Sets

joseph.muscat@um.edu.mt
October 2013

1 Classes

Logic deals with statements and their deductions, and the simplest statements are usually of the type "x is A", where x is an object and A a property. In practice, a property may itself be an object with properties e.g. "roses are red", "red is a color", so that there ought to be no distinction in notation between the two; the two are amalgamated into one concept of "class":

The **class** $\{\, x : x \text{ IS } A \,\}$ corresponds to the property A, and can be thought of as a "collection" of objects x, called its *elements*. For convenience, we will call the class also by the name A to avoid duplicating symbols unnecessarily, because it contains the same information as the "property" A. The statement x IS A (as a property) is rewritten as $x \in A$ (as a class).

The basic construct in logic is $\Rightarrow$; the corresponding structure in classes is the *subset*, and *equality*.

Definition:

Equality of classes	$A = B$	when $\forall x,\ x \text{ IS } A \Leftrightarrow x \text{ IS } B$
Subset	$A \subseteq B$	when $\forall x,\ x \text{ IS } A \ \Rightarrow\ x \text{ IS } B.$

Logic already defines equality of objects and of properties by

$$x = y \ \Leftrightarrow\ \forall C, (x \in C \ \Leftrightarrow\ y \in C)$$
$$A = B \ \Leftrightarrow\ \forall x, (x \in A \ \Leftrightarrow\ x \in B)$$

(Informally, the "entity" $\{\, A : x \text{ IS } A \,\}$ ought to characterize x, just as $\{\, x : x \text{ IS } A \,\}$ characterizes A.) For consistency, if we are to amalgamate the two concepts, we are forced to make the assumption:

Axiom 1 of Extensionality: **Equality** as properties and as objects are equivalent,

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

However, we are immediately faced with a contradiction if we allow every property to act as an object with properties: let $x \text{ IS } A := \text{NOT } (x \text{ IS } x)$ then $A \text{ IS } A \Leftrightarrow \text{NOT } (A \text{ IS } A)$. This is Russel's "barber's paradox" (the barber who shaves everyone who does not shave himself: who shaves the barber?); the self-referential x IS x is similar to references like an "omnipotent who creates an indestructible". The way out of this is to distinguish between two types of classes:

- *Proper classes* are those that cannot have any properties, i.e., substituting a proper class instead of x in "x IS A" may lead to a contradiction; for example, the class $\{\, x : x \notin x \,\}$ is a proper class.
- *Sets* are classes which have properties; i.e., those x such that $x \in A$ for some A (those classes that we can say something about)[1]. Since there is no guarantee that they exist, we need

Axiom 2 of Existence: **There is a set.**

The informal correspondence between classes and properties,

$$A = \{\, x : x \text{ IS } A \,\}, \quad x \in A \Leftrightarrow x \text{ IS } A,$$

is extended to cover the composite properties:

Empty set	FALSE	$\varnothing = \{\,\} := \{\, x : \text{FALSE} \,\}$
Universal class	TRUE	$\Upsilon := \{\, x : \text{TRUE} \,\}$
Complement	NOT (x IS A)	$A^{\mathsf{c}} := \{\, x : x \notin A \,\}$
Intersection	(x IS A) AND (x IS B)	$A \cap B := \{\, x : (x \in A) \text{ AND } (x \in B) \,\}$
	$\forall y,\ (x,y)$ IS A	$\bigcap A = \bigcap_y A_y := \{\, x : \forall y,\ (x,y) \in A \,\}$
Union	(x IS A) OR (x IS B)	$A \cup B := \{\, x : (x \in A) \text{ OR } (x \in B) \,\}$
	$\exists y,\ (x,y)$ IS A	$\bigcup A = \bigcup_y A_y := \{\, x : \exists y,\ (x,y) \in A \,\}$
Singleton	$x = a$	$\{\, a \,\} := \{\, x : x = a \,\}.$

Other definitions:

$$\begin{aligned} A \setminus B &:= \{\, x : (x \text{ IS } A) \text{ BUTNOT } (x \text{ IS } B) \,\} = \{\, x \in A : x \notin B \,\}, \\ A \triangle B &:= (A \setminus B) \cup (B \setminus A) = \{\, x : (x \text{ IS } A) \text{ XOR } (x \text{ IS } B) \,\}, \\ \{\, a, b \,\} &:= \{\, a \,\} \cup \{\, b \,\} = \{\, x : x = a \text{ OR } x = b \,\}, \\ \{\, a, b, c \,\} &:= \{\, a, b \,\} \cup \{\, c \,\} = \{\, x : x = a \text{ OR } x = b \text{ OR } x = c \,\}, \textit{etc.} \end{aligned}$$

We often write $\bigcup_{B \in A} B$ or $\bigcup_{i \in I} A_i$ instead of $\bigcup A$, where I is an indexing set of A; similarly for $\bigcap_{B \in A} B$ or $\bigcap_{i \in I} A_i$ instead of $\bigcap A$.

The logical tautologies can then be written as theorems of classes:

1. $A \subseteq B$ AND $B \subseteq C \ \Rightarrow\ A \subseteq C$
2. $A = A$,
   $A = B \ \Rightarrow\ B = A$,
   $A = B$ AND $B = C \ \Rightarrow\ A = C$

[1] There may also be objects that are not properties, called "ur-elements".

3. $A^{\mathsf{cc}} = A, \quad \varnothing^{\mathsf{c}} = \Upsilon,$
   $\varnothing \subseteq A \subseteq \Upsilon, \quad \Upsilon^{\mathsf{c}} = \varnothing$

4. $A \cup B = B \cup A, \quad A \cup \varnothing = A, \quad A \cup A = A,$
   $A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cup \Upsilon = \Upsilon.$

5. $A \cap B = B \cap A, \quad A \cap \varnothing = \varnothing, \quad A \cap A = A,$
   $A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cap \Upsilon = A.$

6. $A \cap A^{\mathsf{c}} = \varnothing, \quad (A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}},$
   $A \cup A^{\mathsf{c}} = \Upsilon, \quad (A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}.$

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

8. $A \subseteq B \Leftrightarrow A = A \cap B \Leftrightarrow B = A \cup B \Leftrightarrow B^{\mathsf{c}} \subseteq A^{\mathsf{c}}$

9. $A \cap B \subseteq A \subseteq A \cup B$

10. $A \setminus B = A \cap B^{\mathsf{c}} = A \setminus (A \cap B) = (A \cup B) \setminus B;$
    $A \setminus A = \varnothing;\ A \setminus \varnothing = A;$
    $A \subseteq B \Leftrightarrow A \setminus B = \varnothing;$
    $A \setminus (B \setminus C) = (A \setminus B) \cup (A \cap C);$
    $A \setminus (A \setminus B) = A \cap B;$
    $(A \setminus B) \setminus C = A \setminus (B \cup C);$
    $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C);$
    $A = (A \cap B) \cup (A \setminus B)$

11. $A \triangle B = B \triangle A, \quad A \triangle \varnothing = A, \quad A \triangle A = \varnothing$
    $(A \triangle B) \triangle C = A \triangle (B \triangle C), \quad A \triangle \Upsilon = A^{\mathsf{c}},$
    $A \triangle B = (A \cup B) \setminus (A \cap B), \quad A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C).$

12. $\forall x, x \notin \varnothing; \qquad A \neq \varnothing \Leftrightarrow \exists x \in A$
    (since $x \in \varnothing \Leftrightarrow$ FALSE; conversely, $x \notin A \Leftrightarrow (x \in A \Rightarrow \text{FALSE})$, so $\varnothing \subseteq A \subseteq \varnothing$.)

13. For sets, $\{ a \} = \{ b \} \Leftrightarrow a = b, \qquad \{ a \} = \{ b, c \} \Leftrightarrow a = b = c.$

14. $\bigcup \varnothing = \varnothing, \qquad \bigcap \varnothing = \Upsilon, \qquad \bigcup \{ a \} = a = \bigcap \{ a \}$

15. $A \cup B = \bigcup \{ A, B \}, A \cap B = \bigcap \{ A, B \}.$
    (Proof: $x \in A \cup B \Leftrightarrow x \in A$ OR $x \in B \Leftrightarrow \exists y \in \{ A, B \}, x \in y \Leftrightarrow x \in \bigcup \{ A, B \}$.)

16. $A \cap \bigcup_i B_i = \bigcup_i (A \cap B_i) \qquad A \cup \bigcap_i B_i = \bigcap_i (A \cup B_i)$
    $(\bigcap_i A_i)^{\mathsf{c}} = \bigcup_i A_i^{\mathsf{c}} \qquad (\bigcup_i A_i)^{\mathsf{c}} = \bigcap_i A_i^{\mathsf{c}}$
    $\bigcup_i \bigcap_j A_{i,j} \subseteq \bigcap_j \bigcup_i A_{i,j}$
    The reverse inclusion may be false: $(A \cap B) \cup (C \cap D) \subseteq (A \cup C) \cap (B \cup D)$, e.g. $A = D = \{ a \}$, $B = C = \varnothing$.

