Logic and Proofs

1.1 Propositional Logic

Definition: proposition

A *proposition* is a declarative statement that is either true or false, but not both.

Example. The following are propositions:

(1) Columbus is the capital of Ohio. (True)
(2) $2 + 2 = 5$. (False)
(3) A circle with radius r has area πr^2. (True)
(4) The number $\sqrt{2}$ is rational. (False)
(5) Every square is a rectangle. (True)
(6) Every rectangle is a square. (False)

We often refer to statements with letters (typically, $p, q, r, s, \dots$). The *truth value* of a proposition is denoted T (if true) or F (if false). Our goal will be to study how the truth value of propositions varies when combined with *logical operators* and other propositions to form *compound propositions*.

Definition: negation

Let p be a proposition. The *negation of* p, denoted $\neg p$, is the statement "It is not the case that p". The truth value of $\neg p$ is the opposite of p.

We read $\neg p$ as "not p". (Note that some authors use $\sim p$ or $\overline{p}$ to denote $\neg p$.)

Example. Consider the negation of each proposition above:

(1) Columbus is *not* the capital of Ohio. (False)
(2) $2 + 2 \neq 5$. (True)
(3) A circle with radius r *does not have* area πr^2. (False)
(4) The number $\sqrt{2}$ is *not* rational. (True)
(5) Every square is *not* a rectangle. (False)
(6) Every rectangle is *not* a square. (True)

We can represent the relationships between the different possibilities for the truth value of p and its negation in a *truth table*:

p	$\neg p$
T	F
F	T

Definition: conjunction, disjunction

• The *conjunction* of propositions p and q, denoted $p \wedge q$, is the proposition "p and q". The conjunction $p \wedge q$ is true when both p and q are true, and false otherwise.

• The *disjunction* of p and q, denoted $p \vee q$, is the proposition "p or q". The disjunction $p \vee q$ is true when p and q are false, and true otherwise.

Example. (1) The proposition

$$\text{Columbus is the capital of Ohio and } 2 + 2 \neq 5$$

is true because both individual propositions are true, even though the two statements have nothing to do with one another.

(2) The proposition

$$\text{Every square is a rectangle and every rectangle is a square}$$

is false because one of the individual propositions is false.

(3) The proposition

$$\text{The number } \sqrt{2} \text{ is rational or every square is a rectangle}$$

is true because the second statement is true.

(4) The proposition

$$\text{Every rectangle is a square or } 2 + 2 = 5$$

is false because both propositions are false.

Always, *or* in mathematics is defined as above. This is known as the *inclusive or*. Do not confuse this with *exclusive or*, defined below.

Definition: exclusive or

Let p and q be propositions. The *exclusive or* of p and q, denoted $p \oplus q$, is the proposition that is true if exactly one of p and q is true and is false otherwise.

There are four possibilities of pairs (p, q), so the truth tables for $p \wedge q$, $p \vee q$, or $p \oplus q$ have four rows:

p	q	$p \wedge q$	$p \vee q$	$p \oplus q$
T	T	T	T	F
T	F	F	T	T
F	T	F	T	T
F	F	F	F	F

Definition: conditional statement, hypothesis, conclusion

Let p and q be propositions. The *conditional statement* $p \to q$ is the proposition "if p, then q." The conditional statement is false when p is true and q is false, and true otherwise. Given $p \to q$, we call p the *hypothesis* and q the *conclusion*.

Example. Consider the following example by an imaginary politician:

If I am elected, then I will lower taxes.

Only in the instance in which the politician is elected but *does not* lower taxes is the statement clearly false. If the politician is not elected, there is no expectation put on the conclusion.

There are many other ways to phrase a conditional statement. Two are p *only if* q, as well as p *implies* q (see the text for many more examples). Importantly, in this instance we say that p is a *sufficient condition* for q, and we say that q is a *necessary condition* for p. The truth table for a conditional statement is given below.

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

Note in the last two rows, p is *false*, and so the conditional statement is *vacuously true*.

Definition: converse, inverse, contrapositive

Given a conditional statement "if p, then q" $(p \to q)$,

- the *converse* is the statement "if q, then p" $(q \to p)$,
- the *inverse* is the statement "if not p, then not q" $(\neg p \to \neg q)$, and
- the *contrapositive* is the statement "if not q, then not p" $(\neg q \to \neg p)$.

Example. Consider the statement:

I eat oatmeal for breakfast whenever it is Tuesday.

This can be rephrased as $p \to q$:

If it is Tuesday, then I eat oatmeal for breakfast.

Write the converse, inverse, and contrapositive of this statement. We can then express the above three statements as follows:

Converse: If I eat oatmeal for breakfast, then it is Tuesday.
Inverse: If it is not Tuesday, then I do not eat oatmeal.
Contrapositive: If I do not eat oatmeal for breakfast, then it is not Tuesday.

We now look at combining conditional statements.

Definition: biconditional

Let p and q be propositions. The *biconditional statement* $p \leftrightarrow q$ is the proposition "p if and only if q." The binconditional statement $p \leftrightarrow q$ is true when p and q have the same truth values, and is false otherwise.

Example. Consider the statement "You can take the flight if and only if you buy a ticket."

Some use "iff" in place of "if and only if", though this is frowned upon in formal mathematics.

p	q	$p \leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

Example. Show that $p \leftrightarrow q$ has the same truth value as $(p \to q) \wedge (q \to p)$.

(If time) We conclude by noting the precedence of logical operators:

Operator	Precedence
$\neg$	1
$\wedge$	2
$\vee$	3
$\to$	4
$\leftrightarrow$	5

However, especially in cases of disjunction versus conjunction, it is best to make use of parentheses to distinguish hierarchy.

