CDM

Sets

Klaus Sutner

Carnegie Mellon University

10-sets 2017/12/15 23:22

1. Intuitive Set Theory

- Extensionality
- Frege's Ghost
- Products and Sums
- Cantor
- Formalization
- Beyond ZFC

Here is a first attempt at a "definition" of sets.

Definition

A set is an arbitrary collection of objects.

In the words of Georg Cantor:

> *By an "aggregate" we are to understand any collection into a whole M of definite and separate objects m of our intuition or our thought. The objects are called "elements" of M. In signs we express this thus: $M = \{m\}$.*

Cantor's symbolic notation is rather old-fashioned. Nowadays and following G. Peano one would usually write

$$M = \{\, m \mid P(m) \,\}$$

indicating that we wish to collect all objects m that have property P into a set M.

This set formation principle is the core of set theory. As we will see in a while, it also causes major problems.

As it turns out, sets alone (well, plus a bit of logic) suffice as a foundation of more or less all of mathematics and computer science. Set theory provides an extremely powerful and even elegant way to organize and structure any discourse in this domain.

On the face of it, this is a huge surprise: one would suspect that sets are nowhere near powerful enough to express concepts such as natural number, prime, group, field, real number, differentiable function, probability measure, finite state machine, computable function, complexity class, and so on.

The importance of sets can be seen in the fact that Bourbaki chose to dedicate his first volume to sets:

I Set Theory
II Algebra
III Topology
IV Functions of one real variable
V Topological vector spaces
VI Integration
VII Commutative algebra
VIII Lie groups

In many ways, the Bourbaki approach is still the gold standard.

Of course, there is an immediate question: sets of what?

For example, in arithmetic we would like to have sets of natural numbers such as the primes. In calculus we would like to have sets of reals such as intervals. In algebra we are dealing with groups, rings, fields, polynomials, matrices and so on. But in pure set theory none of these objects exist.

One way around this problem is to assume that we are given a collection $\mathcal{U}$ of urelements. We are allowed to form subsets of $\mathcal{U}$, subsets of subsets, combinations with pure sets, and so on.

Different sets of urelements are appropriate for different areas of discourse. The approach is quite practical and is often used tacitly, without any mention of the underlying idea.

As we will see below, urelements are superfluous in the sense that the "stuff of mathematics and computer science" can always be represented in terms of pure sets. Unfortunately, a complete definition of say, the real numbers, in terms of sets only is quite unwieldy.

One winds up with complicated, infinite sets of sets of sets of sets . . . that have at best a tenuous connection to anyone's intuition about what a real number is. But, one can then give a rigorous proof that the reals are complete.

The real purpose of set-theoretic definitions is to provide a precise standard, a solid reference for all the more informal notions that one uses in actual practice.

Higher levels of abstraction are indispensable for any real application, but set theory provides the bedrock foundation.

We can think of a set as a kind of general purpose data structure – a container type. Since we are dealing with arbitrary collections there are several natural operations:

- insert, remove, membership test, ...
- union, intersection, difference, size, ...

Warning: Sets are very different from sequential collections such as lists or arrays:

- Sets are not ordered, there is no first, second, ..., last element.
- There are no multiple elements.

As a matter of fact, sets are notoriously difficult to implement; sequential containers such as lists are much easier to deal with.

Standard notation for finite sets treats them very much like lists, as a sequential container:

$$S = \{a_1, a_2, \ldots, a_{n-1}, a_n\} \qquad \text{set formation}$$
$$\emptyset = \{\} \qquad \text{empty set}$$

But note that this notation is a bit misleading in that we must have

$$\{a, b, c\} = \{b, a, a, c, a, c, b\}$$

Duplicates and order are irrelevant for sets. This is the crux in implementing sets rather than plain lists; we have to make an extra effort to avoid duplication and to ignore order (though elements in a data structure are always ordered in some way, e.g. by the order of insertion which is, of course, extraneous to the actual collection).

There is only one fundamental relation between sets: membership.

Notation

$x \in y$ x is an element of y.

Example

$$5 \in \{2, 3, 5, 8\}$$
$$7 \notin \{2, 3, 5, 8\}$$
$$z \in [0, 1) \iff 0 \leq z < 1$$
$$x \notin \emptyset \quad \text{for any } x$$

As we will see in a moment, for sets even equality can be reduced to membership.

For infinite collections that have a very simple, regular structure we can still use the same notation, augmented by ellipses:

$$\begin{aligned}\text{Even} &= \{\ldots, -2n, \ldots, -2, 0, 2, \ldots, 2n, \ldots\} \\ \text{Primes} &= \{2, 3, 5, 7, \ldots, 1299709, \ldots\}\end{aligned}$$

Note, though, that these definitions depend on the ability of the reader to interpret the dots. As input to a program ellipses are usually not allowed (except in very restricted circumstances such as to denote intervals of integers).

In the examples above, we presumably have the integers as urelements, or we could represented them as pure sets, see below.

To obtain complicated sets we can collect all objects z (all of them sets or urelements) with a certain property $P(z)$ into one set:

$$A = \{\, z \mid P(z) \,\}$$

Very often one selects elements from some larger collection B that has already been constructed.

$$A = \{\, z \in B \mid P(z) \,\}$$

Example

$$\begin{aligned} [n] &= \{\, z \in \mathbb{N} \mid 1 \leq z \leq n \,\} \\ \mathbb{P} &= \{\, z \in \mathbb{N} \mid z \text{ is prime } \} \\ [0,1) &= \{\, z \in \mathbb{R} \mid 0 \leq z < 1 \,\} \\ \mathbb{Q} &= \{\, a/b \mid a, b \in \mathbb{Z}, b > 0 \,\} \end{aligned}$$

There are several simple operations on sets that are useful to construct new sets from given ones.

$$\begin{array}{rll} \text{union} & A \cup B & = \{\, x \mid x \in A \vee x \in B \,\} \\ \text{intersection} & A \cap B & = \{\, x \mid x \in A \wedge x \in B \,\} \\ \text{difference} & A - B & = \{\, x \mid x \in A \wedge x \notin B \,\} \\ \text{symmetric diff.} & A \oplus B & = \{\, x \mid x \in A \oplus x \in B \,\} \end{array}$$

Here $\oplus$ means exclusive or. Incidentally, the symbols $\cup$ and $\cap$ are also due to Peano. So all this boils down to simple propositional logic.

Example

$$\{1,2,3\} \cap \{2,3,4,5\} = \{2,3\}$$
$$\{1,2,3\} \cup \{2,3,4,5\} = \{1,2,3,4,5\}$$
$$\{1,2,3\} \oplus \{2,3,4,5\} = \{1,4,5\}$$

Basic Set Properties

- Associativity
  $x \cup (y \cup z) = (x \cup y) \cup z$ and
  $x \cap (y \cap z) = (x \cap y) \cap z$.
- Commutativity
  $x \cup y = y \cup x$ and $x \cap y = y \cap x$.
- Distributivity
  $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$ and
  $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$.
- Identity
  $x \cup \emptyset = x$ and $x \cap \mathcal{U} = x$.
- Idempotence
  $x \cup x = x$ and $x \cap x = x$.
- Absorption
  $x \cup (x \cap y) = x$ and $x \cap (x \cup y) = x$.

All these properties hold intuitively and one only needs a bit of propositional logic to come up with proofs.

For instance, associativity of intersection and union comes down to associativity of logical conjunction and disjunction.

Exercise

Construct proofs for all the basic set properties.

Complements

In general "the complement of set x" makes no sense: before we can remove the elements of x we have to determine what they should be removed from. In other words, we need to fix some universe $\mathcal{U}$ and consider only $x \subseteq \mathcal{U}$:

$$x^- = \mathcal{U} - x.$$

Now assume $x, y \subseteq \mathcal{U}$ for some fixed universe $\mathcal{U}$. Then we have

- Domination
  $x \cup \mathcal{U} = \mathcal{U}$ and $x \cap \emptyset = \emptyset$.
- Complements
  $x \cup x^- = \mathcal{U}$ and $x \cap x^- = \emptyset$.
- Double Complement (involution):
  $x^{-\,-} = x$.
- De Morgan's Laws:
  $(x \cup y)^- = x^- \cap y^-$ and $(x \cap y)^- = x^- \cup y^-$

These rules should all look eminently familiar: they are essentially the axioms for a Boolean algebra.

