Contents

3 Relations, Orderings, and Functions

Our goal in this chapter is to discuss the basic properties of relations, orderings, and functions along with some of their applications. We begin by examining the very general idea of a relation, which captures the idea of a comparison between two objects. Second, we discuss equivalence relations, which generalize the concepts of equality and modular congruence. Third, we examine partial and total orderings, which generalize the "order relations" of subset (for sets), divisibility (for integers), and the natural ordering of the real numbers. Fourth, we construct a formal definition for a function as a special type of relation, and then discuss various properties of functions including injectivity, surjectivity, function composition, and inverse functions. Finally, we close with a discussion of cardinality, focusing in particular on how the notion of countability allows us (rather unexpectedly!) to show that there are many different sizes of infinite sets.

3.1 Relations

- The idea of a relation is quite simple, and generalizes the idea of a comparison between two objects. Here are some familiar examples of relations that we have already discussed at length:
  - The subset relation $\subseteq$ on a pair of sets.
  - The order relations $\leq$ and $<$ and $\geq$ and $>$ on a pair of integers (or rational numbers, or real numbers).

- The containment relation $\in$ on an element and a set.
- The divisibility relation $|$ on a pair of integers.
- The mod-m congruence relation $\equiv$ on a pair of integers.

- In each of these examples, the relation R captures some information about two objects, and the relation statement $a\,R\,b$ is a proposition that is either true or false.
  - For example, $5 < 3$ is a statement about the two numbers 5 and 3 (it is a false statement, of course).
  - The order of the objects in the relation statement is quite clearly important: for example, $3|6$ is true while $6|3$ is false.
  - Also, the objects in a relation statement need not be drawn from the same universe: in the containment relation $x \in A$, for example, the object x can be anything, while the object A is a set.

- In order to describe a general relation R, then, we could simply list all of the ordered pairs (a, b) for which the relation statement $a\,R\,b$ is true. In fact, we will take this as the definition of a relation!

- Definition: If A and B are sets, we say R is a relation from A to B, written $R : A \to B$, if R is a subset of the Cartesian product $A \times B$. For any $a \in A$ and $b \in B$, we write $a\,R\,b$ if the ordered pair (a, b) is an element of R, and we write $a \not R\, b$ if the ordered pair (a, b) is not an element of R.
  - We think of the statement $a\,R\,b$ as saying the ordered pair (a, b) satisfies the relation R, and we think of $a \not R\, b$ as saying the ordered pair (a, b) does not satisfy the relation R.

- We can recast all of the familiar relations we have encountered already in this language of Cartesian products.

- Example: The relation $R = \leq$ on integers can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : b - a \in \mathbb{Z}_{\geq 0}\}$, which is the set of ordered pairs (a, b) where $b - a$ is a nonnegative integer.
  - Under this definition, we see that $3\,R\,5$ and $4\,R\,13$ because $5-3 = 2$ and $13-4 = 9$ are both nonnegative integers.
  - On the other hand, $2 \not R\, 0$ because $0 - 2 = -2$ is not a nonnegative integer.

- Example: The divisibility relation $R = |$ on integers can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \exists k \in \mathbb{Z} \text{ such that } b = ka\} = \{(a, ka) : a, k \in \mathbb{Z}\}$.
  - Under this definition, we see that $3\,R\,6$ and $4\,R\,20$ because the ordered pairs $(3, 6) = (3, 2 \cdot 3)$ and $(4, 20) = (4, 5 \cdot 4)$ are in the set described above.
  - On the other hand, $2 \not R\, 3$ because $(2, 3)$ is not in the set above.

- Example: The congruence relation $R = \equiv_m$ modulo m can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \exists k \in \mathbb{Z} \text{ such that } b - a = km\} = \{(a, a + km) : a, k \in \mathbb{Z}\}$.
  - Under this definition, if $m = 5$ we see that $3\,R\,18$ and $4\,R\,{-6}$ because the ordered pairs $(3, 18) = (3, 3 + 3 \cdot 5)$ and $(4, -6) = (4, 4 + (-2) \cdot 5)$ are in the set described above.
  - On the other hand, $1 \not R\, 3$ because $(1, 3)$ is not in the set above.

- Example: If A is any set, the identity relation is defined by taking $R = \{(a, a) : a \in A\}$. This is simply the equality relation, in which $a\,R\,b$ precisely when a and b are equal.
  - Under this definition, if $A = \mathbb{R}$ for example, we see that $3\,R\,3$ since $(3, 3)$ is an element of the set R, but $1 \not R\, 3$ and $3 \not R\, \pi$ since $(1, 3)$ and $(3, \pi)$ are not elements of R.

- There are many other things we can also describe using the language of relations.
  - Example: The relation $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \gcd(a, b) = 1\}$ is the "is relatively prime" relation on integers: we have $a\,R\,b$ precisely when a and b are relatively prime.
  - Example: The relation $R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 = y\} = \{(y^2, y) : y \in \mathbb{R}\}$ is the "is a square root of" relation on real numbers: we have $x\,R\,y$ precisely when x is a square root of y (i.e., when $x^2 = y$).

- Example: The relation $R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 = 1\}$ is the "lies on the unit circle" relation on real numbers: we have $x \; R \; y$ precisely when the point (x, y) satisfies the equation $x^2 + y^2 = 1$ (which is to say, when the point lies on the unit circle).
- Example: The relation $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : |b - a| = 1\}$ is the "differs by 1" relation on integers: we have $a \; R \; b$ precisely when a and b differ by 1.

- We can also simply write down arbitrary subsets of ordered pairs to obtain new relations:
- Example: If $A = \{1, 2, 3, 4\}$ and $B = \{1, 3, 5, 7\}$, then some relations are as follows:
  - The relation $R_1 = \{(1, 1), (2, 3), (3, 5), (4, 7)\}$ is a relation from A to B.
  - The relation $R_2 = \{(1, 1), (3, 2), (5, 3), (7, 4)\}$ is a relation from B to A.
  - The relation $R_3 = \{(1, 4), (3, 2), (2, 1)\}$ is a relation from A to A. (We say R_3 is a relation on A.)
  - The relation $R_4 = \{(1, 3), (3, 1), (4, 3)\}$ is a relation from A to A. It is also a relation from A to B.
  - The relation $R_5 = \{(7, 1), (7, 3)\}$ is a relation from B to A. It is also a relation from B to B.
  - The relation $R_6 = \{(1, 1), (3, 3)\}$ is a relation from A to A. It is also a relation from A to B, and from B to A, and from B to B.
  - The relation $R_7 = \{(1, 1), (2, 7), (3, 5), (5, 4)\}$ is a relation but it is not a relation on A or on B, or from A to B, or from B to A.
  - The empty relation $R_8 = \emptyset$ is a relation from A to A, and also from A to B, and from B to A, and from B to B.

- Since relations are merely subsets of a Cartesian product, we can apply any of our set operations to them.
  - For example, if C is a subset of A and D is a subset of B, then if $R_{A,B} : A \to B$ is a relation, we may construct a new relation $R_{C,D} : C \to D$ given by $R_{C,D} = R_{A,B} \cap (C \times D)$; this relation is called the restriction of R to $C \times D$.
  - In the case where R is a relation on A and C is a subset of A, we call $R \cap (C \times C)$ the restriction of R to C, and denote it as $R|_C$.

- Another useful construction is the inverse of a relation, obtained by reversing all of the ordered pairs:
- Definition: If $R : A \to B$ is a relation, then the inverse relation (also sometimes called the converse relation or the transpose relation) $R^{-1} : B \to A$ is defined as $R^{-1} = \{(b, a) : (a, b) \in R\}$, the relation on $B \times A$ consisting of the reverses of all of the ordered pairs in R.
  - Example: If $A = \{1, 2, 3, 4\}$ and $B = \{1, 3, 5, 7\}$, then the inverse of the relation $R_1 = \{(1, 1), (2, 3), (3, 5), (5, 7)\}$ from A to B is the relation $R_1^{-1} = \{(1, 1), (3, 2), (5, 3), (7, 5)\}$ from B to A.
  - Example: If $A = \mathbb{R}$, then the inverse of the relation $R_2 = \leq$ is $R_2^{-1} = \geq$. This follows from the observation that $(a, b) \in R_2$ precisely when $b - a$ is nonnegative, and therefore $(b, a) \in R_2^{-1}$ precisely when $b - a$ is nonnegative (which is to say, when the first element of the ordered pair is greater than or equal to the second element).
  - If $R : A \to B$ is any relation, then it is easy to see that $(R^{-1})^{-1} = R$, since if $(a, b) \in R$ then $(b, a) \in R^{-1}$ so $(a, b) \in (R^{-1})^{-1}$, and vice versa.

- In practice, most of the time we do not explicitly work with the definition of a relation as a set of ordered pairs.
  - Instead, we think of a relation $a \; R \; b$ as a true or false statement that captures some information about a and b, and we usually work using the language of relations rather than subsets of Cartesian products.

3.2 Equivalence Relations

- We now discuss relations that share similar properties to equality.
  - We have already encountered one such relation, namely, modular congruence.
  - The fundamental properties of equality and modular congruence that involve only properties of the relation itself (and not other properties of arithmetic like addition or multiplication) are as follows: for any a, b, c, we have (i) $a = a$, (ii) if $a = b$ then $b = a$, and (iii) if $a = b$ and $b = c$, then $a = c$.

3.2.1 Definition and Examples

- We can easily give general definitions for each of these properties:
- Definitions: If $R : A \to A$ is a relation on the set A, we say R is reflexive when $a \; R \; a$ for all $a \in A$. We say R is symmetric when $a \; R \; b$ implies $b \; R \; a$ for all $a, b \in A$. We say R is transitive when $a \; R \; b$ and $b \; R \; c$ together imply $a \; R \; c$ for all $a, b, c \in A$.
  - In formal language, R is reflexive when $\forall a \in A$, $a \; R \; a$, while R is symmetric when $\forall a \in A \, \forall b \in A$, $(a \; R \; b) \Rightarrow (b \; R \; a)$, and R is transitive when $\forall a \in A \, \forall b \in A \, \forall c \in A$, $[(a \; R \; b) \wedge (b \; R \; c)] \Rightarrow (a \; R \; c)$.
- Here are some examples of relations that variously do and do not possess these three properties:
- Example: Suppose $A = \{1, 2, 3, 4\}$. Some relations on A are as follows:
  - The identity relation $R_1 = \{(1, 1), (2, 2), (3, 3), (4, 4)\}$ is reflexive, symmetric, and transitive. More generally, the identity relation on any set will always be reflexive, symmetric, and transitive.
  - The relation $R_2 = \{(1, 1), (2, 3), (3, 2)\}$ is not reflexive because for example the ordered pair (2,2) is not in R_2. It is symmetric because the reverses of all ordered pairs in R_2 are also in R_2, but it is not transitive because $2 \; R_2 \; 3$ and $3 \; R_2 \; 2$, but $2 \not R_2 \; 2$.
  - The relation $R_3 = \{(1, 1), (1, 2), (2, 1), (2, 2), (2, 4), (3, 3), (4, 2), (4, 4)\}$ is easily seen to be reflexive and symmetric since it contains all ordered pairs (a, a) and also contains the reverse of all its ordered pairs, but it is not transitive because $1 \; R_2 \; 2$ and $2 \; R_2 \; 4$, but $1 \not R_2 \; 4$.
  - The relation $R_4 = \{(1, 2), (2, 4), (1, 4)\}$ is not reflexive because for example it does not contain $(1, 1)$. It is also not symmetric because $1 \; R_4 \; 2$ but $2 \not R_4 \; 1$. However, it is transitive since (observe) the only a, b, c for which $a \; R_4 \; b$ and $b \; R_4 \; c$ are both true is $a = 1$, $b = 2$, and $c = 4$, and in such a case we also have $a \; R_4 \; c$.
  - The relation $R_5 = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), (4, 4)\}$, the $\leq$ relation on A, is clearly reflexive, but it is not symmetric because $1 \; R_5 \; 2$ but $2 \not R_5 \; 1$. It is transitive, although verifying this fact directly using the ordered pair definition is rather tedious.
  - The relation $R_6 = \{(1, 2), (2, 1)\}$ is not reflexive and not transitive, but is symmetric.
  - The relation $R_7 = \{(1, 1), (1, 4), (2, 2), (2, 3), (3, 2), (3, 3), (4, 1), (4, 4)\}$ is reflexive, transitive, and symmetric.
  - The empty relation $R_8 = \emptyset$ is not reflexive, but is symmetric because the conditional statement "for all $a, b \in A$ if $a \; R_8 \; b$ then $b \; R_8 \; a$" is (vacuously) true because the hypothesis is always false. This relation is also transitive, for the same reason.
- Example: The order relation $\leq$ on integers is reflexive and transitive but not symmetric.
  - Recall that we defined $a \leq b$ to mean that $b - a$ is a nonnegative integer, which is to say, an element of the set $\{0, 1, 2, 3, 4, \dots\}$.
  - Then the relation is reflexive because $a \leq a$ (because $a - a = 0$ is nonnegative), and it is transitive because if $a \leq b$ and $b \leq c$ (meaning that $b - a$ and $c - b$ are nonnegative) then $a \leq c$ (because $(c - b) + (b - a) = c - a$ is nonnegative).
  - However, the relation is not symmetric because for example $1 \leq 2$ but $2 \not\leq 1$.

- Remark: The same properties hold for the order relation $\leq$ on rational numbers and real numbers as well, along with the subset relation $\subseteq$ on sets and the divisibility relation $|$ on positive integers. We will return to discuss the general idea of an "order relation" later.

- Example: The implication relation $\Rightarrow$ on logical propositions, where we write $P \Rightarrow Q$ when P implies Q, is reflexive and transitive but not symmetric.
  - Explicitly, the relation is reflexive because $P \Rightarrow P$ for any P, and it is transitive because $P \Rightarrow Q$ and $Q \Rightarrow R$ together imply $P \Rightarrow R$ as one may check with a truth table.
  - However, the relation is not symmetric because $P \Rightarrow Q$ is not the same as its converse $Q \Rightarrow P$.

- Example: If m is any positive integer, the mod-m congruence relation $\equiv_m$ on integers is reflexive, symmetric, and transitive.
  - Recall that we write $a \equiv b \pmod{m}$ when m divides $b - a$. For the purposes of our discussion we will abbreviate this statement as $a \equiv_m b$ for consistency with our notation $a \, R \, b$ for relations.
  - We have (in fact) already shown that this relation is reflexive, symmetric, and transitive as part of our discussion of properties of congruences.
  - To summarize: $a \equiv a \pmod{m}$ because m always divides $a - a = 0$, so $\equiv_m$ is reflexive.
  - Also, if $a \equiv b \pmod{m}$ then $b \equiv a \pmod{m}$: this follows because if m divides $b - a$ then m also divides $-(b - a) = a - b$, so $\equiv_m$ is symmetric.
  - Finally, if $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$: this follows because if m divides $b - a$ and $c - b$ then it also divides $(c - b) + (b - a) = c - a$, so $\equiv_m$ is transitive.

- We now define the general notion of an equivalence relation:

- Definition: If R is a relation on the set A, we say R is an equivalence relation when it is reflexive, symmetric, and transitive.
  - Example: The identity relation on any set A is an equivalence relation. In particular, equality of integers, equality of rational numbers, equality of real numbers, and equality of sets are all equivalence relations.
  - Example: If m is any positive integer, the mod-m congruence relation $\equiv_m$ on integers is an equivalence relation.
  - Non-Example: The subset relation $\subseteq$ is not an equivalence relation since it is not symmetric.
  - Example: The relation $R_7 = \{(1,1), (1,4), (2,2), (2,3), (3,2), (3,3), (4,1), (4,4)\}$ on $A = \{1,2,3,4\}$ from above is an equivalence relation.
  - Example: The relation of having the same birthday (on the set of people) is an equivalence relation: everyone has the same birthday as themselves, if P has the same birthday as Q then Q has the same birthday as P, and if P has the same birthday as Q and Q has the same birthday as R, then P has the same birthday as R.