17. $\{ x : (x \text{ IS } A) \Rightarrow (x \text{ IS } B) \} = A^{\mathsf{c}} \cup B.$

1.1 Defining Sets Properly

The way to form sets from predicates, i.e., sets of elements satisfying a particular property, is to consider

$$\{\, x \in A : x \in B \,\} := A \cap B,$$

where A is a set and B a class (property). So if we let $R := \{\, x \in A : x \notin x \,\}$, then $R \in R \Leftrightarrow R \in A$ AND $R \notin R$, hence $R \notin R$ and $R \notin A$, without contradiction. Incidentally, this shows that for every set A there is a set R not in A; hence the class $\{\, x : x \text{ is a set } \}$ cannot be a set, but is a proper class.

To ensure that $\{\, x \in A : x \in B \,\}$ is a set, we assume:

Axiom 3' of Specification: **A subset of a set is itself a set.**

In particular, (i) $\varnothing$ is a set, since sets exist ($\varnothing = A \cap A^{\mathsf{c}}$), (ii) $A \setminus B$ is a set when A is, (iii) $\bigcap_y A_y$ is a set when any A_x is a set. But A^{c} and $B \setminus A$ (with B proper) need not be sets.

Conversely, if $A \subseteq B$ and A is a proper class, then so is B. In particular, the class of everything, Υ, is a proper class, since there are proper classes.

Axiom 3' essentially states that if we start with a set and remove some elements, the result is still a set. We usually assume a stronger form of this:

Axiom 3 of Replacement: **If the elements x of a non-empty set A are replaced by other sets $f(x)$, the result is a set $\{\, f(x) : x \in A \,\}$.**

Equivalently, $\{\, A_i : i \in I \,\}$ is a set if A_i and I are sets. Note that Axiom 3 is not really one axiom but a *schema* of axioms, one for each f.

It contains as a special case the removal of elements, because this is equivalent to replacing them with an already existing element of A (unless $A = \varnothing$)[2].

Another improper definition of a set is a *circular* one, when a set is defined in terms of itself, as in $A \in A$ or $A \in B \in A$, etc., or an *infinite regress* when there is an unending chain $\ldots \in C \in B \in A$. These improper definitions, in which we feel that we can never know what are the elements of A, are disallowed by requiring:

Axiom 4 of Regularity/Foundation/Well-definition:
For any nonempty set A, it cannot be the case that *all* its elements have elements of A, i.e.,

$$\text{NOT } \exists A \neq \varnothing, \forall x \in A, \exists y \in A, y \in x$$

In particular for any set A, it cannot be the case that $A \in A$ or $A \in B \in A$ etc., $A \ni B \ni C \ni \cdots$, otherwise $\{\, A \,\}$, $\{\, A, B \,\}$, $\ldots$, or $\{\, A, B, C, \ldots \,\}$ contradict the axiom (that these are sets follows from subsequent axioms). Note that, therefore, $\{\, x : x \notin x \,\} = \Upsilon$.

[2] Hence, Axiom 3 does not imply, by itself, that $\varnothing$ is a set; so usually Axiom 3' is dropped and Axiom 2 is replaced by "$\varnothing$ is a set".

1.2 Construction of Sets

Up to now, only $\varnothing$ is guaranteed to be a set. There now follow two axioms about how larger sets can be constructed from given ones.

Axiom 5 of Powers: **When A is a set, so is the set of subsets**

$$2^A := \{\, x : x \subseteq A \,\}$$

Consequences: $2^{\varnothing} = \{\, \varnothing \,\}$ is a set; so applying the axiom of replacement, $\{\, a \,\}$ is a set when a is a set.
$2^{\{\, \varnothing \,\}} = \{\, \varnothing, \{\, \varnothing \,\} \,\}$ is a set; so again, $\{\, a, b \,\}$ is a set when a and b are.
$A = B \Leftrightarrow 2^A = 2^B$.

Axiom 6 of Unions: **When A is a set, so is $\bigcup A$.**

It follows that $A \cup B = \bigcup \{\, A, B \,\}$ is a set when A, B are sets. In particular $A \triangle B$ is a set when A and B are.

Ordered pairs of sets are defined[3] as

$$(a, b) := \{\, \{\, a, 0 \,\}, \{\, b, 1 \,\} \,\}$$

(Here, 0 and 1 are any two distinct markers). It follows that

$$(a, b) = (c, d) \Leftrightarrow a = c \text{ AND } b = d,$$

and so $a \neq b \;\Rightarrow\; (b, a) \neq (a, b)$. Of course, the definition can be extended to (a, b, c) as $(a, (b, c))$ or as $\{\, \{\, a, 0 \,\}, \{\, b, 1 \,\}, \{\, c, 2 \,\} \,\}$
The class of ordered pairs is denoted

$$A \times B := \{\, (a, b) : a \in A \text{ AND } b \in B \,\}$$

We also write $A^2 := A \times A$. $A \times B$ is a set when A and B are, since $A \times B \subseteq 2^{2^{A \cup B \cup \{\, 0,1 \,\}}}$.

$$(A \cup B) \times C = (A \times C) \cup (B \times C),$$
$$(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D),$$
$$\varnothing \times A = \varnothing = A \times \varnothing,$$
$$A \subseteq C \text{ AND } B \subseteq D \;\Rightarrow\; A \times B \subseteq C \times D.$$

[3] A common alternative definition is $\{\, \{\, a \,\}, \{\, a, b \,\} \,\}$; but its disadvantage is that it does not scale up: $(a, b, c) := \{\, \{\, a \,\}, \{\, a, b \,\}, \{\, a, b, c \,\} \,\}$ gives $(a, a, b) = (a, b, b)$.

2 Functions

The class equivalent of a **relation** is $\rho := \{\, (x,y) : \rho_{x,y} \,\}$. Thus $x\ \rho\ y \Leftrightarrow (x,y) \in \rho$. One can generalize to relations with n variables.

A relation on two sets X and Y is a subset of $X \times Y$, denoted by $\rho : X \to Y$.

The *identity* relation is ι, defined as equality $x = y$.

The *inverse* relation is $\rho^{-1} := \{\, (y,x) : x\ \rho\ y \,\} : Y \to X$; then $(\rho^{-1})^{-1} = \rho$.

The *image* of a subset $A \subseteq X$ is the set $\rho A := \{\, y \in Y : \exists x \in A, x\ \rho\ y \,\}$; the image of an element a is the set $\rho\{\, a \,\} = \{\, y \in Y : a\ \rho\ y \,\}$. The *image* of ρ is $\operatorname{im}\rho := \rho X$ (when X is understood); while its *domain* is $\rho^{-1}Y$.

$$\begin{aligned} &A \subseteq B \ \Rightarrow\ \rho A \subseteq \rho B, && \rho\varnothing = \varnothing, \\ &\rho(A \cup B) = \rho A \cup \rho B, && \rho \textstyle\bigcup_i A_i = \bigcup_i \rho A_i, \\ &\rho(A \cap B) \subseteq \rho A \cap \rho B, && \rho \textstyle\bigcap_i A_i \subseteq \bigcap_i \rho A_i. \end{aligned}$$

(The inverse ρ^{-1} is more properly denoted ρ^* because it is not a true inverse: $\rho \circ \rho^{-1}$ need not be ι.)

A relation ρ can be *restricted* $\rho|_{C\times D} := \rho \cap (C \times D)$; relations $\rho : A \to B$ and $\sigma : C \to D$ can be *composed* to give

$$\sigma \circ \rho := \{\, (a,c) : \exists b \in B \cap C, a\ \rho\ b\ \sigma\ c \,\}$$

satisfying

$$\tau \circ (\sigma \circ \rho) = (\tau \circ \sigma) \circ \rho, \quad \rho \circ \iota = \rho = \iota \circ \rho, \quad (\sigma \circ \rho)^{-1} = \rho^{-1} \circ \sigma^{-1}.$$

The union of two relations $\rho, \sigma : X \to Y$, $\rho \cup \sigma$ is also a relation; more generally $\bigcup_i \rho_i$ is a relation.

A **function** is a relation for which the image on any element of the domain X consists of exactly one element, $\forall x \in X, \exists y \in Y, f\{\, x \,\} = \{\, y \,\}$, i.e., the function preserves equality

$$\text{If } (x,y) \in f \text{ and } (x',y') \in f, \text{ then } x = x' \ \Rightarrow\ y = y',$$

equivalently $\forall x \in X, \exists! y \in Y, (x,y) \in f$. We can therefore write $f(x)$ instead of the object y whenever $(x,y) \in f$. Thus

$$x = x' \ \Rightarrow\ f(x) = f(x')$$

The axiom of replacement, written rigorously, states that if A is a set and f a function (with domain and codomain being classes), then $fA = \{\, f(x) : A_x \,\}$ is a set.

