cartesian product* $X \times Y$ of X and Y is the set of all pairs $\langle a, b\rangle$ with $a \in X$ and $b \in Y$. In particular, $X^2 = X \times X$ is the set of all pairs from X.

Now consider a relation on a set, e.g., the $<$-relation on the set $\mathbb{N}$ of natural numbers, and consider the set of all pairs of numbers $\langle n, m\rangle$ where $n < m$, i.e.,

$$R = \{\langle n, m\rangle : n, m \in \mathbb{N} \text{ and } n < m\}.$$

Then there is a close connection between the number n being less than a number m and the corresponding pair $\langle n, m\rangle$ being a member of R, namely, $n < m$ if and only if $\langle n, m\rangle \in R$. In a sense we can consider the set R to *be* the $<$-relation on the set $\mathbb{N}$. In the same way we can construct a subset of $\mathbb{N}^2$ for any relation between numbers. Conversely, given any set of pairs of numbers $S \subseteq \mathbb{N}^2$, there is a corresponding relation between numbers, namely, the relationship n bears to m if and only if $\langle n, m\rangle \in S$. This justifies the following definition:

Definition 2.1 (Binary relation)**.** A *binary relation* on a set X is a subset of X^2. If $R \subseteq X^2$ is a binary relation on X and $x, y \in X$, we write Rxy (or xRy) for $\langle x, y\rangle \in R$.

sfr:rel:set: relations **Example 2.2.** The set $\mathbb{N}^2$ of pairs of natural numbers can be listed in a 2-dimensional matrix like this:

$$\begin{array}{ccccc}
\langle \mathbf{0}, \mathbf{0} \rangle & \langle 0, 1 \rangle & \langle 0, 2 \rangle & \langle 0, 3 \rangle & \dots \\
\langle 1, 0 \rangle & \langle \mathbf{1}, \mathbf{1} \rangle & \langle 1, 2 \rangle & \langle 1, 3 \rangle & \dots \\
\langle 2, 0 \rangle & \langle 2, 1 \rangle & \langle \mathbf{2}, \mathbf{2} \rangle & \langle 2, 3 \rangle & \dots \\
\langle 3, 0 \rangle & \langle 3, 1 \rangle & \langle 3, 2 \rangle & \langle \mathbf{3}, \mathbf{3} \rangle & \dots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

The subset consisting of the pairs lying on the diagonal, i.e.,

$$\{\langle 0, 0 \rangle, \langle 1, 1 \rangle, \langle 2, 2 \rangle, \dots\},$$

is the *identity relation on* $\mathbb{N}$. (Since the identity relation is popular, let's define $\mathrm{Id}_X = \{\langle x, x \rangle : x \in X\}$ for any set X.) The subset of all pairs lying above the diagonal, i.e.,

$$L = \{\langle 0, 1 \rangle, \langle 0, 2 \rangle, \dots, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \dots, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \dots\},$$

is the *less than* relation, i.e., Lnm iff $n < m$. The subset of pairs below the diagonal, i.e.,

$$G = \{\langle 1, 0 \rangle, \langle 2, 0 \rangle, \langle 2, 1 \rangle, \langle 3, 0 \rangle, \langle 3, 1 \rangle, \langle 3, 2 \rangle, \dots\},$$

is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I, $K = L \cup I$, is the *less than or equal to* relation: Knm iff $n \leq m$. Similarly, $H = G \cup I$ is the *greater than or equal to relation.* L, G, K, and H are special kinds of relations called *orders*. L and G have the property that no number bears L or G to itself (i.e., for all n, neither Lnn nor Gnn). Relations with this property are called *irreflexive*, and, if they also happen to be orders, they are called *strict orders.*

Although orders and identity are important and natural relations, it should be emphasized that according to our definition *any* subset of X^2 is a relation on X, regardless of how unnatural or contrived it seems. In particular, $\emptyset$ is a relation on any set (the *empty relation*, which no pair of elements bears), and X^2 itself is a relation on X as well (one which every pair bears), called the *universal relation.* But also something like $E = \{\langle n, m \rangle : n > 5 \text{ or } m \times n \geq 34\}$ counts as a relation. explanation

Problem 2.1. List the elements of the relation $\subseteq$ on the set $\wp(\{a, b, c\})$.

2.2 Special Properties of Relations

sfr:rel:prp: sec Some kinds of relations turn out to be so common that they have been given special names. For instance, $\leq$ and $\subseteq$ both relate their respective domains (say, $\mathbb{N}$ in the case of $\leq$ and $\wp(X)$ in the case of $\subseteq$) in similar ways. To get at exactly how these relations are similar, and how they differ, we categorize them intro

according to some special properties that relations can have. It turns out that (combinations of) some of these special properties are especially important: orders and equivalence relations.

Definition 2.3 (Reflexivity)**.** A relation $R \subseteq X^2$ is *reflexive* iff, for every $x \in X$, Rxx.

Definition 2.4 (Transitivity)**.** A relation $R \subseteq X^2$ is *transitive* iff, whenever Rxy and Ryz, then also Rxz.

Definition 2.5 (Symmetry)**.** A relation $R \subseteq X^2$ is *symmetric* iff, whenever Rxy, then also Ryx.

Definition 2.6 (Anti-symmetry)**.** A relation $R \subseteq X^2$ is *anti-symmetric* iff, whenever both Rxy and Ryx, then $x = y$ (or, in other words: if $x \neq y$ then either $\neg Rxy$ or $\neg Ryx$).

explanation In a symmetric relation, Rxy and Ryx always hold together, or neither holds. In an anti-symmetric relation, the only way for Rxy and Ryx to hold together is if $x = y$. Note that this does not *require* that Rxy and Ryx holds when $x = y$, only that it isn't ruled out. So an anti-symmetric relation can be reflexive, but it is not the case that every anti-symmetric relation is reflexive. Also note that being anti-symmetric and merely not being symmetric are different conditions. In fact, a relation can be both symmetric and anti-symmetric at the same time (e.g., the identity relation is).

Definition 2.7 (Connectivity)**.** A relation $R \subseteq X^2$ is *connected* if for all $x, y \in X$, if $x \neq y$, then either Rxy or Ryx.

Definition 2.8 (Partial order)**.** A relation $R \subseteq X^2$ that is reflexive, transitive, and anti-symmetric is called a *partial order*.

Definition 2.9 (Linear order)**.** A partial order that is also connected is called a *linear order*.

Definition 2.10 (Equivalence relation)**.** A relation $R \subseteq X^2$ that is reflexive, symmetric, and transitive is called an *equivalence relation*. x and y are said to be *R-equivalent* if Rxy.

Moreover, we use *R-equivalence class* to denote a set of elements that are R-equivalent.

Definition 2.11 (Equivalence class)**.** The R-equivalence class containing x, or $[x]_R$, or $[x]$ if R is clear, is defined to be the set { y : Rxy }. x is said to be the *representative* of this R-equivalence class when we write $[x]_R$.

explanation Note that in the above definition, x is said to be the representative only because we use it to denote the class it's in. If we write $[y]$ to denote the same class, then y is the representative. It only depends on how we denote the class.

Problem 2.2. Give examples of relations that are (a) reflexive and symmetric but not transitive, (b) reflexive and anti-symmetric, (c) anti-symmetric, transitive, but not reflexive, and (d) reflexive, symmetric, and transitive. Do not use relations on numbers or sets.

2.3 Orders

sfr:rel:ord:sec Very often we are interested in comparisons between objects, where one object may be less or equal or greater than another in a certain respect. Size is the most obvious example of such a comparative relation, or *order*. But not all such relations are alike in all their properties. For instance, some comparative relations require any two objects to be comparable, others don't. (If they do, we call them *linear* or *total*.) Some include identity (like $\leq$) and some exclude it (like $<$). Let's get some order into all this. explanation

Definition 2.12 (Preorder)**.** A relation which is both reflexive and transitive is called a *preorder*.

Definition 2.13 (Partial order)**.** A preorder which is also anti-symmetric is called a *partial order*.

Definition 2.14 (Linear order)**.** A partial order which is also connected is called a *total order* or *linear order*.

Example 2.15. Every linear order is also a partial order, and every partial order is also a preorder, but the converses don't hold. The universal relation on X is a preorder, since it is reflexive and transitive. But, if X has more than one element, the universal relation is not anti-symmetric, and so not a partial order. For a somewhat less silly example, consider the *no longer than* relation $\preccurlyeq$ on $\mathbb{B}^*$: $x \preccurlyeq y$ iff $\mathrm{len}(x) \leq \mathrm{len}(y)$. This is a preorder (reflexive and transitive), and even connected, but not a partial order, since it is not anti-symmetric. For instance, $01 \preccurlyeq 10$ and $10 \preccurlyeq 01$, but $01 \neq 10$.

The relation of *divisibility without remainder* gives us an example of a partial order which isn't a linear order: for integers n, m, we say n (evenly) divides m, in symbols: $n \mid m$, if there is some k so that $m = kn$. On $\mathbb{N}$, this is a partial order, but not a linear order: for instance, $2 \nmid 3$ and also $3 \nmid 2$. Considered as a relation on $\mathbb{Z}$, divisibility is only a preorder since anti-symmetry fails: $1 \mid -1$ and $-1 \mid 1$ but $1 \neq -1$. Another important partial order is the relation $\subseteq$ on a set of sets.