Example. We will translate the following sentence into propositional logic:

You can see the movie only if you are over 18 years old or you have the permission of a parent.

The three statements here are

m : You can see the movie

e : You are over 18 years old

p : You have the permission of a parent

As a compound conditional statement, this translates to $m \to (e \vee p)$.

Translating a natural language into propositional logic is a necessary component of creating system specifications for hardware and software. When there are several specifications, they must be consistent and not lead to a contradiction.

Example. Consider the following system specifications:

- Whenever the system software is being upgraded, users cannot access the file system.
- If the users can access the file system, then they can save new files.
- If users cannot save new files, then the system software is not being upgraded.

The three statements here are:

p : The system software is being upgraded

q : Users can access the file system

r : Users can save new files

The system specifications translate to $p \to \neg q$, $q \to r$, and $\neg r \to \neg p$.

We could create a truth table and verify that there is choice of truth value for p, q, r so that all three statements are true. We could make a truth table to check this, or use deduction.

Suppose p is true. The first statement requires $\neg q$ to be true, so q is false. Now the last statement requires that $\neg r$ is false, so r is true. If q is false and r is true, then the middle statement is true. Hence, the system is consistent.

Boolean searches use propositional logic.

Example. Suppose we want a web search about hiking in Virginia, but not in West Virginia.

We would search: (HIKING *AND VIRGINIA) NOT WEST*.

Definition: tautology, contradiction, contingency

A compound proposition that is always true is called a *tautology*. A compound proposition that is always false is called a *contradiction*. A compound proposition that is not a tautology or a contradiction is called a *contingency*.

Example. Let p be a proposition. Then $p \vee \neg p$ is a tautology while $p \wedge \neg p$ is a contradiction.

In logical reasoning, it is important to be able to compare logical statements. Oftentimes, we can replace one logical statement with another.

Definition: logically equivalent

The compound propositions p and q are logically equivalent if $p \leftrightarrow q$ is a tautology. We denote logically equivalent compound propositions by $p \equiv q$.

Instead of showing directly that $p \leftrightarrow q$, it is often easier to show that their columns in a truth table are the same.

Example (De Morgan's Laws (for logic))**.** Let p and q be statements. Then

(1) $\neg(p \wedge q) \equiv (\neg p) \vee (\neg q)$
(2) $\neg(p \vee q) \equiv (\neg p) \wedge (\neg q)$

We will verify the second. The first is left as an exercise.

p	q	$p \vee q$	$\neg(p \vee q)$	$(\neg p)$	$(\neg q)$	$(\neg p) \wedge (\neg q)$
T	T	T	F	F	F	F
T	F	T	F	F	T	F
F	T	T	F	T	F	F
F	F	T	T	T	T	T

Comparing the fourth and final columns gives the equivalence.

De Morgan's laws can be useful in negating statements in natural language.

Example. Consider the following statement:

Evelyn likes art and playing the piano

The two statements are

p : Evelyn likes art $\qquad$ q : Evelyn likes playing the piano

Since the negation of $p \wedge q$ is $\neg p \vee \neg q$, then the negation is

Evelyn does not like art, or Evelyn does not like playing the piano

Example (conditional-disjunction equivalence)**.** Show $p \to q \equiv \neg p \vee q$.

p	q	$p \to q$	$(\neg p)$	$(\neg p) \vee q$
T	T	T	F	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Comparing the third and final columns gives the equivalence.

There are many other such rules listed in your textbook (see Section 1.3). Besides De Morgan's laws, you are not expected to know these all by heart, but you should be able to verify all of them. (Display these laws on screen.)

Example (distributive law of disjunction over conjunction)**.** Show that $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$.

p	q	r	$q \wedge r$	$p \vee (q \wedge r)$	$p \vee q$	$p \vee r$	$(p \vee q) \wedge (p \vee r)$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	F	T	F	F
F	F	T	F	F	F	T	F
F	F	F	F	F	F	F	F

Comparing the fifth and final columns gives the equivalence.

1.4, 1.5 Predicates and quantifiers

Quantifiers allow us to make statements involving variables. This is known as *predicate logic*. For example, the mathematical statement "$x > 3$" does not have a truth value *until* we choose a value for x. Similarly, the statement "Team x won the game" is meaningless until we identify "Team x". Predicates are used frequently in programming to decide whether to execute a command. Consider the following line of code: if $x > 0$, then $x := x + 1$.

Definition: propositional function, subject, predicate

A *propositional function* $P(x_1, x_2, \ldots, x_n)$ is a proposition that depends on the variables $x_1, x_2, \ldots, x_n$, known as *subjects*. The property that must be decided for P is called the *predicate*.

Example. If $P(x)$ is the statement that "$x > 3$", then $P(4)$ is true but $P(2)$ is false.

If $Q(x, y)$ denotes the statement "$x = y + 3$", then $Q(4, 1)$ is true but $Q(0, 0)$ is false.

Oftentimes we will want to consider a propositional function not just at one value, but at many or all values in the range of considered values, called the *domain*. Alternatively, we may want to decide if it is true for any value of x. To make such statements, we use quantifiers.

Definition: universal quantifier, counterexample

The universal quantification of $P(x)$, denoted $\forall x P(x)$, is the statement

$$P(x) \text{ for all values of } x \text{ in the domain}$$

Here $\forall$ is called the *universal quantifier*. A value of x in the domain for which $P(x)$ is false is called a *counterexample* to $\forall x P(x)$.

If the domain is empty, then $\forall x P(x)$ is true for any propositional function $P(x)$.

Other common ways to express "for all" are "for every", "for each", and "for arbitrary."