If two functions are equal $f = g$ then they have the same domain X and image Y, and $\forall x \in X, f(x) = g(x)$. Note that the identity relation, restrictions with the same codomain, and composition of functions $X \to Y \subseteq C \to D$ are functions.

$$fA \subseteq B \Leftrightarrow A \subseteq f^{-1}B$$
$$f\bigcup_i A_i = \bigcup_i fA_i \qquad f^{-1}\bigcap_i A_i = \bigcap_i f^{-1}A_i$$
$$f^{-1}A^{\mathsf{c}} = (f^{-1}A)^{\mathsf{c}} \qquad A \subseteq f^{-1}fA, \quad ff^{-1}A \subseteq A$$

Important examples of functions are *projections* $\pi_X : X \times Y \to X$, $(x, y) \mapsto x$, and $\pi_Y : X \times Y \to Y$, $(x, y) \mapsto y$. The *graph* of a function is $G_f := \{\, (x, f(x)) \subseteq X \times Y : x \in X \,\}$.

A function is *1-1 (injective)* when

$$f(x) = f(y) \;\Rightarrow\; x = y \quad (\Leftrightarrow\; f^{-1} \circ f = \iota \;\Leftrightarrow\; fA^{\mathsf{c}} \subseteq (fA)^{\mathsf{c}}),$$

it is *onto (surjective)* when $fX = Y$, i.e.,

$$\forall y \in Y,\; \exists x \in X,\; f(x) = y \quad (\Leftrightarrow\; f \circ f^{-1} = \iota \;\Leftrightarrow\; (fA)^{\mathsf{c}} \subseteq fA^{\mathsf{c}}).$$

If $f \circ g = \iota$ then f must be onto and g 1-1.

The inverse f^{-1} is itself a function when f is both 1-1 and onto (called *bijective*), equivalently $f^{-1} \circ f = \iota_X$, $f \circ f^{-1} = \iota_Y$.

When f is 1-1, $f(A \cap B) = fA \cap fB$; when f and g are 1-1 (resp. onto), then $f \circ g$ is also 1-1 (resp. onto); so the composition of bijective functions is again bijective; so the set of bijective functions with domain and codomain A is closed under composition, called the *permutation group* $A!$.

The set of all functions $f\colon A \to B$ is denoted B^A (it is a set when A and B are, since $B^A \subseteq 2^{A\times B}$); more generally the set

$$\prod_{i\in I} A_i := \{\, f : I \to \bigcup_i A_i \text{ AND } \forall i \in I, f(i) \in A_i \,\}$$

1. If $X = A \cup B$ disjoint, then a function $f : X \to Y$ can be identified with $(f|_A, f|_B)$ and so $Y^{A\cup B}$ with $Y^A \times Y^B$. Ultimately, $f \in Y^X$ is the same as $(f(x))_{x\in X} \in \prod_{x\in X} Y$.
2. A function on $A \times B$ can be written as functions $f_b(a) := f(a, b)$ for each $b \in B$. Thus $C^{A\times B}$ can be identified with $(C^A)^B$.
3. A function to $A \times B$ can be written as a pair $f(x) = (f_A(x), f_B(x))$, so $(A \times B)^C$ is identified with $A^C \times B^C$.

A relation $\approx\colon X \to X$ is called an *equivalence relation* when it is

- *transitive* $x \approx y$ AND $y \approx z \;\Rightarrow\; x \approx z$ (i.e., $\approx \circ \approx \subseteq \approx$);
- *reflexive* $x \approx x$ (i.e., $\iota|_{X^2} \subseteq \approx$); and
- *symmetric* $x \approx y \;\Rightarrow\; y \approx x$ (i.e., $\approx^{-1} = \approx$).

Equivalence relations are in correspondence with partitions: A *partition* of A is a class of subsets $B_i \subseteq A$ such that $A = \bigcup_i B_i$ and $i \neq j \;\Rightarrow\; B_i \cap B_j = \varnothing$.

Proposition 1

Every equivalence relation induces a partition on its domain and vice versa.

PROOF: Suppose ρ is an equivalence relation. Let $[x] := \rho\{\, x \,\}$, called the *equivalence class* of x, and let $P := \{\, [x] : x \in X \,\}$ be the class of equivalence classes. Then $A = \bigcup_{[x]\in P}[x]$ since $y \in A \;\Rightarrow\; y \in [y] \in P$; and $[x] \cap [y] \neq \varnothing \;\Rightarrow\; \exists z, z\rho x$ AND $z\rho y \;\Rightarrow\; x\rho y$, and for any $z, z \in [x] \;\Leftrightarrow\; z\rho x \;\Leftrightarrow\; z\rho y \;\Leftrightarrow\; z \in [y]$ so $[x] = [y]$.

Conversely, if P is a partition of A, let $x \approx y$ be defined by $\exists i, x, y \in A_i \in P$. It is transitive since $x, y \in A_i$ and $y, z \in A_j$ implies $A_i \cap A_j \neq \varnothing$ and so $A_i = A_j$, so $x, z \in A_i$. It is reflexive since $x, x \in A_i$ for some i. And it is obviously symmetric.

□

The class of equivalence classes associated with an equivalence relation $\approx$ is denoted $X/\!\approx$. The *kernel* of a function $f : A \to B$ is the partition on A induced by the equivalence relation $f(x) = f(y)$; the equivalence classes are $f^{-1}(b)$.

$$f \text{ is 1-1} \;\Leftrightarrow\; \ker f = \ker \iota_A \;\Leftrightarrow\; \ker f \circ g = \ker g,$$

$$f \text{ is onto} \;\Leftrightarrow\; \operatorname{im} f = \operatorname{im} \iota_B \;\Leftrightarrow\; \operatorname{im} g \circ f = \operatorname{im} g.$$

An equivalence relation is 'finer' than another, $x \approx_1 y \;\Rightarrow\; x \approx_2 y$, iff its partition is finer than the other, $[x]_1 \subseteq [x]_2$. Moreover, there is a well-defined equivalence relation on $X/\!\approx_1$, $[x]_1 \approx [y]_1 \;\Leftrightarrow\; x \approx_2 y$.

An equivalence relation on X, when restricted to a subset Y, remains an equivalence relation. Its equivalence classes are $[x] \cap Y$. A set of equivalence classes of X corresponds to a subset of X that contains whole equivalence classes.

Equivalence relations on X and Y determine an equivalence relation on $X \times Y$:

$$(x_1, y_1) \approx (x_2, y_2) \;\Leftrightarrow\; (x_1 \approx y_1) \text{ AND } (x_2 \approx y_2)$$

Equivalence classes are $[(x, y)] = [x] \times [y]$, essentially the same as $([x], [y])$.

2.1 Axiom of Choice

Axiom 7 of Choice: **For any sets** $A_i \neq \varnothing$ and I, $\prod_{i\in I} A_i \neq \varnothing$

Equivalently,

For every set A, there is a function $\epsilon : 2^A \to A$ such that $\varnothing \neq B \subseteq A \;\Rightarrow\; \epsilon(B) \in B$;

For every onto function $f\colon X \to Y$ there is a function $g\colon Y \to X$ such that $f \circ g = \iota$.

Proof: For $A \neq \varnothing$, $A^{2^A} \supseteq \prod_{\varnothing \neq B \subseteq A} B \neq \varnothing$, i.e., $\exists \epsilon : 2^A \to A$, such that $\epsilon(B) \in B$. Given $f : X \to Y$, let $g(y) := \epsilon(f^{-1}y)$, so $f \circ g(y) \in ff^{-1}y = \{ y \}$. Given $A_i \neq \varnothing$ (wolog disjoint), let $f : \bigcup_i A_i \to I$ be defined by $f(x) := i$ if $x \in A_i$; then $f \circ g(i) = i$, i.e., $g(i) \in A_i$.

This axiom seems sensible because it disallows collections of sets that have elements that cannot be sampled collectively. Nonetheless it is controversial because it introduces a function that cannot be constructed, and because it has some unexpected consequences (e.g. Banach-Tarski "paradox"); yet these theorems cannot become false when this axiom is not included, they may simply become undecidable. Moreover, there are also paradoxes when the axiom of choice is false, e.g. infinite sets that do not contains countable subsets. A stronger global form of the axiom of choice is: $\exists \epsilon : \Upsilon \to \Upsilon, \epsilon(A) \in A$ for $A \neq \varnothing$.

3 Numbers

Two sets A and B are said to have the same **number** of elements (or are *cardinally equivalent*) when there is a bijective function $f\colon A \to B$. This is an equivalence relation, here denoted by $A \equiv B$.