- Notation: It is very common to use a symbol like $\sim$ to represent an equivalence relation rather than the letter R, simply because the letter R produces expressions that are harder to parse.
  - In our discussion, we will continue to use the letter R because we are still examining basic properties of equivalence relations.

3.2.2 Equivalence Classes

- We saw previously that the residue classes $\overline{a}$ modulo m had a number of fundamental properties. There is a natural extension of this concept to a general equivalence relation:

- Definition: If R is an equivalence relation on the set A, we define the equivalence class of a as $[a] = \{b \in A : a \, R \, b\}$, the set of all elements $b \in A$ that are related to a via R.

- Example: If R is the equality relation on the set A, the equivalence class $[a]$ of the element a is simply the set $\{a\}$ containing a itself, since no other elements of A are related to a.
- Example: If R is the mod-m congruence relation on integers, the equivalence class $[a]$ of the element a is the residue class $\overline{a} = \{b \in \mathbb{Z} : a \equiv b \pmod{m}\}$. We saw earlier that these equivalence classes are $[0]$, $[1]$, ... , $[m-1]$ and that every integer lies in exactly one of these equivalence classes.
- Example: Under the equivalence relation $R_7 = \{(1,1), (1,4), (2,2), (2,3), (3,2), (3,3), (4,1), (4,4)\}$ on $A = \{1,2,3,4\}$, the equivalence classes are $[1] = \{1,4\}$, $[2] = \{2,3\}$, $[3] = \{2,3\}$, and $[4] = \{1,4\}$. Notice that there are two different equivalence classes, namely $[1] = [4] = \{1,4\}$ and $[2] = [3] = \{2,3\}$, and every element of A lies in exactly one of these equivalence classes.
- Example: Under the equivalence relation "having the same birthday" on the set of people, the equivalence class of any person $[P]$ is the set of all people having the same birthday as P. We may alternatively label these equivalence classes by the shared birthday (e.g., January 1, January 2, ... , up through December 31), and from this description, we can see that there are exactly 366 equivalence classes (one for each possible birthday, including February 29) and every person lies in exactly one of these equivalence classes (namely, the one labeled with their birthday).

- Like with the residue classes modulo m, and as suggested by all of the examples above, we can establish some basic properties of equivalence classes:

- Proposition (Properties of Equivalence Classes): Suppose R is an equivalence relation on the set A. Then

1. For any $a \in A$, a is an element of $[a]$.
   - Proof: Since R is reflexive, $a \; R \; a$, so by definition, $a \in [a]$.
2. If $a, b \in A$, then $[a] = [b]$ if and only if $a \; R \; b$.
   - Proof: If $[a] = [b]$, then since $b \in [b]$ by (1) above, this means that b is contained in the residue class $[a]$, meaning that $a \; R \; b$ by definition.
   - Conversely, suppose $a \; R \; b$. If c is any element of the equivalence class $[a]$, then by definition $a \; R \; c$, and so by symmetry $c \; R \; a$.
   - Hence by transitivity applied to $c \; R \; a$ and $a \; R \; b$, we see $c \; R \; b$, or equivalently, $b \; R \; c$.
   - Therefore, c is an element of the equivalence class $[b]$. But since c was arbitrary, this means that $[a]$ is a subset of $[b]$.
   - By the same argument with a and b interchanged, we see that $[b]$ is also a subset of $[a]$, and thus $[a] = [b]$.
3. Two equivalence classes of R on A are either disjoint or identical.
   - Proof: Suppose that $[a]$ and $[b]$ are two equivalence classes of R. If they are disjoint, we are done, so suppose there is some c contained in both: then $a \; R \; c$ and also $b \; R \; c$.
   - By symmetry, $b \; R \; c$ implies $c \; R \; b$, and then by transitivity, we conclude that $a \; R \; b$. Then by property (2), we conclude $[a] = [b]$.
   - Hence the two equivalence classes $[a]$ and $[b]$ are either disjoint or identical, as claimed.
4. There is a unique equivalence class of R on A containing a, namely, $[a]$.
   - Proof: Clearly $[a]$ is an equivalence class of R containing a by property (1) above.
   - On the other hand, by property (3), any other equivalence class containing a must equal $[a]$, so in fact, $[a]$ is the unique equivalence class of R containing a.

- From the results in the proposition, we can see that the equivalence classes are nonempty, pairwise disjoint subsets of A whose union is A. This particular situation is given a name:

- Definition: If A is a set, a partition $\mathcal{P}$ of A is a family of nonempty, pairwise disjoint sets whose union is A. The sets in $\mathcal{P}$ are called parts of the partition.

  - Example: The sets $\{1,5\}$ and $\{2,3,4\}$ yield a partition of $\{1,2,3,4,5\}$; explicitly, we could write $\mathcal{P} = \{\{1,5\},\{2,3,4\}\}$.
  - Example: The sets $\{1\}$, $\{2,3\}$, $\{4,5\}$ yield a different partition of $\{1,2,3,4,5\}$, as do the sets $\{1\}$, $\{2\}$, $\{3\}$, $\{4,5\}$.

- Non-Example: The sets $\{1,2\}$, $\{3,4\}$, and $\{4,5\}$ do not form a partition of $\{1,2,3,4,5\}$ because the sets are not pairwise disjoint: specifically, $\{3,4\}$ and $\{4,5\}$ have the element 4 in common.
- Non-Example: The sets $\{1,2,3\}$ and $\{5\}$ do not form a partition of $\{1,2,3,4,5\}$ because the union of the sets is not all of $\{1,2,3,4,5\}$: the element 4 is missing.
- Example: The sets $\mathbb{Z}_+ = \{1,2,3,\dots\}$, $\{0\}$, and $\mathbb{Z}_- = \{-1,-2,-3,\dots\}$ yield a partition of the integers.

- Our results above show that if R is any equivalence relation on a set A, then the equivalence classes of R yield a partition of A. In fact, the converse of this statement is also true: if we have a partition of A, then it arises as the equivalence classes of an equivalence relation on A.
  - To illustrate the idea, consider the partition $\mathcal{P} = \{\{1,5\},\{2,3,4\}\}$ of $\{1,2,3,4,5\}$, and suppose we had an equivalence relation R with these equivalence classes $\{1,5\}$ and $\{2,3,4\}$.
  - Then R must contain the ordered pairs $(1,1)$, $(2,2)$, $(3,3)$, $(4,4)$, and $(5,5)$ since it is reflexive.
  - Also, R must also contain the pairs $(1,5)$ and $(5,1)$ because 1 and 5 are supposed to lie in the same equivalence class $\{1,5\}$, and likewise R must contain all of the pairs $(2,3)$, $(2,4)$, $(3,2)$, $(3,4)$, $(4,2)$, and $(4,3)$ because 2, 3, and 4 all lie in the same equivalence class.
  - On the other hand, R cannot contain any other pairs than the ones we have listed, because the only remaining ordered pairs involve elements from different parts of the partition, and we cannot include any of those ordered pairs because those elements are required to lie in different equivalence classes.
  - So the only choice is $R = \{(1,1),(1,5),(5,1),(5,5),(2,2),(2,3),(2,4),(3,2),(3,3),(3,4),(4,2),(4,3),(4,4)\}$.
  - Notice here that R is the union of the Cartesian products $\{1,5\}\times\{1,5\}$ and $\{2,3,4\}\times\{2,3,4\}$ of the underlying parts of the partition. From this description, it is quite easy to see that this relation R is indeed an equivalence relation whose equivalence classes are $\{1,5\}$ and $\{2,3,4\}$.
  - Based on this example, we need only collect the important details of this construction and verify that they do work in general.

- Theorem (Equivalence Relations and Partitions): Let A be a set. If R is any equivalence relation on A, then the equivalence classes of R form a partition $\mathcal{P}$ of A. Conversely, if $\mathcal{P}$ is a partition of A, then there exists a unique equivalence relation R on A whose equivalence classes are the sets in $\mathcal{P}$, namely, the equivalence relation $R = \bigcup_{X\in\mathcal{P}}(X\times X)$ consisting of all ordered pairs of elements that are in the same part X of the partition $\mathcal{P}$.
  - Intuitively, the relation R is defined by saying that $a\ R\ b$ when a and b are in the same part of the partition. The choice $R = \bigcup_{X\in\mathcal{P}} X\times X$ is simply a formalization of this idea.
  - Proof: The first statement was shown above, so now suppose $\mathcal{P}$ is a partition of A.
  - Define the relation $R = \bigcup_{X\in\mathcal{P}} X\times X$ consisting of all ordered pairs of elements that are in the same part X of the partition $\mathcal{P}$: we must show that this R is an equivalence relation and that its equivalence classes are the parts of $\mathcal{P}$.
  - First, R is reflexive: for any $a\in A$, by the definition of a partition we must have $a\in X$ for some $X\in\mathcal{P}$. Then the ordered pair (a,a) is an element of $X\times X$, as required.
  - Second, R is symmetric: if $(a,b)\in R$, then by the definition of R as a union, we must have $(a,b)\in X\times X$ for some $X\in\mathcal{P}$. This means $a\in X$ and $b\in X$: then $(b,a)\in X\times X$ also, and so $(b,a)\in R$.
  - Third, R is transitive: if $(a,b)\in R$ and $(b,c)\in R$, then we must have $(a,b)\in X\times X$ and $(b,c)\in Y\times Y$ for some $X,Y\in\mathcal{P}$. This means $a\in X$ and $b\in X$, and also $b\in Y$ and $c\in Y$. Because $\mathcal{P}$ is a partition, since $b\in X$ and $b\in Y$ we must have $X=Y$. Then $a\in X$ and also $c\in X$, so $(a,c)\in X\times X$ and so $(a,c)\in R$.
  - Hence R is an equivalence relation.
  - Now let $a\in A$ and consider the equivalence class $[a]$ of a. Since $\mathcal{P}$ is a partition, $a\in X$ for a unique $X\in\mathcal{P}$. We claim that $[a]=X$.
  - To see this, if $b\in X$, we have $(a,b)\in X\times X$ hence $(a,b)\in R$ hence $a\ R\ b$ hence $b\in[a]$. This shows $X\subseteq[a]$.

- For the other containment, if $b \in [a]$ then $a \ R \ b$ so that $(a,b) \in R$. By the definition of R as a union, this requires $(a,b) \in Y \times Y$ for some $y \in \mathcal{P}$ where $a \in Y$ and $b \in Y$. Since $a \in X$ we must have $Y = X$, so we see $b \in X$. This shows $[a] \subseteq X$, so $[a] = X$ as claimed.
- We conclude that the equivalence classes of R are the same as the parts of $\mathcal{P}$, as required.
- Finally, for uniqueness, if S is another relation with the same property, then for each $X \in \mathcal{P}$, the relation S must contain $X \times X$, hence must contain $R = \bigcup_{X \in \mathcal{P}} X \times X$.
- If S contained any additional ordered pairs, then such an ordered pair would contain elements from two different parts X and Y of the partition, but then $X \cup Y$ would be contained in an equivalence class of S, contrary to hypothesis. Hence we must have $S = R$, so R is unique as claimed.

- From the theorem above, we obtain another way to verify that a relation is an equivalence relation, namely, by checking whether it is obtained from a partition.

3.2.3 Constructions of $\mathbb{Q}$ and Vectors via Equivalence Relations

- We finish our discussion of equivalence relations by discussing the constructions of some fundamental mathematical objects using the language of equivalence relations: the rational numbers $\mathbb{Q}$, and vectors.
  - The rational numbers consist of fractions of the form a/b where a is an integer and b is a nonzero integer, together with the fundamental operations of addition and multiplication, which work via $a/b + c/d = (ad+bc)/(bd)$ and $a/b \cdot c/d = (ac)/(bd)$.
  - However, we must also deal with the fact that there are many equivalent ways of writing the same rational number: for example, $1/2 = 2/4 = 1000/2000 = (-5)/(-10)$.
  - We may neatly accommodate all of these requirements by describing rational numbers in the language of equivalence relations: since $a/b = c/d$ is equivalent to $ad = bc$, we can use this as our starting point.

- Proposition (Construction of $\mathbb{Q}$): Let $S = \mathbb{Z} \times \mathbb{Z}_{\neq 0}$ be the set of ordered pairs (a,b) of integers with $b \neq 0$ and define the relation $\sim$ by saying $(a,b) \sim (c,d)$ precisely when $ad = bc$. Then the following hold:

1. The relation $\sim$ is an equivalence relation on S.
   - Proof: First, $\sim$ is reflexive: clearly we have $(a,b) \sim (a,b)$ because $ab = ab$.
   - Second, $\sim$ is symmetric: clearly we have $(a,b) \sim (b,a)$ because $ab = ba$.
   - Third, $\sim$ is transitive: if $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$, then we know $ad - bc = 0$ and also $cf - de = 0$. Then we see $d(af - be) = f(ad - bc) + b(cf - de) = f \cdot 0 + b \cdot 0 = 0$.
   - Then since $d \neq 0$ since it is a second coordinate, this means $af - be = 0$ so $af = be$, and thus $(a,b) \sim (e,f)$ as required.

If a, b are integers with $b \neq 0$, we now think of a/b as being the equivalence class $[(a,b)]$. We define addition and multiplication on equivalence classes by $[(a,b)] + [(c,d)] = [(ad+bc, bd)]$ and $[(a,b)] \cdot [(c,d)] = [(ac,bd)]$.

2. Addition and multiplication of residue classes are well-defined, in the sense that if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, then $(ad+bc, bd) \sim (a'd' + b'c', b'd')$ and $(ac, bd) \sim (a'c', b'd')$.
   - Proof: Suppose $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, meaning that $ab' = a'b$ and $cd' = c'd$.
   - First, to show $(ad+bc, bd) \sim (a'd'+b'c', b'd')$, by definition we need $(ad+bc)(b'd') = (bd)(a'd'+b'c')$.
   - Then using $ab' = a'b$ and $cd' = c'd$, we can write $(ad+bc)(b'd') = ab'dd' + cd'bb' = a'bdd' + c'dbb' = (bd)(a'd' + b'c')$, as required.
   - Second, to show $(ac, bd) \sim (a'c', b'd')$, by definition we need $(ac)(b'd') = (bd)(a'c')$.
   - Using $ab' = a'b$ and $cd' = c'd$ then gives $(ac)(b'd') = (ab')(cd') = (a'b)(c'd) = (bd)(a'c')$, as required.
3. Addition is associative, commutative, possesses an identity $[(0,1)]$, and every element $[(a,b)]$ has an additive inverse $[(-a,b)]$.
   - Proof: Each of these follows by writing out the corresponding property and checking that the two quantities are equivalent under $\sim$.

- Associative: We have $([(a,b)]+[(c,d)])+[(e,f)] = [(ad+bc,bd)]+[(e,f)] = [(adf+bcf)+bde, bdf)]$ while $[(a,b)]+([(c,d)]+[(e,f)]) = [(a,b)]+[(cf+de,df)] = [(adf+(bcf+bde),bdf)]$, which is the same because the integers $(adf+bcf)+bde$ and $adf+(bcf+bde)$ are equal.
- Commutative: We have $[(a,b)]+[(c,d)] = [(ad+bc,bd)]$ and $[(c,d)]+[(a,b)] = [(cb+da,db)]$, which are the same since $ad+bc = cb+da$ and $bd = db$.
- Identity: We have $[(a,b)]+[(0,1)] = [(a\cdot 1+b\cdot 0, b\cdot 1)] = [(a,b)]$.
- Inverses: We have $[(a,b)]+[(-a,b)] = [(a\cdot b+(-a)\cdot b, b\cdot b)] = [(0,b^2)]$. But $(0,b^2)\sim(0,1)$ so $[(0,b^2)] = [(0,1)]$ which is the additive identity.