Notice that the examples L and G from Example 2.2, although we said there that they were called "strict orders," are not linear orders even though they are connected (they are not reflexive). But there is a close connection, as we will see momentarily.

Definition 2.16 (Irreflexivity)**.** A relation R on X is called *irreflexive* if, for all $x \in X$, $\neg Rxx$.

Definition 2.17 (Asymmetry)**.** A relation R on X is called *asymmetric* if for no pair $x, y \in X$ we have Rxy and Ryx.

Definition 2.18 (Strict order)**.** A *strict order* is a relation which is irreflexive, asymmetric, and transitive.

Definition 2.19 (Strict linear order)**.** A strict order which is also connected is called a *strict linear order.*

A strict order on X can be turned into a partial order by adding the diagonal Id_X, i.e., adding all the pairs $\langle x, x\rangle$. (This is called the *reflexive closure* of R.) Conversely, starting from a partial order, one can get a strict order by removing Id_X.

Proposition 2.20. *sfr:rel:ord: strict-partial*

1. *If R is a strict (linear) order on X, then $R^+ = R \cup \mathrm{Id}_X$ is a partial order (linear order).*

2. *If R is a partial order (linear order) on X, then $R^- = R \setminus \mathrm{Id}_X$ is a strict (linear) order.*

Proof. 1. Suppose R is a strict order, i.e., $R \subseteq X^2$ and R is irreflexive, asymmetric, and transitive. Let $R^+ = R \cup \mathrm{Id}_X$. We have to show that R^+ is reflexive, antisymmetric, and transitive.

R^+ is clearly reflexive, since for all $x \in X$, $\langle x, x\rangle \in \mathrm{Id}_X \subseteq R^+$.

To show R^+ is antisymmetric, suppose R^+xy and R^+yx, i.e., $\langle x, y\rangle$ and $\langle y, x\rangle \in R^+$, and $x \neq y$. Since $\langle x, y\rangle \in R \cup \mathrm{Id}_X$, but $\langle x, y\rangle \notin \mathrm{Id}_X$, we must have $\langle x, y\rangle \in R$, i.e., Rxy. Similarly we get that Ryx. But this contradicts the assumption that R is asymmetric.

Now suppose that R^+xy and R^+yz. If both $\langle x, y\rangle \in R$ and $\langle y, z\rangle \in R$, it follows that $\langle x, z\rangle \in R$ since R is transitive. Otherwise, either $\langle x, y\rangle \in \mathrm{Id}_X$, i.e., $x = y$, or $\langle y, z\rangle \in \mathrm{Id}_X$, i.e., $y = z$. In the first case, we have that R^+yz by assumption, $x = y$, hence R^+xz. Similarly in the second case. In either case, R^+xz, thus, R^+ is also transitive.

If R is connected, then for all $x \neq y$, either Rxy or Ryx, i.e., either $\langle x, y\rangle \in R$ or $\langle y, x\rangle \in R$. Since $R \subseteq R^+$, this remains true of R^+, so R^+ is connected as well.

2. Exercise.

□

Problem 2.3. Complete the proof of Proposition 2.20, i.e., prove that if R is a partial order on X, then $R^- = R \setminus \mathrm{Id}_X$ is a strict order.

Example 2.21. $\leq$ is the linear order corresponding to the strict linear order $<$. $\subseteq$ is the partial order corresponding to the strict order $\subsetneq$.

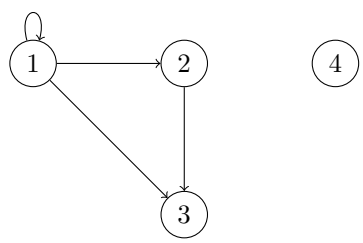
2.4 Graphs

sfr:rel:grp:sec A *graph* is a diagram in which points—called "nodes" or "vertices" (plural of "vertex")—are connected by edges. Graphs are a ubiquitous tool in discrete mathematics and in computer science. They are incredibly useful for representing, and visualizing, relationships and structures, from concrete things like networks of various kinds to abstract structures such as the possible outcomes of decisions. There are many different kinds of graphs in the literature which differ, e.g., according to whether the edges are directed or not, have labels or not, whether there can be edges from a node to the same node, multiple edges between the same nodes, etc. *Directed graphs* have a special connection to relations.

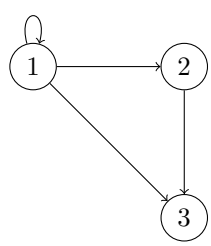
Definition 2.22 (Directed graph)**.** A *directed graph* $G = \langle V, E\rangle$ is a set of *vertices* V and a set of *edges* $E \subseteq V^2$.

According to our definition, a graph just is a set together with a relation on that set. Of course, when talking about graphs, it's only natural to expect that they are graphically represented: we can draw a graph by connecting two vertices v_1 and v_2 by an arrow iff $\langle v_1, v_2\rangle \in E$. The only difference between a relation by itself and a graph is that a graph specifies the set of vertices, i.e., a graph may have isolated vertices. The important point, however, is that every relation R on a set X can be seen as a directed graph $\langle X, R\rangle$, and conversely, a directed graph $\langle V, E\rangle$ can be seen as a relation $E \subseteq V^2$ with the set V explicitly specified. explanation

Example 2.23. The graph $\langle V, E\rangle$ with $V = \{1, 2, 3, 4\}$ and $E = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 3\rangle\}$ looks like this:



This is a different graph than $\langle V', E\rangle$ with $V' = \{1, 2, 3\}$, which looks like this:



Problem 2.4. Consider the less-than-or-equal-to relation $\leq$ on the set $\{1, 2, 3, 4\}$ as a graph and draw the corresponding diagram.

2.5 Operations on Relations

It is often useful to modify or combine relations. We've already used the union of relations above (which is just the union of two relations considered as sets of pairs). Here are some other ways: sfr:rel:ops:sec

Definition 2.24. Let R, $S \subseteq X^2$ be relations and Y a set.

1. The *inverse* R^{-1} of R is $R^{-1} = \{\langle y, x\rangle : \langle x, y\rangle \in R\}$.
2. The *relative product* $R \mid S$ of R and S is
$$(R \mid S) = \{\langle x, z\rangle : \text{for some } y, Rxy \text{ and } Syz\}$$
3. The *restriction* $R \restriction Y$ of R to Y is $R \cap Y^2$
4. The *application* $R[Y]$ of R to Y is
$$R[Y] = \{y : \text{for some } x \in Y, Rxy\}$$

Example 2.25. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on $\mathbb{Z}$, i.e., the set of pairs $\langle x, y\rangle$ where $x + 1 = y$, for $x, y \in \mathbb{Z}$. Sxy holds iff y is the successor of x.

1. The inverse S^{-1} of S is the predecessor relation, i.e., $S^{-1}xy$ iff $x - 1 = y$.
2. The relative product $S \mid S$ is the relation x bears to y if $x + 2 = y$.
3. The restriction of S to $\mathbb{N}$ is the successor relation on $\mathbb{N}$.
4. The application of S to a set, e.g., $S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition 2.26 (Transitive closure)**.** The *transitive closure* R^+ of a relation $R \subseteq X^2$ is $R^+ = \bigcup_{i=1}^{\infty} R^i$ where $R^1 = R$ and $R^{i+1} = R^i \mid R$.

The *reflexive transitive closure* of R is $R^* = R^+ \cup \mathrm{Id}_X$.

Example 2.27. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So R^*xy iff for some $i \geq 1$, $x + i = y$. In other words, S^+xy iff $x < y$ (and R^*xy iff $x \leq y$).

Problem 2.5. Show that the transitive closure of R is in fact transitive.

Chapter 3

Functions

3.1 Basics

sfr:fun:bas:sec A *function* is a mapping which pairs each object of a given set with a single partner in another set. For instance, the operation of adding 1 defines a function: each number n is paired with a unique number $n+1$. More generally, functions may take pairs, triples, etc., of inputs and returns some kind of output. Many functions are familiar to us from basic arithmetic. For instance, addition and multiplication are functions. They take in two numbers and return a third. In this mathematical, abstract sense, a function is a *black box*: what matters is only what output is paired with what input, not the method for calculating the output. explanation

Definition 3.1 (Function)**.** A *function* $f\colon X \to Y$ is a mapping of each element of X to an element of Y. We call X the *domain* of f and Y the *codomain* of f. The elements of X are called inputs or *arguments* of f, and the element of Y that is paired with an argument x by f is called the *value of* f for argument x, written $f(x)$.

The *range* $\operatorname{ran}(f)$ of f is the subset of the codomain consisting of the values of f for some argument; $\operatorname{ran}(f) = \{f(x) : x \in X\}$.

Example 3.2. Multiplication takes pairs of natural numbers as inputs and maps them to natural numbers as outputs, so goes from $\mathbb{N} \times \mathbb{N}$ (the domain) to $\mathbb{N}$ (the codomain). As it turns out, the range is also $\mathbb{N}$, since every $n \in \mathbb{N}$ is $n \times 1$.