A set A is said to have *less* elements than B when A is *embedded* in B, denoted $A \lesssim B$, i.e., there is a subset $C \subseteq B$ such that $A \equiv C \subseteq B$; this is equivalent to saying that

(i) there is a 1-1 function $f\colon A \to B$ (take $C := fA$), i.e., the *pigeon-hole principle*: if $A \not\lesssim B$ then any function $A \to B$ is not 1-1; if you fit 10 pigeons in 7 pigeon-holes at least one pigeon-hole must be shared; or

(ii) there is an onto function $g\colon B \to A$ (Proof: If $f : A \to B$ is 1-1, then $g(b) := \begin{cases} a & \text{if } f(a) = b, \\ a_0 & \text{if } b \notin \operatorname{im} f \end{cases}$; if $g : B \to A$ is onto, let f be a (1-1) choice function $f(a) \in g^{-1}(a)$).

Clearly, $A \lesssim A$, and $A \lesssim B \lesssim C \;\Rightarrow\; A \lesssim C$. Also, $A \subseteq B \;\Rightarrow\; A \lesssim B$, and $A \equiv fA$.

Proposition 2

Either A has less elements than B or vice-versa.

PROOF: Assuming $A, B \neq \varnothing$, and the Hausdorff Maximality Principle (a consequence of the Axiom of Choice, see Ordered Spaces), then the set of 1-1 functions with domain in A and image in B has a maximal element f, when ordered by inclusion. Were both the domain and image of f strictly less than A, B, then it can be extended by some (a, b) and remain 1-1. Thus either $f : A \to B$ is 1-1, or $f^{-1} : B \to A$ is 1-1.

□

The following states that if A has at most as many elements as B, and B as many as A, then they have the same number of elements:

Proposition 3

If $A \lesssim B \lesssim A$ then $A \equiv B$.

That is, if $f\colon A \to B$ and $g\colon B \to A$ are both 1-1, then A, B are numerically equivalent.

PROOF: Let $F\colon 2^A \to 2^A$ be defined by $F(C) := g(f(C)^\mathsf{c})^\mathsf{c}$. It is increasing ($C_1 \subseteq C_2 \;\Rightarrow\; F(C_1) \subseteq F(C_2)$), hence has a fixed point (see Complete lattices), i.e., a set D such that $g(f(D)^\mathsf{c})^\mathsf{c} = D$. Therefore $g(f(D)^\mathsf{c}) = D^\mathsf{c}$, so one can define a new map $h : A \to B$ by

$$h(x) := \begin{cases} f(x) & x \in D, \\ g^{-1}(x) & x \in D^\mathsf{c}. \end{cases}$$

It is 1-1 and onto since $g^{-1}D^\mathsf{c} = f(D)^\mathsf{c}$.

□

Corollary: If $A \lesssim B \lesssim C \lesssim A$, then $A \equiv B$.

Thus we can form equivalence classes of cardinally equivalent sets, and any representative set can be taken to be its (cardinal) **number $\boldsymbol{n}$**. More specifically, such a set can be chosen to be an ordinal number (see Order). Note that the class of numbers is not a set (else $2^\mathcal{N} \in \mathcal{N}$); neither need be each equivalence class (since $x \mapsto \{\, x \,\}$, $\Upsilon \to [1]$ is 1-1).

Moreover for disjoint sets A_i with cardinal numbers $\boldsymbol{n}_i$, we define (these are well-defined)

$$\sum_i \boldsymbol{n}_i := [\bigcup_i A_i], \qquad \prod_i \boldsymbol{n}_i := [\prod_i A_i], \qquad \boldsymbol{n}^{\boldsymbol{m}} := [A^B].$$

We also write $\boldsymbol{n} \leqslant \boldsymbol{m}$ when $A \lesssim B$.

Proposition 4

A function $f\colon A \to 2^A$ on sets cannot be onto, nor $f\colon 2^A \to A$ be 1-1:

$$\boldsymbol{n} < 2^{\boldsymbol{n}}$$

PROOF: Suppose $f\colon A \to 2^A$ is onto. Then $B := \{\, x \in A : x \notin f(x) \,\}$ is a subset of A. So there must be an element $y \in A$ such that $f(y) = B$, but $y \in B \Leftrightarrow y \notin f(y) \Leftrightarrow y \notin B$.

□

This was the first indication to Cantor that there was something wrong with sets as he defined them, since $2^\Upsilon = \Upsilon$.

One can form a sequence of cardinally inequivalent numbers as follows: for any set A, let $A^+ := A \cup \{\, A \,\}$:

$$\begin{aligned}
0 &:= \varnothing = \{\,\} && =\{\} \\
1 &:= 0^+ = \{\, 0 \,\} && =\{\{\}\} \\
2 &:= 1^+ = \{\, 0, 1 \,\} && =\{\{\},\{\{\}\}\} \\
3 &:= 2^+ = \{\, 0, 1, 2 \,\} && =\{\{\},\{\{\}\},\{\{\},\{\{\}\}\}\} \\
4 &:= 3^+ = \{\, 0, 1, 2, 3 \,\} && =\{\{\},\{\{\}\},\{\{\},\{\{\}\}\},\{\{\},\{\{\}\},\{\{\},\{\{\}\}\}\}\} \\
&\ldots
\end{aligned}$$

It will be shown shortly that these sets are inequivalent, so it is clear that we would need more and more symbols to denote all these numbers. We normally adopt a *number system* to extend these numerals and denote larger numbers: in the decimal system, after 9, we write 10, then 11, etc.; in the binary system, we write 10 after 1, then 11, 100, 101, etc.

The numbers constructed this way are called the **natural numbers**. Any set that is numerically equivalent to one of them is called **finite**. To say this better, we need to form the set of natural numbers, but how do we characterize them?

A class is said to be *inductive* when $n \in A \;\Rightarrow\; n^+ \in A$. The intersection of inductive classes is itself inductive, so one can generate an inductive class starting from any set, as $\mathrm{Ind}(A) := \bigcap\{\, B : \text{inductive}, A \subseteq B \,\}$. Then the class of natural numbers is the inductive class generated by 0,

$$\mathbb{N} := \mathrm{Ind}(\{\, 0 \,\}) = \bigcap\{\, A : A \text{ inductive}, 0 \in A \,\}$$

The numbers defined above, namely 0, $1 = 0^+$, 2, etc. are in $\mathbb{N}$, since it is inductive; and there are no other since $\mathbb{N}$ is the least such class.

The next axiom is controversial in that it accepts that "actual" infinity is logically consistent.

Axiom 2' of Infinity: $\mathbb{N}$ **is a set**

Equivalently, there exists an inductive set; this can replace the axiom that there exists a set. By replacement, $\{\, a_0, a_1, \ldots \,\}$ is also a set when a_i are sets. The sets that are

(i) numerically equivalent to $\mathbb{N}$ are called *countably infinite*, with their number sometimes denoted by ω.

(ii) numerically less than $\mathbb{N}$ are called *finite*, or

(iii) numerically more than $\mathbb{N}$ are called *uncountable*.

Theorem 5

(Peano's Axioms)

1. $0 \in \mathbb{N}$
2. $\forall n \in \mathbb{N}, n^+ \in \mathbb{N}$
3. $\forall n \in \mathbb{N}, n^+ \neq 0$
4. $\forall m, n \in \mathbb{N}, m^+ = n^+ \Leftrightarrow m = n$
5. $0 \in A$ AND $\forall n \in A, n^+ \in A \Rightarrow \mathbb{N} \subseteq A$ **(induction)**

PROOF: (1) and (2) say that $\mathbb{N}$ is inductive, generated from 0, (5) that it is the smallest inductive set. (3) follows from the definition of n^+. (4) follows from $A \cup \{ A \} = B \cup \{ B \} \Leftrightarrow A = B$, which uses the regularity axiom.

□

(Note: by the Löwenheim-Skolem theorem, there is an uncountable model satisfying these axioms. A model is a set with a relation $\in$ defined on it that satisfies the set axioms. Any model M has a countable sub-model N such that M and N have exactly the same true statements.)

Proposition 6

Countable Well-Ordering

Every non-empty $A \subseteq \mathbb{N}$ has a least element.

PROOF: Suppose A has no least element, and let $B := \mathbb{N} \setminus A$; then $0 \notin A$, so $0 \in B$ and $\{ 0, \ldots, n \} \subseteq B \Rightarrow \{ 0, \ldots, n, n^+ \} \subseteq B$ else n^+ would be the least element of A; hence $B = \mathbb{N}$ by induction and $A = \varnothing$.

□

Note that $\mathbb{N}$ is well-ordered using the relation $\in$; in particular, distinct natural numbers satisfy $m \in n$ OR $n \in m$.

(If the universal set were $\mathbb{N}$, then the axioms of replacement, regularity and choice are theorems and there is no need for the axiom of infinity.)

The natural numbers are indeed cardinally inequivalent.

PROOF: Suppose $f : n^+ \to m^+$ is a bijection; note that $n \equiv 0$ implies $n = 0$ as the only set cardinally equivalent to $\varnothing$. Pick the least n, m with such

a bijection. Then let $g : n \to m$, $g(r) := \begin{cases} f(n) & \text{when } r = f^{-1}(m) \\ f(r) & o/w \end{cases}$; it remains 1-1 and onto, so $n \equiv m$. By induction, $n = m$ so $n^+ = m^+$.