4. Multiplication is commutative, associative, possesses an identity $[(1,1)]$, every element $[(a,b)]\neq[(0,1)]$ has a multiplicative inverse $[(b,a)]$, and distributes over addition.
   - Proof: These all follow similarly to (3).
   - Associative: We have $([(a,b)]\cdot[(c,d)])\cdot[(e,f)] = [(ac,bd)]\cdot[(e,f)] = [(ace,bdf)] = [(a,b)]\cdot[(ce,df)] = [(a,b)]\cdot([(c,d)]\cdot[(e,f)])$.
   - Commutative: We have $[(a,b)]\cdot[(c,d)] = [(ac,bd)] = [(ca,db)] = [(c,d)]\cdot[(a,b)]$.
   - Identity: We have $[(a,b)]\cdot[(1,1)] = [(a\cdot 1, b\cdot 1)] = [(a,b)]$.
   - Inverses: We have $[(a,b)]\cdot[(b,a)] = [(ab,ab)]$. But $(ab,ab)\sim(1,1)$ so $[(ab,ab)] = [(1,1)]$ which is the multiplicative identity.
   - Finally, for the distributive law we simply multiply out like above: $[(a,b)]\cdot([(c,d)]+[(e,f)]) = [(a,b)]\cdot[(cf+de,df)] = [(acf+ade,bdf)]$, and also $[(a,b)]\cdot[(c,d)]+[(a,b)]\cdot[(e,f)] = [(ac,bd)]+[(ae,bf)] = [(abcf+abde,b^2df)]$.
   - Then since $(acf+ade,bdf)\sim(abcd+abde,b^2df)$, the corresponding equivalence classes are equal.
5. The set $\mathbb{Q}$ of equivalence classes $[(a,b)]$ in S forms a field under the operations $+$ and $\cdot$.
   - Proof: This is just a rephrasing of (3) and (4) put together.

- Using similar ideas, we can formalize the construction of vectors (often simply described as "arrows", where any two arrows that have the same length and point in the same direction are considered equivalent) using equivalence classes:
- Example (Vectors): A directed line segment in the plane (or 3-space) is given by drawing an arrow from its starting point P to its ending point Q.
  - Let R be the relation of translation (on the set of directed line segments): we write $S_1\ R\ S_2$ if the directed line segment S_1 can be translated to obtain the directed line segment S_2.
  - It is easy to see from this geometric description that R is an equivalence relation. The equivalence classes of directed line segments under R are called vectors.
  - Because there is a unique element in each equivalence class whose starting point is the origin, we may label each equivalence class with the endpoint of this unique vector. Thus, for example, the vector $\langle 1,2\rangle$ is the equivalence class of directed line segments, one of which starts at the origin $(0,0)$ and ends at the point $(1,2)$. Another directed segment in the same equivalence class $\langle 1,2\rangle$ is the vector starting at $(5,0)$ and ending at $(6,2)$.

3.3 Orderings

- We now discuss relations that generalize the properties of the order relation $\leq$ on real numbers (and also rational numbers and integers) and the subset relation $\subseteq$ on sets.

3.3.1 Partial and Total Orderings

- As we have already seen, both relations $\leq$ and $\subseteq$ satisfy some of the properties of an equivalence relation: specifically, they are both reflexive and transitive.
  - However, neither of these relations is symmetric: in fact, the only time when $a\leq b$ and $b\leq a$ are both true is when $a=b$; similarly, the only time when $A\subseteq B$ and $B\subseteq A$ are both true is when $A=B$.

- This latter property is (almost) the opposite of being symmetric, and is given a name accordingly:

- Definition: If R is a relation on the set A, then R is antisymmetric when $a\ R\ b$ and $b\ R\ a$ together imply $a = b$.
  - In formal language, R is antisymmetric when $\forall a \in A\, \forall b \in B$, $[(a\ R\ b) \wedge (b\ R\ a)] \Rightarrow (a = b)$.
  - Example: The order relation $\leq$ on real numbers is antisymmetric, because $a \leq b$ and $b \leq a$ implies $a = b$. (In fact, these are equivalent.)
  - Example: The subset relation $\subseteq$ on sets is antisymmetric, because $A \subseteq B$ and $B \subseteq A$ implies $A = B$. (In fact, these are equivalent.)
  - Example: The identity relation R on A is antisymmetric, since the only time that $a\ R\ b$ is true is when $a = b$.
  - Notice that the identity relation on A is both symmetric and antisymmetric. In particular, despite what may be suggested by the terminology, "antisymmetric" does not mean the same thing as "not symmetric", and "symmetric" does not mean the same thing as "not antisymmetric".

- Both of these relations involve the idea of one object being "at least as big" as another, so we would like to find a way to describe this concept in the abstract language of relations.
  - If R is a generic relation in which $a\ R\ b$ means that b is at least as big as a, then certainly we should demand that $a\ R\ a$ so that R is reflexive (since a is at least as big as itself).
  - We would also want R to be transitive, since if c is at least as big as b and b is at least as big as a, then c should be at least as big as a.
  - Finally, antisymmetry is also a natural condition: the only situation in which we would like b to be at least as big as a and a to be at least as big as b is when $a = b$.
  - These are the conditions we will require for an order relation.

- Definition: The relation R on a set A is called a partial ordering of A (or partial order on A) if R is reflexive, antisymmetric, and transitive.
  - Example: The less-than-or-equal-to relation $\leq$ on real numbers (or rational numbers, or integers) is a partial ordering, as is the subset relation $\subseteq$ on sets. In each case we can view the relation $x \leq y$ as expressing the idea that y is "at least as big" as x.
  - Example: The greater-than-or-equal-to relation $\geq$ on real numbers (or rational numbers, or integers) is also a partial ordering. Although this may seem strange at first, we can think of this relation $x \geq y$, which is the inverse relation of $x \leq y$, as capturing the idea that y is "at least as small" as x, in parallel to how $x \leq y$ captures the idea that y is "at least as big" as x.
  - Example: The relation $R_9 = \{(1,1),(1,2),(2,2),(3,3),(3,4),(4,4)\}$ on the set $\{1,2,3,4\}$ is a partial ordering. It is easy to see that R_9 is reflexive (it contains all pairs (a,a)) and antisymmetric (it does not contain both (a,b) and (b,a) for any $a \neq b$), and it is a straightforward check to see it is also transitive.
  - Non-Example: The divisibility relation $|$ on the set of all integers is not a partial ordering: although it is reflexive and transitive, it is not antisymmetric because for example $1|(-1)$ and $(-1)|1$, but $-1 \neq 1$.
  - Example: The divisibility relation $|$ on the set of positive integers is a partial ordering: it is reflexive and transitive, and is also symmetric because if a and b are positive with $a|b$ and $b|a$, then $a = b$ (since $a|b$ implies $a \leq b$ for a, b positive, so $a|b$ and $b|a$ give $a \leq b$ and $b \leq a$ so that $a = b$).
  - It is not hard to see that if S is a subset of A, then the restriction of a partial ordering on A to S yields a partial ordering on S. Hence, for example, the divisibility relation $|$ is also a partial ordering on the set of positive even integers.

- Example: Show that the relation R_{10} is a partial ordering on all finite strings of digits, where $a\ R_{10}\ b$ when the string b contains the string a (consecutively, in the same order) somewhere inside of it.
  - To illustrate this relation, note that $123\ R_{10}\ 412390$ because the second string contains the first one (as its second through fourth digits) but $123\ \not R_{10}\ 31213$ because the second string does not have "123" in it anywhere.

- This relation is reflexive (any string contains itself), antisymmetric (if two strings each contain one other, they would have to be the same length and identical), and transitive (if c contains b and b contains a, then c contains a since a is located inside the string for b). Hence it is a partial ordering, as claimed.
- Remark: Interestingly, this relation is not a partial ordering if we allow infinite strings of digits, since it is no longer antisymmetric: for example, the alternating strings $12121212\ldots$ and $21212121\ldots$ each contain the other, but they are not equal.

- We use the term "partial ordering" because a partial order on A gives us a way of comparing some, but not necessarily all, pairs of elements of A.
  - For example, if R is the subset relation, then for $A = \{1, 2\}$ and $B = \{3\}$, we cannot compare A to B using R, because $A \not\subseteq B$ and also $B \not\subseteq A$.
  - If R is the divisibility relation on positive integers, then we cannot compare 2 to 3, since $2 \nmid 3$ and $3 \nmid 2$.
  - Likewise, for the relation R_9 on $\{1, 2, 3, 4\}$ we cannot compare 1 to 3, because neither of the ordered pairs $(1, 3)$ and $(3, 1)$ is in R_9.
  - Similarly, for the relation R_{10} on strings of digits, we cannot compare 123 to 4567, because neither string contains the other.
  - However, for some of the order relations we have listed, it is possible to compare any two elements in the set: for example, for any two real numbers a and b, it is true that either $a \leq b$ or $b \leq a$ (or both, in which case $a = b$).
  - This situation is important enough that we give it a name:

- Definition: If R is a partial ordering on A such that for any $a, b \in A$ at least one of $a\ R\ b$ and $b\ R\ a$ is true[1], we call R a total ordering (or linear ordering) on A.
  - Example: The order relation $\leq$ on real numbers (or rational numbers, or integers) is a total ordering.
  - Example: The standard dictionary ordering on the letters of the alphabet (namely: a, b, c, ... , z) where we write $L_1 \leq L_2$ if L_2 is after L_1 in the alphabet, is a total ordering.
  - Example: The divisibility relation on the set $\{1, 2, 4, 8, 16, \ldots\}$ of powers of 2 is a total ordering, since it is clearly a partial ordering, and for any two powers of 2, one of them must divide the other.

- Notice that if R is a total ordering then since R is antisymmetric, we see that for any a, b with $a \neq b$, exactly one of $a\ R\ b$ and $b\ R\ a$ is true.
  - Thus, we may think of R as allowing us to compare any two unequal elements of A to identify which one is "bigger".
  - Given a total ordering, we can also imagine arranging all of the elements of A "in order" along a line (whence the name linear ordering); indeed, for the ordering $\leq$ on the real numbers, this is precisely the so-called "number line".
  - Like with partial orderings, the restriction of a total ordering to a subset S of A is a total ordering on S.

- Notation: Because partial orderings behave so much like the $\leq$ relation on real numbers, it is very common to use a similar symbol, such as $\preceq$, or even just the $\leq$ symbol itself, to represent a generic partial ordering.
  - In our discussion, we will continue to use the letter R because we are still examining basic properties of orderings.

[1] For a general relation R, the condition that $a\ R\ b$ or $b\ R\ a$ is true is called the connex property.

3.3.2 Smallest, Largest, Minimal, and Maximal Elements

- In many contexts when we are working with partial orderings, important properties are often attached to extremal elements (i.e., elements that are the largest, or smallest, with respect to the ordering).
    - For example, the greatest common divisor of two integers is (per its name) the greatest integer d such that $d|a$ and $d|b$.
    - Similarly, the union of two subsets A and B is the smallest set C such that $A \subseteq C$ and $B \subseteq C$.
- We can make these notions precise as follows:
- Definition: If R is a partial ordering on a set A, we say that an element $x \in A$ is a smallest element (or least element) of A with respect to R when $x \; R \; a$ for all $a \in A$. We say $x \in R$ is a largest element (or greatest element) of A with respect to R if $a \; R \; x$ for all $a \in A$.
    - When the relation R is clear from the context, we usually refer to these elements just as the "smallest element of A" or "largest element of A", respectively. Also, as we will show shortly, a smallest element (or largest element) is necessarily unique, so we may refer to *the* smallest element rather than merely *a* smallest element.
    - Example: If U is a universal set and R is the subset relation on $\mathcal{P}(U)$, the smallest element of $\mathcal{P}(U)$ is the empty set $\emptyset$ and the largest element of $\mathcal{P}(U)$ is U itself.
    - Example: If $A = \{1, 2, 5, 10\}$and R is the divisibility relation, the smallest element of A is 1 and the largest element of A is 10.
    - Smallest and largest elements need not exist with respect to a partial or even a total ordering:
    - Example: If $A = \{2, 3, 4, 5, 6, 7\}$ and R is the divisibility relation, then A has no smallest element since no element in A divides all elements in A, and also A has no largest element since no element of A is divisible by all elements in A.
    - Example: If A is the set of positive integers and R is the total ordering $\leq$, the smallest element of A is 1, but there is no largest element of A. (No integer n is largest, since $n + 1$ is always larger.)
    - Example: If A is the set of positive real numbers and R is the total ordering $\leq$, then A has no smallest element and no largest element. (No positive real number a can be smallest or largest, since $a/2$ is always smaller and $2a$ is always larger.)
- Closely related to smallest and largest elements are minimal and maximal elements:
- Definition: If R is a partial ordering on a set A, we say that an element $x \in A$ is a minimal element of A with respect to R (or just minimal) when $y \; R \; x$ implies $y = x$, and we say $x \in A$ is a maximal element of A with respect to R (or just maximal) when $x \; R \; y$ implies $y = x$.
    - The idea is that a minimal element has no other elements below it, while a maximal element has no other elements above it, with respect to the ordering.
    - Example: If U is a universal set and R is the subset relation on $\mathcal{P}(U)$, the empty set $\emptyset$ is the only minimal element, and U is the only maximal element. A nonempty set is not minimal since the empty set is always contained in it, while a proper subset of U is not maximal since it is always contained in U.
    - As we will prove in a moment, smallest elements are always minimal and largest elements are always maximal, but minimal elements need not be smallest and maximal elements need not be largest.
    - Example: If $A = \{2, 3, 4, 5, 6, 7\}$ and R is the divisibility relation, then 2, 3, and 5 are all minimal elements, since no other element of A divides any of them. Also, the elements 4, 5, 6, and 7 are all maximal elements, since they do not divide any other element of A. Note in particular that 5 is both minimal and maximal.
    - Example: If A is the collection of subsets of $\{2, 3, 4, 5, 6, 7\}$ that contain 2 or contain 3 but do not contain both 6 and 7, and R is the subset relation, then $\{2\}$ and $\{3\}$ are both minimal elements, since no other element of A is a subset of either one. Also, the sets $\{2, 3, 4, 5, 6\}$ and $\{2, 3, 4, 5, 7\}$ are both maximal, since no other element of A contains either one.
    - Minimal and maximal elements also need not exist at all, even for total orderings:

- Example: If A is the set of positive integers and R is the total ordering $\leq$, then 1 is the only minimal element of A, and there is no maximal element of A. (No integer n can be maximal, since $n+1$ is always larger.)
- Example: If A is the set of positive real numbers and R is the total ordering $\leq$, then A has no minimal element and no maximal element. (No positive real number r can be minimal or maximal, since $r/2$ is always smaller and $2r$ is always larger.)