Multiplication is a function because it pairs each input—each pair of natural numbers—with a single output: $\times\colon \mathbb{N}^2 \to \mathbb{N}$. By contrast, the square root operation applied to the domain $\mathbb{N}$ is not functional, since each positive integer n has two square roots: $\sqrt{n}$ and $-\sqrt{n}$. We can make it functional by only returning the positive square root: $\sqrt{}\colon \mathbb{N} \to \mathbb{R}$. The relation that pairs each student in a class with their final grade is a function—no student can get two different final grades in the same class. The relation that pairs each student in explanation

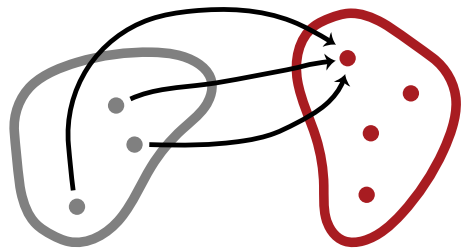


Figure 3.1: A function is a mapping of each element of one set to an element of another. An arrow points from an argument in the domain to the corresponding value in the codomain.

a class with their parents is not a function—generally each student will have at least two parents.

We can define functions by specifying in some precise way what the value of the function is for every possible argment. Different ways of doing this are by giving a formula, describing a method for computing the value, or listing the values for each argument. However functions are defined, we must make sure that for each argment we specify one, and only one, value.

Example 3.3. Let $f\colon \mathbb{N} \to \mathbb{N}$ be defined such that $f(x) = x + 1$. This is a definition that specifies f as a function which takes in natural numbers and outputs natural numbers. It tells us that, given a natural number x, f will output its successor $x + 1$. In this case, the codomain $\mathbb{N}$ is not the range of f, since the natural number 0 is not the successor of any natural number. The range of f is the set of all positive integers, $\mathbb{Z}^+$.

Example 3.4. Let $g\colon \mathbb{N} \to \mathbb{N}$ be defined such that $g(x) = x + 2 - 1$. This tells us that g is a function which takes in natural numbers and outputs natural numbers. Given a natural number n, g will output the predecessor of the successor of the successor of x, i.e., $x + 1$. Despite their different definitions, g and f are the same function.

explanation

Functions f and g defined above are the same because for any natural number x, $x + 2 - 1 = x + 1$. f and g pair each natural number with the same output. The definitions for f and g specify the same mapping by means of different equations, and so count as the same function.

Example 3.5. We can also define functions by cases. For instance, we could define $h\colon \mathbb{N} \to \mathbb{N}$ by

$$h(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

Since every natural number is either even or odd, the output of this function will always be a natural number. Just remember that if you define a function by cases, every possible input must fall into exactly one case. In some cases, this will require a a proof that the cases are exhaustive and exclusive.

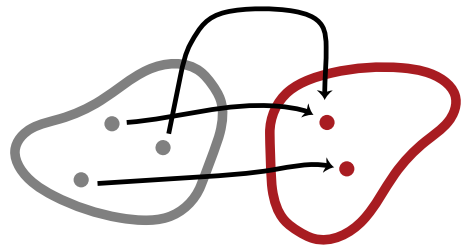


Figure 3.2: A surjective function has every element of the codomain as a value.

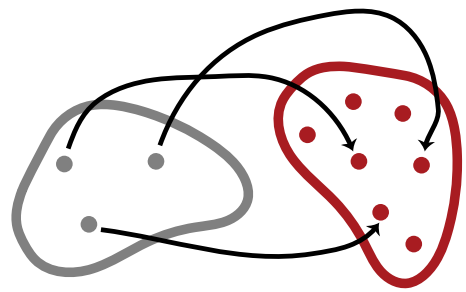


Figure 3.3: An injective function never maps two different arguments to the same value.

3.2 Kinds of Functions

sfr:fun:kin:sec

Definition 3.6 (Surjective function)**.** A function $f\colon X \to Y$ is *surjective* iff Y is also the range of f, i.e., for every $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$.

If you want to show that a function is surjective, then you need to show that every object in the codomain is the output of the function given some input or other. explanation

Definition 3.7 (Injective function)**.** A function $f\colon X \to Y$ is *injective* iff for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$.

Any function pairs each possible input with a unique output. An injective function has a unique input for each possible output. If you want to show that a function f is injective, you need to show that for any elements x and x' of the domain, if $f(x) = f(x')$, then $x = x'$. explanation

An example of a function which is neither injective, nor surjective, is the constant function $f\colon \mathbb{N} \to \mathbb{N}$ where $f(x) = 1$.

An example of a function which is both injective and surjective is the identity function $f\colon \mathbb{N} \to \mathbb{N}$ where $f(x) = x$.

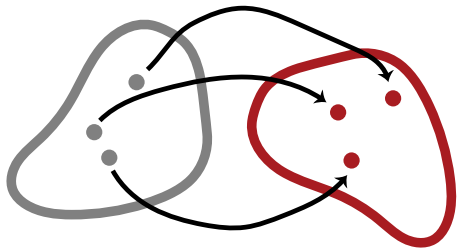


Figure 3.4: A bijective function uniquely pairs the elements of the codomain with those of the domain.

The successor function $f\colon \mathbb{N} \to \mathbb{N}$ where $f(x) = x + 1$ is injective, but not surjective.

The function

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

is surjective, but not injective.

Definition 3.8 (Bijection)**.** A function $f\colon X \to Y$ is *bijective* iff it is both surjective and injective. We call such a function a *bijection* from X to Y (or between X and Y).

3.3 Inverses of Functions

explanation One obvious question about functions is whether a given mapping can be "reversed." For instance, the successor function $f(x) = x + 1$ can be reversed in the sense that the function $g(y) = y - 1$ "undoes" what f does. But we must be careful: While the definition of g defines a function $\mathbb{Z} \to \mathbb{Z}$, it does not define a function $\mathbb{N} \to \mathbb{N}$ ($g(0) \notin \mathbb{N}$). So even in simple cases, it is not quite obvious if functions can be reversed, and that it may depend on the domain and codomain. Let's give a precise definition. sfr:fun:inv:sec

Definition 3.9. A function $g\colon Y \to X$ is an *inverse* of a function $f\colon X \to Y$ if $f(g(y)) = y$ and $g(f(x)) = x$ for all $x \in X$ and $y \in Y$.

explanation When do functions have inverses? A good candidate for an inverse of $f\colon X \to Y$ is $g\colon Y \to X$ "defined by"

$$g(y) = \text{“the” } x \text{ such that } f(x) = y.$$

The scare quotes around "defined by" suggest that this is not a definition. At least, it is not in general. For in order for this definition to specify a function, there has to be one and only one x such that $f(x) = y$—the output of g has to be uniquely specified. Moreover, it has to be specified for every $y \in Y$. If there

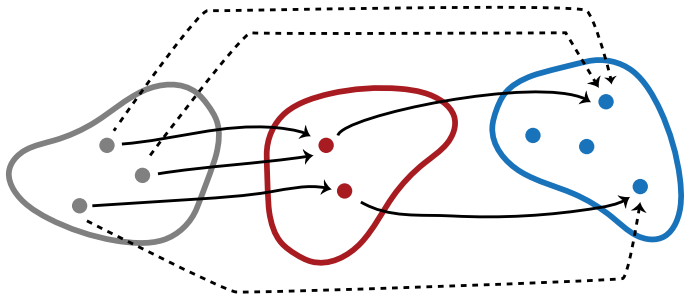


Figure 3.5: The composition $g \circ f$ of two functions f and g.

are x_1 and $x_2 \in X$ with $x_1 \neq x_2$ but $f(x_1) = f(x_2)$, then $g(y)$ would not be uniquely specified for $y = f(x_1) = f(x_2)$. And if there is no x at all such that $f(x) = y$, then $g(y)$ is not specified at all. In other words, for g to be defined, f has to be injective and surjective.

Proposition 3.10. *If $f\colon X \to Y$ is bijective, f has a unique inverse $f^{-1}\colon Y \to X$.*

Proof. Exercise. □

Problem 3.1. Show that if f is bijective, an inverse g of f exists, i.e., define such a g, show that it is a function, and show that it is an inverse of f, i.e., $f(g(y)) = y$ and $g(f(x)) = x$ for all $x \in X$ and $y \in Y$.

Problem 3.2. Show that if $f\colon X \to Y$ has an inverse g, then f is bijective.

Problem 3.3. Show that if $g\colon Y \to X$ and $g'\colon Y \to X$ are inverses of $f\colon X \to Y$, then $g = g'$, i.e., for all $y \in Y$, $g(y) = g'(y)$.

3.4 Composition of Functions

sfr:fun:cmp:sec We have already seen that the inverse f^{-1} of a bijective function f is itself a function. It is also possible to compose functions f and g to define a new function by first applying f and then g. Of course, this is only possible if the ranges and domains match, i.e., the range of f must be a subset of the domain of g. explanation

Definition 3.11 (Composition)**.** Let $f\colon X \to Y$ and $g\colon Y \to Z$. The *composition* of f with g is the function $(g \circ f)\colon X \to Z$, where $(g \circ f)(x) = g(f(x))$.

The function $(g \circ f)\colon X \to Z$ pairs each member of X with a member of Z. We specify which member of Z a member of X is paired with as follows—given an input $x \in X$, first apply the function f to x, which will output some $y \in Y$. Then apply the function g to y, which will output some $z \in Z$. explanation

Example 3.12. Consider the functions $f(x) = x + 1$, and $g(x) = 2x$. What function do you get when you compose these two? $(g \circ f)(x) = g(f(x))$. So that means for every natural number you give this function, you first add one, and then you multiply the result by two. So their composition is $(g \circ f)(x) = 2(x+1)$.