□

In particular, $n < n^+$.

Proposition 7

A set A is infinite	$\Leftrightarrow \exists B \subset A, \lvert A\rvert = \lvert B\rvert$
	$\Leftrightarrow \exists B \subseteq A, \lvert B\rvert = \lvert\mathbb{N}\rvert$
	$\Leftrightarrow \lvert\mathbb{N}\rvert \leqslant \lvert A\rvert$
	$\Leftrightarrow \forall n \in \mathbb{N}, n \leqslant \lvert A\rvert$

PROOF: (needs the axiom of choice) By definition, A is infinite when for all $n \in \mathbb{N}$, $|A| \neq n$. $0 \leqslant |A|$ since any function $0 \to A$ is 1-1; if $n \leqslant |A|$ then there is a 1-1 function $f\colon n \to A$, but not onto (else A is finite); let $f(n^+) = x \notin fn$, then $f\colon n^+ \to A$ is still 1-1; by induction, $n \leqslant |A|$ for all $n \in \mathbb{N}$. Select distinct elements $a_i \in A$; if $n \leqslant |A|$ for all n, then $\{ a_0, \ldots, a_n \}$ is not all of A, hence there is a 1-1 mapping $f : \mathbb{N} \to A$, $n \mapsto a_n$. In particular $\mathbb{N} \equiv B := f\mathbb{N} \subseteq A$. Now $\mathbb{N} \equiv \mathbb{N} \setminus 0$ (using $n \mapsto n^+$), so $B = f\mathbb{N} \equiv B \setminus \{ b_0 \} =: C$; thus $A = B \cup B^{\mathsf{c}} \equiv C \cup B^{\mathsf{c}} \subset A$. Clearly $B \subset 0 = \varnothing$ is impossible; if $B \subset n \equiv B$ then $n \notin B$, so $B \cup \{ n \} \subset n \cup \{ n \} = n^+ \equiv B \cup \{ n \}$; by induction, $B \subset n \equiv B$ is false.

□

In particular $\mathbb{N}$ is the smallest infinite set (in cardinality). Any unbounded subset of $\mathbb{N}$ is countably infinite.

3.1 Computable Sets

A predicate statement is said to be *computable* (or *recursive*) when there is a (terminating) algorithm that decides whether x IS A is true or not. Sets which correspond to computable properties are called *computable sets*; since the input of an algorithm is from a countable set, computable sets are usually taken to be subsets of $\mathbb{N}$; e.g. $\varnothing$, $\mathbb{N}$, finite sets, evens, primes. If A, B are computable, then so are A^{c}, $A \cup B$, $A \cap B$, $A \times B$. In fact, the number of computable subsets of $\mathbb{N}$ can only be countable, so most subsets are uncomputable.

A predicate is said to be *semi-computable* (or *recursively enumerable*) when there is an algorithm that confirms that x IS A is true, but may run forever if false; it can check $n \in A$, but not necessarily $n \notin A$. So a set is *semi-computable* when there is an algorithm that lists the elements (check the next $n \in A$ every unit of time; at any stage there are a finite number of algorithm instances).

A is then computable when both A and A^{c} are semi-computable. If A, B are semi-computable then so are $A \cup B$, $A \cap B$, $A \times B$. There are semi-computable sets that are not computable, e.g. the Halting problem = the set of algorithms that halt with input 0; also, the set of provable statements (see the discussion after Tarski's theorem).

Proposition 8

The set of computable subsets of $\mathbb{N}$ is countable but is not computable.

PROOF: Algorithms are finite sequences from a finite set of symbols, so there are at most a countable number of them and can be listed. Suppose $A_n(m)$ is a computable list of computable subsets, i.e., $(n, m) \in A$ is computable. Let $B := \{\, 1 - A_n(n) : n \in \mathbb{N} \,\}$; it is a computable set (because the list is computable) that is not found in the list $(\forall n, B \neq A_n)$.

□

There is no algorithm that decides whether any given set is computable or not; or that decides whether two computable sets are the same or not.

Suppose A_n is a list of algorithms with inputs in $\mathbb{N}$; among these is the "universal Turing machine" algorithm $M : n \mapsto A_n$ (yes, programs can be created mechanically!). The set of terminating algorithms in this list is not computable, otherwise its complement, the set of non-terminating algorithms would also be computable, a contradiction (by Cantor's diagonal argument). Thus there is no terminating algorithm which decides whether any other algorithm halts or not (Turing's theorem).

A function $f\colon \mathbb{N} \to \mathbb{N}$ (or subsets) is *computable* when there is an algorithm that calculates $f(n)$ for each $n \in \mathbb{N}$. The set of computable functions is countable; interestingly, the largest number that can be produced by a program n-words long grows faster than all computable functions. Can refine the concept of $A \leqslant B$ for countably infinite sets by letting $A \leqslant B$ mean there is a computable function $f\colon A \to B$, or even finer with f a computable 1-1 function (or a computable onto function $f\colon B \to A$). It turns out that the Halting problem is the largest computable countable set in this sense.

Proposition 9

Recursive functions

Functions on $\mathbb{N}$ can be defined uniquely by specifying $f(0)$, and $f(n^+)$ in terms of $m \leqslant n$.

PROOF: Let $f(0) := a$, $f(n^+) := F(f(n), n)$ where $f\colon A \times \mathbb{N} \to A$ is a function. Consider the functions $f_k : \{0, \ldots, k\} \to A$ defined by $f_0(0) := a, f_k(n^+) := F(f_k(n), n)$ for $n < k$. Then the domain of f_k lies in that of f_{k+1} and $f_{k+1}(n) = f_k(n)$ for $n \leqslant k$ (by induction on k). So can define the extension $f := \bigcup_{k=0}^{\infty} f_k$, whose domain is $\mathbb{N}$, and $f(0) = f_0(0) = a, f(n^+) = f_{n+1}(n^+) = F(f_{n+1}(n), n) = F(f(n), n)$.

If h were another such function, then $h(0) = a = f(0), h(n^+) = f(n^+)$ so that $f = h$ by induction.

The proof also holds for recursion of the type $f(n^+) = F(f(0), \ldots, f(n), n)$. □

3.2 Set arithmetic

The following statement cannot be proved or disproved using the previous axioms:

Axiom of Continuum: **The smallest number strictly greater than $\mathbb{N}$ is $2^{\mathbb{N}}$, and more generally, for $\boldsymbol{n} \notin \mathbb{N}$, the smallest number strictly greater than $\boldsymbol{n}$ is $2^{\boldsymbol{n}}$,**

$$\boldsymbol{m} > \boldsymbol{n} \notin \mathbb{N} \text{ AND } (\boldsymbol{k} > \boldsymbol{n} \;\Rightarrow\; \boldsymbol{k} \geqslant \boldsymbol{m}) \;\Rightarrow\; \boldsymbol{m} = 2^{\boldsymbol{n}}.$$

(This axiom can be strengthened further to the *Axiom of Constructibility*: every set can be constructed from $\mathbb{N}$ and its subsets and unions, i.e., every set belongs to an L_α where $L_0 = \mathbb{N}$, $L_{\alpha^+} = 2^{L_\alpha}$, $L_{\lim \alpha} = \bigcup_\alpha L_\alpha$.

Also *axiom of limitation of size*: every class is a set unless it is cardinally equivalent to Υ (implies the axioms of replacement, and choice because even Υ can be well-ordered.))

The union, product, and power of sets become the following operations of numbers:

Proposition 10

Addition

$$\boldsymbol{m} + \boldsymbol{n} = \begin{cases} \boldsymbol{m} & n = 0 \\ (\boldsymbol{m} + k)^+ & \boldsymbol{n} = k^+ \\ \max(\boldsymbol{m}, \boldsymbol{n}) & \boldsymbol{m} \notin \mathbb{N} \text{ OR } \boldsymbol{n} \notin \mathbb{N} \end{cases}$$

For example, $1{+}1 = 1{+}0^+ = (1{+}0)^+ = 1^+ = 2$, and $\omega{+}\omega = \max(\omega, \omega) = \omega$.

Proposition 11

Multiplication

$$\boldsymbol{mn} = \begin{cases} 0 & \boldsymbol{n} = 0 \\ 1 & \boldsymbol{m} = 1 \text{ OR } \boldsymbol{n} = 0 \\ \boldsymbol{m}k + \boldsymbol{m} & \boldsymbol{n} = k^+ \\ \max(\boldsymbol{m}, \boldsymbol{n}) & \boldsymbol{m} \notin \mathbb{N} \text{ OR } \boldsymbol{n} \notin \mathbb{N} \end{cases}$$

For example, $2\omega = \omega = \omega^2$.