Of course, this is no coincidence: Fix some universe $\mathcal{U}$ and interpret $+$ by $\cup$, $\cdot$ by $\cap$, and $\overline{x}$ as $x^- = \mathcal{U} - x$. We obtain a structure

$$\langle \mathfrak{P}(\mathcal{U}), \cup, \cap, ^-, \emptyset, \mathcal{U} \rangle$$

Lemma

The powerset of $\mathcal{U}$ with the operations union, intersection and complement forms a Boolean algebra.

Inquisitive minds might wonder why true/false should behave just like all subsets of a fixed universe $\mathcal{U}$. We'll come up with a fairly good explanation in a while.

Definition

Let X be a set (intended: a set of sets).

$$\bigcup X = \{\, z \mid \exists\, x\, (z \in x \wedge x \in X)\,\}$$
$$\bigcap X = \{\, z \mid \forall\, x\, (x \in X \Rightarrow z \in x)\,\}$$

These definitions are a bit easier to read if we write

$$\bigcup X = \{\, z \mid \exists\, x\ \in X (z \in x)\,\}$$
$$\bigcap X = \{\, z \mid \forall\, x \in X\, (z \in x)\,\}$$

Exercise

Show that $\bigcup\{a, b\} = a \cup b$ *and* $\bigcap\{a, b\} = a \cap b$.

Example: Generating Subgroups

A typical example of the use of the general intersection operator is to guarantee the existence of a subgroup generated by some elements $A \subseteq G$, where G is some group.

In this case let

$$X = \{ H \subseteq G \mid A \subseteq H, H \text{ subgroup } \}$$

Then X is not empty since certainly $G \in X$.

But subgroups are closed under intersection, so $\bigcap X$ must be the subgroup we are after.

Note that from a certain perspective this argument is a bit circular: $\bigcap X \in X$, so we are using an object to define itself. This is called an impredicative definition.

Unary union/intersection is particularly useful for indexed families of sets:

$$X = \{ x_i \mid i \in I \}: \qquad \bigcup X = \bigcup_{i \in I} x_i$$

where I is some index set, typically $[n]$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ or some such.

Example

$$\bigcup_{\varepsilon > 0} [0, 1 - \varepsilon] = [0, 1) \subseteq \mathbb{R}$$

$$\bigcap_{\varepsilon > 0} [0, 1 + \varepsilon] = [0, 1] \subseteq \mathbb{R}$$

- Intuitive Set Theory

2. Extensionality

- Frege's Ghost
- Products and Sums
- Cantor
- Formalization
- Beyond ZFC

Two sets are considered to be the same iff they contain precisely the same elements.

$$x = y \iff \forall\, z\, (z \in x \Leftrightarrow z \in y)$$

The importance of this principle was first recognized by Leibniz and is enshrined in his *principium identitatis indiscernibilium*: if we cannot tell two entities apart they are identical. Later G. Frege incorporated this principle in his system as the infamous Axiom V, a decision that would cause him major headaches.

It is a consequence of Extensionality that, in sets, order is irrelevant and there are no multiple occurrences.

For example, $\{1, 2, 3\} = \{3, 2, 1, 2, 1, 3\}$ simply because the elements of both sets are the same.

A more subtle point is the following: the description of the set is also irrelevant, all that matters are the actual elements.

Here is an example of two sets of natural numbers:

$$A = \{1, 2\}$$
$$B = \{\, n \in \mathbb{N}^+ \mid x^n + y^n = z^n \text{ has solution in } \mathbb{N}^+ \,\}$$

Then $A = B$, but this is Fermat's Last "theorem" and requires a very complicated, non-elementary proof (at least at present no one knows of a short, elementary proof and there are good reasons to believe that none exists).

So equality of sets can be exceedingly complicated even when one of the sets in question is finite.

Definition

x is a subset of y if

$$x \subseteq y \text{ if } \forall z\,(z \in x \Rightarrow z \in y).$$

x is a proper subset of y, $x \subset y$ if $x \subseteq y \wedge x \neq y$

Thus $x = y \iff x \subseteq y \wedge y \subseteq x$.

Note that some authors write $x \subset y$ to indicate that x is a subset of y. We always indicate the weak version of an order by adding a line of sorts to the less-than symbols.

Example

For any set S we always have $\emptyset \subseteq S$.

Example

We have $\mathbb{N} \supset 2\mathbb{N} \supset 4\mathbb{N} \supset \ldots$, so the subset relation is not well-founded.

Example

For arithmetic types we have $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

The last example may seem blindingly obvious and boring, but it is actually quite tricky. There are many reasonable ways to define these arithmetic sets of numbers so that the inclusions do not hold.

To see where this problem comes from, suppose we have the natural numbers as urelements and we also have addition on them. We can define the integers as follows: Let ρ be the equivalence relation on $\mathbb{N} \times \mathbb{N}$ given by

$$(a, b)\ \rho\ (c, d) \iff a + d = c + b.$$

Then the integers can be defined as the quotient structure

$$\mathbb{Z} = \mathbb{N} \times \mathbb{N}/\rho.$$

Of course, according to this definition $\mathbb{Z}$ is not a subset of $\mathbb{N}$.

But we can identify $\mathbb{N}$ with such a subset: the $n \geq 0$ is identified with

$$\{\, (a, b) \in \mathbb{N} \times \mathbb{N} \mid a = b + n \,\}.$$

Painful, but very precise. And quite similar to what happens in programming.

We can easily express the Induction Principle on $\mathbb{N}$ using subsets:

Lemma

Let $A \subseteq \mathbb{N}$. Suppose $0 \in A$ and for all x: $x \in A \Rightarrow (x+1) \in A$. Then $A = \mathbb{N}$.

Proof. Suppose otherwise, so $A \neq \mathbb{N}$. So there exists an x, $x \notin A$. Pick the least $a \notin A$. Clearly $a \neq 0$. But then $a - 1 \in A$, contradicting the second hypothesis.

□

Of course, this is but the tip of an iceberg: many other sets are also generated by applying certain operations (here: the successor function) to given primitive elements (here: 0). They all obey a similar induction principle.

Definition

The powerset of a set is the set of all its subsets.

Notation:

$$\mathfrak{P}(y) = \{\, x \mid x \subseteq y \,\}$$

Example

$$\mathfrak{P}(\emptyset) = \{\emptyset\}$$
$$\mathfrak{P}(\{a\}) = \{\emptyset, \{a\}\}$$
$$\mathfrak{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$$

Powersets appear to be much larger than the original sets.

Lemma

If A has n elements, then $\mathfrak{P}(A)$ has 2^n elements.

Proof. By induction, $n = 0$ is obvious.

So suppose $A = \{a_1, a_2, \ldots, a_{n-1}, a_n\}$, $n \geq 1$, and let $x \subseteq A$.

Case 1: $a_n \notin x$.

So x is a subset of $\{a_1, a_2, \ldots, a_{n-1}\}$. There are 2^{n-1} such subsets by IH.

Case 2: $a_n \in x$.

Then $x = \{a_n\} \cup y$ where y is a subset of $\{a_1, a_2, \ldots, a_{n-1}\}$. Again, there are 2^{n-1} such subsets by IH.

Total: $2^{n-1} + 2^{n-1} = 2^n$. □

Much more interesting and much harder is the question: What happens when x is infinite?

For example, what is the size of $\mathfrak{P}(\mathbb{N})$?

As in the finite case it turns out that $\mathfrak{P}(\mathbb{N})$ is much larger than $\mathbb{N}$ but the technical details are quite involved; see the section on cardinality below.

Exercise

Show that symmetric difference is associative:
$A \oplus (B \oplus C) = (A \oplus B) \oplus C$.

Exercise

Show that $A \subseteq B$ *implies that* $A \oplus (B - A) = B$.

Exercise

Show that $A, B \subseteq C$ *implies that* $(C - A) \oplus B = C - (A \oplus B)$.

Exercise

Show that $A \subseteq B$ *iff* $B - (B - A) = A$.

Exercise

Show that $\mathfrak{P}(A \cap B) = \mathfrak{P}(A) \cap \mathfrak{P}(B)$. *How about union instead of intersection?*

- Intuitive Set Theory
- Extensionality

3. Frege's Ghost

- Products and Sums
- Cantor
- Formalization
- Beyond ZFC

You wake up one morning to find that Frege's Ghost (see below for the real Frege) has ruined your favorite C++ compiler: it now supports the language F++ (Frege-plus-plus) which has no data types other than set. You urgently need to implement a SAT solver for a research project.

Is there any way one could to this in F++?

Surprisingly, the answer is a resounding **Yes**.