Problem 3.4. Show that if $f\colon X \to Y$ and $g\colon Y \to Z$ are both injective, then $g \circ f\colon X \to Z$ is injective.

Problem 3.5. Show that if $f\colon X \to Y$ and $g\colon Y \to Z$ are both surjective, then $g \circ f\colon X \to Z$ is surjective.

3.5 Isomorphism

explanation An *isomorphism* is a bijection that preserves the structure of the sets it relates, where structure is a matter of the relationships that obtain between the elements of the sets. Consider the following two sets $X = \{1, 2, 3\}$ and $Y = \{4, 5, 6\}$. These sets are both structured by the relations successor, less than, and greater than. An isomorphism between the two sets is a bijection that preserves those structures. So a bijective function $f\colon X \to Y$ is an isomorphism if, $i < j$ iff $f(i) < f(j)$, $i > j$ iff $f(i) > f(j)$, and j is the successor of i iff $f(j)$ is the successor of $f(i)$. sfr:fun:iso:sec

Definition 3.13 (Isomorphism)**.** Let U be the pair $\langle X, R\rangle$ and V be the pair $\langle Y, S\rangle$ such that X and Y are sets and R and S are relations on X and Y respectively. A bijection f from X to Y is an *isomorphism* from U to V iff it preserves the relational structure, that is, for any x_1 and x_2 in X, $\langle x_1, x_2\rangle \in R$ iff $\langle f(x_1), f(x_2)\rangle \in S$.

Example 3.14. Consider the following two sets $X = \{1, 2, 3\}$ and $Y = \{4, 5, 6\}$, and the relations less than and greater than. The function $f\colon X \to Y$ where $f(x) = 7 - x$ is an isomorphism between $\langle X, <\rangle$ and $\langle Y, >\rangle$.

3.6 Partial Functions

explanation It is sometimes useful to relax the definition of function so that it is not required that the output of the function is defined for all possible inputs. Such mappings are called *partial functions*. sfr:fun:par:sec

Definition 3.15. A *partial function* $f\colon X \nrightarrow Y$ is a mapping which assigns to every element of X at most one element of Y. If f assigns an element of Y to $x \in X$, we say $f(x)$ is *defined*, and otherwise *undefined*. If $f(x)$ is defined, we write $f(x)\downarrow$, otherwise $f(x)\uparrow$. The *domain* of a partial function f is the subset of X where it is defined, i.e., $\mathrm{dom}(f) = \{x : f(x)\downarrow\}$.

Example 3.16. Every function $f\colon X \to Y$ is also a partial function. Partial functions that are defined everywhere on X—i.e., what we so far have simply called a function—are also called *total* functions.

Example 3.17. The partial function $f\colon \mathbb{R} \nrightarrow \mathbb{R}$ given by $f(x) = 1/x$ is undefined for $x = 0$, and defined everywhere else.

Problem 3.6. Given $f\colon X \nrightarrow Y$, define the partial function $g\colon Y \nrightarrow X$ by: for any $y \in Y$, if there is a unique $x \in X$ such that $f(x) = y$, then $g(y) = x$; otherwise $g(y) \uparrow$. Show that if f is injective, then $g(f(x)) = x$ for all $x \in \mathrm{dom}(f)$, and $f(g(y)) = y$ for all $y \in \mathrm{ran}(f)$.

3.7 Functions and Relations

sfr:fun:rel:sec A function which maps elements of X to elements of Y obviously defines a relation between X and Y, namely the relation which holds between x and y iff $f(x) = y$. In fact, we might even—if we are interested in reducing the building blocks of mathematics for instance—*identify* the function f with this relation, i.e., with a set of pairs. This then raises the question: which relations define functions in this way? explanation

Definition 3.18 (Graph of a function)**.** Let $f\colon X \nrightarrow Y$ be a partial function. The *graph* of f is the relation $R_f \subseteq X \times Y$ defined by

$$R_f = \{\langle x, y\rangle : f(x) = y\}.$$

Proposition 3.19. *Suppose $R \subseteq X \times Y$ has the property that whenever Rxy and Rxy' then $y = y'$. Then R is the graph of the partial function $f\colon X \nrightarrow Y$ defined by: if there is a y such that Rxy, then $f(x) = y$, otherwise $f(x) \uparrow$. If R is also* serial*, i.e., for each $x \in X$ there is a $y \in Y$ such that Rxy, then f is total.*

Proof. Suppose there is a y such that Rxy. If there were another $y' \neq y$ such that Rxy', the condition on R would be violated. Hence, if there is a y such that Rxy, that y is unique, and so f is well-defined. Obviously, $R_f = R$ and f is total if R is serial. □

Problem 3.7. Suppose $f\colon X \to Y$ and $g\colon Y \to Z$. Show that the graph of $(g \circ f)$ is $R_f \mid R_g$.

Chapter 4

The Size of Sets

4.1 Introduction

When Georg Cantor developed set theory in the 1870s, his interest was in part to make palatable the idea of an infinite collection—an actual infinity, as the medievals would say. Key to this rehabilitation of the notion of the infinite was a way to assign sizes—"cardinalities"—to sets. The cardinality of a finite set is just a natural number, e.g., $\emptyset$ has cardinality 0, and a set containing five things has cardinality 5. But what about infinite sets? Do they all have the same cardinality, ∞? It turns out, they do not. sfr:siz:int:sec

The first important idea here is that of an enumeration. We can list every finite set by listing all its elements. For some infinite sets, we can also list all their elements if we allow the list itself to be infinite. Such sets are called enumerable. Cantor's surprising result was that some infinite sets are not enumerable.

4.2 Enumerable Sets

One way of specifying a finite set is by listing its elements. But conversely, since there are only finitely many elements in a set, every finite set can be enumerated. By this we mean: its elements can be put into a list (a list with a beginning, where each element of the list other than the first has a unique predecessor). Some infinite sets can also be enumerated, such as the set of positive integers. sfr:siz:enm:sec

Definition 4.1 (Enumeration)**.** Informally, an *enumeration* of a set X is a list (possibly infinite) of elements of X such that every element of X appears on the list at some finite position. If X has an enumeration, then X is said to be *enumerable*. If X is enumerable and infinite, we say X is denumerable.

explanation A couple of points about enumerations:

1. We count as enumerations only lists which have a beginning and in which every element other than the first has a single element immediately pre-

ceding it. In other words, there are only finitely many elements between the first element of the list and any other element. In particular, this means that every element of an enumeration has a finite position: the first element has position 1, the second position 2, etc.

2. We can have different enumerations of the same set X which differ by the order in which the elements appear: 4, 1, 25, 16, 9 enumerates the (set of the) first five square numbers just as well as 1, 4, 9, 16, 25 does.

3. Redundant enumerations are still enumerations: 1, 1, 2, 2, 3, 3, ... enumerates the same set as 1, 2, 3, ... does.

4. Order and redundancy *do* matter when we specify an enumeration: we can enumerate the positive integers beginning with 1, 2, 3, 1, ..., but the pattern is easier to see when enumerated in the standard way as 1, 2, 3, 4, ...

5. Enumerations must have a beginning: ..., 3, 2, 1 is not an enumeration of the positive integers because it has no first element. To see how this follows from the informal definition, ask yourself, "at what position in the list does the number 76 appear?"

6. The following is not an enumeration of the positive integers: 1, 3, 5, ..., 2, 4, 6, ... The problem is that the even numbers occur at places $\infty + 1$, $\infty + 2$, $\infty + 3$, rather than at finite positions.

7. Lists may be gappy: 2, −, 4, −, 6, −, ... enumerates the even positive integers.

8. The empty set is enumerable: it is enumerated by the empty list!

Proposition 4.2. *If X has an enumeration, it has an enumeration without gaps or repetitions.*

Proof. Suppose X has an enumeration $x_1, x_2, \ldots$ in which each x_i is an element of X or a gap. We can remove repetitions from an enumeration by replacing repeated elements by gaps. For instance, we can turn the enumeration into a new one in which x'_i is x_i if x_i is an element of X that is not among $x_1, \ldots, x_{i-1}$ or is − if it is. We can remove gaps by closing up the elements in the list. To make precise what "closing up" amounts to is a bit difficult to describe. Roughly, it means that we can generate a new enumeration $x''_1, x''_2, \ldots$, where each x''_i is the first element in the enumeration $x'_1, x'_2, \ldots$ after x''_{i-1} (if there is one). □

The last argument shows that in order to get a good handle on enumerations and enumerable sets and to prove things about them, we need a more precise definition. The following provides it.

Definition 4.3 (Enumeration)**.** An *enumeration* of a set X is any surjective function $f\colon \mathbb{Z}^+ \to X$.

explanation Let's convince ourselves that the formal definition and the informal definition using a possibly gappy, possibly infinite list are equivalent. A surjective function (partial or total) from $\mathbb{Z}^+$ to a set X enumerates X. Such a function determines an enumeration as defined informally above: the list $f(1)$, $f(2)$, $f(3)$, ... Since f is surjective, every element of X is guaranteed to be the value of $f(n)$ for some $n \in \mathbb{Z}^+$. Hence, every element of X appears at some finite position in the list. Since the function may not be injective, the list may be redundant, but that is acceptable (as noted above).