Example. Let $P(x)$ be the statement "$x + 1 > x$" where the domain is all real numbers. This is true for all x, so the universal quantifier $\forall x P(x)$ is true.

Let $Q(x)$ be the statement "$x < 2$" where the domain is all real numbers. Since $Q(3)$ is false, then $x = 3$ is a counterexample for $\forall x Q(x)$. Hence $\forall x Q(x)$ is false. However, if we restrict the domain to negative integers, then $\forall x Q(x)$ is true.

Definition: existential quantifier

The existential quantification of $P(x)$, denoted $\exists x P(x)$, is the proposition

There exists an element x in the domain such that $P(x)$

Here $\exists$ is called the *existential quantifier*.

Note that this statement requires only existence, not uniqueness. We might use instead of "there exists" the phrases "for some", "for at least one", or "there is."

Example. Let $P(x)$ denote the statement $x > 3$ where the domain consists of all real numbers. Then $P(4)$ is true so $\exists x P(x)$ is true.

Let $Q(x)$ denote the statement $x = x + 1$ where the domain consists of all real numbers. Then no x satisfies $Q(x)$, so $\exists x Q(x)$ is false.

Definition: uniqueness quantifier

The notation $\exists! x P(x)$ is the proposition

There exists a unique x in the domain such that $P(x)$

Here $\exists!$ is called the *uniqueness quantifier*.

Example. Let $P(x)$ denote the statement $x - 1 = 0$ where the domain consists of all real numbers. Then $P(1)$ is true but $P(x)$ is false for all $x \neq 1$. So $\exists! x P(x)$ is true.

Let $Q(x)$ denote the statement $x^2 = 1$ where the domain consists of all real numbers. Then $Q(1)$ and $Q(-1)$ are both true, so $\exists! x Q(x)$ is false.

In a natural way, we can combine our quantifiers with our logical operators. Note that $\forall$ and $\exists$ have higher precedence than those from propositional calculus.

Definition: logical equivalence

Statements involving predicates and quantifiers are logically equivalent if and only if they have the same truth value no matter which predicates are substituted into these statements and which domain of discourse is used for the variables in these propositional functions.

Example. We will show that $\forall x(P(x) \wedge Q(x))$ and $\forall P(x) \wedge \forall Q(x)$ are logically equivalent.

Assume we have a common domain and let a be in that domain. Then $P(a) \wedge Q(a)$ is true. But this implies that $P(a)$ is true and that $Q(a)$ is true. But our choice of a was arbitrary. So, $\forall x P(x)$ is true and $\forall x Q(x)$ is true. Thus, $\forall P(x) \wedge \forall Q(x)$ is true.

Now suppose that $\forall P(x) \wedge \forall Q(x)$ is true. Thus, $\forall P(x)$ is true and $\forall Q(x)$ is true. Hence, if a is in the domain, then $P(a)$ is true and $Q(a)$ is true. Thus, $P(a) \wedge Q(a)$ is true. Since this holds for all a in the domain, then $\forall x(P(x) \wedge Q(x))$ is true.

Example (De Morgan's Laws (for propositional functions))**.** Let p and q be statements. Then

(1) $\neg(\forall x P(x)) \equiv \exists x, \neg P(x)$
(2) $\neg(\exists x P(x)) \equiv \forall x, \neg P(x)$

We will prove the first and leave the second as an exercise.

Suppose the statement $\neg\forall x P(x)$ is true. This implies that $\forall x P(x)$ is false, which implies that there is some counterexample to $\forall x P(x)$. Let y be the counterexample. That is, $P(y)$ is false. But this implies that $\neg P(y)$ is true. So $\exists x \neg P(x)$ is true. The converse is similar.

Translating sentences from natural language using predicates and quantifiers can be nontrivial.

Example. Translate each sentence into a statement using quantifiers.

(1) Someone in your class has seen the movie Borat.
    There is a student x in this class having the property that x has seen Borat.
(2) All the students in this class go to Miami.
    For every student x in this class, x goes to Miami.
(3) There is an integer that squares to 9.
    There exists an integer n having the property that $n^2 = 9$.
    (Note the domain of this statement is all integers, not all real numbers.)
(4) Every integer that is not odd is even.
    Let $E(n)$ be the propositional function that is true when n is even. Let $O(n)$ be the propositional function that is true when n is odd.

$$\forall n \in \mathbb{Z}, \neg O(n) \to E(n)$$

The statement $n \in \mathbb{Z}$ means that n is *an element of* the set $\mathbb{Z}$, which is the set of integers. We will discuss this notation more in coming sections.

We can combine two or more quantifiers to make more complex statements.

Example. Let $P(x, y)$ be the statement $x + y = 0$ and consider the quantification $\forall x \exists y P(x, y)$ where the domain is all real numbers. Translate to English and determine the truth value of the statement.

This is the proposition

For all real numbers x, there exists a real number y such that $x + y = 0$

This is considered a *nested quantifier* because we could set $Q(x)$ to be the statement $\exists y P(x, y)$ and then the statement above becomes $\forall x Q(x)$.

This statement is true if we can pick an arbitrary element x and show that $Q(x)$ is true. But now for $Q(x)$ to be true, we have to show that there exists some y that makes $P(x, y)$ true. Obviously picking $y = -x$ makes this true.

On the other hand, the quantification $\exists y \forall x P(x, y)$ is *false*. Therefore, one may not in general interchange a universal quantifier with that of an existential quantifier.

Example. Let $P(x, y)$ be the statemement $x+y = y+x$ and consider the quantification $\forall x \forall y P(x, y)$ where the domain is all real numbers. Translate to English and determine the truth value of the statement.

The quantification is the statement

For all real numbers x, for all real numbers y, $x + y = y + x$

This is the commutative law of addition for real numbers (so it is true).