Of course, we have to assume that some control structures still work, and that the compiler provides some basic operations on sets. Given such minimal support, we can implement integers and lists as pure sets and from there any finite discrete object (such as a formula in CNF).

This is just the tip of an iceberg, in mathematics and computer science "everything" can be implemented as a set.

We will not bother to explain in detail how F++ works, here is an outline. Built-in functions are

- constant empty set $\emptyset$,
- unordered pairs $\{x, y\}$,
- difference $x - y$,
- unary union $\bigcup x$, unary intersection $\bigcap x$,
- Boolean operations for membership and equality.

Needless to say, we have variables, assignments, if-then-else, composition, looping and recursion (whatever that may mean exactly).

For example, we can express binary union in F_{++} as

$$x \cup y = \bigcup \{x, y\}$$

The successor function $S(x) = x \cup \{x\}$ is easy: $S(x) = x \cup \{x, x\}$.

Recursion in $\mathrm{F}{+}{+}$ takes the following form:

$$f(x, \boldsymbol{y}) = g(x, \boldsymbol{y}, \bigcup\{\, f(z, \boldsymbol{y}) \mid z \in x \,\}$$

In other words, we exploit the fact that the membership relation is

well-founded, so we can recurse

For example, we can define transitive closure as

$$\mathsf{TC}(x) = x \cup \bigcup\{\, \mathsf{TC}(z) \mid z \in x \,\}$$

First let us deal with the issue of representing lists as sets.

Definition

The (Kuratowski) pair of x and y is

$$\pi(x, y) = \{\{x\}, \{x, y\}\}$$

Lemma

$\pi(u, v) = \pi(x, y)$ *implies* $u = x$ *and* $v = y$.

Careful, we are dealing with sets: $\pi(x, x) = \{\{x\}\}$. By the lemma, there are unpairing functions π_1 and π_2 such that

$$\pi_1(\pi(u, v)) = u \qquad \pi_2(\pi(u, v)) = v.$$

Let $z = \pi(u, v) = \{\{u\}, \{u, v\}\}$

Note that always $\bigcup z = \{u, v\}$ and $\bigcap z = \{u\}$, a singleton.

It follows that $\bigcup z - \bigcap z \neq \emptyset$ iff $u \neq v$.

This is the key observation that allows us to define unpairing functions as a little programming exercise in $\mathrm{F}{+}{+}$.

$$\pi_1(z) = \bigcup\bigcap z$$

$$\pi_2(z) = \begin{cases} \bigcup(\bigcup z - \bigcap z) & \text{if } \bigcup z - \bigcap z \neq \emptyset, \\ \pi_1(z) & \text{otherwise.} \end{cases}$$

We leave it as an exercise to show that

$$\pi_i(\pi(x_1, x_2)) = x_i$$

Thus, from the programming perspective, both the constructor Kuratowski pair and the two corresponding destructors are easily implemented in $\mathrm{F{+}{+}}$.

The lemma justifies the notation (u, v) instead of the more pedantic $\pi(u, v)$: the only thing that really matters about pairs is the unpairing functions, there is no loss of information in the pairing operation.

Exercise

Prove the lemma and check that everything can be implemented in $\mathrm{F}{+}{+}$.

Exercise

Find another way to implement pairs as sets.

Exercise

Does the attempt $(x, y) = \{x, \{x, y\}\}$ *succeed?*

Pairs can be extended to lists, here usually called n-tuples, by induction. For $n \geq 3$ set

$$(a_1, \ldots, a_n) = (a_1, (a_2, \ldots, a_n)).$$

Lemma

$(a_1, \ldots, a_n) = (b_1, \ldots, b_n)$ *implies* $a_i = b_i$.

Note, though, that we cannot conclude that $(a_1, \ldots, a_n) = (b_1, \ldots, b_m)$ implies $n = m$.

Exercise

Explain what goes wrong and how to fix it. (Assume we already know how to define natural numbers.)

Exercise

Find an alternative way to implement lists as functions.

Here is a critical application of Kuratowski pairs.

Definition

The Cartesian Product of two sets A and B is

$$A \times B = \{ (a, b) \mid a \in A, b \in B \}$$

Again we can extend the pairing operation by induction to $n > 2$ sets:

$$A_1 \times A_2 \times \ldots \times A_n = A_1 \times (A_2 \times \ldots \times A_{n-1} \times A_n)$$

Example

In geometry, the plane can be thought of as $\mathbb{R} \times \mathbb{R}$. And ordinary 3-space is $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

We can now define a pure binary relation as follows:

$$\mathrm{Rel}(\rho) :\Leftrightarrow \forall\, x \in \rho\, \exists\, u, v\, (x = (u, v))$$

In other words, a pure relation is a set $\rho \subseteq A \times B$ where A and B are both sets. Here A is the domain of ρ and B is the codomain, in symbols $\mathsf{dom}(\rho)$ and $\mathsf{cod}(\rho)$.

Similarly we can define the support and range of a relation

$$\mathrm{spt}(\rho) \coloneqq \{\, x \mid \exists\, y\, ((x, y) \in \rho)\,\} \subseteq \mathsf{dom}(\rho)$$
$$\mathrm{rng}(\rho) \coloneqq \{\, y \mid \exists\, x\, ((x, y) \in \rho)\,\} \subseteq \mathsf{cod}(\rho)$$

A binary relation is a triple $R = \langle\, \rho, A, B \,\rangle$ where ρ is a pure binary relation and $\mathrm{dom}(\rho) \subseteq A$ and $\mathrm{cod}(\rho) \subseteq B$.

In this context ρ is also called the graph of R.

This definition may seem a bit overblown; after all, the interesting part of R is ρ, the domain and codomain seem superfluous.

Alas, in many branches of mathematics (in particular algebra and category theory), it is critical to supply this additional information; pure relations are just not that useful in the end.

Notation warning: some authors refer to our pure relations (or graphs) as relations.

The next step is to define partial functions. Suppose $R = \langle\, \rho, A, B \,\rangle$ is a binary relation. R is single-valued if

$$\forall\, a \in A, b_1, b_2 \in B\, ((a, b_1) \in \rho \wedge (a, b_1) \in \rho \Rightarrow b_1 = b_2)$$

$$\begin{aligned} \mathrm{PFct}(f) :\Leftrightarrow\ & \mathrm{Rel}(f) \wedge \\ & \forall\, a, b_1, b_2\, ((a, b_1) \in R \wedge (a, b_1) \in R \Rightarrow b_1 = b_2) \end{aligned}$$

The condition here is referred to as single-valuedness.

To get a function, we insist that the relation is total:

$$\begin{aligned} \mathrm{Fct}(f) :\Leftrightarrow\ & \mathrm{PFct}(f) \wedge \mathsf{dom}(f) = \\ & \forall\, a, b_1, b_2\, ((a, b_1) \in R \wedge (a, b_1) \in R \Rightarrow b_1 = b_2) \end{aligned}$$

Injections and surjections

$$\begin{aligned}\mathrm{Inj}(f, A, B) &:\Leftrightarrow \mathrm{Fct}(f, A, B) \wedge \\ &\qquad \forall\, a_1, a_2 \in A, b \in B\, ((a_1, b) \in f \wedge (a_2, b) \in R \Rightarrow a_1 = a_2)\\ \mathrm{Surj}(f, A, B) &:\Leftrightarrow \mathrm{Fct}(f, A, B) \wedge \forall\, b \in B\, \exists\, a \in A\, ((a, b) \in R)\end{aligned}$$

Then bijections

$$\mathrm{Bij}(f, A, B) :\Leftrightarrow \mathrm{Inj}(f, A, B) \wedge \mathrm{Surj}(f, A, B)$$

Now we can tackle issues of cardinality and so forth.

$$\begin{aligned}\mathrm{Fct}(f, A, B) &:\Leftrightarrow \mathrm{Rel}(f, A, B) \wedge \mathrm{dom}(R) = A \wedge \\ &\qquad \forall\, a \in A, b_1, b_2 \in B\, ((a, b_1) \in R \wedge (a, b_1) \in R \Rightarrow b_1 = b_2)\end{aligned}$$

Notation Warning

As

Can we implement natural numbers in $\mathrm{F}{+}{+}$?

No problem. We can use the successor function $S(x) = x \cup \{x\}$ from above to represent a natural number n by a set $\underline{n}$ as follows:

$$\underline{0} \rightsquigarrow \emptyset$$
$$\underline{n} \rightsquigarrow \underbrace{S(S(\ldots S(}_{n}\emptyset)\ldots))$$

So, $\underline{3} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ represents the number 3.