On the other hand, given a list that enumerates all elements of X, we can define a surjective function $f\colon \mathbb{Z}^+ \to X$ by letting $f(n)$ be the nth element of the list that is not a gap, or the final element of the list if there is no nth element. There is one case in which this does not produce a surjective function: if X is empty, and hence the list is empty. So, every non-empty list determines a surjective function $f\colon \mathbb{Z}^+ \to X$.

Definition 4.4. A set X is enumerable iff it is empty or has an enumeration. sfr:siz:enm: defn:enumerable

Example 4.5. A function enumerating the positive integers ($\mathbb{Z}^+$) is simply the identity function given by $f(n) = n$. A function enumerating the natural numbers $\mathbb{N}$ is the function $g(n) = n - 1$.

Problem 4.1. According to Definition 4.4, a set X is enumerable iff $X = \emptyset$ or there is a surjective $f\colon \mathbb{Z}^+ \to X$. It is also possible to define "enumerable set" precisely by: a set is enumerable iff there is an injective function $g\colon X \to \mathbb{Z}^+$. Show that the definitions are equivalent, i.e., show that there is an injective function $g\colon X \to \mathbb{Z}^+$ iff either $X = \emptyset$ or there is a surjective $f\colon \mathbb{Z}^+ \to X$.

Example 4.6. The functions $f\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ and $g\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ given by

$$f(n) = 2n \text{ and}$$
$$g(n) = 2n + 1$$

enumerate the even positive integers and the odd positive integers, respectively. However, neither function is an enumeration of $\mathbb{Z}^+$, since neither is surjective.

Problem 4.2. Define an enumeration of the positive squares 4, 9, 16, ...

Example 4.7. The function $f(n) = (-1)^n \lceil \frac{(n-1)}{2} \rceil$ (where $\lceil x \rceil$ denotes the *ceiling* function, which rounds x up to the nearest integer) enumerates the set of integers $\mathbb{Z}$. Notice how f generates the values of $\mathbb{Z}$ by "hopping" back and forth between positive and negative integers:

$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	...
$-\lceil \frac{0}{2} \rceil$	$\lceil \frac{1}{2} \rceil$	$-\lceil \frac{2}{2} \rceil$	$\lceil \frac{3}{2} \rceil$	$-\lceil \frac{4}{2} \rceil$	$\lceil \frac{5}{2} \rceil$	$-\lceil \frac{6}{2} \rceil$	...
0	1	-1	2	-2	3	...	

You can also think of f as defined by cases as follows:

$$f(n) = \begin{cases} 0 & \text{if } n = 1 \\ n/2 & \text{if } n \text{ is even} \\ -(n-1)/2 & \text{if } n \text{ is odd and } > 1 \end{cases}$$

Problem 4.3. Show that if X and Y are enumerable, so is $X \cup Y$.

Problem 4.4. Show by induction on n that if X_1, X_2, ..., X_n are all enumerable, so is $X_1 \cup \cdots \cup X_n$.

That is fine for "easy" sets. What about the set of, say, pairs of positive integers? explanation

$$\mathbb{Z}^+ \times \mathbb{Z}^+ = \{\langle n, m\rangle : n, m \in \mathbb{Z}^+\}$$

We can organize the pairs of positive integers in an *array*, such as the following:

	1	**2**	**3**	**4**	...
1	$\langle 1,1\rangle$	$\langle 1,2\rangle$	$\langle 1,3\rangle$	$\langle 1,4\rangle$	...
2	$\langle 2,1\rangle$	$\langle 2,2\rangle$	$\langle 2,3\rangle$	$\langle 2,4\rangle$	...
3	$\langle 3,1\rangle$	$\langle 3,2\rangle$	$\langle 3,3\rangle$	$\langle 3,4\rangle$	...
4	$\langle 4,1\rangle$	$\langle 4,2\rangle$	$\langle 4,3\rangle$	$\langle 4,4\rangle$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Clearly, every ordered pair in $\mathbb{Z}^+ \times \mathbb{Z}^+$ will appear exactly once in the array. In particular, $\langle n, m\rangle$ will appear in the nth column and mth row. But how do we organize the elements of such an array into a one-way list? The pattern in the array below demonstrates one way to do this:

1	2	4	7	...
3	5	8	...	...
6	9	...	...	...
10	...	...	...	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

This pattern is called *Cantor's zig-zag method*. Other patterns are perfectly permissible, as long as they "zig-zag" through every cell of the array. By Cantor's zig-zag method, the enumeration for $\mathbb{Z}^+ \times \mathbb{Z}^+$ according to this scheme would be:

$$\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \langle 1,3\rangle, \langle 2,2\rangle, \langle 3,1\rangle, \langle 1,4\rangle, \langle 2,3\rangle, \langle 3,2\rangle, \langle 4,1\rangle, \ldots$$

What ought we do about enumerating, say, the set of ordered triples of positive integers?

$$\mathbb{Z}^+ \times \mathbb{Z}^+ \times \mathbb{Z}^+ = \{\langle n, m, k\rangle : n, m, k \in \mathbb{Z}^+\}$$

We can think of $\mathbb{Z}^+ \times \mathbb{Z}^+ \times \mathbb{Z}^+$ as the Cartesian product of $\mathbb{Z}^+ \times \mathbb{Z}^+$ and $\mathbb{Z}^+$, that is,

$$(\mathbb{Z}^+)^3 = (\mathbb{Z}^+ \times \mathbb{Z}^+) \times \mathbb{Z}^+ = \{\langle\langle n, m\rangle, k\rangle : \langle n, m\rangle \in \mathbb{Z}^+ \times \mathbb{Z}^+, k \in \mathbb{Z}^+\}$$

and thus we can enumerate $(\mathbb{Z}^+)^3$ with an array by labelling one axis with the enumeration of $\mathbb{Z}^+$, and the other axis with the enumeration of $(\mathbb{Z}^+)^2$:

	1	**2**	**3**	**4**	...
$\langle \mathbf{1}, \mathbf{1}\rangle$	$\langle 1,1,1\rangle$	$\langle 1,1,2\rangle$	$\langle 1,1,3\rangle$	$\langle 1,1,4\rangle$	...
$\langle \mathbf{1}, \mathbf{2}\rangle$	$\langle 1,2,1\rangle$	$\langle 1,2,2\rangle$	$\langle 1,2,3\rangle$	$\langle 1,2,4\rangle$	...
$\langle \mathbf{2}, \mathbf{1}\rangle$	$\langle 2,1,1\rangle$	$\langle 2,1,2\rangle$	$\langle 2,1,3\rangle$	$\langle 2,1,4\rangle$	...
$\langle \mathbf{1}, \mathbf{3}\rangle$	$\langle 1,3,1\rangle$	$\langle 1,3,2\rangle$	$\langle 1,3,3\rangle$	$\langle 1,3,4\rangle$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Thus, by using a method like Cantor's zig-zag method, we may similarly obtain an enumeration of $(\mathbb{Z}^+)^3$.

Cantor's zig-zag method makes the enumerability of $(\mathbb{Z}^+)^2$ (and analogously, $(\mathbb{Z}^+)^3$, etc.) visually evident. Following the zig-zag line in the array and counting the places, we can tell that $\langle 2, 3\rangle$ is at place 8, but specifying the inverse $g\colon (\mathbb{Z}^+)^2 \to \mathbb{Z}^+$ of the zig-zag enumeration such that

$$g(\langle 1,1\rangle) = 1, \quad g(\langle 1,2\rangle) = 2, \quad g(\langle 2,1\rangle) = 3, \quad \ldots \quad g(\langle 2,3\rangle) = 8, \quad \ldots$$

would be helpful. To calculate the position of each pair in the enumeration, we can use the function below. (The exact derivation of the function is somewhat messy, so we are skipping it here.)

$$g(n, m) = \frac{(n + m - 2)(n + m - 1)}{2} + n$$

Accordingly, the pair $\langle 2, 3\rangle$ is in position $((2+3-2)(2+3-1)/2)+2 = (3\cdot 4/2)+2 = (12/2)+2 = 8$; pair $\langle 3, 7\rangle$ is in position $((3+7-2)(3+7-1)/2)+3 = 39$.

Functions like g above, i.e., inverses of enumerations of sets of pairs, are called *pairing functions*.

Definition 4.8 (Pairing function)**.** A function $f\colon X \times Y \to \mathbb{Z}^+$ is an arithmetical *pairing function* if f is total and injective. We also say that f *encodes* $X \times Y$, and that for $f(\langle x, y\rangle) = n$, n is the *code* for $\langle x, y\rangle$.

explanation

The idea is that we can use such functions to encode, e.g., pairs of positive integers in $\mathbb{Z}^+$, or, in other words, represent pairs of positive integers as positive integers. Using the inverse of the pairing function, we can *decode* the integer, i.e., find out which pair of positive integers is represented.