- We have various properties of smallest, largest, minimal, and maximal elements:
- Proposition (Properties of Extremal Elements): Suppose R is a partial ordering on a set A. Then the following hold:

1. There is at most one smallest element of A and at most one largest element of A.
   - Proof: Suppose x and y are both smallest elements of A. Then $x \; R \; y$ (since x is smallest) and $y \; R \; x$ (since y is smallest), hence by antisymmetry $x = y$.
   - A very similar argument holds for largest elements.
2. If A is finite and nonempty, then A has at least one minimal element and at least one maximal element.
   - Proof: Induct on the number of elements n of A. The base case $n = 1$ is trivial, since if $A = \{x\}$ then x is both minimal and maximal.
   - For the inductive step, suppose that $n \geq 2$ and any partial ordering on a set with n elements has a minimal element and a maximal element, and let $A = \{x_1, \ldots, x_n, x_{n+1}\}$ have $n+1$ elements.
   - Consider the relation R' given by restricting R to $A' = \{x_1, \ldots, x_n\}$. By the inductive hypothesis, there is some element $y \in A'$ that is minimal in A' under R', meaning that the only element x_i with $1 \leq i \leq n$ with $x_i \; R \; y$ is y itself.
   - Now consider x_{n+1}. If $x_{n+1} \not R \; y$ then y is minimal in A, since now the only element x_i with $1 \leq i \leq n+1$ with $x_i \; R \; y$ is y itself.
   - Otherwise, if $x_{n+1} \; R \; y$ then we claim x_{n+1} is minimal in A. To see this suppose there were some $z \neq x_{n+1}$ with $z \; R \; x_{n+1}$: then by transitivity we would have $z \; R \; y$, but this is impossible because $z \in A'$ and y is minimal in A' and thus $z = y$, but then $x_{n+1} \; R \; y$ and $z \; R \; x_{n+1}$ together would imply $y = x_{n+1}$, which is impossible because $y \in A'$ and $x_{n+1} \notin A'$.
   - Thus we see in either case that A has a minimal element. By a similar argument A also has a maximal element, and so the result holds for all finite sets by induction.
3. If A is finite and nonempty and R is a total ordering, then A has a unique smallest element and a unique largest element.
   - Proof: The existence of these elements follows by an argument similar to (2), while the uniqueness follows from (1).
4. If $x \in A$ is smallest then x is the unique minimal element of A, and if $x \in A$ is largest then x is the unique maximal element of A.
   - Proof: Suppose x is a smallest element of A. Then x is minimal because if $y \; R \; x$ then since x is smallest we also have $x \; R \; y$ so by antisymmetry we would have $y = x$.
   - Additionally, if z is some other minimal element, then since x is smallest we have $x \; R \; z$, but since z is minimal this implies $z = x$: thus, x is the unique minimal element.
   - A similar argument (with all of the directions reversed) establishes the corresponding result for largest elements.
5. If R is a total ordering and $x \in A$ is a minimal element of A, then x is the smallest element of A. Likewise, if $x \in A$ is a maximal element of A, then x is the largest element of A. In particular, a total ordering has at most one minimal element and one maximal element.
   - Proof: Suppose R is a total ordering and x is minimal. Then for any $y \in A$ we either have $y \; R \; x$ or $x \; R \; y$.
   - But since x is minimal, $y \; R \; x$ can only happen when $y = x$. So, if $y \neq x$ we must have $x \; R \; y$, meaning that x is the smallest element of A.

- A similar argument establishes the corresponding result for maximal and largest elements. The last statement then follows immediately from (1).

- We remark that item (5) in the proposition above is essentially the converse of item (4) – if x is a unique minimal element of A then x is the smallest element of A – but it has an extra hypothesis: namely, that R is a total ordering.
  - In fact this extra hypothesis is necessary, although it is not so easy to write down a counterexample (i.e., a set with a minimal element but no smallest element) since (3) implies such a set A must be infinite.
  - Here is a partial ordering with a unique minimal element but no smallest element: take A to be the set of positive real numbers along with an extra number "x" under the usual ordering $\leq$, where we also declare $x \leq x$ but otherwise x is not comparable to any of the positive real numbers. Then x is the unique minimal element of A (since $y \leq x$ is only true when $y = x$) but x is not the smallest element of A, since for example $x \not\leq 1$.

- We close our discussion here with some examples of smallest, largest, minimal, and maximal elements that are of concrete interest.
  - Example: If A is the set of positive common divisors of two positive integers a and b and R is the divisibility relation, the largest element of A under R is the greatest common divisor $\gcd(a, b)$.
  - Example: If A is the set of positive common multiples of two positive integers a and b and R is the divisibility relation, the smallest element of A under R is the least common multiple $\text{lcm}(a, b)$.
  - Example: If $\mathcal{F}$ is the collection of sets that are simultaneously subsets of the sets B and C, and R is the subset relation, the largest element of $\mathcal{F}$ under R is the intersection $B \cap C$.
  - Example: If $\mathcal{F}$ is the collection of sets each containing all of the elements of the two sets B and C, and R is the subset relation, the smallest element of $\mathcal{F}$ under R is the union $B \cup C$.
  - Example: If A is the set of real numbers of the form x^2 for some $x \in \mathbb{R}$, and R is the relation $\leq$, then the smallest element of A is the number 0. (We usually express this statement in this simpler form: if $x \in \mathbb{R}$ then $x^2 \geq 0$. This seemingly trivial inequality has surprisingly many applications in establishing other inequalities.)

3.4 Functions

- We now discuss how to formalize the idea of a function using the language of relations.

3.4.1 Definition and Examples

- The idea of a function is already quite familiar: to each element of its domain, a function f associates a unique value in its range.
  - More explicitly, we write $f(a) = b$ to indicate that the value of f at the element a is equal to b.
  - We can then view f as a relation by saying that $f(a) = b$ precisely when $(a, b) \in f$.
  - The requirement that f is defined on every element of its domain means that for all $a \in A$, where A is the domain of f, there exists some value b in some other set B such that $(a, b) \in f$. Furthermore, because f is well-defined, there is only one such element b.
  - We can summarize all of this as follows:

- Definition: If A and B are sets, a function (or map) from A to B is a relation $f : A \to B$ such that for every $a \in A$ there exists a unique $b \in B$ with $(a, b) \in f$, and in such an event we write $f(a) = b$. The set A is called the domain of f and the set B is called the target (or codomain) of f.
  - We emphasize that the domain and target are part of the definition of a function. Two functions are equal when their domains are equal, their targets are equal, and their underlying sets of ordered pairs are equal.

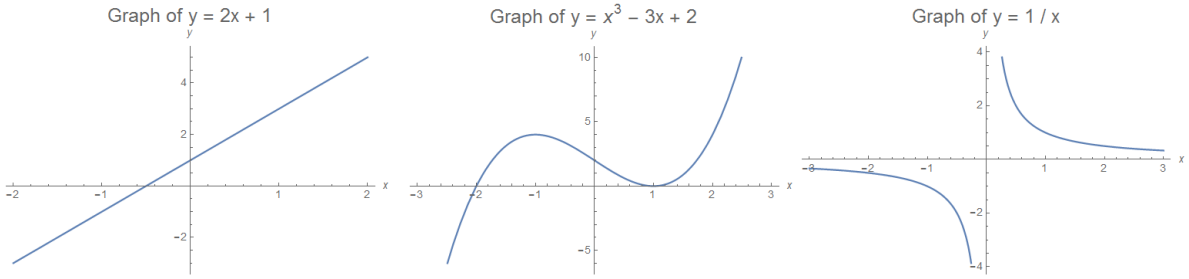
- Example: Some functions from $\{1,2,3,4\}$ to $\{1,2,3,4\}$ are $f_1 = \{(1,2), (2,3), (3,1), (4,4)\}$, $f_2 = \{(1,1), (2,3), (3,2), (4,2)\}$, and $f_3 = \{(1,4), (2,3), (3,2), (4,1)\}$. We have, for example, $f_1(1) = 2$, $f_1(3) = 1$, $f_2(4) = 2$, and $f_3(1) = 4$.
- Example: One function from $\{a,b,c\}$ to $\{31,37\}$ is given by $f = \{(a,31), (b,31), (c,37)\}$. For this function, $f(a) = 31$, $f(b) = 31$, and $f(c) = 37$.
- Non-Example: The relation $R : \{1,2,3\} \to \{1,2,3,4\}$ given by $R = \{(1,1), (1,2), (2,2), (3,1)\}$ is not a function because it is not well-defined on the element 1 (since it contains the ordered pairs $(1,1)$ and $(1,2)$).
- Example: If T is the set of triangles in the Cartesian plane, then there is a function $f_4 : T \to \mathbb{R}$ where $f_4(\triangle)$ is the area of the triangle $\triangle$. Every triangle has a well-defined area, and this area is an element of the target set $\mathbb{R}$.
- Example: If S is the set of integers greater than 1, then there is a function $f_5 : S \to \mathbb{Z}$ where $f_5(n)$ is the smallest prime number dividing n. For example, we have $f_5(100) = 2$ and $f_5(33) = 3$.
- Example: If A is the set of all capital cities and B is the set of all countries, then there is a function $l : A \to B$ where $l(C)$ is the country of which C is the capital. (In order for this to be a well-defined function, we observe that no city is the capital of more than one country.)
- Example: If A is any set, the identity function $i_A : A \to A$ is the function with $i_A(a) = a$ for all $a \in A$. Note that this definition is still well-posed when A is the empty set: in this case i_A is the empty function consisting of no ordered pairs at all.
- Non-Example: If S is the set of all people, then the relation $R : S \to S$, consisting of all ordered pairs (P,Q) where P is a parent of Q, is not a function: there exist some people P that are the parent of more than one person, and for such people there is not a unique value to call $R(P)$.
- Example: If S is the set of all people, consider the relation $R : S \to \mathcal{P}(S)$ consisting of all ordered pairs (P,Q) where Q is the set of all children of P. Then R is a function, because to each person in S there is associated a unique element of $\mathcal{P}(S)$, namely, the set of all children of P. This set may be empty or contain more than one person, but in all cases it is well-defined and unique.

- Many functions (and most of the functions we typically work with) can be defined by a general rule or description, such as the function $f_3 : \{1,2,3,4\} \to \{1,2,3,4\}$ above: explicitly, we can see that $f_3(n) = 5 - n$ for all $n \in \{1,2,3,4\}$.
  - We typically abbreviate such a definition by merely writing $f_3(n) = 5 - n$ with the implicit assumption that this rule is valid for all n in the domain of f_3, which in this case is $\{1,2,3,4\}$.
  - Example: Some examples of functions from $\mathbb{R}$ to $\mathbb{R}$ that can be defined in this way are the squaring function $p(x) = x^2$, the sine function $s(x) = \sin(x)$, and the absolute value function $a(x) = |x| = \begin{cases} x & \text{for } x \geq 0 \\ -x & \text{for } x < 0 \end{cases}$.
  - When defining a function in this way, it is very important to ensure that the definition is unambiguous and well-defined.
  - For example, although it may seem valid to define a function $f : \mathbb{Q} \to \mathbb{Z}$ by saying $f(a/b) = a$ for any $a/b \in \mathbb{Q}$, this definition does not actually yield a well-defined function: notice that, per the rule given, we would have $f(1/2) = 1$ while $f(2/4) = 2$, but $1/2 = 2/4$ as rational numbers. (One way to fix this definition would be to specify that a/b must be in lowest terms, and also to clarify what happens with negative elements of the domain.)

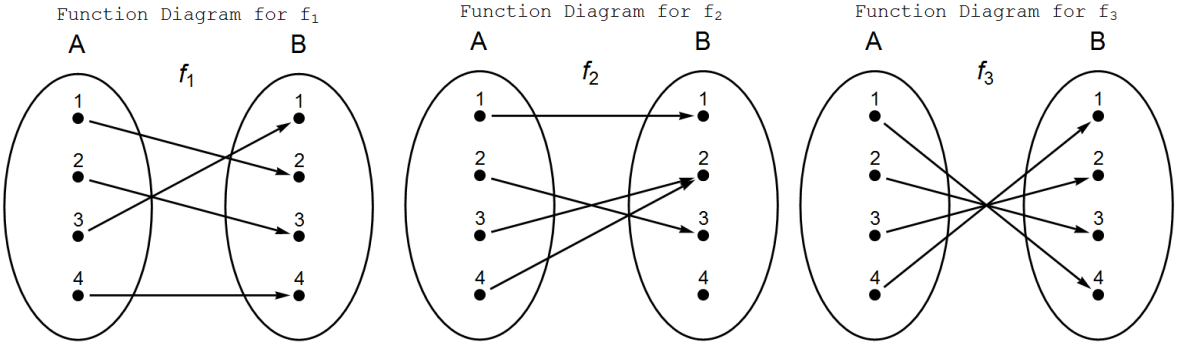
- It is crucial to specify the domain and target when we define a function via a rule in this manner; otherwise, the definition can be ambiguous.
  - To illustrate why, consider the functions $g_1 : \mathbb{R} \to \mathbb{R}$ with $g_1(x) = x^2$ and $g_2 : \mathbb{Z} \to \mathbb{Z}$ with $g_2(x) = x^2$.
  - The functions g_1 and g_2 are (seemingly) defined by the same rule, but they are different functions because their underlying sets of ordered pairs are different: notice for example that $(1/2, 1/4) \in g_1$, but $(1/2, 1/4) \notin g_2$.

- It is often very helpful to represent functions geometrically.

- For functions from (a subset of) $\mathbb{R}$ to (a subset of) $\mathbb{R}$ we may draw the graph of a function f, which consists of all points (x, y) in the Cartesian plane such that $(x, y) \in f$.[2]
- If the domain is unbounded (i.e., contains points arbitrarily far from 0) we can of course only draw a portion of the graph.
- Here are some examples of graphs of functions:



- For functions $f : A \to B$ defined on finite sets, or sets that do not consist of real numbers, the graph is typically either not useful, or not possible to draw sensibly. For this reason we also use "relation diagrams", in which we represent the sets A and B as collections of points and draw an arrow from $a \in A$ to $b \in B$ whenever $f(a) = b$.
- Here are function diagrams for $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$, and $f_3 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ from $A = \{1, 2, 3, 4\}$ to $B = \{1, 2, 3, 4\}$:



- An important property of a function is its set of "output values":
- Definition: If $f : A \to B$ is a function, the set of elements $b \in B$ for which there exists at least one $a \in A$ with $f(a) = b$ is called the image (or range) of f.
  - Terminology: Some authors use the word "range" as a synonym for "codomain", while others use it as synonym for "image". We will avoid using the word "range" for this reason.
  - Example: For the functions on $\{1, 2, 3, 4\}$ given by $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$, and $f_3 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$, the image of f_1 is $\{1, 2, 3, 4\}$, the image of f_2 is $\{1, 2, 3\}$, and the image of f_3 is $\{1, 2, 3, 4\}$.
  - The image of a function $f : A \to B$ is always a subset of the target set B, but need not be equal: for example, the image of f_2 above is only the set $\{1, 2, 3\}$ even though the target set is $\{1, 2, 3, 4\}$.
  - Example: The image of the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ is the set $\mathbb{R}_{\geq 0}$ of nonnegative real numbers.
- Since we view functions as relations, all of the operations we can perform with relations can also be performed on functions. One important operation is that of restricting a function to a smaller domain; since this operation on functions is particularly useful, we (re-)record the definition explicitly:
- Definition: If C is a subset of A and $f : A \to B$ is a function, the restriction of f to the domain C, denoted $f|_C$, is the function $f|_C : C \to B$ given by $f|_C = f \cap (C \times B)$.

[2] In fact, if $R : A \to B$ where A and B are both subsets of $\mathbb{R}$, we may actually draw the graph of the relation R, consisting of all points $(x, y) \in R$, although we will not need to invoke this idea.