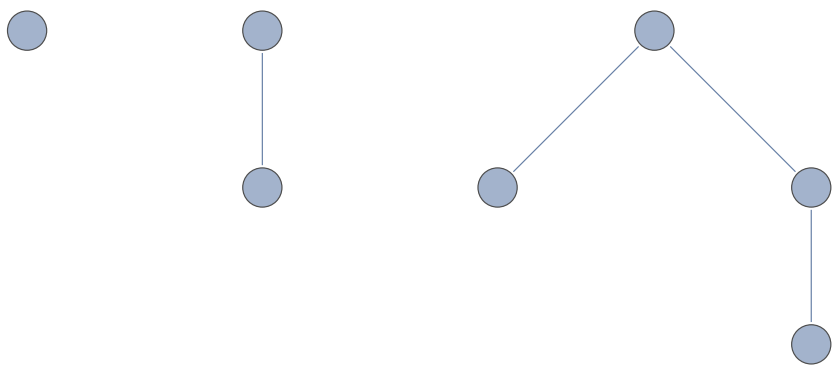
Definition

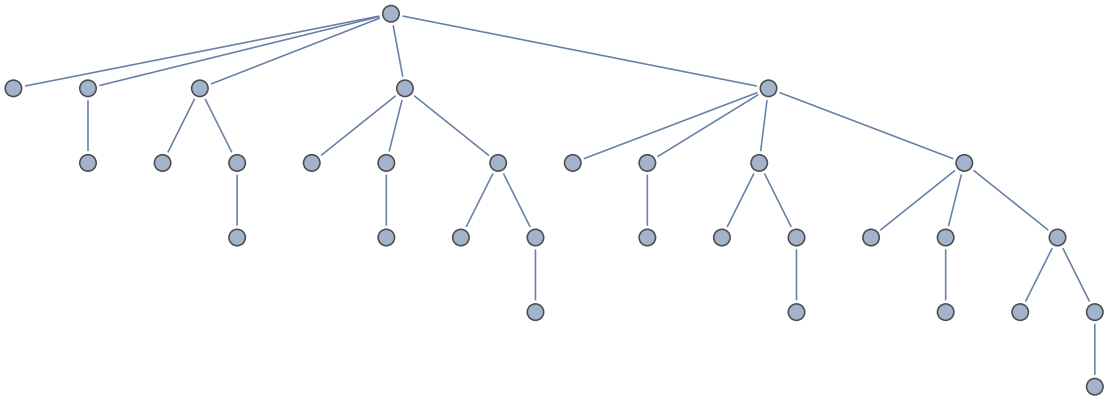
These sets are the (finite) von Neumann ordinals N_n.

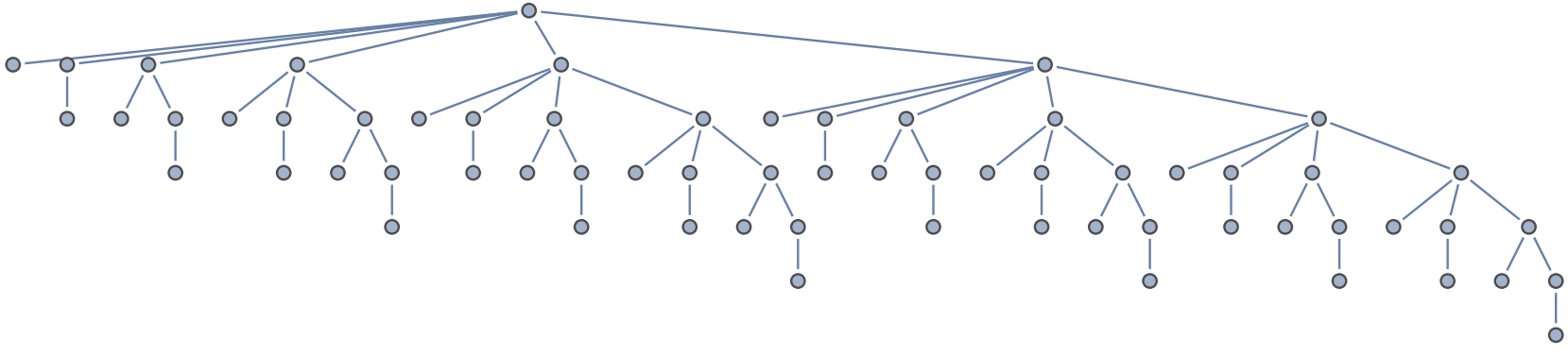
There are also infinite von Neumann ordinals, more on this later.

Note that we can represent any pure set by a tree: the leaves are instances of the empty set, and edges indicate membership (set on top, elements one level down). The internal nodes are the compound sets used to construct the main set which is represented by the root.

Here are the (boring) pictures for N_0, N_1 and N_2.







The previous pictures are indicative of one serious flaw of the set-theoretic approach to mathematical life that we already mentioned: while it is possible to express (essentially) all objects of mathematical discourse as pure sets, the sets in question are often mind-numbingly complicated.

The "pure set" approach has to be tempered with good intuition, and/or devices such as urelements, otherwise one quickly gets lost in the details.

In particularly in the case of arithmetic we could simply assume that $\mathbb{N} = \{0, 1, \ldots, n, \ldots\}$ is given as a set of urelements.

But if we like, we can replace the urelement n by the von Neumann ordinal N_n and deal with pure sets instead.

This is no more than a change in the level of abstraction; the bread and butter of any computer scientist.

A data type by itself is useless, we need operations. But note that

$$N_n = \{N_0, N_1, \ldots, N_{n-1}\}$$

so $N_x \in N_y$ iff $x < y$ and we can handle order.

We can also compute maxima: $N_x \cup N_y = N_{\max(x,y)}$.

More interestingly, we can define addition:

$$\begin{aligned} \mathsf{add}(x, N_0) &= x \\ \mathsf{add}(x, S(N_y)) &= S(\mathsf{add}(x, N_y)) \end{aligned}$$

To see this, first note that $\bigcup$ implements the predecessor function for our von Neumann numbers.

It is not hard to check that F++ can handle the recursive definition for add.

Why not simply represent natural numbers as deeply nested empty sets:

$$\underline{0} \rightsquigarrow \emptyset$$
$$\underline{n} \rightsquigarrow \underbrace{\{\{\ldots\{}_{n}\emptyset\}\ldots\}\}$$

In other words, $M_0 = \emptyset$ and $M_{n+1} = \{M_n\}$.

There is nothing fundamentally wrong with this approach (which is due to Zermelo), we could still implement order – but not as elegantly as with von Neumann ordinals.

The real problem is that this approach does not generalize gracefully to infinite numbers (which is important to e.g. to prove termination of nested recursions). The von Neumann ordinals can be generalized very nicely and naturally to transfinite numbers.

Frege proposed a more ingenious way to capture the naturals. The brilliant idea is to associate n with the collection of all sets of size n.

$$\underline{0} \rightsquigarrow \{\emptyset\}$$
$$\underline{n} \rightsquigarrow \{\, x \mid \exists\, y \in \underline{n-1}, a \notin y\, (x = y \cup \{a\})\,\}$$

Alas, this approach produces proper classes, not sets, and can't be used directly in this form: a class is a collection of objects that is so large that we cannot quite treat it the same way as ordinary sets (e.g., there is no way we can assign a cardinality to a class).

This problem can be fixed by selecting only a few sets of cardinality n to represent n (the first to appear in some natural hierarchy of sets).

The existence of proper classes is a technical difficulty that arises in many formalizations of set theory: some perfectly reasonable collections such as the collection of all sets, all vector spaces, all cardinals, all total orders, and so, are too large to be classified as "sets."

There are ways to deal with these problems systematically; see below a description of von Neumann-Bernays-Gödel set theory. In a nutshell, one allows for classes whose elements are sets, but classes cannot be nested themselves. Formation is unrestricted

$$X = \{\, x \mid \varphi(x) \,\}$$

but x is required to range over sets whereas X may be a class (and φ cannot quantify over class variables).

For example, Russell's paradoxical $\{\, x \mid x \notin x \,\}$ is a class, so no contradiction ensues.

For those interested in computing with sets (aka logicians) we point out that F_{++} is actually quite weak.

For example, the function

$$x \mapsto |x|$$

that associates a set with its cardinality cannot be computed in F_{++}.

Take a look at the paper by Jensen and Karp, "Primitive Recursive Set Functions" for lots of information about F_{++} or Barwise "Admissible Sets and Structures" for other ways to define computation on sets.