There are other enumerations of $(\mathbb{Z}^+)^2$ that make it easier to figure out what their inverses are. Here is one. Instead of visualizing the enumeration in an array, start with the list of positive integers associated with (initially) empty spaces. Imagine filling these spaces successively with pairs $\langle n, m\rangle$ as

follow. Starting with the pairs that have 1 in the first place (i.e., pairs $\langle 1, m\rangle$), put the first (i.e., $\langle 1, 1\rangle$) in the first empty place, then skip an empty space, put the second (i.e., $\langle 1, 2\rangle$) in the next empty place, skip one again, and so forth. The (incomplete) beginning of our enumeration now looks like this

$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	$f(8)$	$f(9)$	$f(10)$	$\dots$
$\langle 1, 1\rangle$		$\langle 1, 2\rangle$		$\langle 1, 3\rangle$		$\langle 1, 4\rangle$		$\langle 1, 5\rangle$		$\dots$

Repeat this with pairs $\langle 2, m\rangle$ for the place that still remain empty, again skipping every other empty place:

$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	$f(8)$	$f(9)$	$f(10)$	$\dots$
$\langle 1, 1\rangle$	$\langle 2, 1\rangle$	$\langle 1, 2\rangle$		$\langle 1, 3\rangle$	$\langle 2, 2\rangle$	$\langle 1, 4\rangle$		$\langle 1, 5\rangle$	$\langle 2, 3\rangle$	$\dots$

Enter pairs $\langle 3, m\rangle$, $\langle 4, m\rangle$, etc., in the same way. Our completed enumeration thus starts like this:

$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	$f(8)$	$f(9)$	$f(10)$	$\dots$
$\langle 1, 1\rangle$	$\langle 2, 1\rangle$	$\langle 1, 2\rangle$	$\langle 3, 1\rangle$	$\langle 1, 3\rangle$	$\langle 2, 2\rangle$	$\langle 1, 4\rangle$	$\langle 4, 1\rangle$	$\langle 1, 5\rangle$	$\langle 2, 3\rangle$	$\dots$

If we number the cells in the array above according to this enumeration, we will not find a neat zig-zag line, but this arrangement:

	1	**2**	**3**	**4**	**5**	**6**	$\dots$
1	1	3	5	7	9	11	$\dots$
2	2	6	10	14	18	$\dots$	$\dots$
3	4	12	20	28	$\dots$	$\dots$	$\dots$
4	8	24	40	$\dots$	$\dots$	$\dots$	$\dots$
5	16	48	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
6	32	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We can see that the pairs in the first row are in the odd numbered places of our enumeration, i.e., pair $\langle 1, m\rangle$ is in place $2m - 1$; pairs in the second row, $\langle 1, m\rangle$, are in places whose number is the double of an odd number, specifically, $2 \cdot (2m - 1)$; pairs in the third row, $\langle 1, m\rangle$, are in places whose number is four times an odd number, $4 \cdot (2m - 1)$; and so on. The factors of $(2m - 1)$ for each row, $1, 2, 4, 8, \dots$, are powers of 2: $2^0, 2^1, 2^2, 2^3, \dots$ In fact, the relevant exponent is one less than the first member of the pair in question. Thus, for pair $\langle n, m\rangle$ the factor is $n-1$. This gives us the general formula: $2^{n-1} \cdot (2m-1)$, and hence:

Example 4.9. The function $f\colon (\mathbb{Z}^+)^2 \to \mathbb{Z}^+$ given by

$$h(n, m) = 2^{n-1}(2m - 1)$$

is a pairing function for the set of pairs of positive integers $(\mathbb{Z}^+)^2$.

explanation Accordingly, in our second enumeration of $(\mathbb{Z}^+)^2$, the pair $\langle 2, 3\rangle$ is in position $2^{2-1} \cdot (2 \cdot 3 - 1) = 2 \cdot 5 = 10$; pair $\langle 3, 7\rangle$ is in position $2^{3-1} \cdot (2 \cdot 7 - 1) = 52$.

Another common pairing function that encodes $(\mathbb{Z}^+)^2$ is the following:

Example 4.10. The function $f\colon (\mathbb{Z}^+)^2 \to \mathbb{Z}^+$ given by

$$j(n, m) = 2^n 3^m$$

is a pairing function for the set of pairs of positive integers $(\mathbb{Z}^+)^2$.

explanation j is injective, but nor surjective. That means the inverse of j is a partial, surjective function, and hence an enumeration of $(\mathbb{Z}^+)^2$. (Exercise.)

Problem 4.5. Give an enumeration of the set of all positive rational numbers. (A positive rational number is one that can be written as a fraction n/m with $n, m \in \mathbb{Z}^+$).

Problem 4.6. Show that $\mathbb{Q}$ is enumerable. (A rational number is one that can be written as a fraction z/m with $z \in \mathbb{Z}$, $m \in \mathbb{Z}^+$).

Problem 4.7. Define an enumeration of $\mathbb{B}^*$.

Problem 4.8. Recall from your introductory logic course that each possible truth table expresses a truth function. In other words, the truth functions are all functions from $\mathbb{B}^k \to \mathbb{B}$ for some k. Prove that the set of all truth functions is enumerable.

Problem 4.9. Show that the set of all finite subsets of an arbitrary infinite enumerable set is enumerable.

Problem 4.10. A set of positive integers is said to be *cofinite* iff it is the complement of a finite set of positive integers. Let I be the set that contains all the finite and cofinite sets of positive integers. Show that I is enumerable.

Problem 4.11. Show that the enumerable union of enumerable sets is enumerable. That is, whenever X_1, X_2, ... are sets, and each X_i is enumerable, then the union $\bigcup_{i=1}^{\infty} X_i$ of all of them is also enumerable.

Problem 4.12. Let $f : X \times Y \to \mathbb{Z}^+$ be an arbitrary pairing function. Show that the inverse of f is an enumeration of $X \times Y$.

Problem 4.13. Specify a function that encodes $\mathbb{N}^3$.

4.3 Non-enumerable Sets

Some sets, such as the set $\mathbb{Z}^+$ of positive integers, are infinite. So far we've seen examples of infinite sets which were all enumerable. However, there are also infinite sets which do not have this property. Such sets are called *non-enumerable*. sfr:siz:nen:sec

First of all, it is perhaps already surprising that there are non-enumerable sets. For any enumerable set X there is a surjective function $f\colon \mathbb{Z}^+ \to X$. If a set is non-enumerable there is no such function. That is, no function mapping the infinitely many elements of $\mathbb{Z}^+$ to X can exhaust all of X. So there are "more" elements of X than the infinitely many positive integers.

How would one prove that a set is non-enumerable? You have to show that no such surjective function can exist. Equivalently, you have to show that the elements of X cannot be enumerated in a one way infinite list. The best way to do this is to show that every list of elements of X must leave at least one element out; or that no function $f\colon \mathbb{Z}^+ \to X$ can be surjective. We can do this using Cantor's *diagonal method*. Given a list of elements of X, say, $x_1, x_2, \ldots$, we construct another element of X which, by its construction, cannot possibly be on that list.

Our first example is the set $\mathbb{B}^\omega$ of all infinite, non-gappy sequences of 0's and 1's.

sfr:siz:nen: thm-nonenum-bin-omega

Theorem 4.11. *$\mathbb{B}^\omega$ is non-enumerable.*

Proof. Suppose, by way of contradiction, that $\mathbb{B}^\omega$ is enumerable, i.e., suppose that there is a list $s_1, s_2, s_3, s_4, \ldots$ of all elements of $\mathbb{B}^\omega$. Each of these s_i is itself an infinite sequence of 0's and 1's. Let's call the j-th element of the i-th sequence in this list $s_i(j)$. Then the i-th sequence s_i is

$$s_i(1), s_i(2), s_i(3), \ldots$$

We may arrange this list, and the elements of each sequence s_i in it, in an array:

	1	2	3	4	...
1	$\mathbf{s_1(1)}$	$s_1(2)$	$s_1(3)$	$s_1(4)$	...
2	$s_2(1)$	$\mathbf{s_2(2)}$	$s_2(3)$	$s_2(4)$	...
3	$s_3(1)$	$s_3(2)$	$\mathbf{s_3(3)}$	$s_3(4)$	...
4	$s_4(1)$	$s_4(2)$	$s_4(3)$	$\mathbf{s_4(4)}$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

The labels down the side give the number of the sequence in the list $s_1, s_2, \ldots$; the numbers across the top label the elements of the individual sequences. For instance, $s_1(1)$ is a name for whatever number, a 0 or a 1, is the first element in the sequence s_1, and so on.

Now we construct an infinite sequence, $\overline{s}$, of 0's and 1's which cannot possibly be on this list. The definition of $\overline{s}$ will depend on the list $s_1, s_2, \ldots$. Any infinite list of infinite sequences of 0's and 1's gives rise to an infinite sequence $\overline{s}$ which is guaranteed to not appear on the list.

To define $\overline{s}$, we specify what all its elements are, i.e., we specify $\overline{s}(n)$ for all $n \in \mathbb{Z}^+$. We do this by reading down the diagonal of the array above (hence the name "diagonal method") and then changing every 1 to a 0 and every 1

to a 0. More abstractly, we define $\overline{s}(n)$ to be 0 or 1 according to whether the n-th element of the diagonal, $s_n(n)$, is 1 or 0.

$$\overline{s}(n) = \begin{cases} 1 & \text{if } s_n(n) = 0 \\ 0 & \text{if } s_n(n) = 1. \end{cases}$$

If you like formulas better than definitions by cases, you could also define $\overline{s}(n) = 1 - s_n(n)$.