Note that the truth value is equivalent to that $\forall y \forall x P(x, y)$, so the order of two universal quantifiers may be interchanged.

Example. Consider the statement,

Every real number except zero has a multiplicative inverse

State this in English and symbolically using quantifiers. Determine the truth value.

Stated another way

For every real number x, if $x \neq 0$, then there exists a real number y such that $xy = 1$

We can state this using quantifiers as

$$\forall x((x \neq 0) \rightarrow \exists y(xy = 1))$$

This statement is false. A counterexample is $x = 0$.

1.7 Introduction to Proofs

In mathematical writing, a *theorem* is a statement that can be shown to be true. Typically, a *proposition* is a less important theorem (but this is often a matter of taste). A *lemma* is typically a smaller statement used in proving a more important proof. A *corollary* is a consequence of another theorem. A *conjecture* is a proposed theorem (without proof) based on some established evidence.

Theorems and other statements are verified using a *proof*, which is a logical argument (typically written in natural language) that establishes the validity of the statement. Proofs rely on axioms, such as the laws of real numbers, as well as other theorems that have already been justified. In this section we will discuss some methods of proof as well as consider some simple examples.

Theorems are typically stated as conditionals (or biconditionals), but are not always stated in terms of universal quantifiers, so it is important for approaching a proof to be able to translate.

To prove a conditional $p \to q$, one can use *direct proof* in which p is assumed to be true, then use logical inference to establish that q is true.

Definition: even, odd, parity

An integer n is *even* if there exists an integer k such that $n = 2k$, and n is *odd* if there exists an integer k such that $n = 2k + 1$. Two integers have the *same parity* when both are even or both are odd, and they have *opposite parity* when one is even and one is odd.

Example. Prove: "If n is an odd integer, then n^2 is odd."

Restated as a conditional, the theorem says

For all integers n, if n is odd, then n^2 is odd

This statement is of the form $\forall n(P(n) \to Q(n))$, where $P(n)$ is the statement n is odd and $Q(n)$ is the statement that n^2 is odd. The domain is the set of integers. So, we begin by choosing an *arbitrary* integer n and assuming $P(n)$ is true. We then try to establish that $Q(n)$ is true.

Proof. Suppose n is an odd integer. By the definition of odd integers, $n = 2k + 1$ for some integer k. By algebra, $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$. Since $2k^2 + 2k$ is another integer, then n^2 is odd by the definition of odd integers. □

Example. Prove that the sum of two even integers is even.

Before we begin, we translate as a conditional:

$$\text{If } m \text{ and } n \text{ are even integers, then } m+n \text{ is an even integer}$$

Let $P(n)$ be the statement that n is even. We are trying to prove

$$\forall m \forall n((P(m) \wedge P(n)) \to P(m+n))$$

So, we choose *arbitrary* integers m and n and assume both are even. Then try to conclude that $P(m+n)$ is even.

Proof. Suppose m and n are even integers. Then $m = 2k$ and $n = 2\ell$ for integers k and ℓ. (Note we did not choose k for both since that would imply $m = n$ and this is not a hypothesis.) Then $m + n = 2k + 2\ell = 2(k + \ell)$, which is even by definition. □

Definition: perfect square

An integer n is a *perfect square* if $n = a^2$ for some integer a.

Example. Prove: "If m and n are both perfect squares, then nm is a perfect square."

Proof. Suppose that m and n are perfect squares. By definition, $m = a^2$ and $n = b^2$ for integers a and b. Then by algebra, $mn = a^2b^2 = (ab)^2$. Since ab is another integer, then mn is a perfect square by definition. □

Definition: rational, irrational

The real number r is *rational* if there exist integers p and q with $q \neq 0$ such that $r = p/q$. A real number that is not rational is called *irrational*.

Example. Prove: "The product of two rational numbers is rational."

Stated as a conditional:

$$\text{If } r \text{ and } s \text{ are rational numbers, then } rs \text{ is rational}$$

Proof. Suppose that r and s are rational. By definition, $r = p/q$ and $s = p'/q'$ for integers p, q, p', q' with $q, q' \neq 0$. Then $rs = (pp')/(qq')$. The product of two integers is an integer. Since $q, q' \neq 0$, then $qq' \neq 0$. Thus, rs is rational by definition. □

A *proof by contraposition* uses the fact that $p \to q$ is logically equivalent to $\neg q \to \neg p$. In contrast to direct proof, we assume q is false and try to prove that p is false.

Example. Prove that if a is an integer and a^2 is even, then a is even.

We will instead prove the statement

$$\text{If } a \text{ is odd (i.e., not even), then } a^2 \text{ is odd}$$

Proof. Suppose that a is odd. Then $a = 2k + 1$ for some integer k. By algebra, $a^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, so a^2 is odd. □

Example. Prove that if x, y, and z are integers and $x + y + z$ is odd, then at least one of x, y, and z is odd.

The negation of the conclusion is that all three are even. So we use proof by contraposition.

Proof. Suppose x, y, z are even. Then $x = 2a$, $y = 2b$, and $z = 2c$ for integers a, b, c. Then by algebra, $x + y + z = 2a + 2b + 2c = 2(a + b + c)$ which is even by definition. □

Sometimes theorems are "easy" to prove, and sometimes they are simply not true at all!

Example. Prove: "If n is an integer with $10 \leq n \leq 15$ is a perfect square, then n is also a perfect cube."

Recall that the statement $p \to q$ is *vacuously true* when p is true. Since there are no perfect squares between 10 and 15, the given statement is vacuously true.

A common mistake in proof writing is to *assume the conclusion* of the statement. Another, which is similar is called *circular reasoning*, also called *begging the question*.