- Intuitive Set Theory
- Extensionality
- Frege's Ghost

4. Products and Sums

- Cantor
- Formalization
- Beyond ZFC

Here is a more serious application of Kuratowski pairs.

Definition

The Cartesian Product of two sets A and B is

$$A \times B = \{ (a, b) \mid a \in A, b \in B \}$$

Again we can extend the pairing operation by induction to n sets:

$$A_1 \times A_2 \times \ldots \times A_n = A_1 \times (A_2 \times \ldots \times A_{n-1} \times A_n)$$

Example

In geometry, the plane can be thought of as $\mathbb{R} \times \mathbb{R}$. And ordinary 3-space is $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

We can now define a binary relation from A to B as any set $R \subseteq A \times B$.

$$\mathrm{Rel}(R, A, B) :\Leftrightarrow R \subseteq A \times B$$

Similarly we can define the domain of a relation from A to B as

$$\mathrm{dom}(R) := \{\, a \in A \mid \exists\, b \in B\, ((a, b) \in R) \,\}$$

The next step is to define functions

$$\mathrm{Fct}(f, A, B) :\Leftrightarrow \mathrm{Rel}(f, A, B) \wedge \mathrm{dom}(R) = A \wedge \\ \forall\, a \in A, b_1, b_2 \in B\, ((a, b_1) \in R \wedge (a, b_1) \in R \Rightarrow b_1 = b_2)$$

Injections and surjections

$$\begin{aligned}\mathrm{Inj}(f, A, B) &:\Leftrightarrow \mathrm{Fct}(f, A, B) \wedge \\ &\qquad \forall\, a_1, a_2 \in A, b \in B\, ((a_1, b) \in f \wedge (a_2, b) \in R \Rightarrow a_1 = a_2) \\ \mathrm{Surj}(f, A, B) &:\Leftrightarrow \mathrm{Fct}(f, A, B) \wedge \forall\, b \in B\, \exists\, a \in A\, ((a, b) \in R)\end{aligned}$$

Then bijections

$$\mathrm{Bij}(f, A, B) :\Leftrightarrow \mathrm{Inj}(f, A, B) \wedge \mathrm{Surj}(f, A, B)$$

Now we can tackle issues of cardinality and so forth.

One might wonder what a Cartesian product of length 0 or 1 would be. In analogy to multiplication we would expect something like

$$\prod_{i=1}^{0} A_i = \mathbf{1}$$

$$\prod_{i=1}^{1} A_i = A_1$$

The question is what should $\mathbf{1}$ be? We are dealing with sets, not numbers.

Another Nuisance

There is another problem. According to our definition of tuples

$$\prod_{i=1}^{3} A_i = A_1 \times (A_2 \times A_3)$$

which is not the same as

$$(A_1 \times A_2) \times A_3$$

Our Cartesian product operation is not associative. This seems quite awkward, it should not matter which which sets we group together first.

Exercise

Come up with a common sense solution for these problems.

One often needs to refer to a collection of sets that are "numbered" by some index set. It turns out that it is best to keep things general and allow arbitrary index sets.

Definition

Let I be a set, the index set. A family of sets, indexed by I is a map A with domain I.

Notation: $(A_i)_{i\in I}$ to indicate that $A(i) = A_i$.

This should be perfectly familiar when I is one of the usual choices for index sets: $[n] = \{1, 2, \ldots, n\}$, $(n) = \{0, 1, \ldots, n-1\}$, $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{R}$.

Families usually appear in conjunction with infinitary operations. For example, we may consider the family of closed real intervals $A_n = [0, 1/n] \subseteq \mathbb{R}$ for $n \geq 1$. Then $\bigcap_n A_n = \{0\}$. On the other hand, for the open intervals $A_n = (0, 1/n) \subseteq \mathbb{R}$ we have $\bigcap_n A_n = \emptyset$.

We know how to form the Cartesian product of a finite number of sets. Can we generalize this to a product of a whole family of sets?

Definition

Let $(A_i)_{i\in I}$ be a family of sets and set $A = \bigcup_i A_i$. The Cartesian product of this family is defined by

$$\prod_{i\in I} A_i = \{\, \alpha : I \to A \mid \alpha(i) \in A_i \,\}.$$

Thus, $\prod_{i\in I} A_i$ consists of all I-indexed sequences of elements in $\bigcup_i A_i$ subject to the condition that the ith element has to be taken from A_i.

For example, if $I = \mathbb{N}$ and $A_i = \mathbf{2}$ then $\prod_{i\in\mathbb{N}} A_i$ is just the collection of all infinite binary sequences. This would often be written simply as $\prod_{\mathbb{N}} \mathbf{2}$.

Likewise, all real sequences (a central notion in analysis) can be obtained as $\prod_{\mathbb{N}} \mathbb{R}$.

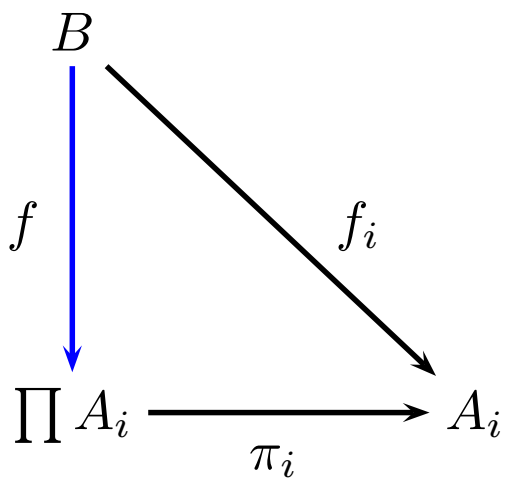
So what's so special about products? It turns out the the crucial property of products $A = \prod A_i$ can be summarized rather succinctly. Let

$$\pi_j : \prod A_i \to A_j \qquad \pi_j(\alpha) = \alpha(j)$$

be the projection map onto the jth component.

Theorem (Universal Property)

Given a set B and a family of functions $f_i : B \to A_i$ there is a unique function $f : B \to \prod A_i$ such that $\pi_i \circ f = f_i$ for all $i \in I$.

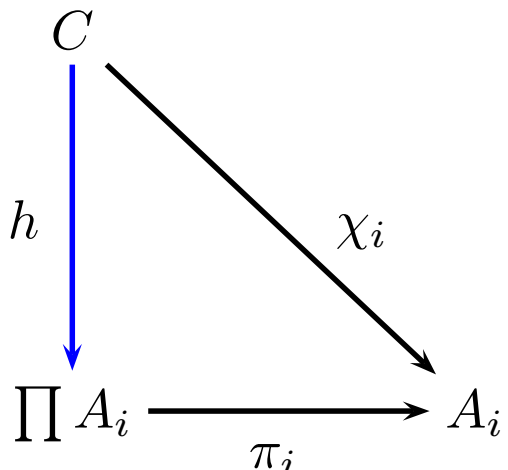


This theorem is easy to prove: just set $f(b)(i) = f_i(b)$.

What's more important is that any other set C that pretends to be a product must already be the same as $\prod A_i$. By this we mean the following.

Theorem (Uniqueness)

Suppose a set C together with a family of maps $\chi_j : C \to A_j$ satisfies the same universality property as $\prod A_i$. Then there exists a bijection $h : C \to \prod A_i$ such that



The Uniqueness theorem may seem a bit abstract, but what it means concretely is that it does not matter exactly how we code up the product, any other reasonable coding will be essentially the same.

Reasonable here means that the alternative product would have to satisfy the same universal property. The bijection simply associates every element in our product with a corresponding element in the alternative product.

This should sound eminently familiar to the computer science student: an array of length n, a list of length n and a record with n fields are all the "same" in some sense, though, of course, there are important differences.

Exercise

Show how to reconcile our two definitions of Cartesian product for index sets of size 2. How does this pertain to Cartesian products of arbitrary finite length?

There is an analogous result for disjoint unions

$$\sum A_i := \{\, (a, i) \mid a \in A_i \,\}$$

together with the injections $\iota_j : A_j \to \sum A_i$.

Exercise

Explain what the universal property of $\sum A_i$ is.

Exercise

Show that $\sum A_i$ is unique up to bijections.

- Intuitive Set Theory
- Extensionality
- Frege's Ghost
- Products and Sums

5. Cantor

- Formalization
- Beyond ZFC

Definition

The size of a set is called its cardinality.

Needless to say, this is not much of a definition. We'll have more to say about this later, for the time being use your intuition.