- The ordered pairs in $f|_C$ are precisely those of the form (c, b) where $c \in C$ and $(c, b) \in f$: we can think of $f|_C$ as the function obtained by "throwing away" the information about the values on f on the elements of A not in C.
- Example: For $f : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$ with $f = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, the restriction of f to the domain $\{1, 3\}$ is the function $g : \{1, 3\} \to \{1, 2, 3, 4\}$ with $g = \{(1, 2), (3, 1)\}$.
- In the particular situation where f is defined using a rule, we simply use the same rule for $f|_C$ on the smaller domain C.
- Example: For $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$, we may restrict f to the positive real numbers to obtain a new function $g : \mathbb{R}_+ \to \mathbb{R}$ defined by $g(x) = x^2$.

- In some situations we can also restrict, or enlarge, the target set of a function.
  - Indeed, if $f : A \to B$ is a function with image $\text{im}(f)$, then we also have a function $g : A \to \text{im}(f)$ given by the same collection of ordered pairs, whose target set is now $\text{im}(f)$.
  - More generally, if C is any set with $\text{im}(f) \subseteq C$, we may also view the same collection of ordered pairs as yielding a function $h : A \to C$.
  - It is a matter of taste whether to consider this function h as being "the same as" f, since its underlying collection of ordered pairs, domain, and image are the same as f's. In practice, it is common to view this function as being equivalent to f, since it carries the same information.
  - However, we have adopted the convention that the domain and target are parts of the definition of a function, and so we would not consider h to be equal to f, since its target set is different.

3.4.2 Function Composition

- We now discuss ways of constructing new functions from other functions, of which the most fundamental is function composition.
  - Informally, if f and g are functions, the notation $f(g(x))$ is used to symbolize the result of applying f to the value $g(x)$. This operation is well-defined provided that the image of g is a subset of the domain of f.
  - We use the notation $f \circ g$ to refer to the composite function itself, so that $(f \circ g)(x) = f(g(x))$.
  - We may formalize this as follows:

- Definition: Let $g : A \to B$ and $f : B \to C$ be functions. Then the composite function $f \circ g : A \to C$ is defined by taking $(f \circ g)(a) = f(g(a))$ for all $a \in A$.
  - More explicitly, the ordered pairs in $f \circ g$ are those pairs $(a, c) \in A \times C$ for which there exists a $b \in B$ with $(a, b) \in g$ (so that $g(a) = b$) and with $(b, c) \in f$ (so that $f(b) = c$).
  - In symbolic language, $f \circ g = \{(a, c) \in A \times C : \exists b \in B, [(a, b) \in g)] \wedge [(b, c) \in f]\}$.

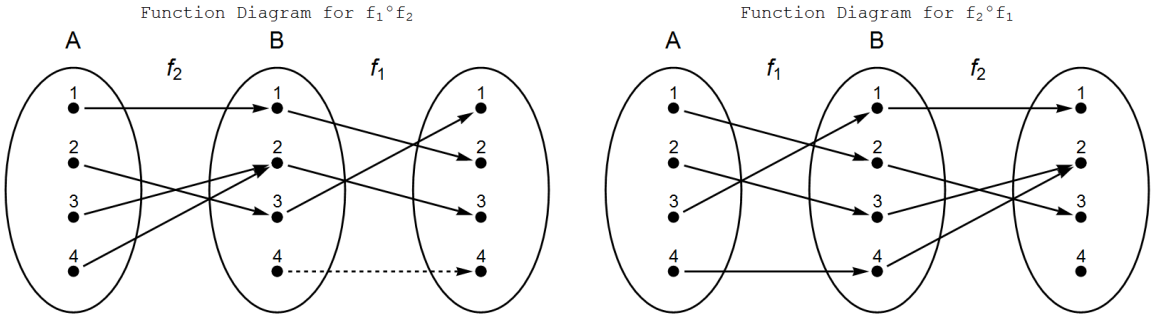
- In practice, if f and g are both described by rules, it is easiest to find compositions using the definition $(f \circ g)(a) = f(g(a))$.

- Example: Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be the functions $f(x) = x^2$ and $g(x) = 2x + 1$. Find $f \circ g$, $g \circ f$, $f \circ f$, and $g \circ g$.
  - We have $(f \circ g)(x) = f(g(x)) = f(2x+1) = (2x+1)^2$, and similarly $(g \circ f)(x) = g(f(x)) = g(x^2) = 2x^2+1$.
  - Also, $(f \circ f)(x) = f(f(x)) = f(x^2) = x^4$, and $(g \circ g)(x) = g(g(x)) = g(2x + 1) = 4x + 3$.

- Notice that the result of function composition depends on the order of the functions: in general, it will be the case that $f \circ g$ and $g \circ f$ are completely unrelated functions.
  - Indeed, depending on the domains and images of f and g, it is quite possible that one of $f \circ g$ is defined while the other is not.

- For example, suppose $f : \{1, 2\} \to \{a, b\}$ has $f(1) = a$ and $f(2) = b$, and $g : \{a, b\} \to \{3, 4\}$ has $g(a) = 3$ and $g(b) = 4$.
- Then the composite function $g \circ f$ exists and is a function from $\{1, 2\}$ to $\{3, 4\}$, where, specifically, we have $(g \circ f)(1) = g(f(1)) = g(a) = 3$, and $(g \circ f)(2) = g(f(2)) = g(b) = 4$.
- However, the composite function $f \circ g$ does not exist: the only possible elements in the domain are the elements in the domain of g, but if we try to evaluate $(f \circ g)(a)$, for example, we would have $(f \circ g)(a) = f(g(a)) = f(3)$, and this expression does not make sense because 3 is not in the domain of f. Similarly, $(f \circ g)(b) = f(g(b)) = f(4)$ also does not make sense.

- If f and g are given as sets of ordered pairs, we can use function diagrams to visualize and evaluate compositions: we draw the diagrams for the two functions together, and then follow the arrows from left to right.
  - For example, for the functions $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$ and $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$ on $\{1, 2, 3, 4\}$, here are composition diagrams for $f_1 \circ f_2$ and $f_2 \circ f_1$:



  - By following the arrows from left to right, we can see that if $g = f_1 \circ f_2$, then $g(1) = 2$, $g(2) = 1$, $g(3) = 3$, and $g(4) = 3$. Similarly, for $h = f_2 \circ f_1$, we have $h(1) = 3$, $h(2) = 2$, $h(3) = 1$, and $h(4) = 2$.

- As we have seen, function composition is not commutative. However, composition does satisfy some other algebraic properties:

- Proposition (Properties of Composition): Suppose A, B, C, D are sets.

1. Function composition is associative: If $f : C \to D$, $g : B \to C$, and $h : A \to B$ are any functions then $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are equal as functions from A to D.
   - Proof: Observe first that the domain of both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ is A, and the target of both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ is D.
   - Now let $a \in A$. Then by definition we have $[(f \circ g) \circ h](a) = [(f \circ g)](h(a)) = f(g(h(a)))$, and we also have $[f \circ (g \circ h)](a) = f[(g \circ h)(a)] = f(g(h(a)))$.
   - Since these two quantities are equal, we see $[(f \circ g) \circ h](a) = [f \circ (g \circ h)](a)$ for all $a \in A$.
   - Hence the functions $(f \circ g) \circ h$ and $f \circ (g \circ h)$ have the same domain and target, and take the same value at every element of their common domain, so they are the same function.
2. The identity function behaves as a left and right identity: For any $f : A \to B$, $f \circ i_A = f$ and $i_B \circ f = f$.
   - Proof: Observe that the domain of $f \circ i_A$ is A and the target is B, the same as for f.
   - Then for any $a \in A$ we have $(f \circ i_A)(a) = f(i_A(a)) = f(a)$, and so we see $f \circ i_A$ and f take the same value at every point of their shared domain. Hence they are equal as functions.
   - In the same way, the domain of $i_B \circ f$ is A and the target is B, the same as for f.
   - Then for any $a \in A$ we have $(i_B \circ f)(a) = i_B(f(a)) = f(a)$, and so we see $i_B \circ f$ and f take the same value at every point of their shared domain. Hence they are equal as functions.

3.4.3 Inverses of Functions, One-to-One and Onto Functions

- Next we examine inverses of functions.

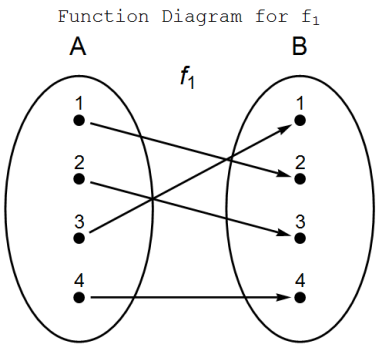
- Under the common interpretation of a function f as a "machine" that operates on an input value to produce an output value, the inverse f^{-1} would correspond to a machine that inverts this process, taking an output value of f and giving the corresponding input value.
- In particular, if $f : A \to B$, then we would like to have $f^{-1} : B \to A$, and on the level of ordered pairs, if $(a, b) \in f$, then we would like $(b, a) \in f^{-1}$.
- Indeed, we have already defined an object with this exact property, namely, the inverse relation to f.
- However, if $f : A \to B$ is an arbitrary function, the inverse relation f^{-1} need not be a function from B to A.
- For example, suppose $f : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ is the function with $f(1) = 2$, $f(2) = 4$, and $f(3) = 2$, so that as a set of ordered pairs, $f = \{(1, 2), (2, 4), (3, 2)\}$.
- Then the inverse relation is $f^{-1} = \{(2, 1), (4, 2), (2, 3)\} = \{(2, 1), (2, 3), (4, 2)\}$. However, f^{-1} is not a function (on any domain) because it contains the ordered pairs $(2, 1)$ and $(2, 3)$, meaning that f^{-1} is not well-defined on the element 2.
- It is easy to identify the difficulty here: the problem is that f maps both 1 and 3 to 2, so we cannot assign a unique value to $f^{-1}(2)$ since we want it to equal both 1 and 3.
- As another example, suppose $g : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ is the function with $g(1) = 2$, $g(2) = 4$, and $g(3) = 1$.
- Then $g = \{(1, 2), (2, 4), (3, 1)\}$ so $g^{-1} = \{(2, 1), (4, 2), (1, 3)\} = \{(1, 3), (2, 1), (4, 2)\}$. We can see that g^{-1} is indeed a function, but it is a function from $\{1, 2, 4\} \to \{1, 2, 3\}$, not a function from $\{1, 2, 3, 4\} \to \{1, 2, 3\}$.
- In this case, we see that the inverse relation to $g : A \to B$ is not a function $g^{-1} : B \to A$ from B to A, but rather a function $g^{-1} : \text{im}(g) \to A$ from the image of g to A.
- We can clarify this behavior by identifying the precise characteristics of the functions that cause these behaviors:

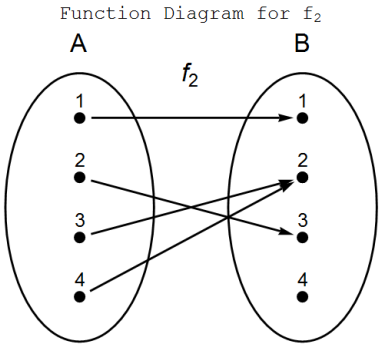
- Definition: The function $f : A \to B$ is one-to-one (or injective) if for any $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ implies $a_1 = a_2$.
  - Equivalently, $f : A \to B$ is one-to-one when $a_1 \neq a_2$ implies $f(a_1) \neq f(a_2)$, which is the same as saying that f maps unequal elements in its domain to unequal elements in its image.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x - 4$ is one-to-one, because $f(a_1) = f(a_2)$ implies $3a_1 - 4 = 3a_2 - 4$, and this only occurs when $a_1 = a_2$.
  - Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not one-to-one, because $f(2) = 4 = f(-2)$.
  - Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is one-to-one, because $f(a_1) = f(a_2)$ implies $2a_1 = 2a_2$, which only occurs for $a_1 = a_2$.
  - Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \sin(x)$ is not one-to-one, because $f(0) = 0 = f(\pi)$.
  - Example: If $A \subseteq B$, then the inclusion map $\iota : A \to B$ given by $\iota(a) = a$ for all $a \in A$ is one-to-one.
  - We will remark that the property of being one-to-one depends on the domain of f, although not on the target set. For example, the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ is not one-to-one, but its restriction $f|_{\mathbb{R}+}$ to the positive real numbers is one-to-one.
  - In general, any restriction of a one-to-one function to a smaller domain will still be one-to-one, since if $C \subseteq A$ then if $c_1, c_2 \in C$ with $f|_C(c_1) = f|_C(c_2)$ then by definition $f(c_1) = f(c_2)$ and so $c_1 = c_2$.

- Definition: The function $f : A \to B$ is onto (or surjective) if $\text{im}(f) = B$.
  - Equivalently, $f : A \to B$ is onto when for any $b \in B$, there exists an $a \in A$ with $f(a) = b$.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x - 4$ is onto, because for any $b \in \mathbb{R}$, there exists an $a \in \mathbb{R}$ with $f(a) = b$, namely, $a = (b + 4)/3$, as can be found by solving the equation $3a - 4 = b$ for a.
  - Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not onto, because there is no $a \in \mathbb{R}$ such that $f(a) = -1$.

- Non-Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is not onto, because there is no $a \in \mathbb{Z}$ with $2a = 1$.
- Example: The function $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(x) = e^x$ is onto, because for any $b \in \mathbb{R}_+$, there exists an $a \in \mathbb{R}$ with $f(a) = b$, namely, $a = \ln(b)$, since in such a case we have $f(\ln(b)) = e^{\ln(b)} = b$.
- We will remark that the property of being onto requires explicitly knowing the target set for f. Every function is surjective onto its image, but is not surjective onto any (strictly) larger set.

- Using function diagrams, it is easy to see visually whether a function is one-to-one or onto:





- For the functions shown above from $A = \{1, 2, 3, 4\}$ to $B = \{1, 2, 3, 4\}$, we can see that f_1 is one-to-one since no two arrows land at the same point in the target, and f_1 is onto since every point in the target has at least one arrow pointing to it.
- On the other hand, f_2 is not one-to-one because it has two arrows pointing to 2, and it is not onto because it has no arrow pointing to 4.

- We can now establish the precise relationship between being one-to-one (or onto) and the existence of an inverse function:

- Proposition (One-to-One, Onto, and Inverses): Suppose $f : A \to B$ is a function.

1. The inverse relation f^{-1} is a function (from $\text{im}(f)$ to A) if and only if f is one-to-one.
   - Proof: Note that f^{-1} is a function precisely when $(c, a) \in f^{-1}$ and $(c, b) \in f^{-1}$ implies $a = b$.
   - This condition is equivalent to saying that if $(a, c) \in f$ and $(b, c) \in f$ then $a = b$, which is in turn equivalent to saying that if $f(a) = c = f(b)$ then $a = b$. But this last condition is precisely the same as saying f is one-to-one.
2. If $f^{-1} : B \to A$ is a function, then $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.
   - Proof: For the first statement, note that $f^{-1} \circ f$ is a function from A to A.
   - Now let $a \in A$ be arbitrary and set $b = f(a) \in B$. Then $(a, b) \in f$ so $(b, a) \in f^{-1}$, meaning that $f^{-1}(b) = a$.
   - Now we compute $(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a$ by the above.
   - But since a was arbitrary, and $f^{-1} \circ f$ and i_A have the same domain and target and take the same values for all $a \in A$, they are equal as functions.
   - The argument to see that $f \circ f^{-1} = i_B$ is similar: as above note $f \circ f^{-1}$ and i_B have the same domain and target.
   - Now let $b \in B$ be arbitrary and set $a = f^{-1}(b) \in A$. Then $(b, a) \in f^{-1}$ and so $(a, b) \in f$.
   - We compute $(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b$, so since b was arbitrary, $f \circ f^{-1}$ and i_B are equal as functions.
3. If there exists a function $g : B \to A$ such that $g \circ f = i_A$, then f is one-to-one.
   - Proof: Suppose $g : B \to A$ has $g \circ f = i_A$ and that $f(a_1) = f(a_2)$.
   - Then $a_1 = i_A(a_1) = (g \circ f)(a_1) = g(f(a_1)) = g(f(a_2)) = (g \circ f)(a_2) = i_A(a_2) = a_2$, so f is one-to-one.
4. If there exists a function $g : B \to A$ such that $f \circ g = i_B$, then f is onto.
   - Proof: Suppose $g : B \to A$ has $f \circ g = i_B$ and let $b \in B$ be arbitrary.