Example. Find the mistake in the following "proof" of "If n^2 is even, then n is even."

Proof. Suppose n^2 is even. Then $n^2 = 2k$ for some integer k. Let $n = 2\ell$ for some integer ℓ. Then n is even by definition. □

This "proof" is incorrect because instead of reasoning *why* n is even, we simply concluded (out of nowhere!) and equivalent statement for n being even.

Proof by contradiction (if time)

Suppose we want to prove a statement p is true (not necessarily a conditional). Perhaps we can find a contradiction q (i.e., $q = r \wedge \neg r$) such that $\neg p \to q$ is true. Since q is false, then this implies that $\neg p$ is false, so p is true.

Example. Prove that $\sqrt{2}$ is irrational.

Let p be the statement "$\sqrt{2}$ is irrational". We assume p is false and try to arrive at a contradiction.

Proof. Suppose $\sqrt{2}$ is *not* irrational, so $\sqrt{2}$ is rational. Then by definition, $\sqrt{2} = a/b$ for integers a and b, $b \neq 0$. Assume a and b have no common factors (that is, a/b is expressed in lowest terms). Squaring both sides of $\sqrt{2} = a/b$ gives $2 = a^2/b^2$. Hence, $2b^2 = a^2$. This implies a^2 is even, so a is even. That is, $a = 2c$ for some integer c. Then we have $2b^2 = 4c^2$, so $b^2 = 2c^2$. That is, b is also even. But then a and b have a common factor (of 2), a contradiction. □

In the proof we assumed $\neg p$, or the statement that $\sqrt{2} = a/b$ is rational, and proved q="a and b have a common factor" is true. But q is false, so $\neg p$ is false. That is, p is true.

Here is another classical example. Recall that an integer $n > 1$ is *prime* if its only factors are 1 and itself. (We do *not* consider 1 to be a prime.)

Example. Prove that there are infinitely many primes.

Suppose there are finitely many primes. We list them $p_1, p_2, \ldots, p_n$. The product $p = p_1 p_2 \cdots p_n$ is divisible by each of the p_i, so $p + 1$ is not divisible by any. Hence, $p + 1$ is prime, a contradiction.

Sets and Functions

2.1 Sets

Definition: Set,elements

A *set* is an unordered collection of distinct objects, called *elements*.

A set is said to *contain* its elements. We write $a \in A$ to say that a is an element of the set A. Conversely, we write $a \notin A$ to say that a is *not* an element of the set A.

When a set is small, we can list the elements directly using braces. For example, the set of vowels in the English language could be denoted $V = \{a, e, i, o, u\}$. However, recall that a set is by definition unordered. So it is not important which order we list the vowels in.

When a set is too big to list its elements, we list *some* of the elements then use ellipses $(\ldots)$ to indicate that the pattern continues. For examples, the postive integers less than 100 could be written $\{1, 2, 3, \ldots, 99\}$. It is important that the pattern is clear to the reader. If we had only written $\{1, \ldots, 99\}$, then we could mean all of the numbers between 1 and 99, or just the odds.

Example. Important sets to know are the following:

(1) the natural numbers $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
(2) the integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
(3) the real numbers, $\mathbb{R}$
(4) the complex numbers, $\mathbb{C}$

We use a superscript $+$ to denote the positive elements in any of these sets. So the positive real numbers is denoted $\mathbb{R}^+$.

Another way to list a set is to use *set builder* notation. This has the form

$$\{x \in D \mid x \text{ has property } P\}$$

where D is the domain from which we can choose x. If the domain is understood by context, then it is often omitted.

Example. Write the set $\{1, 3, 5, 7\}$ and the set of rational numbers in set builder notation.

The set $\{1, 3, 5, 7\}$ could be written in set builder notation as

$$\{x \in \mathbb{Z}^+ \mid x \text{ is odd and } x < 10\}.$$

The *rational numbers* can be written as

$$\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0 \right\}.$$

We will also use interval notation for subsets of $\mathbb{R}$:

$$[a,b] = \{x \mid a \leq x \leq b\} \qquad [a,b) = \{x \mid a \leq x < b\}$$
$$(a,b] = \{x \mid a < x \leq b\} \qquad (a,b) = \{x \mid a < x < b\}.$$

Definition: Subset, equal sets, proper subset

The set A is a *subset* of a set B, denoted $A \subseteq B$, if and only if every element of A is also an element of B. Two sets A and B are *equal*, denoted $A = B$, if they have the same elements. We say A is a *proper subset* of B, denoted $A \subset B$, if A is a subset of B and $A \neq B$.

In logical form, $A \subseteq B$ is equivalent to the quantification $\forall x(x \in A \to x \in B)$. Hence, to show that A is a subset of B, it suffices to choose an arbitrary element of A and justify why it is in B. Conversely, to show that A is *not* a subset of B (denoted $A \not\subseteq B$) we need only find one element in A that is not in B.

In logical form, we say A and B are equal if and only if $\forall x(x \in A \leftrightarrow x \in B)$. To show that sets A and B are equal, it suffices to show that $A \subseteq B$ and $B \subseteq A$. That is, we have the equivalence,

$$\forall x(x \in A \leftrightarrow x \in B) \equiv \forall x(x \in A \to x \in B) \wedge \forall x(x \in B \to x \in A)$$

Definition: Empty set, singleton set

The set with no elements is called the *empty set*, denoted $\emptyset$. A set with only one element is called a *singleton set*.

Do not confuse $\emptyset$ with $\{\emptyset\}$, which is a singleton set containing one element, the empty set.

Theorem 1: Common subsets

The empty set is a subset of every set. A set is always a subset of itself.