At least for finite sets it is easy to make sense out of this idea: just count the elements. So

$$S = \{a_1, a_2, \ldots, a_{n-1}, a_n\}$$

has cardinality n.

Of course, we have tacitly assumed here that the representation of S in the curly braces does not have repetitions; all the a_i must be distinct.

We write $|S| = n$ or sometimes $\text{card}(S) = n$ for the cardinal number of S.

Example

$$|S| = 0 \iff S = \emptyset$$
$$|[n]| = n$$
$$|[n] \times [m]| = n \cdot m$$
$$|[n] \times [n] \times [n]| = n^3$$

It can be amazingly difficult to determine the sizes of various finite sets; the field of combinatorics has developed a rich collection of tools for this purpose.

Suppose A and B are finite. What is

$$|A \cup B| = ???$$

Not just $|A| + |B|$: that only works when $A \cap B = \emptyset$.

In general we must correct the over-count:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Exercise

Find the right expression for $|A \cup B \cup C|$.

Exercise

Generalize to n *sets* $A_1, \ldots, A_n$.

- Intuitive Set Theory
- Extensionality
- Frege's Ghost
- Products and Sums
- Cantor

6. Formalization

- Beyond ZFC

So far so good. It looks like we can handle sets just fine informally, using common sense.

In mathematics, sets were first studied systematically by Georg Cantor (1845-1918). Surprisingly, he was lead to the study of set during his wok on Fourier analysis.

Cantor found lots of strange properties of infinite sets, and at the time many mathematicians disagreed about the nature of infinite sets.

This led Gottlob Frege (1848-1925) to propose an axiom system for sets, which was supposed to settle all debate once and for all.

His system is a bit different from its modern counterparts (see the lecture on Formal Systems). To simplify matters we will reinterpret it slightly and express the axioms in modern terminology.

Here is a modern re-write of G. Frege's axiom V which states that two sets are equal if, and only if, they contain the same elements.

Since Frege also assumes that every concept P is associated with an "extension", essentially the set $\{ z \mid P(z) \}$ of all objects with property P, we also have a powerful set formation axiom as below.

- **Extensionality**

$$x = y \text{ if } \forall z \, (z \in x \iff z \in y)$$

- **Formation**
  For any property $P(z)$:

$$\exists x \, \forall z \, (z \in x \iff P(z))$$

For us, the quantifiers all range over the collection of all sets. Note that by Extensionality the set x in Formation is unique.

It would be a slightly tedious to carry out all the details, but one could give a completely set-theoretic treatment of, say, calculus, using just these two axioms.

Starting from the von Neumann ordinals as representations of the natural numbers one proceeds in stages and constructs the integers, the rationals and the reals.

Functions on the reals are then sets of pairs of reals, and so on and so on.

Frege's two axioms (and a lot of stamina) are the only tools needed to do this.

Exercise

Find a way to express integers in terms of sets of von Neumann ordinals.

Unfortunately, there is a fatal flaw in Frege's system: his axioms are inconsistent. Bertrand Russell pointed this out in a letter to Frege in 1902 (just when Frege was getting ready to publish the second volume of his Grundgesetze).

Since Frege was trying to build a uniform foundation for mathematics this is a bit of a catastrophe.

But even if one takes a more casual approach to set theory there were warning signs.

Let V be the collection of all sets: by comprehension we can form

$$V = \{ x \mid x = x \}$$

Then V has smaller cardinality than $\mathfrak{P}(V)$.

But $\mathfrak{P}(V) \subseteq V$, so its cardinality is at most the cardinality of V, contradiction.

Let On be the collection of all ordinals.

Then On is well-ordered with respect to $\in$ and must have some ordertype Ω, an ordinal.

But then $\Omega \in On$ and thus corresponds to the ordertype of a strict initial segment of On, contradiction.

Russell considers the following set, whose existence is guaranteed by comprehension.

$$S = \{ z \mid z \notin z \}$$

Looks strange, but why not?

But then both $S \in S$ and $S \notin S$ lead to a contradiction. For example, if $S \in S$ then by definition of S, $S \notin S$.

We are sunk.

How serious is this, really? Well, it nearly killed Frege . . . He mentioned Russell's result in an appendix, and presented a supposed remedy (which Russell originally accepted, but later both he and Frege realized that the remedy did not work, it shrinks the universe to one point).

However, unless you decide to become a logician (and thus permanently unemployable), you will never encounter inconsistent set constructions.

There are ways to fix this problem, but they are a bit technical, and not needed for standard applications. As long as you apply common sense to set formation, you'll be OK.

Here is a glimpse at a system of axioms that works, and that has become the de facto standard: Zermelo-Fraenkel set theory.

It is a truly remarkable feature of this system that it is powerful enough to express most of mathematics in a clean and precise way. Yet, it uses only a handful of axioms (well, really schemata) that are fairly easy to accept intuitively.

One way to avoid Russell's paradox is to write down a list of carefully designed axioms that describe all properties of sets and are very conservative when it comes to set formation.

They keep Frege's Extensionality axiom.

$$x = y \iff \forall\, z\ (z \in x \Leftrightarrow z \in y)$$

And add a number of modest set-existence axioms: axioms that say sets with certain narrow properties exist – not like Frege's sledgehammer Formation axiom.

E.g., there is an axiom that says "the empty set exists".

Then there is one for unordered pairs $\{x, y\}$, for union, and so on.

Needless to say, this is much more tedious than Frege's Formation, but it gets us around Russell.

We'll just look at a few, the others are on the web.

Empty Set

$$\exists u \forall z \, (z \notin u)$$

Unordered Pair

$$\exists u \forall z \, (z \in u \Leftrightarrow z = x \vee z = y)$$

Union

$$\exists u \forall z \, (z \in u \Leftrightarrow \exists y \, (y \in x \wedge z \in y))$$

Power Set

$$\exists u \forall z \, (z \in u \Leftrightarrow z \subseteq x)$$

So Unordered Pair says: $\{x, y\}$ exists.

Comprehension (Aussonderungsaxiom)

$$\exists\, u \forall\, z\ (z \in u \Leftrightarrow z \in x \wedge P(z))$$

Infinity

$$\exists\, u\, (\emptyset \in u \wedge \forall\, z\ (z \in u \Rightarrow S(z) \in u))$$

For the Infinity axiom we have used the standard successor function defined by $S(x) = x \cup \{x\}$.

The last axiom deals with a applying functions to a set of elements. A formula $Q(x, y)$ with two free variables is called functional if

$$\forall\, w, u, v\ (Q(w, u) \wedge Q(w, v) \Rightarrow u = v)$$

For any functional property Q we have the following axiom:

Replacement Axiom

$$\exists\, u \forall\, z\ (z \in u \Leftrightarrow \exists\, w\, (w \in x \wedge Q(w, z)))$$

By Replacement we can apply a function to a set, and get back another set. This axiom does not follow from the previous ones, it closes a critical gap.

And it is not an axiom, it is an axiom schema.

All the axioms mentioned so far are fairly natural and easy to defend.

Here is one that is more problematic:

Foundation

$$x \neq \emptyset \Rightarrow \exists\, u \in x\, (u \cap x = \emptyset)$$

Foundation essentially precludes situations such as $x = \{x\}$. Given Foundation, one can construct a hierarchical model of sets.

Here is another plausible and eminently useful axiom that is not so easy to defend: the Axiom of Choice (AC).

The axiom was introduced by Ernst Zermelo in 1904 to give a construction of a well-order of the reals.

Zermelo's idea was not too well-received at the time, but it has become mainstream by now: the Axiom of Choice is indispensable for many arguments in mathematics. For example, it ensures that every vector space has a basis.

Alas, as we will see, it also has some nearly absurd consequences.

Suppose we have a collection $(A_i)_{i \in I}$ of non-empty sets that are pairwise disjoint: $i \neq j \Rightarrow A_i \cap A_j = \emptyset$.

We would like to construct a choice set C for A_i: a set that selects exactly one element from each A_i.

$$C \cap A_i = \{a_i\} \quad \text{for all } i.$$

We can also think of this as a choice function, a map

$$C : I \to \bigcup A_i \qquad C(i) \in A_i$$

(in which case we can omit the disjointness condition).

If I is finite, or if the sets A_i carry a nice structure such as being subsets of $\mathbb{N}$ then (AC) is not needed: the other axioms guarantee the existence of choice sets/functions.

But in general the existence of a choice set/function does not follow from the other axioms.

In many ways (AC) is so natural that one feels it ought simply to be incorporated in the standard system of set theory (some logicians would furiously disagree).