- Then $b = i_B(b) = (f \circ g)(b) = f(g(b))$, meaning that if we set $a = g(b)$, then we have $f(a) = b$, so f is onto.

- By combining all of these observations we can give several equivalent characterizations of when a function has an inverse function:

- Theorem (Inverse Functions): Suppose $f : A \to B$ is a function. Then the following are equivalent:
  1. f is one-to-one and onto.
  2. f^{-1} is a function from B to A.
  3. There exists a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$.

  - Proof: We show that (1) implies (2), that (2) implies (3), and that (3) implies (1). This is sufficient because the other implications (such as (1) implies (3)) follow from these three because logical implication is transitive as we have previously noted.
  - (1) $\Rightarrow$ (2): If f is one-to-one, then f^{-1} is a function from $\text{im}(f)$ to A by result (1) from the proposition above. If f is also onto, then $\text{im}(f) = B$, and so f^{-1} is a function from B to A.
  - (2) $\Rightarrow$ (3): If f^{-1} is a function from B to A, then simply take $g = f^{-1}$; by result (2) from the proposition above, $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$ as required.
  - (3) $\Rightarrow$ (1): If there exists a function $g : B \to A$ such that $g \circ f = i_A$, then by result (3) from the proposition above, we see f is one-to-one. If g also has the property that $f \circ g = i_B$, then by result (4) from the proposition above, we see f is also onto.

- We can also deduce that (when it exists) the inverse function is the unique two-sided inverse of f:

- Corollary (Uniqueness of Inverse): Suppose $f : A \to B$ and $g : B \to A$ are functions such that $g \circ f = i_A$ and $f \circ g = i_B$. Then $g = f^{-1}$.

  - Proof: If there exists such a function g, then by the theorem above, f^{-1} is a function from B to A and it satisfies the same properties as g.
  - Then by the basic properties of function composition, we can write $g = i_A \circ g = (f^{-1} \circ f) \circ g = f^{-1} \circ (f \circ g) = f^{-1} \circ i_B = f^{-1}$, as required.

- The actual calculation of the inverse function, when it exists, is trivial when f is described as a list of ordered pairs, since f^{-1} is obtained simply by reversing all of the pairs.

  - When f is described as a rule (typically, for functions written algebraically), to find the inverse we simply solve the equation $y = f(x)$ for x in terms of y: this will give $x = f^{-1}(y)$.

- Example: Verify that the function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = 3x - 2$ is invertible and find its inverse function.

  - To show that h is one-to-one, notice that $h(a) = h(b)$ is the same as $3a - 2 = 3b - 2$, and this can easily be rearranged to obtain $a = b$.
  - To find h^{-1}, we solve $y = 3x - 2$ for x in terms of y. We obtain $x = \dfrac{y+2}{3}$, so $h^{-1}(y) = \dfrac{y+2}{3}$. Note that this calculation also shows that h is onto.

- In the example above, notice h is a composite function: h scales its argument by 3 and then subtracts 2.

  - Its inverse function reverses each of these operations in the opposite order: namely, h^{-1} first adds 2 and then divides its argument by 3.
  - The observation in this example holds in general:

- Proposition (Properties of Inverses): If $f : B \to C$ and $g : A \to B$ are invertible functions, then so are f^{-1} and $f \circ g$, and $(f^{-1})^{-1} = f$ and $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

  - Proof: By our theorem on invertible functions, to show two functions are inverses we need only verify that composing them in either order yields the appropriate identity function.

- For f and f^{-1} we have $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$, meaning that f fills the role of the inverse function $(f^{-1})^{-1}$. So f^{-1} is invertible and its inverse is f, as claimed.
- For $f \circ g$ and $g^{-1} \circ f^{-1}$, first observe that $[f \circ g] \circ [g^{-1} \circ f^{-1}] = f \circ [g \circ g^{-1}] \circ f^{-1} = f \circ i_B \circ f^{-1} = f \circ f^{-1} = i_C$.
- Likewise, $[g^{-1} \circ f^{-1}] \circ [f \circ g] = g^{-1} \circ [f^{-1} \circ f] \circ g = g^{-1} \circ i_B \circ g = g^{-1} \circ g = i_A$.
- Hence $f \circ g$ is invertible and its inverse is $g^{-1} \circ f^{-1}$, as claimed.

3.4.4 Bijections

- As we have already seen, functions that are both one-to-one and onto have convenient properties. We now analyze these functions in a bit more detail.
- Definition: A function that is both one-to-one and onto is called a bijection.
  - From our results on inverses, $f : A \to B$ is equivalently a bijection when it has an inverse function $f^{-1} : B \to A$.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x - 4$ is a bijection because it is both one-to-one and onto.
  - Example: The function $g : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ given by $g(x) = \tan(x)$ is a bijection because it is both one-to-one and onto.
  - Non-Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 3n-4$ is not a bijection: although it is one-to-one, it is not onto since there is no n with $f(n) = 0$.
  - Non-Example: The function $h : [0, \pi] \to [-1, 1]$ given by $g(x) = \sin(x)$ is not a bijection: although it is onto, it is not one-to-one since for example $h(0) = h(\pi)$.
  - Non-Example: The function $k : \mathbb{Q} \backslash \{0\} \to \mathbb{Q}$ given by $k(x) = 1/x$ is not a bijection: although it is one-to-one, it is not onto because there is no x for which $k(x) = 0$.
- If $f : A \to B$ is a bijection, it establishes a one-to-one correspondence between the elements of A and the elements of B: to each element $a \in A$, f associates a unique element of B, namely $f(a)$, and to each element $b \in B$, f associates a unique element of A, namely $f^{-1}(b)$.
  - We may think of f as being a "relabeling": if we relabel the elements of the set A by applying f to them, then the result is the set B.
  - In general, if there exists a bijection $f : A \to B$, we say that A and B are in one-to-one correspondence. This property is, in fact, an equivalence relation:
- Proposition (One-to-One Correspondences): Suppose A, B, and C are sets.
  1. The identity function $i_A : A \to A$ is a bijection from A to A.
     - Proof: The identity function is self-evidently one-to-one and onto (alternatively, it is its own inverse).
  2. If $f : A \to B$ is a bijection, then its inverse $f^{-1} : B \to A$ is also a bijection.
     - Proof: If $f : A \to B$ is a bijection, then from our results on inverses we know that $f^{-1} : B \to A$ is a function.
     - Furthermore, since $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$, we see that f^{-1} is invertible with inverse f, and therefore $f^{-1} : B \to A$ is also a bijection.
  3. If $f : B \to C$ and $g : A \to B$ are bijections, then $f \circ g : A \to C$ is also a bijection.
     - Proof: If $f : B \to C$ and $g : A \to B$ are bijections, then from our results on inverses we know that $f^{-1} : B \to A$ and $g^{-1} : C \to B$ are functions.
     - Also, we know that $f \circ g : A \to C$ is invertible with inverse $(f \circ g)^{-1} = g^{-1} \circ f^{-1} : C \to A$, so it is also a bijection.
  4. The relation on sets defined by $A \sim B$ when there exists a bijection $f : A \to B$ is an equivalence relation.
     - Proof: This follows immediately from (1)-(3): (1) shows reflexivity, (2) shows symmetry, and (3) shows transitivity.

5. If $f : A \to B$ is a bijection and $g : C \to D$ is a bijection, then the "product map" $f \times g : (A \times C) \to (B \times D)$ given by $(f \times g)(a, c) = (f(a), g(c))$ is also a bijection.
    - Proof: If $(f \times g)(a_1, c_1) = (f \times g)(a_2, c_2)$ then by definition $(f(a_1), g(c_1)) = (f(a_2), g(c_2))$ which is the same as saying $f(a_1) = f(a_2)$ and $g(c_1) = g(c_2)$.
    - Then because f and g are both one-to-one, we see $a_1 = a_2$ and $c_1 = c_2$, so $(a_1, c_1) = (a_2, c_2)$. Hence $f \times g$ is one-to-one.
    - Also, if $b \in B$ and $d \in D$, then because f and g are both onto, there exist $a \in A$ and $c \in C$ with $f(a) = b$ and $g(c) = d$. Then $(f \times g)(a, c) = (f(a), g(c)) = (b, d)$, so $f \times g$ is onto and thus a bijection.

- Example: Determine whether $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is a bijection.
    - Although f is onto, f is not one-to-one since for example $f(-1) = 1 = f(1)$, so f is not a bijection.

- Example: Determine whether $f : \mathbb{R}_+ \to \mathbb{R}_+$ given by $f(x) = x^2$ is a bijection.
    - This function is a bijection: it is one-to-one since $x^2 = y^2$ with x, y positive can only occur for $x = y$, and it is onto since every positive real number has a positive real square root.
    - Equivalently, we could observe that f has an inverse function $f^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ given by $f^{-1}(x) = \sqrt{x}$.

- Example: Show that there is a bijection between $\mathbb{R}$ and $\mathbb{R}_+$.
    - We claim that the exponential function $f : \mathbb{R} \to \mathbb{R}_+$ with $f(x) = e^x$ is a bijection. To see this, simply observe f has an inverse function $f^{-1} : \mathbb{R}_+ \to \mathbb{R}$ given by the natural logarithm $f^{-1}(x) = \ln x$.

3.5 Cardinality and Countability

- In the previous section, we discussed how bijections provide a one-to-one correspondence between the elements of the domain with the elements of the target. Our goal now is to use bijections to give a formal treatment of the notion of cardinality, which provides a way of measuring the "size" of a set, with a primary focus on infinite sets.

- Quite surprisingly, as we will show, there are actually many different cardinalities that infinite sets can have.

3.5.1 Cardinality

- We start by observing that the process of counting the elements of a finite set A is the same as labeling the elements of A with the positive integers $1, 2, 3, \ldots, n$.
    - By our interpretation of a bijection as a relabeling, this is the same as giving a bijection between A and the set $\{1, 2, 3, \ldots, n\}$.
    - We can use this idea to give a formal definition of the cardinality of a finite set:

- Definition: If A is a set and n is a nonnegative integer, we say the cardinality of A is n (written $\#A = n$) if there exists a bijection between A and the set $\{1, 2, 3, \ldots, n\}$. If there exists an integer n such that the cardinality of A is n, we say A is a finite set, and otherwise we say A is an infinite set.
    - We take the usual convention that if $n = 0$ the set written as $\{1, 2, 3, \ldots, n\}$ means the empty set, and so the cardinality of $\emptyset$ is 0.
    - We must verify that this definition is well-posed, in the sense that for any finite set A, there is a unique positive integer n for which there exists a bijection between A and $\{1, 2, 3, \ldots, n\}$.
    - If there were bijections between A and $\{1, 2, 3, \ldots, n\}$, and also between A and $\{1, 2, 3, \ldots, m\}$, then since one-to-one correspondence is an equivalence relation, this would give a bijection between $\{1, 2, 3, \ldots, n\}$ and $\{1, 2, 3, \ldots, m\}$.

- However, such a bijection cannot exist unless $m = n$, as can be tediously[3] verified with induction.

- We have various other basic properties of cardinality:

- Proposition (Properties of Cardinality): Suppose A and B are sets.

1. If A is finite and $B \subseteq A$, then B is finite, and $\#B \leq \#A$ with equality if and only if $B = A$.
   - Proof: Induction on $n = \#A$. The base case $n = 0$ is trivial, since in that case $A = B = \emptyset$ so $\#A = \#B = 0$, and we have equality.
   - For the inductive step, suppose $\#A = n$ with $n \geq 1$. If $B = A$ the result is trivial so suppose B is a proper subset of A.
   - Since $\#A = n$ there exists a bijection $f : A \to \{1, 2, \ldots, n\}$. Then the set $f(B) = \{f(b) : b \in B\}$ is a proper subset of $\{1, 2, \ldots, n\}$ since B is a proper subset of A and f is a bijection. Restricting f to $f|_B$ yields a bijection of B with this proper subset, which must have cardinality k for some $k < n$. Then by relabeling the elements of this subset as $\{1, 2, \ldots, k\}$ we see that $\#B = k < n = \#A$, as required.
2. If A and B are finite and disjoint, then $\#(A \cup B) = \#A + \#B$.
   - Proof: Suppose $\#A = n$ and $\#B = m$ and let $f : A \to \{1, 2, \ldots, n\}$ and $g : B \to \{1, 2, \ldots, m\}$ be bijections. Then the function $h : \{1, 2, \ldots, m+n\} \to A \cup B$ with $h(k) = \begin{cases} f(k) & \text{for } 1 \leq k \leq m \\ g(k-n) & \text{for } m+1 \leq k \leq m+n \end{cases}$ is also a bijection, so $\#(A \cup B) = m + n = \#A + \#B$.
3. If A is finite, then for any B we have $\#(A \backslash B) = \#A - \#(A \cap B)$.
   - Proof: By (1) we see that $A \backslash B$ and $A \cap B$ are finite since they are both subsets of A. Since they are also disjoint and have union A, by (2) we have $\#A = \#(A \backslash B) + \#(A \cap B)$ which yields the desired result immediately.
4. If A and B are finite, then $\#(A \cup B) = \#A + \#B - \#(A \cap B)$.
   - Proof: Let $C = A \backslash B = \{x \in A : x \notin B\}$ and observe that $C \cup B = A \cup B$ and that C and B are disjoint. Then by (2) and (3) we have $\#(A \cup B) = \#(C \cup B) = \#C + \#B = \#A + \#B - \#(A \cap B)$ as claimed.
   - Remark: This result generalizes inductively to larger unions, yielding a general statement that is known as the inclusion-exclusion formula. For example, for three sets one obtains $\#(A \cup B \cup C) = \#A + \#B + \#C - \#(A \cap B) - \#(A \cap C) - \#(B \cap C) + \#(A \cap B \cap C)$.
5. If A and B are finite, then $\#(A \times B) = \#A \cdot \#B$.
   - Proof: Suppose $\#A = n$ and $\#B = m$ and let $f : \{1, 2, \ldots, n\} \to A$ and $g : \{1, 2, \ldots, m\} \to B$ be bijections. Then the function $h : \{1, 2, \ldots, mn\} \to A \times B$ defined by taking $h(a + n(b-1)) = (f(a), g(b))$ for $1 \leq a \leq n$ and $1 \leq b \leq m$ is also a bijection, so $\#(A \times B) = mn = \#A \cdot \#B$ as claimed.
   - Remark: This result generalizes inductively to larger Cartesian products. For example, for three sets one obtains $\#(A \times B \times C) = \#A \cdot \#B \cdot \#C$.
6. If A is infinite and $A \subseteq B$, then B is infinite. In particular, $A \cup C$ is infinite precisely when A or C is infinite.
   - Proof: Suppose $A \subseteq B$. By (1), if B is finite, then A is finite, so taking the contrapositive yields that if A is infinite, then B is infinite.
   - For the second part, if A or C is infinite, then since each is a subset of the union $A \cup C$, the union is infinite. Otherwise, if both A and C are finite, then by (4) so is $A \cup C$.