1. The countable union and finite product of countable sets is countable.

2. For $\boldsymbol{m}$ infinite, $\boldsymbol{m}^2 = \boldsymbol{m}$.

   Proof: Well-order $\boldsymbol{m}$, then can well-order $\boldsymbol{m}^2$ using $(x, y) < (a, b)$ by testing $\max(x, y) < \max(a, b)$ then $x < a$ then $y < b$ in order. Now $\boldsymbol{m}^2 = \boldsymbol{m}$ is true for countably infinite sets; suppose that for $\boldsymbol{n} < \boldsymbol{m}$, we have $\boldsymbol{n}^2 = \boldsymbol{n}$, and suppose that $\boldsymbol{m}^2 > \boldsymbol{m}$; then $\exists (a, b) \in \boldsymbol{m}^2$ such that $X := \{\, (x, y) \in \boldsymbol{m}^2 : (x, y) < (a, b) \,\}$ has cardinality $\geqslant \boldsymbol{m}$; let $c := \max(a, b)+1$, then $\forall (x, y) \in X, \max(x, y) \leqslant \max(a, b) < c$, so $X \subseteq c^2$; but $|c| \leqslant \boldsymbol{n}$ for some $\boldsymbol{n} < \boldsymbol{m}$; so $|X| \leqslant |c|^2 \leqslant \boldsymbol{n}^2 \leqslant \boldsymbol{n}$ by induction principle, so that $|X| \leqslant \boldsymbol{n} < \boldsymbol{m}$ a contradiction; hence $\boldsymbol{m}^2 \leqslant \boldsymbol{m}$; so apply transfinite induction.

   Now, $\boldsymbol{m} \leqslant \boldsymbol{m} + \boldsymbol{n} \leqslant 2\boldsymbol{m} \leqslant \boldsymbol{m}^2 = \boldsymbol{m}$, $\boldsymbol{m} \leqslant \boldsymbol{mn} \leqslant \boldsymbol{m}^2 = \boldsymbol{m}$.

3. More generally,
$$\sum_{i<\boldsymbol{n}} \boldsymbol{m}_i = \boldsymbol{n} \sup\{\, \boldsymbol{m}_i : i < \boldsymbol{n} \,\}.$$
   For example $\sum_{i<\boldsymbol{n}} \boldsymbol{m} = \boldsymbol{nm}$, $\sum_{n\in\mathbb{N}} n = \omega^2 = \omega$, $\sum_{i<\boldsymbol{n}} 1 = \boldsymbol{n}$.

Proposition 12

Power

$$\boldsymbol{m}^{\boldsymbol{n}} = \begin{cases} 0 & \boldsymbol{m} = 0, \boldsymbol{n} \neq 0 \\ 1 & \boldsymbol{m} = 1 \text{ OR } \boldsymbol{n} = 0 \\ \boldsymbol{m}^k\boldsymbol{m} & \boldsymbol{n} = k^+ \\ \boldsymbol{m} & \boldsymbol{n} < \theta, \boldsymbol{m} = \lim_{\nu\to\theta} \alpha_\nu \\ 2^{\boldsymbol{m}} & \theta \leqslant \boldsymbol{n} \leqslant \boldsymbol{m} \\ 2^{\boldsymbol{n}} & 0, 1 \neq \boldsymbol{m} \leqslant \boldsymbol{n} \notin \mathbb{N} \end{cases}$$

Proof: $2^{\boldsymbol{n}} \leqslant \boldsymbol{m}^{\boldsymbol{n}} \leqslant (2^{\boldsymbol{m}})^{\boldsymbol{n}} = 2^{\boldsymbol{mn}} = 2^{\boldsymbol{n}}$.
For example, $\omega^\omega = 2^\omega$. Moreover, even for infinite numbers,

Proposition 13

Rules of Arithmetic

$0 \leqslant \boldsymbol{m}$
$\boldsymbol{m} \leqslant \boldsymbol{n}$ OR $\boldsymbol{n} \leqslant \boldsymbol{m}$
$\boldsymbol{m} \leqslant \boldsymbol{n}$ AND $\boldsymbol{n} \leqslant \boldsymbol{m} \;\Rightarrow\; \boldsymbol{m} = \boldsymbol{n}$

$\boldsymbol{m} + \boldsymbol{n} = \boldsymbol{n} + \boldsymbol{m}$
$\boldsymbol{m} + (\boldsymbol{n} + \boldsymbol{k}) = (\boldsymbol{m} + \boldsymbol{n}) + \boldsymbol{k}$ **(also true for infinite sums)**
$\boldsymbol{m} \leqslant \boldsymbol{n} \;\Rightarrow\; \boldsymbol{m} + \boldsymbol{k} \leqslant \boldsymbol{n} + \boldsymbol{k}$ **(also true for infinite sums i.e.,** $\boldsymbol{m}_i \leqslant \boldsymbol{n}_i \;\Rightarrow\; \sum_i \boldsymbol{m}_i \leqslant \sum_i \boldsymbol{n}_i$**)**

$\boldsymbol{mn} = \boldsymbol{nm}$
$\boldsymbol{m}(\boldsymbol{nk}) = (\boldsymbol{mn})\boldsymbol{k}$ **(also true for infinite products)**
$1\boldsymbol{m} = \boldsymbol{m}$**;** $\boldsymbol{m}(\boldsymbol{n} + \boldsymbol{k}) = \boldsymbol{mn} + \boldsymbol{mk}$ **(also** $\boldsymbol{m}\sum_i \boldsymbol{n}_i = \sum_i \boldsymbol{mn}_i$**)**
$\boldsymbol{m} \leqslant \boldsymbol{n} \;\Rightarrow\; \boldsymbol{mk} \leqslant \boldsymbol{nk}$**,**
$\boldsymbol{m}_i \leqslant \boldsymbol{n}_i \;\Rightarrow\; \prod_i \boldsymbol{m}_i \leqslant \prod_i \boldsymbol{n}_i$

$\boldsymbol{m}^{\boldsymbol{n}+\boldsymbol{k}} = \boldsymbol{m}^{\boldsymbol{n}}\boldsymbol{m}^{\boldsymbol{k}}$**, also** $\boldsymbol{m}^{\sum_i \boldsymbol{n}_i} = \prod_i \boldsymbol{m}^{\boldsymbol{n}_i}$ **(hence** $\prod_{i<\boldsymbol{n}} \boldsymbol{m} = \boldsymbol{m}^{\boldsymbol{n}}$**)**
$\boldsymbol{m}^{\boldsymbol{nk}} = (\boldsymbol{m}^{\boldsymbol{n}})^{\boldsymbol{k}}$
$\boldsymbol{m} \leqslant \boldsymbol{n} \;\Rightarrow\; \boldsymbol{m}^{\boldsymbol{k}} \leqslant \boldsymbol{n}^{\boldsymbol{k}}$**, and** $\boldsymbol{k}^{\boldsymbol{m}} \leqslant \boldsymbol{k}^{\boldsymbol{n}}$ **(for** $\boldsymbol{k} \neq 0$**)**
$(\prod_i \boldsymbol{m}_i)^{\boldsymbol{n}} = \prod_i \boldsymbol{m}_i^{\boldsymbol{n}}$**.**

1. If $\boldsymbol{m} < \boldsymbol{n}$, both infinite, then $\boldsymbol{n} - \boldsymbol{m} = \boldsymbol{n}$ (must use axiom of choice).
2. $\prod_{n\in\mathbb{N}} n = 2^{\omega}$
   (since $2^{\omega} = \prod_{n\in\mathbb{N}} 2 \leqslant \prod_n n \leqslant \prod_n \omega = \omega^{\omega} = 2^{\omega}$).
3. If $\boldsymbol{m}_i < \boldsymbol{n}_i$ then $\sum_i \boldsymbol{m}_i < \prod_i \boldsymbol{n}_i$ (generalization of Cantor's theorem).

The first few cardinal numbers are

$$0, 1, 2, 3, \ldots, \omega, \boldsymbol{c} := \omega^{\omega}, \boldsymbol{f} := \boldsymbol{c}^{\boldsymbol{c}}, \ldots, \aleph_{\omega}, \aleph_{\omega}^{\aleph_{\omega}}, \ldots, \aleph_{2\omega}, \ldots, \aleph_{\aleph_{\aleph}} \quad, \ldots$$
$$\ddots$$

The building blocks of sets are 0 and 1; every other set can be obtained from their union (but all non-empty sets require a sum that involves the same cardinality, $\sum_{i<n} 1 = n$). In terms of products, a set that can be reduced as $\prod_{i\in I} A_i$ with I and A_i of smaller cardinality, is called *composite*, the other irreducible ones are the *primes* $0, 1, 2, 3, 5, \ldots, \omega, \ldots$

Skolem's "Paradox": There are uncountable sets; but Henkin's theorem in logic shows that there must be a model of set theory which is countable i.e., the uncountable sets (e.g. 2^{ω}) in this model are countable! The resolution is that, although the uncountable sets are indeed countable, there are no bijective "countable" functions that can count these sets.