Definition

Zermelo-Fraenkel set theory ZF is defined by the axioms above. ZFC is ZF plus the Axiom of Choice.

It is good practice to point out whenever a result depends on (AC) rather than just use it tacitly. For example, the famous Nielson-Schreier theorem that states that a subgroup of a free group is again free requires (AC), as does the "fact" that the additive group of the reals is isomorphic to the additive group of the complex numbers.

Note that the axiomatic approach to set theory makes no attempt whatsoever to explain the nature of sets or the membership relation at any deeper level.

All we have is a first-order theory ZF or ZFC in a language with just a single binary relation symbol. The axioms are set up to reflect certain properties of sets that we consider self-evident. This is certainly true for the basic axioms, but Choice and Foundation can engender lengthy discussions.

The interesting point is that a relatively small number of relatively simple axioms (or schemata) seem to capture almost everything there is to say about sets. We have a very compact and tight description of a rather complicated subject.

At any rate, here is another important application of Choice, a positive answer to the question whether we can perform induction on the real numbers.

In order to do induction we need to well-order $\mathbb{R}$. Clearly, the standard order does not work; it fails already on the integers or the positive rationals.

But how do we define a well-order? It was Zermelo's idea to use the Axiom of Choice to define an abstract ordering on the reals that is provably a well-order. To this end, pick a choice function

$$f : \mathfrak{P}(\mathbb{R}) \to \mathbb{R}$$

such that $\emptyset \neq X \subseteq \mathbb{R} \Rightarrow f(X) \in X$. Then enumerate the reals according to

$$\begin{aligned} r_0 &= f(\mathbb{R}) \\ r_1 &= f(\mathbb{R} - \{r_0\}) \\ r_2 &= f(\mathbb{R} - \{r_0, r_1\}) \\ &\dots \end{aligned}$$

Repeating this process we obtain a well-order $r_0 < r_1 < \ldots < r_n < \ldots$

But the "..." here is really tricky: since there are more reals than natural numbers we need transfinite induction. After we have exhausted all stages $n \in \mathbb{N}$ we continue:

$$\begin{aligned} r_\omega &= f(\mathbb{R} - \{\, r_i \mid i \in \mathbb{N} \,\}) \\ r_{\omega+1} &= f(\mathbb{R} - \{\, r_i \mid i \in \mathbb{N} \,\} - \{r_\omega\}) \\ r_{\omega+2} &= f(\mathbb{R} - \{\, r_i \mid i \in \mathbb{N} \,\} - \{r_\omega, r_{\omega+1}\}) \\ &\ldots \end{aligned}$$

Note the ω hiding in the subscript. This is an ordinal number, an extension of the natural numbers for purposes of enumeration. There is a nice definition of these objects in terms of von Neumann ordinals. See the notes on ordinals and cardinals for details.

If (AC) is useful, and makes intuitive sense, why not just adopt it and not make a big fuss about it?

Because (AC) also has a dark side, a few strange consequences, and a few extremely bizarre consequences.

First off, (AC) implies that there are sets of reals that fail to be Lebesgue measurable. This is certainly a bit counter-intuitive; it is not clear what exactly should prevent us from assigning a measure to an arbitrary set of reals.

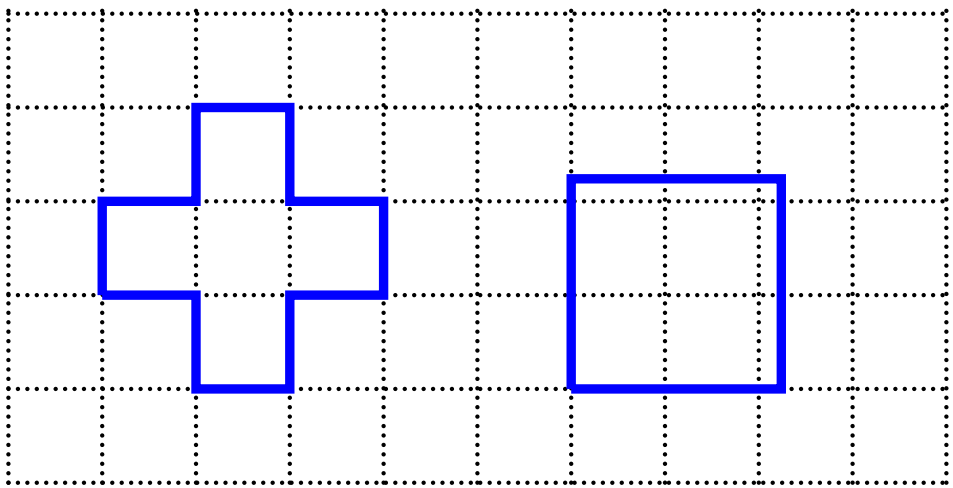
Worse, a result by F. Hausdorff from 1914 states that one can partition a sphere (after removing countably many points) into three parts A, B and C such that all three pieces are congruent, and are also congruent to $B \cup C$.

This sounds more like a paradox that a theorem.

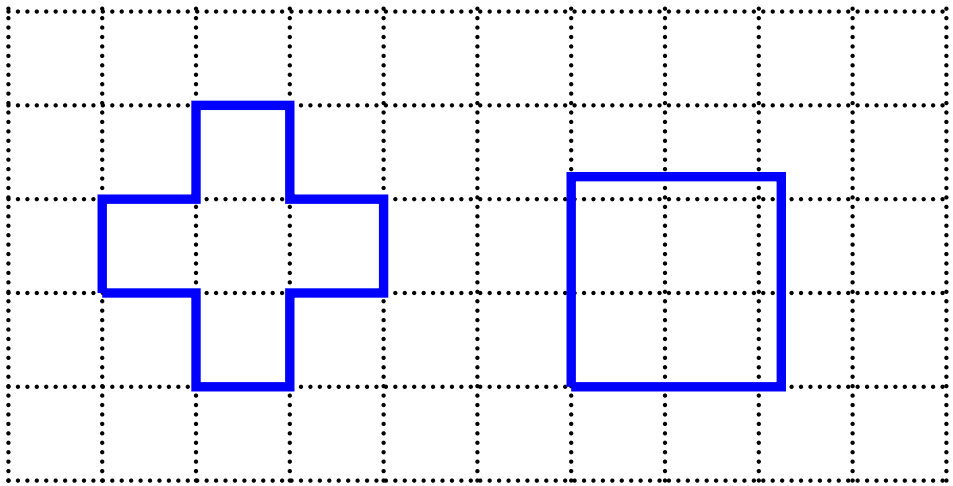
By comparison, consider the following entirely reasonable and unsurprising result in the plane.

Theorem (Bolyai-Gerwin Theorem)

Let P and Q be two polygons of equal area. Then P can be partitioned into finitely many triangles that can be reassembled to form Q.



Take a pair of (mental) scissors, and find a simple decomposition of the cross on the left whose pieces can be rearranged to form of the square on the right.



The fewer cuts, the better.

Of course, the restriction to triangles is severe. Also, having another dimension helps. Still, the next theorem which builds on Hausdorff's result is truly wild and sounds positively like a serious error. Needless to say, its proof requires the Axiom of Choice.

Theorem (Banach-Tarski Paradox)

The unit sphere can be decomposed into finitely many pieces, that can be reassembled to form a sphere of radius 2.

The pieces in the Banach-Tarski decomposition are very strange and cannot be visualized. In particular we cannot assign qualities such as "volume" to these pieces: otherwise we would immediately have a contradiction.

So, this result is a correct theorem of ZFC, but it is very, very counter-intuitive. In a sense, this is much worse than Cantor's problems with cardinality.

We would like to find a d-dimensional measure, a map $\mu : \mathfrak{P}(\mathbb{R}^d) \to \mathbb{R}_0^+$ such that

- Any two equidecomposable sets $A, B \subseteq \mathbb{R}^d$ have the same measure: $\mu(A) = \mu(B)$.
- The measure is additive on disjoint sets: $A \cap B = \emptyset$ implies $\mu(A \cup B) = \mu(A) + \mu(B)$.
- The measure is normalized: $\mu([0,1]^d) = 1$.

Theorem (Hausdorff 1914)

Measures do not exist for dimensions $d \geq 3$.

Theorem (Banach 1923)

Assuming the Axiom of Choice, measures do exist for dimensions $d = 1, 2$.

Theorem (Gödel 1938)

ZFC is equiconsistent with ZF.

Theorem (Cohen 1963)

(AC) is independent of ZF.

Theorem (Solovay 1970)

There is a model of ZF where every set of reals is measurable.