Proof. Let S be a set. We claim $\emptyset \subseteq S$. That is, we wish to show that $\forall x(x \in \emptyset \to x \in S)$. For any x, the statement $x \in \emptyset$ is false because $\emptyset$ contains no elements. Thus, $x \in \emptyset \to x \in S$ is true.

Next we claim $S \subseteq S$. That is, we claim $\forall x(x \in S \to x \in S)$. Rewriting the conditional we have $x \in S \to x \in S \equiv \neg(x \in S) \vee (x \in S)$, which is a tautology. $\square$

Definition: Finite set, infinite set, cardinality

Let S be a set. If there are exactly n distinct elements in S, where N is a nonnegative integer, then S is a *finite set* of *cardinality* n. We denote the cardinality of S by $|S|$. A set that is not finite is said to be *infinite*.

Example. Let A be the set of odd positive integers less than 10, then A is finite and $|A| = 5$. On the other hand, the set of positive integers is finite. The empty set has $|\emptyset| = 0$.

Definition: Power set

Given a set S, the *power set* of S is the set of all subsets of S, denoted $\mathcal{P}(S)$.

Example. Let $S = \{0, 1, 2\}$. Then

$$\mathcal{P}(S) = \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$$

Recall that an ordered n-tuple $(a_1, a_2, \ldots, a_n)$ is the ordered collection that has a_1 as its first element, a_2 as its second, etc. This is in contrast to a set, where order does not matter.

Definition: Cartesian product

Let A and B be sets. The *Cartesian product* of A and B, denoted $A \times B$ is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$.

In set builder notation, $A \times B = \{(a, b) \mid a \in A \wedge b \in B\}$. We can of course extend the notion of the Cartesian product to several sets $A_1, A_2, \ldots, A_n$:

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i \text{ for } i = 1, 2, \ldots, n\}.$$

Example. The Cartesian product of $A = \{1, 2\}$ and $B = \{a, b, c\}$ is

$$A \times B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

2.2 Set Operations

We now study operations on sets, which are closely related to logical operators.

Definition: Union, intersection, complement

Let A and B be sets. The *union* of A and B, denoted $A \cup B$, is the set that contains those elements that are either in A or in B, or in both. The *intersection* of A and B, denoted $A \cap B$ is the set that contains those elements that are either in both A and B. The *difference* of A and B, denoted $A - B$, is the set containing those elements that are in A but not in B.

Note that $A - B$ is sometimes denoted $A \backslash B$. In set builder notation, these are

$$A \cup B = \{x : x \in A \vee x \in B\}, \quad A \cap B = \{x : x \in A \wedge x \in B\}, \quad A - B = \{x : x \in A \wedge x \notin B\}$$

We can also represent these using *Venn diagrams*:

Example. Let $A = \{1, 3, 5\}$ and $B = \{1, 2, 3\}$.

Then $A \cup B = \{1, 2, 3, 5\}$, $A \cap B = \{1, 3\}$, $A - B = \{2\}$.

We can extend union and intersection to several (even infinitely many) sets:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n \qquad \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

A *universal set* for sets A and B is any set containing both of them. In most cases, there will be a natural choice for U. For example, if A and B are both sets of integers, then we may take $U = \mathbb{Z}$.

Definition: Complement

Let A and be a subset of a universal set U. The *complement of* A (in U) is $\overline{A} = U - A$.

If A and B are subsets of a universal set U, then $A - B = A \cap \overline{B}$.

Example. If U is the English alphabet and V is the set of vowels, then $\overline{V}$ is the set of consonants.

We now turn to verifying several set identities, which is similar to verifying logical equivalences.

Example. We will prove the first De Morgan's law for set identities: $\overline{A \cap B} = \overline{A} \cup \overline{B}$:

$$\begin{aligned}\overline{A \cap B} &= \{x \mid x \notin (A \cap B)\} = \{x \mid \neg(x \in (A \cap B))\} = \{x \mid \neg(x \in A \wedge x \in B)\} \\ &= \{x \mid \neg(x \in A) \vee \neg(x \in B)\} = \{x \mid x \notin A \vee x \notin B\} = \{x \mid x \in \overline{A} \vee x \in \overline{B}\} \\ &= \{x \mid x \in \overline{A} \cup \overline{B}\} = \overline{A} \cup \overline{B}.\end{aligned}$$

Alternatively, to prove two sets are equal, we can show that each side is a subset of the other side.

Example. Prove the distributive law for union: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Proof. Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$, so $x \in (A \cup B) \cap (A \cup C)$. If $x \in B \cap C$, then $x \in B$ so $x \in A \cup B$, and $x \in C$ so $x \in A \cup C$. Again we have $x \in (A \cup B) \cap (A \cup C)$. In either case we have $A \cup (B \cap C) \in (A \cup B) \cap (A \cup C)$.

Suppose $y \in (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$. Suppose $y \in A$, then $y \in A \cup (B \cap C)$. Now suppose $y \notin A$. Since $y \in A \cup B$, then $y \in B$. Similarly, since $y \in A \cup C$, then $y \in C$. Thus, $y \in B \cap C$ so again we have $y \in A \cup (B \cap C)$. In either case we have $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$. $\square$

The next result gives a general way to count the number of elements in a union.

Inclusion-Exclusion Principle

Let A and B be sets. Then $|A \cup B| = |A| + |B| - |A \cap B|$.

2.3 Functions

Definition: Function

Let A and B be nonempty sets. A *function* f from A to B, denoted $f : A \to B$, is an assignment of exactly one element of B to each element of A.

If $a \in A$, then $f(a) = b$ where $b \in B$ is the unique element assigned by the function f to $a \in A$.