3.3 Generalizations of Sets

A **multiset** allows repetition of identical elements, e.g. $\{\!\{a,a,a,b\}\!\}$ (4 elements), such as a wallet of money. All of set theory can be extended to multisets:

$$\begin{aligned}
\{\!\{a,a,b\}\!\} \subseteq \{\!\{a,a,a,b,b\}\!\}, \quad & \{\!\{a,a,b\}\!\} \not\subseteq \{\!\{a,b,b\}\!\} \\
\{\!\{a,a,a,b\}\!\} \cup \{\!\{a,b,b\}\!\} &= \{\!\{a,a,a,a,b,b,b\}\!\} \\
\{\!\{a,a,a,b\}\!\} \cap \{\!\{a,b,b\}\!\} &= \{\!\{a,b\}\!\} \\
\{\!\{a,a,a,b\}\!\} \setminus \{\!\{a,b,b\}\!\} &= \{\!\{a,a\}\!\} \\
\{\!\{a,a,b\}\!\} \times \{\!\{b,b\}\!\} &= \{\!\{(a,b),(a,b),(a,b),(a,b),(b,b),(b,b)\}\!\} \\
|\{\!\{a,a,a,b\}\!\}| &= 4
\end{aligned}$$

Multisets can be implemented in set theory by attaching distinct labels to identical elements, e.g. $\{\!\{a_1,a_2,a_3,b_1,b_2\}\!\}$; or what amounts to the same thing, as a function $f : X \to \mathbb{N}$, e.g. $f : a \mapsto 3, b \mapsto 2, c \mapsto 0, \ldots$; so $A \cup B$, $A \cap B$, $A \times B$ correspond respectively to the functions $x \mapsto f_A(x) + f_B(x)$, $x \mapsto f_A \wedge f_B(x) = \min(f_A(x), f_B(x))$, $(x,y) \mapsto f_A(x) f_B(y)$.

Furthermore, an **anti-set** allows *anti-objects*, where objects a and anti-objects a^* annihilate when grouped together, e.g.

$$\{\!\{a,a^*,b,b^*,b^*\}\!\} = \{\!\{b^*\}\!\},\ \{\!\{a,b\}\!\}^* = \{\!\{a^*,b^*\}\!\},\ \varnothing^* = \varnothing.$$

They can be implemented rigorously as functions $X \to \mathbb{Z}$.

4 Combinatorics

Finding the cardinality of sets ("counting") is one of the most important aspects of mathematics.

Proposition 14

Sylvester's Principle:

Let $N(A_1,\ldots,A_n)$ denote the number of ways that N objects satisfy the conditions A_1 to A_n. Then

$$\begin{aligned}
&N(\text{NOT } A_1,\ldots,\text{NOT } A_n) \\
&= N - \textstyle\sum_i N(A_i) + \sum_{i,j} N(A_i,A_j) + \cdots + (-1)^n N(A_1,\ldots,A_n)
\end{aligned}$$

(Note: if the order of N objects is important, then consider the $N!$ ordered lists of objects.)

Let X be a set of objects x, each having a list of parameters $|x| = (|x|_1,\ldots,|x|_r) \in \mathbb{N}^r$; and suppose objects can be *combined* together $(x,y) \mapsto xy$ so that each parameter adds up: $|xy| = |x| + |y|$. The main parameter for finite objects is

usually its size (cardinality). Subsets of X can also be combined as $AB := \{\, ab : a \in A, b \in B \,\}$. So A^n represents the set of ordered combinations of n objects from A (possibly repeated). The set 2^X with this operation and disjoint union $(+)$ forms a semi-ring (with $0 = \varnothing$ and $1 = \{\, \varnothing \,\}$) that can be embedded in an algebra over $\mathbb{C}$.

The ordered combination of an arbitrary number of objects from A (with replacement) can be written as:

$$\mathrm{Seq}(A) := \bigcup_{n\in\mathbb{N}} A^n = 1 + A + A^2 + \cdots = (1 - A)^{-1}$$

Proof: $1 + A\,\mathrm{Seq}(A) = \mathrm{Seq}(A)$. More generally, the combination of specific numbers of objects is $\mathrm{Seq}_I(A) := \bigcup_{n\in I} A^n$.

The unordered combination of objects from A with replacement (multisets) is:

$$\mathrm{MSets}(A) := \prod_{a\in A} \mathrm{Seq}\{\, a \,\}$$

The unordered combination of objects from A without replacement is:

$$\mathrm{Sets}(A) := \prod_{a\in A} 2_a = \prod_{a\in A} \{\, \varnothing, a \,\}$$

Each subset $A \subseteq X$ gives rise to a "generating" array of dimension r, where $a_{\boldsymbol{n}}$ counts the number of objects with parameter set $\boldsymbol{n}$,

$$\boldsymbol{a} = \sum_{x\in A} \boldsymbol{e}_{|x|} = \sum_{\boldsymbol{n}\geqslant \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{e}_{\boldsymbol{n}}$$

For example,

Null object set	$1 = \{\, \varnothing \,\}$	$\boldsymbol{e}_0 = (1, 0, 0, \ldots)$
Node	$Z := \{\, \bullet \,\}$	$\boldsymbol{e}_1 = (0, 1, 0, 0, \ldots)$
	$2_a := \{\, \varnothing, a \,\}$	$(1, 1, 0, \ldots)$
	$\mathbb{N} = \mathrm{Seq}(Z)$	$\boldsymbol{1} = (1, 1, 1, \ldots)$

If a_n is the number of ways of having objects in A with parameter n, and b_m the number of ways of having objects from B with parameter m, then $(a_n) * (b_m) := (a_0 b_0, a_1 b_0 + a_0 b_1, \ldots)$ gives the number of ways of combining objects from AB, having parameter $n + m$. For example, there is only one way of getting n as a sum of 1s, namely $n = 1 + \cdots + 1$; the number of ways of getting m as a sum of 3s can be listed as $(1, 0, 0, 1, 0, 0, 1, \ldots)$ (i.e., only one way if a multiple of 3). Then the number of ways of getting n using either 1 or 3 is $(1, 1, \ldots) * (1, 0, 0, 1, \ldots) = (1, 1, 1, 2, 2, 2, \ldots)$.

The mapping $A \mapsto \boldsymbol{a}$ is an algebra-morphism $2^X \to \mathbb{N}^{\mathbb{N}^r}$,

$$A + B \mapsto \boldsymbol{a} + \boldsymbol{b}, \qquad AB \mapsto \boldsymbol{a} * \boldsymbol{b}, \qquad A \circ B \mapsto \boldsymbol{a} \circ \boldsymbol{b}$$

since

$$\sum_{x\in A+B}\boldsymbol{e}_{|x|}=\sum_{x\in A}\boldsymbol{e}_{|x|}+\sum_{x\in B}\boldsymbol{e}_{|x|}=\boldsymbol{a}+\boldsymbol{b}$$

$$\sum_{xy\in AB}\boldsymbol{e}_{|xy|}=\sum_{x\in A,y\in B}\boldsymbol{e}_{|x|+|y|}=\sum_{n\geqslant 0}\sum_{0\leqslant m\leqslant n}a_m b_{n-m}\boldsymbol{e}_n=\boldsymbol{a}*\boldsymbol{b}$$

$A\circ B$ is defined as the set in which each object of B is replaced by any object in A, and $\boldsymbol{a}\circ\boldsymbol{b}:=\sum_{\boldsymbol{n}\geqslant\boldsymbol{0}}a_{\boldsymbol{n}}\boldsymbol{b}^{\boldsymbol{n}}=a_0\boldsymbol{e}_0+a_1\boldsymbol{b}_1+\cdots$.

Note that there is another morphism $\mathbb{N}^r\to C^\omega(\mathbb{C}^r)$, $\boldsymbol{a}\mapsto a$, where $a(\boldsymbol{z}):=\sum_{\boldsymbol{n}}a_{\boldsymbol{n}}\boldsymbol{z}^{\boldsymbol{n}}$ (generating function). For any subset A,

Number of objects	$N=\sum_{x\in A}1=\sum_{\boldsymbol{n}}a_{\boldsymbol{n}}=a(\boldsymbol{1})$
Sum of parameter	$\sum_{x\in A}\lvert x\rvert_1=\sum_n na_{\boldsymbol{n}}=\sum_{\boldsymbol{n}}na_{n,\boldsymbol{k}}=\partial_1 a(\boldsymbol{1})$
Mean of parameter	$\mu=a'(\boldsymbol{1})/a(\boldsymbol{1})$
Variance of parameter	$\sigma^2=\sum_{x\in A}\frac{(\lvert x\rvert_1-\mu)^2}{N}=\sum_n\frac{(n-\mu)^2a_n}{N}=\frac{a''(1)}{a(1)}+\mu-\mu^2$