So the Banach-Tarski paradox depends crucially on the Axiom of Choice.

- Intuitive Set Theory
- Extensionality
- Frege's Ghost
- Products and Sums
- Cantor
- Formalization

7. Beyond ZFC

Texts on set theory often contain comments along the lines of

> One can check that essentially all of modern mathematics can be constructed in ZFC.

Note the hedge: essentially.

The problem is that, on rare occasions, one would like to have sets that are so huge that ZFC cannot prove their existence. Category theory is a notorious example.

Huge collections of objects may be entirely reasonable, but cannot be dealt with easily in ZF.

For example, V, the collection of all sets is not a set.

The same is true for the collection of all ordinals or all cardinals.

And we cannot say $\emptyset^- = V$.

To deal with these issues elegantly one needs a slightly more complicated ontology: one has to admit (proper) classes such as V in addition to plain sets.

In the following we use uppercase letters X, Y, Z and so on to denote variables ranging over classes.

We briefly describe a system due to von Neumann, Bernays andGödel, usually abbreviated as NBG (some people prefer to remember the order by thinking of "no bloody good").

We keep Extensionality

$$X = Y \iff \forall\, Z\, (Z \in X \Leftrightarrow Z \in Y)$$

The key idea is that a proper class cannot be an element of a class, only sets can be elements. So we can define X to be a set via

$$M(X) :\Leftrightarrow \exists\, Y\, (X \in Y)$$

We use lowercase letters x, y, z ... for sets. So $\forall\, x\, \varphi(x)$ is shorthand for $\forall\, X\, (M(X) \Rightarrow \varphi(X))$.

Beyond Extensionality we keep the following axioms, but for sets only:

- Empty Set
- Pairing
- Union
- Power Set
- Infinity

As usual one writes $\{x, y\}$ for the unordered pair, (x, y) for the Kuratowski pair and likewise for n-tuples, $\mathfrak{P}(x)$ for the power set.

Again, x and y are sets in this context, not proper classes.

It is fairly customary at this point to assume Foundation, which needs to be phrased in terms of classes in this setting:

$$X \neq \emptyset \Rightarrow \exists\, x\, (x \in X \wedge x \cap X = \emptyset)$$

Again, Foundation is somewhat less fundamental than the other axioms. Omitting this axiom causes some technical problems but is not a fiasco.

Here is a more interesting new axiom: we can use unrestricted comprehension as in naive set theory to form classes (but not necessarily sets).

Let $\varphi(x)$ be a formula with free variable x that contains no quantifiers ranging over classes. Then we can collect all sets satisfying φ into a class:

$$\exists\, X\, \forall\, x\, (x \in X \Leftrightarrow \varphi(x))$$

Note that this is an axiom schema rather than a single axiom.

It allows us, for example, to form the class On of all ordinals.

The idea behind proper classes is that they are prevented from being a set by the fact that they contain too many elements. The following axiom allows one to get some mileage out of dealing with a proper class: if a class X fails to be a set there is a surjection from it onto V.

$$\neg M(X) \Leftrightarrow \exists\, F\, (\mathrm{Surj}(F) \wedge \mathrm{spt}(F) = X \wedge \mathrm{rng}(F) = V)$$

This axiom is quite powerful, e.g., Global Choice follows from it.

Consider the class On of all ordinals. Clearly On is a proper class, so by the axiom there is a surjection $On \to V$.

But then there must also be on injection $V \to On$, which injection provides a well-order on all of V.

We can exploit this well-order to construct a global choice function, essentially by setting

$$C(x) = \min\big(y \mid y \in x \big)$$

Checking the list of axioms of NBG we see that there is only one schema: Comprehension.

It is tempting to ask whether one could replace Comprehension by a finite list of axioms.

Note that a similar attempt fails in the framework of ZF, there is no finite axiomatization.

But the presence of classes makes it possible to avoid the Comprehension schema and replace it by handful of simple (albeit somewhat technical) axioms.

$$\exists X \forall u, v \, (\langle u, v \rangle \in X \iff u \in v)$$

$$\forall X, Y \exists Z \forall u \, (u \in Z \iff u \in X \wedge u \in Y)$$

$$\forall X \exists Y \forall u \, (u \in Y \iff u \notin X)$$

$$\forall X \exists Y \forall u \, (u \in Y \iff \exists v \, (\langle u, v \rangle \in X))$$

$$\forall X \exists Y \forall u, v \, (\langle u, v \rangle \in Y \iff u \in X))$$

$$\forall X \exists Y \forall u, v, w \, (\langle u, v, w \rangle \in Y \iff \langle v, w, u \rangle \in X)$$

$$\forall X \exists Y \forall u, v, w \, (\langle u, v, w \rangle \in Y \iff \langle u, w, v \rangle \in X)$$

The first axiom says that there is a class the encodes the $\in$ relation on sets (via Kuratowski pairs).

The second axiom, together with extensionality, allows us to introduce an intersection operation $\cap$.

The third provides a universal complementation operation.

The fourth is more interesting: if we think of the class X there as coding a relation, then the axiom provides a domain operator dom.

The fifth allows us to produce relations with a given domain.

The last two axioms make it possible to rearrange terms in an n-tuple.

The axioms may seem rather weak, but they conspire to establish the following theorem.

Suppose $\varphi(\boldsymbol{X}, \boldsymbol{Y})$ is a formula with free variables as indicated in which only sets are quantified over.

Theorem (Class Existence)

Given the alternative axioms, it is provable that

$$\exists\, Z\, \forall\, \boldsymbol{x}\, (\langle\, x_1, \ldots, x_n \,\rangle \in Z \iff \varphi(\boldsymbol{x}, \boldsymbol{Y}))$$

In other words, Class Comprehension is now a theorem and we really have a finite axiomatization of NBG.

We still need some more set existence axioms: union, power set and comprehension.

$$\exists\, u \forall\, z\, (z \in u \Rightarrow \exists\, y\, (y \in x \wedge z \in y))$$

$$\exists\, u \forall\, z\, (z \in u \Rightarrow z \subseteq x)$$

$$\exists\, u \forall\, z\, (z \in u \Rightarrow z \in x \wedge z \in Y)$$

Note that we already have Class Existence, so the Y in the last axiom is relatively easy to get a hold of.

With also need a suitable modification of Replacement.

It is a labor of love to write out a definition of a formula $\mathrm{Fct}(X)$ that expresses the assertion that X is a class representing a function (a single-valued class of pairs).

$$\mathrm{Fct}(X) \Rightarrow \forall\, x\, \exists\, y\, \forall\, u\, (u \in y \Leftrightarrow \exists\, v\, (\langle\, v, y\,\rangle\ \in X \wedge v \in x)$$

To avoid the trivial model, we also add Infinity.

Needless to say, one can optionally also include Choice and Regularity.

One naturally wonders what the difference between ZF and NBG in the end is.

Obviously, the existence of class variables makes it easier to deal with ordinals, cardinals, transfinite induction and the like. It is the case, though, that NBG is a conservative extension of ZF: every formula of ZF provable in NBG is already provable in ZF alone.

So why bother? The huge difference is that NBG is finitely axiomatized (check, no schemata anywhere, just plain axioms). On the other hand it is known that ZF does not admit finite axiomatizations.

Another way to accommodate the demands of category theory is to consider only sets, but to adopt a strong set existence axiom (unprovable in ZF).

Keep axioms for extensionality, singletons, pairs and unions, as well as foundation and replacement.

Then add Tarski's Axiom: for every set x we have

$$\exists\, U \left(x \in U \wedge \forall\, z \in U\, (\mathcal{P}(z) \subseteq U \wedge \mathcal{P}(z) \in U) \wedge \right.$$
$$\left. \forall\, z \in \mathcal{P}(U)\, (|z| < |U| \Rightarrow z \in U) \right)$$

Every set x belongs to a Grothendieck universe U.

This may seem a bit peculiar, but one can show that all the following "axioms" are consequences of the Tarski-Grothendiek axiom:

- Infinity
- Power set
- Choice

More importanly, one now can establish the existence of inaccessible cardinals and prove the consistency of ZF in TG.

- Intuitive set theory leads to paradoxes, but the problematic constructions are far removed from "real world" applications.
- Axiomatic set theory can deal with paradoxes, but at the cost of added technical difficulties.
- In practice, intuitive set theory, augmented with a bit of axiomatics, serves as a solid foundation of mathematics and computer science.
- In particular elementary concepts such as relations, functions, computable functions and so on can all be defined within set theory.