Clearly $\overline{s}$ is a non-gappy infinite sequence of 0's and 1's, since it is just the mirror sequence to the sequence of 0's and 1's that appear on the diagonal of our array. So $\overline{s}$ is an element of $\mathbb{B}^\omega$. But it cannot be on the list s_1, s_2, ... Why not?

It can't be the first sequence in the list, s_1, because it differs from s_1 in the first element. Whatever $s_1(1)$ is, we defined $\overline{s}(1)$ to be the opposite. It can't be the second sequence in the list, because $\overline{s}$ differs from s_2 in the second element: if $s_2(2)$ is 0, $\overline{s}(2)$ is 1, and vice versa. And so on.

More precisely: if $\overline{s}$ were on the list, there would be some k so that $\overline{s} = s_k$. Two sequences are identical iff they agree at every place, i.e., for any n, $\overline{s}(n) = s_k(n)$. So in particular, taking $n = k$ as a special case, $\overline{s}(k) = s_k(k)$ would have to hold. $s_k(k)$ is either 0 or 1. If it is 0 then $\overline{s}(k)$ must be 1—that's how we defined $\overline{s}$. But if $s_k(k) = 1$ then, again because of the way we defined $\overline{s}$, $\overline{s}(k) = 0$. In either case $\overline{s}(k) \neq s_k(k)$.

We started by assuming that there is a list of elements of $\mathbb{B}^\omega$, s_1, s_2, ... From this list we constructed a sequence $\overline{s}$ which we proved cannot be on the list. But it definitely is a sequence of 0's and 1's if all the s_i are sequences of 0's and 1's, i.e., $\overline{s} \in \mathbb{B}^\omega$. This shows in particular that there can be no list of *all* elements of $\mathbb{B}^\omega$, since for any such list we could also construct a sequence $\overline{s}$ guaranteed to not be on the list, so the assumption that there is a list of all sequences in $\mathbb{B}^\omega$ leads to a contradiction. □

explanation

This proof method is called "diagonalization" because it uses the diagonal of the array to define $\overline{s}$. Diagonalization need not involve the presence of an array: we can show that sets are not enumerable by using a similar idea even when no array and no actual diagonal is involved.

Theorem 4.12. *$\wp(\mathbb{Z}^+)$ is not enumerable.* *sfr:siz:nen: thm-nonenum-pownat*

Proof. We proceed in the same way, by showing that for every list of subsets of $\mathbb{Z}^+$ there is a subset of $\mathbb{Z}^+$ which cannot be on the list. Suppose the following is a given list of subsets of $\mathbb{Z}^+$:

$$Z_1, Z_2, Z_3, \ldots$$

We now define a set $\overline{Z}$ such that for any $n \in \mathbb{Z}^+$, $n \in \overline{Z}$ iff $n \notin Z_n$:

$$\overline{Z} = \{n \in \mathbb{Z}^+ : n \notin Z_n\}$$

$\overline{Z}$ is clearly a set of positive integers, since by assumption each Z_n is, and thus $\overline{Z} \in \wp(\mathbb{Z}^+)$. But $\overline{Z}$ cannot be on the list. To show this, we'll establish that for each $k \in \mathbb{Z}^+$, $\overline{Z} \neq Z_k$.

So let $k \in \mathbb{Z}^+$ be arbitrary. We've defined $\overline{Z}$ so that for any $n \in \mathbb{Z}^+$, $n \in \overline{Z}$ iff $n \notin Z_n$. In particular, taking $n = k$, $k \in \overline{Z}$ iff $k \notin Z_k$. But this shows that $\overline{Z} \neq Z_k$, since k is an element of one but not the other, and so $\overline{Z}$ and Z_k have different elements. Since k was arbitrary, $\overline{Z}$ is not on the list $Z_1, Z_2, \ldots$ □

The preceding proof did not mention a diagonal, but you can think of it as involving a diagonal if you picture it this way: Imagine the sets $Z_1, Z_2, \ldots$, written in an array, where each element $j \in Z_i$ is listed in the j-th column. Say the first four sets on that list are $\{1, 2, 3, \ldots\}$, $\{2, 4, 6, \ldots\}$, $\{1, 2, 5\}$, and $\{3, 4, 5, \ldots\}$. Then the array would begin with explanation

$$\begin{array}{lcccccccc}
Z_1 = \{ & \mathbf{1}, & 2, & 3, & 4, & 5, & 6, & \ldots\} \\
Z_2 = \{ & & \mathbf{2}, & & 4, & & 6, & \ldots\} \\
Z_3 = \{ & 1, & 2, & & & 5 & & \} \\
Z_4 = \{ & & & 3, & \mathbf{4}, & 5, & 6, & \ldots\} \\
\vdots & & & & & \ddots & &
\end{array}$$

Then $\overline{Z}$ is the set obtained by going down the diagonal, leaving out any numbers that appear along the diagonal and include those j where the array has a gap in the j-th row/column. In the above case, we would leave out 1 and 2, include 3, leave out 4, etc.

Problem 4.14. Show that $\wp(\mathbb{N})$ is non-enumerable by a diagonal argument.

Problem 4.15. Show that the set of functions $f\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ is non-enumerable by an explicit diagonal argument. That is, show that if $f_1, f_2, \ldots$, is a list of functions and each $f_i\colon \mathbb{Z}^+ \to \mathbb{Z}^+$, then there is some $\overline{f}\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ not on this list.

4.4 Reduction

sfr:siz:red:sec We showed $\wp(\mathbb{Z}^+)$ to be non-enumerable by a diagonalization argument. We already had a proof that $\mathbb{B}^\omega$, the set of all infinite sequences of 0s and 1s, is non-enumerable. Here's another way we can prove that $\wp(\mathbb{Z}^+)$ is non-enumerable: Show that *if* $\wp(\mathbb{Z}^+)$ *is enumerable then* $\mathbb{B}^\omega$ *is also enumerable*. Since we know $\mathbb{B}^\omega$ is not enumerable, $\wp(\mathbb{Z}^+)$ can't be either. This is called *reducing* one problem to another—in this case, we reduce the problem of enumerating $\mathbb{B}^\omega$ to the problem of enumerating $\wp(\mathbb{Z}^+)$. A solution to the latter—an enumeration of $\wp(\mathbb{Z}^+)$—would yield a solution to the former—an enumeration of $\mathbb{B}^\omega$.

How do we reduce the problem of enumerating a set Y to that of enumerating a set X? We provide a way of turning an enumeration of X into an enumeration of Y. The easiest way to do that is to define a surjective function $f\colon X \to Y$. If $x_1, x_2, \ldots$ enumerates X, then $f(x_1), f(x_2), \ldots$ would enumerate Y. In our case, we are looking for a surjective function $f\colon \wp(\mathbb{Z}^+) \to \mathbb{B}^\omega$.

Problem 4.16. Show that if there is an injective function $g\colon Y \to X$, and Y is non-enumerable, then so is X. Do this by showing how you can use g to turn an enumeration of X into one of Y.

Proof of Theorem 4.12 by reduction. Suppose that $\wp(\mathbb{Z}^+)$ were enumerable, and thus that there is an enumeration of it, Z_1, Z_2, Z_3, ...

Define the function $f\colon \wp(\mathbb{Z}^+) \to \mathbb{B}^\omega$ by letting $f(Z)$ be the sequence s_k such that $s_k(n) = 1$ iff $n \in Z$, and $s_k(n) = 0$ otherwise. This clearly defines a function, since whenever $Z \subseteq \mathbb{Z}^+$, any $n \in \mathbb{Z}^+$ either is an element of Z or isn't. For instance, the set $2\mathbb{Z}^+ = \{2, 4, 6, \dots\}$ of positive even numbers gets mapped to the sequence $010101\dots$, the empty set gets mapped to $0000\dots$ and the set $\mathbb{Z}^+$ itself to $1111\dots$.

It also is surjective: Every sequence of 0s and 1s corresponds to some set of positive integers, namely the one which has as its members those integers corresponding to the places where the sequence has 1s. More precisely, suppose $s \in \mathbb{B}^\omega$. Define $Z \subseteq \mathbb{Z}^+$ by:

$$Z = \{n \in \mathbb{Z}^+ : s(n) = 1\}$$

Then $f(Z) = s$, as can be verified by consulting the definition of f.

Now consider the list

$$f(Z_1), f(Z_2), f(Z_3), \dots$$

Since f is surjective, every member of $\mathbb{B}^\omega$ must appear as a value of f for some argument, and so must appear on the list. This list must therefore enumerate all of $\mathbb{B}^\omega$.