- The number of *permutations* of a set is $n!$ For finite numbers it equals $n(n-1)\cdots 2.1$.
- Number of ways of choosing m from n objects $=:\binom{n}{m}$; for finite numbers $\binom{n}{m}=\frac{n!}{m!(n-m)!}$. It satisfies $\binom{n}{m}=\binom{n-1}{m}+\binom{n-1}{m-1}$ (=ways of including or not a specific object). For example, number of ways of choosing m from n objects with repetition (multisets) is $\binom{n+m-1}{m-1}$ (= ways with $n,m-1$ plus those with $n-1,m$); it also equals the number of natural number solutions of $x_1+\cdots+x_m=n$.
- Number of length-n binary strings $=2^n$: $X=1+\{\circ,\bullet\}X$, so $x=1+2zx$, $x=(1-2z)^{-1}$. Similarly, binary strings without 00 satisfy $X=\{\varnothing,\circ\}(1+\{\bullet\}X)$, so $x=(1+z)(1+zx)$.
- Number of subsets of $\{1,\ldots,n\}$ not containing any consecutive numbers = Fibonacci numbers $=\sum_k\binom{n+1-k}{k}$ (since $N_{n+1}=N_n+N_{n-1}$ depending on whether the subset does not/does contain $n+1$).
- Number of ways of placing n brackets correctly $(()(()))$ = Catalan number $C_n:=\frac{1}{n+1}\binom{2n}{n}=\binom{2n}{n}-\binom{2n}{n-1}$; set satisfies $X=1+ZX^2$, so $C_n=C_0C_{n-1}+C_1C_{n-2}+\cdots+C_{n-1}C_0$.
- Number of ways or placing 1s in a string of n 0s is 2^{n-1} (decompositions of n): $X=\mathrm{Seq}(\mathbb{N}^+)=(1-Z\mathbb{N})^{-1}$. Similarly, decompositions of n into k parts is $X=(\mathbb{N}^+)^k$.

- Number of ways of partitioning $n = p+q$ where $p \in P \subseteq \mathbb{N}$ and $q \in Q \subseteq \mathbb{N}$ - $\operatorname{Seq}_P(Z)\operatorname{Seq}_Q(Z)$, so form the convolution $1_P * 1_Q$; more generally for $n = \sum_i p_i$ with $p_i \in P_i$, form $\prod_i \operatorname{Seq}_{P_i}(Z)$ and the convolution product $\prod_i 1_{P_i}$.

  Number of ways of partitioning n into 1s, 2s, 5s, 10s, 20s, and 50s: $X = \operatorname{MSets}\{1,2,5,10,20,50\} = (1-Z)^{-1}(1-Z^2)^{-1}(1-Z^5)^{-1}(1-Z^{10})^{-1}(1-Z^{20})^{-1}(1-Z^{50})^{-1}$.

  To count the number of ways that $n = pq$ where $p \in P \subseteq \mathbb{N}$ and $q \in Q \subseteq \mathbb{N}$ - the Dirichlet convolution $a*b(n) := \sum_{ij=n} a_i b_j$ gives the required number of ways.

- Number of ways of partitioning n into disjoint parts of size $n_1, \ldots, n_k$ (so $n_1 + \cdots + n_k = n$), is $\binom{n}{n_1, \cdots, n_k} := \dfrac{n!}{n_1! \cdots n_k!}$;

- Number of ways of partitioning n is $P(n)$ for $X = \operatorname{MSets}(\mathbb{N}^+) = (1-Z)^{-1}(1-Z^2)^{-1}\cdots$; it satisfies $P(n) = \sum_k P_k(n)$ where $P_k(n)$ is the number of ways of partitioning n in k parts,

  $$P_k(n) = P_k(n-k) + \cdots + P_1(n-k), P_k(n) = 0 \text{ for } k > n, P_k(k) = 1.$$

  For $n = 1, 2, \ldots$, $P(n) = 1, 2, 3, 5, 11, 14, 22, \ldots$

  The number of ways of partitioning n, taking the order into consideration, is $\sum_k \binom{n-1}{k-1}$; the number of ways of partitioning into unequal parts $\operatorname{Sets}(\mathbb{N}^+) = (1+Z)(1+Z^2)\cdots$ (=number of ways of partitioning into odd parts).

- Number of ordered rooted trees: satisfy $X = Z\operatorname{Seq}(X)$, equivalently $X = Z + X^2$, so $T_{n+1} = T_n T_1 + \cdots + T_1 T_n$, $T_1 = 1$ ($X = \frac{1}{2}(1+\sqrt{1-4Z}) = 1 - Z - Z^2 - 2Z^3 - 5Z^4 + \cdots$). Similarly, set of binary trees $X = Z(1+X^2)$.

- Number of unordered rooted trees: $X = Z\operatorname{MSets}(X)$.

- Number of ways of placing n brackets on a string of length $n+1$: $X = Z + X^2(1-X)^{-1}$. Similarly, unordered trees with n leaves, $X = Z + \operatorname{Sets}_{\geqslant 2}(X)$.

Labeled Objects can be combined (taking labels to be disjoint); so can subsets $AB = \{xy : x \in A, y \in B\}$. The representing "exponential generating" sequence is now $\boldsymbol{a} = \sum_{x \in A} \frac{1}{|x|!}\boldsymbol{e}_{|x|} = (A_n/n!)$, where A_n is the number of labeled objects of size n. There is again a morphism, $A \mapsto \boldsymbol{a}$ with $AB \mapsto \boldsymbol{a} * \boldsymbol{b}$ (since $\sum_{x\in A} \frac{z^{|x|}}{|x|!} \sum_{y \in B} \frac{z^{|y|}}{|y|!} = \sum_{n,m} \binom{n+m}{n} \frac{z^{n+m}}{(n+m)!} = \sum_{xy \in AB} \binom{|xy|}{|x|} \frac{z^{|xy|}}{|xy|!}$).

Permutations of labeled objects from A:	$\operatorname{Seq}(A) = 1 + A + A^2 + \cdots = (1-A)^{-1}$
Sets of labeled objects from A:	$\operatorname{Sets}(A) = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = e^A$
Cycles (directed) of labeled objects from A:	$\operatorname{Cycles}(A) = A + \frac{A^2}{2} + \frac{A^3}{3} + \cdots = \ln(1-A)^{-1}$.
Undirected cycles:	$\frac{1}{2}(A + \frac{1}{2}A^2 + \operatorname{Cycles}(A))$.

1. The permutations of n objects is $\text{Seq}(Z)$; number of directed n-cycles is $(n-1)!$ since $\text{Cycles}(Z) = Z + Z^2/2 + \cdots$.

   Derangements of n objects, i.e., permutations that move every object, are given by $\text{Sets}(\text{Cycles}_{>1}(Z)) = e^{\ln(1-Z)^{-1}-Z} = e^{-Z}(1-Z)^{-1}$. Similarly, involutions (only transpositions) are $\text{Sets}(\text{Cycles}_{\{1,2\}}(Z)) = e^{Z+Z^2/2}$; permutations with only 3-cycles or 5-cycles are $e^{Z^3/3+Z^5/5}$.

2. Number of strings from an alphabet of m letters is m^n: $\text{Sets}(Z)^m = e^{mZ}$ where each set gives the positions of the corresponding letter, e.g. $BCACAA$ corresponds to $\{3,5,6\}\{1\}\{2,4\}$. Similarly, strings where no letter appears twice $= \{\varnothing, \bullet\}^m$; strings where each letter appears at least once $= \text{Sets}_{>0}(Z)^m = (e^Z-1)^m$; non-isomorphic strings from an infinite alphabet is $\text{Seq}(\text{Sets}(Z)) = (1-e^Z)^{-1}$.

   Adding a parameter tracking the number of some specific letter, gives a generating function $p(z,w) = (1-(m-1+w)z)^{-1}$.

3. Number of surjections $= \text{Seq}(\text{Sets}_{>0}(Z)) = (2-e^Z)^{-1}$.

4. Ordered rooted labeled trees, $X = Z\,\text{Seq}(X)$; unordered rooted labeled trees, $X = Z\,\text{Sets}(X)$.

Theorem 15

Rota's theorem

If $g\colon 2^X \to \mathbb{N}$ **and** $f(A) := \sum_{B\subseteq A} g(B)$**, then** $g(A) = \sum_{B\subseteq A}(-1)^{|A|-|B|} f(B)$**.**

Theorem 16

Ramsey's theorem

Let the subsets of $\{1,\ldots,n\}$ **of size** r **be divided disjointly into** $\mathcal{A}$ **and** $\mathcal{B}$**. Then** $\forall p,q \geqslant r, \exists N_r(p,q), n > N_r(p,q) \Rightarrow \exists A : |A| = p$ **whose subsets of size** r **are all in** $\mathcal{A}$ **or** $\exists B : |B| = q$ **whose subsets of size** r **are all in** $\mathcal{B}$**.**

(in most cases of r,p,q, $N_r(p,q)$ is not known).

Given subsets $A_i \subseteq X$, then one can find distinct $x_i \in A_i \Leftrightarrow$ any k subsets contain more than k elements.