[3] For completeness: without loss of generality assume $n \leq m$, and induct on n. The base case $n = 0$ follows by observing that the only function from the empty set is the empty function (with image the empty set) so necessarily $m = 0$ also.

For the inductive step, assume that having a bijection from $\{1, 2, 3, \ldots, n\}$ to $\{1, 2, 3, \ldots, m\}$ for $m = k$ implies $n = k$, and suppose we have a bijection from $\{1, 2, 3, \ldots, n\}$ to $\{1, 2, 3, \ldots, m\}$ where $m = k+1$. If we have a bijection $f : \{1, 2, \ldots, k+1\} \to \{1, 2, \ldots, n\}$, then let $g = f|_{\{1,2,\ldots k\}}$ be the restriction of f to $\{1, 2, \ldots, k\}$ and observe that the image of g is the set $\{1, 2, \ldots, n\}$ with one element removed. So then g is a bijection between $\{1, 2, \ldots, k\}$ and its image, which (by relabeling) is in turn in bijection with the set $\{1, 2, \ldots, n-1\}$. Hence by the inductive hypothesis, we see $k = n - 1$, and so $m = k + 1 = n$ as claimed.

7. If A is infinite and B is nonempty, then $A \times B$ is infinite.
   - We remark that $A \times \emptyset = \emptyset$, so the hypothesis that B be nonempty is needed here for $A \times B$ to be infinite.
   - Proof: Suppose A is infinite and $x \in B$. Then $A \times B$ contains the subset $A \times \{x\}$, which is in bijection with the infinite set A hence is also infinite. Then by (6), we see $A \times B$ is infinite.
8. If $f : A \to B$ is one-to-one, then $\#A = \#\text{im}(f)$.
   - Proof: Observe that f gives a bijection of A with $\text{im}(f)$, since $f : A \to \text{im}(f)$ is one-to-one by hypothesis, and is also onto by definition of $\text{im}(f)$.
   - Since bijections preserve cardinality that means $\#\text{im}(f) = \#A$, as claimed.
9. If A and B are finite and have same cardinality, then a function $f : A \to B$ is one-to-one if and only if it is onto, if and only if it is a bijection.
   - Proof: Suppose $f : A \to B$ is one-to-one and $\#A = \#B$. Then by (8), f is a bijection of A with $\text{im}(f)$, so $\#\text{im}(f) = \#A = \#B$. But then because B is finite, and $\text{im}(f) \subseteq B$, we have $\text{im}(f) = B$ by (1). Hence f is onto.
   - Conversely, suppose $f : A \to B$ is onto, and define $S_b = \{a \in A : f(a) = b\}$ for each $b \in B$. Then the S_b are pairwise disjoint sets (if $x \in S_b \cap S_{b'}$ then $b = f(x) = b'$) such that $A = \cup_{b \in B} S_b$ (since any $a \in A$ lies in $S_{f(a)}$) and $\#S_b \geq 1$ for each $b \in B$ (since f is onto).
   - Then by repeatedly applying (2) we see that $\#A = \#S_1 + \#S_2 + \cdots + \#S_{\#B}$ and by summing we also have $\#B \leq \#S_1 + \cdots + \#S_{\#B}$ since each of the sizes is at least 1.
   - But then we have $\#A = \#B \leq \#S_1 + \cdots + \#S_{\#B} = \#A$, meaning that we must have equality in the middle, and so $\#S_b = 1$ for each $b \in B$. That means f is one-to-one.
   - So we deduce that f is one-to-one if and only if f is onto. This means either condition is equivalent to both, which is to say, either condition is equivalent to saying f is a bijection.

- We can use these results to establish two fundamental counting principles, as follows:
  - ("Addition Principle") When choosing among n disjoint options labeled 1 through n, if option i has a_i possible outcomes for each $1 \leq i \leq n$, then the total number of possible outcomes is $a_1 + a_2 + \cdots + a_n$.
  - To illustrate the addition principle, if a restaurant offers 5 main courses with chicken, 6 main courses with beef, and 12 vegetarian main courses, then (presuming no course is counted twice) the total possible number of main courses is $5 + 6 + 12 = 23$.
  - The addition principle can be justified using our results about cardinalities of unions of disjoint sets: if A_i corresponds to the set of outcomes of option i, then the union $A_1 \cup A_2 \cup \cdots \cup A_n$ corresponds to a single choice of one outcome from one of the A_i. Then because all of the different options are disjoint, the number of such choices is $\#(A_1 \cup A_2 \cup \cdots \cup A_n) = \#A_1 + \#A_2 + \cdots + \#A_n$ by repeatedly applying (2).
  - ("Multiplication Principle") When making a sequence of n independent choices, if step i has b_i possible outcomes for each $1 \leq i \leq n$, then the total number of possible collections of choices is $b_1 \cdot b_2 \cdot \cdots \cdot b_n$.
  - To illustrate the multiplication principle, if a fair coin is tossed (2 possible outcomes) and then a fair 6-sided die is rolled (6 possible outcomes), the total number of possible results of flipping a coin and then rolling a die is $2 \cdot 6 = 12$.
  - The multiplication principle follows from our results about cardinalities of Cartesian products: if B_i corresponds to the set of outcomes of choice i, then the elements of the Cartesian product $B_1 \times B_2 \times \cdots \times B_n$ correspond to ordered n-tuples of outcomes, one for each choice. The number of such n-tuples is $\#(B_1 \times B_2 \times \cdots \times B_n) = \#B_1 \cdot \#B_2 \cdot \cdots \cdot \#B_n$ by repeatedly applying (5).

- By employing these principles appropriately, we can solve a variety of basic counting problems.

- Example: Determine the number of possible outcomes from rolling a 6-sided die 5 times in a row.
  - Each individual roll has 6 possible outcomes. Thus, by the multiplication principle, the number of possible sequences of 5 rolls is $6^5 = \boxed{7776}$.

- Example: Determine the number of subsets of the set $\{1, 2, \ldots, n\}$.

- We may characterize a subset S of $\{1, 2, \ldots, n\}$ by listing, for each $k \in \{1, 2, \ldots, n\}$, whether $k \in S$ or $k \notin S$.
- By the multiplication principle, the number of possible ways of making this sequence of n choices is $\boxed{2^n}$.

- Example: If $\#A = n$ and $\#B = m$, find the total number of functions $f : A \to B$.
  - If $A = \{a_1, a_2, \ldots, a_n\}$, then a function $f : A \to B$ is characterized by the values $f(a_1)$, $f(a_2)$, ... , $f(a_n)$.
  - Since $\#B = m$, there are m possible choices for each of the n values $f(a_1)$, $f(a_2)$, ... , $f(a_n)$.
  - Since all such choices are allowed, the total number of functions is therefore $\boxed{m^n}$.

- Example: Find the number of positive integer divisors of 90000.
  - Note that $90000 = 2^4 3^2 5^4$, so any positive integer divisor must have the form $2^a 3^b 5^c$ where $a \in \{0, 1, 2, 3, 4\}$, $b \in \{0, 1, 2\}$, and $c \in \{0, 1, 2, 3, 4\}$.
  - On the other hand, every such integer is a divisor, and so since there are 5 choices for a, 3 for b, and 5 for c, there are $5 \cdot 3 \cdot 5 = \boxed{75}$ divisors in total.
  - Remark: In the same way, one may see that $n = 2^{n_2} 3^{n_3} 5^{n_5} \cdots$ has a total of $(n_2 + 1)(n_3 + 1)(n_5 + 1) \cdots$ positive integer divisors.

3.5.2 Countable and Uncountable Sets

- Because we have defined cardinality in terms of bijections, and the property of being in a one-to-one correspondence is an equivalence relation on sets, we see that there is a bijection between two finite sets if and only if they have the same cardinality.
  - This gives us an alternative way to view cardinalities, namely, as representing the equivalence classes of sets under the relation of being in one-to-one correspondence.
  - For example, one equivalence class contains the sets $\{1, 2\}$, $\{1, 5\}$, $\{22, \pi\}$, $\{A, B\}$, $\{\star, \text{potato}\}$, ... , since any two of these sets are in one-to-one correspondence with one another. This equivalence class may be thought of as being the collection of all sets of cardinality 2.
  - The advantage of this approach to cardinality is that it also extends to infinite sets:

- Definition: We say two sets are equinumerous (or equipollent) if there exists a bijection between them.
  - Example: The sets $\{1, 2, 3\}$ and $\{a, b, Q\}$ are equinumerous because there exists a bijection between them, namely, the function $f = \{(1, a), (2, b), (3, Q)\}$.
  - Example: The sets $\mathbb{Z}$ and $2\mathbb{Z}$ (the even integers) are equinumerous because there exists a bijection between them, namely, the function $f : \mathbb{Z} \to 2\mathbb{Z}$ given by $f(n) = 2n$ (it is easy to see that f is one-to-one and onto).
  - We think of two equinumerous sets as having the same cardinality: from our observations above, this interpretation agrees with the definition of cardinality for finite sets.
  - It is somewhat strange to think of the set of even integers as having the same cardinality as the set of all integers, because the set of even integers is a proper subset of the set of all integers (indeed, in some sense[4] only "half" of all integers are even). But this is the type of statement we must accept if we are to give any sensible definition for the cardinality of an infinite set that behaves well under set operations.

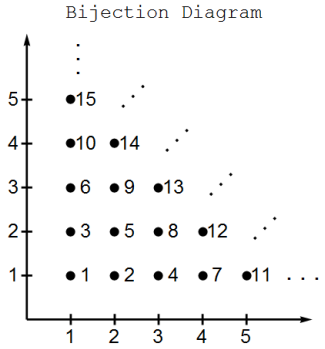
[4] One may make precise the idea that half of all integers are even by noting that if E is the set of even integers, then the limit $\lim_{N\to\infty} \frac{\#[E \cap \{-N, \ldots, N\}]}{\#[\{-N, \ldots, N\}]}$ is equal to $\frac{1}{2}$. Equivalently, the proportion of the integers in $\{-N, -N+1, \ldots, N-1, N\}$ that are even approaches $1/2$ as $N \to \infty$). In general, if S is a subset of the integers, its "natural density" is defined as the limit $\lim_{N\to\infty} \frac{\#[S \cap \{-N, \ldots, N\}]}{\#[\{-N, \ldots, N\}]}$, if the limit exists; note that there do exist sets whose natural density is undefined, such as the set of integers with leading digit 1 (in base 10).

- Example: The sets $\mathbb{Z}$ and $\mathbb{Z}_{>0}$ (the positive integers) are equinumerous, because the function $f : \mathbb{Z} \to \mathbb{Z}_{>0}$ given by $f(n) = \begin{cases} 2n+2 & \text{if } n \geq 0 \\ -2n+1 & \text{if } n < 0 \end{cases}$ is a bijection, since it maps the nonnegative integers to the even positive integers and it maps the negative integers to the odd positive integers.
- Example: The sets $\mathbb{Z}_{>0}$ (the positive integers) and the set S of perfect squares are equinumerous, because the function $f : \mathbb{Z}_{>0} \to S$ given by $f(n) = (n-1)^2$ is a bijection.

- As we have noted above, counting elements of a set is the same as assigning positive integer labels to the elements of the set, which is in turn the same as creating a bijection with a subset of the positive integers.
- Definition: If S is a set, we say S is countable if there exists a bijection between S and a subset of the positive integers, and we say S is countably infinite if S is countable and infinite. If S is not countable, we say S is uncountable.
  - By definition, any finite set is countable since it can be put in bijection with the set $\{1, 2, 3, \ldots, n\}$ where n is its cardinality.
- Proposition (Properties of Countability): The following are true:

1. If S is a countably infinite subset of the positive integers, there exists a bijection between S and $\mathbb{Z}_{>0}$.
   - Intuitively, we can just define the bijection by mapping 1 to the smallest element of S, 2 to the second smallest, and so forth.
   - Proof: By the well ordering axiom, since S is nonempty it has a smallest element a_1.
   - Since S is infinite, $S \backslash \{a_1\}$ is also infinite hence nonempty, so it has a smallest element $a_2 > a_1$.
   - By a trivial induction, we may continue this process for each positive integer $n \geq 1$ to construct $a_n > a_{n-1} > \cdots > a_1$ where $S \backslash \{a_1, \ldots, a_n\}$ is infinite and has all elements greater than a_n. Since the a_i are all distinct positive integers in increasing order, we also see that $a_n \geq n$ for each n.
   - Setting $f(n) = a_n$ then yields a one-to-one function $f : \mathbb{Z}_{>0} \to S$. But f is also onto, since any $k \in S$ will be the smallest element of $S \backslash \{1, 2, \ldots, k-1\}$ hence necessarily is among the values $f(1), \ldots, f(k)$.
2. More generally, any subset of a countable set is countable.
   - Proof: Suppose A is countable and $B \subseteq A$. Then by definition there is a bijection $f : A \to Z$ with a subset Z of the positive integers.
   - The restriction $f|_B$ is a then bijection from B to $\text{im}(f|_B) \subseteq Z$, which is also a subset of the positive integers.
   - Hence there is a bijection from B to a subset of the positive integers, so B is countable.
3. A nonempty set S is countable if and only if there exists an onto function $f : \mathbb{Z}_{>0} \to S$.
   - The utility of this result is that it provides an easier way to establish countability, since onto maps are less restrictive and thus easier to construct than bijections.
   - Proof: Suppose S is nonempty. If there exists an onto function $f : \mathbb{Z}_{>0} \to S$, let n_x be the smallest positive integer such that $f(n_x) = x$. (Note that this integer necessarily exists by applying the well-ordering axiom to the set of integers f maps to x which is nonempty since f is onto.)
   - Then for $A = \{n_x : x \in S\}$, we see that $f|_A$ is a bijection (since it is onto and also one-to-one) with the subset A of $\mathbb{Z}_{>0}$ with S, so S is countable.
   - Conversely, suppose S is countable and nonempty, so that there exists a bijection $g : A \to S$ where A is a subset of the positive integers. Let $x \in S$ (here is where we are using the fact that S is nonempty), and then define $f : \mathbb{Z}_{>0} \to S$ via $f(n) = \begin{cases} g(n) & \text{if } n \in A \\ x & \text{if } n \notin A \end{cases}$.
   - Clearly f is onto since it contains the image of g (which is A), so there exists an onto function $f : \mathbb{Z}_{>0} \to S$ as claimed.
4. The Cartesian product of two countable sets is countable.

- Proof: Since the product map of two bijections is a bijection on the respective Cartesian products, and a subset of a countable set is countable by (2) above, it is enough to prove that the Cartesian product $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ is countable.
- We give an explicit bijection $f : \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ by labeling the points in "diagonal stripes" as shown in the diagram below:



- More explicitly, the bijection is given by $f(a,b) = \dfrac{(a+b)(a+b-1)}{2} - a + 1$ for positive integers a and b.
- It is a straightforward induction on b to see that this labeling is correct on all of the points with $a = 1$: then increasing a by 1 and decreasing b by 1 decreases f by exactly 1 (since $a + b$ is not changed), so the labeling is also correct on all of the diagonal stripes.
- Thus, $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ is countable, hence so is the Cartesian product of any two countable sets.