So if $\wp(\mathbb{Z}^+)$ were enumerable, $\mathbb{B}^\omega$ would be enumerable. But $\mathbb{B}^\omega$ is non-enumerable (Theorem 4.11). Hence $\wp(\mathbb{Z}^+)$ is non-enumerable. □

explanation

It is easy to be confused about the direction the reduction goes in. For instance, a surjective function $g\colon \mathbb{B}^\omega \to X$ does *not* establish that X is non-enumerable. (Consider $g\colon \mathbb{B}^\omega \to \mathbb{B}$ defined by $g(s) = s(1)$, the function that maps a sequence of 0's and 1's to its first element. It is surjective, because some sequences start with 0 and some start with 1. But $\mathbb{B}$ is finite.) Note also that the function f must be surjective, or otherwise the argument does not go through: $f(x_1)$, $f(x_2)$, ... would then not be guaranteed to include all the elements of Y. For instance, $h\colon \mathbb{Z}^+ \to \mathbb{B}^\omega$ defined by

$$h(n) = \underbrace{000\dots0}_{n\ 0\text{'s}}$$

is a function, but $\mathbb{Z}^+$ is enumerable.

Problem 4.17. Show that the set of all *sets of* pairs of positive integers is non-enumerable by a reduction argument.

Problem 4.18. Show that $\mathbb{N}^\omega$, the set of infinite sequences of natural numbers, is non-enumerable by a reduction argument.

Problem 4.19. Let P be the set of functions from the set of positive integers to the set $\{0\}$, and let Q be the set of *partial* functions from the set of positive integers to the set $\{0\}$. Show that P is enumerable and Q is not. (Hint: reduce the problem of enumerating $\mathbb{B}^\omega$ to enumerating Q).

Problem 4.20. Let S be the set of all surjective functions from the set of positive integers to the set $\{0,1\}$, i.e., S consists of all surjective $f\colon \mathbb{Z}^+ \to \mathbb{B}$. Show that S is non-enumerable.

Problem 4.21. Show that the set $\mathbb{R}$ of all real numbers is non-enumerable.

4.5 Equinumerous Sets

sfr:set:equ:sec We have an intuitive notion of "size" of sets, which works fine for finite sets. But what about infinite sets? If we want to come up with a formal way of comparing the sizes of two sets of *any* size, it is a good idea to start with defining when sets are the same size. Let's say sets of the same size are *equinumerous*. We want the formal notion of equinumerosity to correspond with our intuitive notion of "same size," hence the formal notion ought to satisfy the following properties: intro

Reflexivity: Every set is equinumerous with itself.

Symmetry: For any sets X and Y, if X is equinumerous with Y, then Y is equinumerous with X.

Transitivity: For any sets X, Y, and Z, if X is equinumerous with Y and Y is equinumerous with Z, then X is equinumerous with Z.

In other words, we want equinumerosity to be an *equivalence relation*.

Definition 4.13. A set X is *equinumerous* with a set Y, $X \approx Y$, if and only if there is a bijective $f\colon X \to Y$.

Proposition 4.14. *Equinumerosity defines an equivalence relation.*

Proof. Let X, Y, and Z be sets.

Reflexivity: Using the identity map $1_X\colon X \to X$, where $1_X(x) = x$ for all $x \in X$, we see that X is equinumerous with itself (clearly, 1_X is bijective).

Symmetry: Suppose that X is equinumerous with Y. Then there is a bijective $f\colon X \to Y$. Since f is bijective, its inverse f^{-1} exists and also bijective. Hence, $f^{-1}\colon Y \to X$ is a bijective function from Y to X, so Y is also equinumerous with X.

Transitivity: Suppose that X is equinumerous with Y via the bijective function $f\colon X \to Y$ and that Y is equinumerous with Z via the bijective function $g\colon Y \to Z$. Then the composition of $g \circ f\colon X \to Z$ is bijective, and X is thus equinumerous with Z.

Therefore, equinumerosity is an equivalence relation. □

Theorem 4.15. *Suppose X and Y are equinumerous. Then X is enumerable if and only if Y is.*

Proof. Let X and Y be equinumerous. Suppose that X is enumerable. Then either $X = \emptyset$ or there is a surjective function $f\colon \mathbb{Z}^+ \to X$. Since X and Y are equinumerous, there is a bijective $g\colon X \to Y$. If $X = \emptyset$, then $Y = \emptyset$ also (otherwise there would be an element $y \in Y$ but no $x \in X$ with $g(x) = y$). If, on the other hand, $f\colon \mathbb{Z}^+ \to X$ is surjective, then $g \circ f\colon \mathbb{Z}^+ \to Y$ is surjective. To see this, let $y \in Y$. Since g is surjective, there is an $x \in X$ such that $g(x) = y$. Since f is surjective, there is an $n \in \mathbb{Z}^+$ such that $f(n) = x$. Hence,

$$(g \circ f)(n) = g(f(n)) = g(x) = y$$

and thus $g \circ f$ is surjective. We have that $g \circ f$ is an enumeration of Y, and so Y is enumerable. □

Problem 4.22. Show that if X is equinumerous with U and and Y is equinumerous with V, and the intersections $X \cap Y$ and $U \cap V$ are empty, then the unions $X \cup Y$ and $U \cup V$ are equinumerous.

Problem 4.23. Show that if X is infinite and enumerable, then it is equinumerous with the positive integers $\mathbb{Z}^+$.

4.6 Comparing Sizes of Sets

explanation Just like we were able to make precise when two sets have the same size in a way that also accounts for the size of infinite sets, we can also compare the sizes of sets in a precise way. Our definition of "is smaller than (or equinumerous)" will require, instead of a bijection between the sets, a total injective function from the first set to the second. If such a function exists, the size of the first set is less than or equal to the size of the second. Intuitively, an injective function from one set to another guarantees that the range of the function has at least as many elements as the domain, since no two elements of the domain map to the same element of the range. sfr:siz:car:sec

Definition 4.16. X is *no larger than* Y, $X \preceq Y$, if and only if there is an injective function $f\colon X \to Y$.

Theorem 4.17 (Schröder-Bernstein)**.** *Let X and Y be sets. If $X \preceq Y$ and $Y \preceq X$, then $X \approx Y$.*

explanation In other words, if there is a total injective function from X to Y, and if there is a total injective function from Y back to X, then there is a total bijection from X to Y. Sometimes, it can be difficult to think of a bijection between two equinumerous sets, so the Schröder-Bernstein theorem allows us to break the comparison down into cases so we only have to think of an injection from

the first to the second, and vice-versa. The Schröder-Bernstein theorem, apart from being convenient, justifies the act of discussing the "sizes" of sets, for it tells us that set cardinalities have the familiar anti-symmetric property that numbers have.

Definition 4.18. X is *smaller than* Y, $X \prec Y$, if and only if there is an injective function $f\colon X \to Y$ but no bijective $g\colon X \to Y$.

sfr:siz:car:thm:cantor **Theorem 4.19** (Cantor)**.** *For all X, $X \prec \wp(X)$.*

Proof. The function $f\colon X \to \wp(X)$ that maps any $x \in X$ to its singleton $\{x\}$ is injective, since if $x \neq y$ then also $f(x) = \{x\} \neq \{y\} = f(y)$.

There cannot be a surjective function $g\colon X \to \wp(X)$, let alone a bijective one. For suppose that $g\colon X \to \wp(X)$. Since g is total, every $x \in X$ is mapped to a subset $g(x) \subseteq X$. We show that g cannot be surjective. To do this, we define a subset $Y \subseteq X$ which by definition cannot be in the range of g. Let

$$\overline{Y} = \{x \in X : x \notin g(x)\}.$$

Since $g(x)$ is defined for all $x \in X$, $\overline{Y}$ is clearly a well-defined subset of X. But, it cannot be in the range of g. Let $x \in X$ be arbitrary, we show that $\overline{Y} \neq g(x)$. If $x \in g(x)$, then it does not satisfy $x \notin g(x)$, and so by the definition of $\overline{Y}$, we have $x \notin \overline{Y}$. If $x \in \overline{Y}$, it must satisfy the defining property of $\overline{Y}$, i.e., $x \notin g(x)$. Since x was arbitrary this shows that for each $x \in X$, $x \in g(x)$ iff $x \notin \overline{Y}$, and so $g(x) \neq \overline{Y}$. So $\overline{Y}$ cannot be in the range of g, contradicting the assumption that g is surjective. □

It's instructive to compare the proof of Theorem 4.19 to that of Theorem 4.12. There we showed that for any list $Z_1, Z_2, \ldots,$ of subsets of $\mathbb{Z}^+$ one can construct a set $\overline{Z}$ of numbers guaranteed not to be on the list. It was guaranteed not to be on the list because, for every $n \in \mathbb{Z}^+$, $n \in Z_n$ iff $n \notin \overline{Z}$. This way, there is always some number that is an element of one of Z_n and $\overline{Z}$ but not the other. We follow the same idea here, except the indices n are now elements of X instead of $\mathbb{Z}^+$. The set $\overline{Y}$ is defined so that it is different from $g(x)$ for each $x \in X$, because $x \in g(x)$ iff $x \notin \overline{Y}$. Again, there is always an element of X which is an element of one of $g(x)$ and $\overline{Y}$ but not the other. And just as $\overline{Z}$ therefore cannot be on the list $Z_1, Z_2, \ldots, \overline{Y}$ cannot be in the range of g. explanation

Problem 4.24. Show that there cannot be an injective function $g\colon \wp(X) \to X$, for any set X. Hint: Suppose $g\colon \wp(X) \to X$ is injective. Then for each $x \in X$ there is at most one $Y \subseteq X$ such that $g(Y) = x$. Define a set $\overline{Y}$ such that for every $x \in X$, $g(\overline{Y}) \neq x$.

Photo Credits

Bibliography