Definition: domain, codomain, range/image

If f is a function from A to B, we say A is the *domain* of f and B is the *codomain* of f. If $f(a) = b$, then we say b is the *image* of a and a is a *preimage* of b. The *range* (or *image*) of f is the set of all images of elements of A.

We were careful to say that b is *the* image but a is *a* preimage, as preimages need not be unique.

Example (Examples of functions)**.** 1) Define g to be the function that assigns to each student in a class their current grade. The domain is the set of students in the class and the codomain is the set $\{A, B, C, D, F\}$. So $g(\text{Bob}) = F$, because Bob has never shown up for this class.

2) Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by $f(x) = x^2$. The domain of f is $\mathbb{Z}$ and the codomain is $\mathbb{Z}$. The image of f is the set of all integers that are perfect squares $\{0, 1, 4, 9, \dots\}$. (A function is *real-valued* if its codomain is $\mathbb{R}$, and *integer-valued* if its codomain is $\mathbb{Z}$.)

3) (Pointwise addition and multiplication) Let f_1 and f_2 be functions from A to $\mathbb{R}$. Then $f_1 + f_2$ and $f_1 f_2$ are also functions from A to $\mathbb{R}$ defined for all $x \in A$ by

$$(f_1 + f_2)(x) = f_1(x) + f_2(x), \qquad (f_1 f_2)(x) = f_1(x) f_2(x).$$

We now generalize the notion of image of a function.

Definition: Image of a subset

Let $f : A \to B$ be a function and let $S \subseteq A$. The *image* of S under f, denoted $f(S)$, is the subset B that consists of the images of the elements of S.

In set-builder notation, $f(S) = \{t \mid \exists s \in S(t = f(s))\} = \{f(s) \mid s \in S\}$. The image of the function f is then $f(A) = \{f(a) : a \in A\}$.

Definition: Graph of a function

Let $f : A \to B$ be a function. The *graph* of f is the subset $\{(a, b) : a \in A, b = f(a)\}$.

Definition: Injective, surjective, bijective

Let $f : A \to B$ be a function.

• The function f is said to be *one-to-one* (or *injective*) if and only if $f(a) = f(b)$ implies $a = b$ for all $a, b \in A$.

• The function f is said to be *onto* (or *surjective*) if and only if for every $b \in B$ there exists an element $a \in A$ such that $f(a) = b$.

• The function f is said to be a *one-to-one correspondence* (or *bijective*) if it is both one-to-one and onto.

Example (Diagrams)**.**

Example. (1) Let $f : \mathbb{Z} \to \mathbb{Z}$ be the function given by $f(n) = 2n$.

Suppose $f(m) = f(n)$ for integers m, n. Then $2m = 2n$, so $m = n$ (by algebra). Thus, f is 1-1.

However, f is not onto because the image consists only of even integers. In particular, if $f(n) = 1$, then $2n = 1$, so $n = \frac{1}{2} \notin \mathbb{Z}$.

(2) Let $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ be the function given by $f(m, n) = 2m - n$.

Since $f(1, 0) = 2 = f(0, -2)$ and $(1, 0) \neq (0, -2)$, then f is not 1-1.

If $z \in \mathbb{Z}$, then $f(0, -z) = 2(0) - (-z) = z$, so f is onto. (Note also that $f(z, z) = 2z - z = z$, but it is only necessary to produce one preimage of an arbitrary element.)

(3) Let $h : \mathbb{R} \to \mathbb{R}$ be given by $h(x) = 2x + 1$.

If $h(x) = h(y)$ for $x, y \in \mathbb{R}$, then $2x + 1 = 2y + 1$ and by algebra we have $x = y$. Thus, h is 1-1.

If $r \in \mathbb{R}$, then $h\left(\frac{1}{2}(r - 1)\right) = r$, so h is onto. Thus, h is a one-to-one correspondence.

(4) Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then f is not one-to-one or onto. However, if we restrict the domain $f : \mathbb{R}^+ \to \mathbb{R}^+$ then it is bijective.

Suppose $f(a) = f(b)$ for $a, b \in \mathbb{R}^+$. Then $a^2 = b^2$. Taking square roots and noting that the domain consists only of positive real numbers we have $a = b$. So, f is 1-1.

Let $r \in \mathbb{R}^+$. Note that $\sqrt{r} \in \mathbb{R}^+$ and $f(\sqrt{r}) = (\sqrt{r})^2 = r$. Hence, f is onto.

The natural way to combine functions is to use composition, but we must be careful with domains.

Definition: Composition of functions

Let $g : A \to B$ and $f : B \to C$ be functions. The *composition* $f \circ g : A \to C$ is defined by the rule $(f \circ g)(a) = f(g(a)) \quad$ for all $a \in A$.

(Draw composition diagram)

Even when $f \circ g$ is defined then $g \circ f$ may not be when $C \neq A$. Furthermore, even when both are defined, generally we have that $f \circ g \neq g \circ f$. (That is, function composition is *noncommutative*.)

Example (Composition)**.** Consider $f(x) = x^2$ and $g(x) = x + 1$ (both functions defined $\mathbb{R} \to \mathbb{R}$).

The notion of composition is inherently tied to that of invertibility of functions.

Definition: Inverse function

Let $f : A \to B$ be a bijective function. The *inverse function* of f, denoted f^{-1} is the function that assigns to an element $b \in B$ the unique element $a \in A$ such that $f(a) = b$.

If $f : A \to B$ is bijective with inverse function $f^{-1} : B \to A$, then for any $a \in A$ and $b \in B$ we have $f^{-1}(b) = a$ when $f(a) = b$. So,

$$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = a \qquad (f \circ f^{-1})(b) = f(f^{-1}(b)) = b.$$

Thus, $f^{-1} \circ f = \iota_A$ and $f \circ f^{-1} = \iota_B$.