5. The union of two countable sets is countable.
   - Proof: Suppose A and B are countable. If either A or B is empty then the union is just the other of the two sets, so the result is trivial.
   - Now assume both sets are nonempty. Then by (3) there exist onto functions $f_A : \mathbb{Z}_{>0} \to A$ and $f_B : \mathbb{Z}_{>0} \to B$.
   - Now define the function $f : \mathbb{Z}_{>0} \to A \cup B$ via $f(n) = \begin{cases} f_A(\frac{n+1}{2}) & \text{if } n \text{ is odd} \\ f_B(\frac{n}{2}) & \text{if } n \text{ is even} \end{cases}$.
   - Then f is onto, since its image contains each value $f_A(k) = f(2k)$ and $f_B(k) = f(2k-1)$ for each positive integer k. Hence by (3) again we see that $A \cup B$ is countable.
6. More generally, a countable union of countable sets is countable: if I is a countable indexing set and S_i is a countable set for each $i \in I$, then $\bigcup_{i \in I} S_i$ is countable.
   - Proof: If any S_i is empty we may simply discard it without affecting the union, so suppose each S_i is nonempty. Additionally, if I is finite, then an easy induction using (5) shows that $\bigcup_{i \in I} S_i$ is countable.
   - So assume that I is infinite. Then by (1) there is a bijection $f : \mathbb{Z}_{>0} \to I$ and the positive integers, so by setting $T_j = S_{f(j)}$ for each positive integer j, we are reduced to showing that $\bigcup_{j=1}^{\infty} T_j$ is countable.
   - By (3), for each $j \geq 1$ there exists an onto function $f_j : \mathbb{Z}_{>0} \to T_j$. Now define the function $g : \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \bigcup_{j=1}^{\infty} T_j$ via $g(a,b) = f_a(b)$. Then g is onto, since its image contains $\text{im}(f_j) = T_j$ for each j.
   - Finally, since $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ is countable by (5), composing a bijection $h : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ with g yields an onto map $h \circ g : \mathbb{Z}_{>0} \to \bigcup_{j=1}^{\infty} T_j$, so by (3) we see that $\bigcup_{j=1}^{\infty} T_j$ is countable.
7. (Cantor) The set of rational numbers $\mathbb{Q}$ is countable.
   - Proof 1: For $\mathbb{Q}$, associate the rational number a/b in lowest terms with $b > 0$ to the ordered pair (a,b) in the Cartesian product $\mathbb{Z} \times \mathbb{Z}$. This yields a bijection between $\mathbb{Q}$ and a subset of $\mathbb{Z} \times \mathbb{Z}$.
   - Then since $\mathbb{Z} \times \mathbb{Z}$ is countable by (4) above, and any subset of a countable set is countable by (2) above, we conclude $\mathbb{Q}$ is countable, as claimed.
   - Proof 2: By definition $\mathbb{Q}$ is the union of the countable sets $S_n = \dfrac{1}{n}\mathbb{Z} = \{\ldots, -\dfrac{2}{n}, -\dfrac{1}{n}, 0, \dfrac{1}{n}, \dfrac{2}{n}, \dfrac{3}{n}, \ldots\}$ for integers $n \geq 1$. By (6), a countable union of countable sets is countable, so $\mathbb{Q}$ is countable.

- Remark: It is also possible to show that $\mathbb{Z} \times \mathbb{Z}$ is countable directly by labeling the points in "spirals" outward from the origin. The countability of $\mathbb{Q}$ can also be established using this method, where we label the points (a, b) in spirals, where a/b is a rational number in lowest terms.
- Remark: Another way to show that $\mathbb{Q}$ is countable is first to observe that the rational numbers between 0 and 1 are countable, by simply listing them first in order of increasing denominators and then in order of increasing numerators, skipping terms already listed: $\{\frac{0}{1}, \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \dots\}$. Then we can obtain any rational number merely by including reciprocals and negatives (and negative reciprocals) after each term in the list above: $\{\frac{0}{1}, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \frac{1}{3}, -\frac{1}{3}, \frac{3}{1}, -\frac{3}{1}, \dots\}$.

- So far we have only given examples of sets that are countable. However, not every set is countable:
- Theorem (Cardinality of Power Set): If S is any set, finite or infinite, then there does not exist a bijection between S and its power set $\mathcal{P}(S)$. In particular, the power set $\mathcal{P}(\mathbb{Z}_{>0})$ is uncountable.
  - Proof: Suppose $f : S \to \mathcal{P}(S)$ is any function. We will show that f cannot be onto, so in particular, f cannot be a bijection.
  - Let $A = \{a \in S : a \notin f(a)\}$ be the collection of elements of S that are not an element of their image under f. We claim that A is not in the image of f.
  - For any $s \in S$, either $s \in A$ or $s \notin A$.
  - If $s \in A$, then by definition of A, $s \notin f(s)$. Hence $f(s) \neq A$ because s is an element of A but not $f(s)$.
  - If $s \notin A$, then by definition of A, $s \in f(s)$. Hence $f(s) \neq A$, because s is an element of $f(s)$ but not A.
  - In either case, $f(s) \neq A$. Since this holds for every $s \in S$, we conclude $A \notin \text{im}(f)$. Hence f is not onto, so (in particular) is not a bijection.
  - Remark: Compare this argument to our analysis of Russell's paradox, in which we established that there is no set of all sets. It uses the same technique of considering sets whose (image) does not contain itself.

- It is also true that the set $\mathbb{R}$ of real numbers is uncountable, as first established by Cantor in 1874:
- Theorem (Uncountability of $\mathbb{R}$): The set $\mathbb{R}$ of real numbers is uncountable. In fact, the set of real numbers in the interval $[0, 1]$ is uncountable.
  - In this proof we will use a few basic facts about decimal expansions of real numbers; in particular, recall that every real number has a decimal expansion, and some real numbers have two decimal representations, such as $1.000\cdots = 0.999\ldots$. More specifically, the real numbers with two decimal expansions are the ones of the form $n/10^k$ where n and k are integers: one representation ends in an infinite string of 0s while the other ends in an infinite string of 9s.
  - Proof: By way of contradiction suppose that the set of real numbers in $[0, 1]$ is countable. Then we may list the elements as $r_1, r_2, r_3, \ldots$.
  - Arrange the decimal expansions of these real numbers in an array as follows:

$$\begin{aligned} r_1 &= 0.d_{1,1}d_{2,1}d_{3,1}d_{4,1}\ldots \\ r_2 &= 0.d_{1,2}d_{2,2}d_{3,2}d_{4,2}\ldots \\ r_3 &= 0.d_{1,3}d_{2,3}d_{3,3}d_{4,3}\ldots \\ r_4 &= 0.d_{1,4}d_{2,4}d_{3,3}d_{4,4}\ldots \\ \vdots &\ \ \vdots\ \ \vdots \end{aligned}$$

  - Now we construct a real number in $[0, 1]$ that cannot be equal to any of the numbers r_1, r_2, r_3, r_4 using the "diagonal" digits $d_{i,i}$: if $d_{i,i} = 1$, set $e_i = 2$, and if $d_{i,i} = 2$, set $e_i = 1$.
  - We claim the real number $\alpha = 0.e_1e_2e_3e_4\ldots$ cannot be equal to any of the numbers r_i.
  - To see this, first observe that for any i, the ith decimal digit of α differs from the ith decimal digit of r_i. Then because α cannot have two decimal representations and its representation cannot be equal to any decimal expansion of any r_i, we conclude that $\alpha \in [0, 1]$ is a real number not equal to any r_i.
  - This is a contradiction, and therefore the set of real numbers in $[0, 1]$ is countable.
  - Then $\mathbb{R}$ must be uncountable also, since otherwise $[0, 1]$ would be a subset of a countable set and thus countable itself.
  - Remark: This type of argument, first given by Cantor, is known as a diagonalization argument.

3.5.3 Infinite Cardinalities

- We now discuss some other results about infinite sets. We have seen above that there are at least two different "sizes" of infinite sets (namely, countably infinite and uncountably infinite) but in fact there are more:
- Proposition (Infinite Cardinals): There exists an infinite sequence of infinite sets $S_1, S_2, S_3, \ldots$, no two of which are equinumerous.
    - Another way to interpret this result is that there are infinitely many different infinite cardinalities, or more informally, there are infinitely many different infinities.
    - Proof: As we have shown, there does not exist an onto map from a set to its power set.
    - Hence if we take $S_1 = \mathbb{Z}_{>0}$, and define $S_n = \mathcal{P}(S_{n-1})$ for each $n \geq 2$, then any map from S_i to S_j with $i < j$ cannot be onto, since an appropriate restriction would necessarily yield an onto map from S_i to $S_{i+1} = \mathcal{P}(S_i)$.
    - This means in particular that no two of the infinite sets $S_1, S_2, S_3, \ldots$ are equinumerous, as required.
- By definition, two sets have the same cardinality if there is a one-to-one correspondence between them. But it is also natural to want to compare sets of different cardinalities, which we may do using one-to-one functions:
- Definition: If A and B are sets, we say A is dominated by B, written $A \precsim B$, if there exists a one-to-one function $f : A \to B$.
    - The motivation for this definition is the observation that if $f : A \to B$ is one-to-one, then f is a bijection from A to $\text{im}(f) \subseteq B$, and so A is in bijection with a subset of B. This is a reasonable way to capture the idea that B has "at least as many" elements as A.
    - Example: $\{1, 2, 3\} \precsim \{a, p, q, s\}$ because there exists a one-to-one function $f : \{1, 2, 3\} \to \{a, p, q, s\}$, such as $f = \{(1, a), (2, p), (3, s)\}$.
    - Example: $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \precsim \mathbb{Z}$ because there exists a one-to-one function $f : \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{Z}$, namely the explicit map we constructed that gives a bijection of $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ with $\mathbb{Z}_{>0}$.
- Note that we have used the symbol $\precsim$, which suggests that this relation should behave like a partial ordering.
    - Reflexivity follows immediately, because the identity function from A to itself is one-to-one, so $A \precsim A$.
    - Transitivity is also straightforward: if $A \precsim B$ and $B \precsim C$, then there exist one-to-one functions $f : B \to C$ and $g : A \to B$. Then it is straightforward to check that $f \circ g : A \to C$ is also one-to-one, whence $A \precsim C$.
    - However, this relation is not antisymmetric: there are examples of sets A and B with $A \precsim B$ and $B \precsim A$ but with $A \neq B$. For example, $\{1, 2\} \precsim \{a, b\}$ and $\{a, b\} \precsim \{1, 2\}$, and also $\mathbb{Z} \precsim \mathbb{Q}$ and $\mathbb{Q} \precsim \mathbb{Z}$.
    - However, these examples do suggest that if $A \precsim B$ and $B \precsim A$, then A and B are equinumerous, in which case the relation $\precsim$ is antisymmetric when viewed on cardinalities (i.e., on equivalence classes of equinumerous sets). This turns out to be true, but not so easy to prove:
- Theorem (Cantor-Schröder-Bernstein): Suppose A and B are sets such that there exists an injection from A to B and an injection from B to A. Then there exists a bijection between A and B.
    - The proof of this theorem is somewhat involved, but the overall idea is to consider the one-to-one maps $f : A \to B$ and $g : B \to A$. If f is onto then we are done.
    - Otherwise, we glue together part of f with part of the surjective map $g^{-1} : \text{im}(g) \to B$ to create a one-to-one map $h : A \to B$ that also takes on the values in B that were missing from $\text{im}(f)$. Rather than motivating the construction further, we simply give the proof.
    - Proof: Suppose $f : A \to B$ and $g : B \to A$ are one-to-one. Then g has an inverse function $g^{-1} : \text{im}(g) \to B$ whose image is B.
    - Now define a sequence of sets $A_1, A_2, A_3, \ldots$ recursively: take $A_1 = A \backslash \text{im}(g)$, and for each $n \geq 2$, take $A_n = g(f(A_{n-1})) = \{g(f(a)) : a \in A_{n-1}\}$.
    - Also define $X = \bigcup_{n \geq 1} A_n$ and $Y = A \backslash X$, and finally define $h : A \to B$ via $h(a) = \begin{cases} f(a) & \text{if } a \in X \\ g^{-1}(a) & \text{if } a \in Y \end{cases}$.

- Observe that h is well-defined because X and Y are disjoint by definition, and also that if $a \in Y$ (so that $a \notin X$) then by definition $a \notin A_1$, so $a \in \text{im}(g)$ and thus $g^{-1}(a)$ makes sense.
- To show that h is one-to-one, suppose $h(a_1) = h(a_2)$.
- If $a_1, a_2 \in X$ then we would have $f(a_1) = f(a_2)$, but since f is one-to-one, we see $a_1 = a_2$. Likewise, if $a_1, a_2 \in Y$ then we would have $g^{-1}(a_1) = g^{-1}(a_2)$, and then applying g yields $a_1 = a_2$.
- For the remaining case assume without loss of generality that $a_1 \in X$ and $a_2 \in Y$. Then we would have $f(a_1) = g^{-1}(a_2)$, implying $g(f(a_1)) = a_2$, but this would mean $a_2 \in g(f(X)) = X$, which is a contradiction. Hence this case cannot occur, and so $a_1 = a_2$ in all cases, meaning that h is one-to-one.
- To show that h is onto, let $b \in B$: then $g(b) \in A$.
- If $g(b) \in Y$, then $h(g(b)) = g^{-1}(g(b)) = b$, so $b \in \text{im}(h)$.
- If $g(b) \in X$, then by definition of X as a union we have $g(b) \in A_n$ for some n.
- In particular since $g(b) \in \text{im}(g)$ we have $n \neq 1$. This means $g(b) \in g(f(A_{n-1}))$, meaning that for some $a \in A_{n-1}$ we have $g(b) = g(f(a))$.
- But then since g is one-to-one this implies $b = f(a) = h(a)$ since $a \in A_{n-1} \subseteq X$, and so we also have $b \in \text{im}(h)$ in this case.
- Hence $b \in \text{im}(h)$ in either cases, so h is onto. Thus, h is a bijection as required.

- The Cantor-Schröder-Bernstein theorem shows that the relation $\precsim$ is a partial ordering on cardinalities.
  - A natural followup question is whether this relation is actually a *total* ordering on cardinalities.
  - Equivalently, we are asking whether any two sets are always comparable under $\precsim$, which is to say, given any two sets, does there necessarily exist an injection from one the other?
  - It turns out that the answer relies on a foundational axiom of set theory known as the axiom of choice, which (in one formulation) states that the Cartesian product of an arbitrary collection of nonempty sets is nonempty.
  - If the axiom of choice is accepted, it can be shown that $\precsim$ is a total ordering on sets: in fact, it is actually true that the axiom of choice is *equivalent* to the statement that $\precsim$ is a total ordering on sets.

- In this formulation (the Cartesian product of an arbitrary collection of nonempty sets is nonempty), the axiom of choice seems like a natural assumption to make, and it is generally accepted by most mathematicians in practical work.
  - There exist many other equivalent formulations of the axiom of choice, some of which seem fairly natural, and others which are less so.
  - Another statement equivalent to the axiom of choice is called Zorn's lemma, which states that every nonempty partially-ordered set having the property that any totally ordered subset has an upper bound (an element greater than or equal to every element of the subset) has a maximal element (an element such that no element is greater than it).
  - A third equivalent to the axiom of choice (familiar to students who have studied linear algebra) is the statement that every vector space has a basis.
  - A fourth equivalent to the axiom of choice is called the well-ordering principle, which states that every set admits a well-ordering (a total ordering in which every nonempty subset has a smallest element).
  - This fact was one of our axioms [N3] for the definition of the integers. However, it is much less intuitive to ask what a well-ordering on the set $\mathbb{R}$ would look like: the usual total ordering $\leq$ is not a well-ordering, because there are many sets, like the open interval $(0, 1)$ or even $\mathbb{R}$ itself, that have no smallest element under $\leq$.
  - It has also been proven that the axiom of choice is independent of the standard Zermelo-Fraenkel axioms of set theory, in the sense that the axioms are consistent provided the axiom of choice is accepted if and only if the axioms are consistent provided the axiom of choice is rejected.

Well, you're at the end of my handout. Hope it was helpful.