Example (Inverses)**.** (1) Let $f : \mathbb{Z} \to \mathbb{Z}$ be given by $f(x) = x + 1$. This function is bijective and so it has an inverse. If we choose $y \in \mathbb{Z}$, then $f(y - 1) = (y - 1) + 1 = y$. Hence, the inverse function of f is $f^{-1} : \mathbb{Z} \to \mathbb{Z}$ given by $f^{-1}(y) = y - 1$.

(2) The inverse of $g : \mathbb{R}^+ \to \mathbb{R}^+$ given by $g(x) = x^2$ has inverse function $g^{-1}(y) = \sqrt{y}$.

(3) The function $h : \mathbb{R}^+ \to \mathbb{R}^+$ given by $h(x) = e^x$ has inverse function $h^{-1} : \mathbb{R}^+ \to \mathbb{R}^+$ given by $h^{-1}(y) = \ln(y)$.

2.4 Sequences

We normally think of a sequence as an (infinite) ordered set. However, it will be convenient to describe them more generally.

Definition: Sequence

A *sequence* is a function $f : D \to S$ where D is a subset of $\mathbb{Z}$ and S is a set. We call $a_n = f(n)$ for $n \in D$ a *term* of the sequence and denote the sequence by $\{a_n\}$ for $N \in D$.

Example. Write the first five terms of the sequence $\{a_n\}$ where $a_n = 1/n$.

The sequence begins $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$

Definition: Geometric progression

Let $a, r \in \mathbb{R}$. A *geometric progression* (or *geometric sequence*) is a sequence of the form

$$a, ar, ar^2, \ldots, ar^n, \ldots.$$

We call a the *initial term* and r the *common ratio*.

The function in a geometric progression is $f(x) = ar^x$. Our indexing set is $\mathbb{N}$ (starts with $n = 0$).

Example. The sequence $\{b_n\}$ with $b_n = \frac{1}{2}(-2)^n$ is a geometric progression ($a = 1/2$, $r = (-2)^n$). The sequence that begins, $\frac{1}{2}, -1, 2, -4, \ldots$

Example. Suppose we have a geometric progression whose first term is $4! = 64$ and third term is 96. Thus, $a_0 = a = 64$ and $a_3 = ar^2 = 96$. Thus, $64r^2 = 96$ so $r^2 = 3/2$. Hence, $r = \sqrt{3/2}$.

Definition: Arithmetic progression

Let $a, d \in \mathbb{R}$. An *arithmetic progression* (or *arithmetic sequence*) is a sequence of the form

$$a, a + d, a + 2d, \ldots, a + nd, \ldots.$$

We call a the *initial term* and d the *common difference*.

Example. The sequence $\{s_n\}$ with $s_n = 2 - 5n$ is an arithmetic progression ($a = 2$, $d = -5$). The sequence begins with $2, -3, -8, -13, \ldots$.

Example. Suppose an arithmetic progression has first term 5 and fourth term 14. That is, $a_0 = 5$ and $a_3 = 14$ (careful with indexing). Then we have $5 + d(3) = 14$. Solving gives $d = 3$. Hence, the arithmetic progression is $\{a_n\}$ with $a_n = 5 + 3n$.

Many important sequences can be defined recursively.

Definition: Recurrence relation

A *recurrence relation* for the sequence $\{a_n\}$ is an equation that expresses a_n in terms of one or more of the previous terms of the sequence. A sequence is called a *solution* of a recurrence relation if its terms satisfy the recurrence relation.

We call the first few terms necessary to start the recurrence relation the *initial conditions*.

Example. (1) Let $\{a_n\}$ be a sequence that satisfies the recurrence relation $a_n = a_{n-1} + 2$ and suppose $a_0 = 5$. Then the sequence begins $5, 7, 9, 11, \ldots$

(2) Recall that for a positive integer n, the *factorial* of n is $n! = n \cdot (n-1) \cdot (n-2) \ldots 1$. The sequence $\{a_n\}$ where $a_n = n!$ can be defined recursively by $a_1 = 1$ and $a_n = na_{n-1}$ for $n > 1$.

(3) Suppose we have a sequence $\{a_n\}$ where $a_0 = 3$ and for $n > 1$, $a_n = 2^{a_{n-1}}$. Then

$$a_1 = 2^{a_0} = 2^1 = 2 \quad \text{and} \quad a_2 = 2^{a_1} = 2^2 = 4.$$

(4) The *Fibonacci sequence* $\{f_n\}$ is defined by the initial conditions $f_0 = 0$, $f_1 = 1$, and satisfies the recurrence relation $f_n = f_{n-1} + f_{n-2}$. The sequence begins $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$

Some recurrence relations can be resolved to determine a *closed formula* for the terms a_n that does not depend on earlier terms in the sequence.

Example. Consider our sequence that satisfies the recurrence relation $a_n = a_{n-1} + 2$ and suppose $a_0 = 5$. We could simply write down the first few terms and see that $a_n = 5 + 2n$.

Given a sequence, we can add the terms[1]. The notation we will use for this is called *summation notation* (or *sigma notation*). Suppose we have the following terms in a sequence $\{a_n\}$: $a_m, a_{m+1}, \ldots, a_n$. We express the sum as

$$\sum_{j=m}^{n} a_j = a_m + a_{m+1} + \cdots + a_n.$$

The variable j is the *index of summation*, m is the *lower limit*, and n is the *upper limit*.

Example. Suppose we have the sequence $\{a_n\}$ where $a_n = n^2$. The sum of the first 5 terms is

$$\sum_{j=1}^{5} j^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 1 + 4 + 9 + 16 + 25 = 55.$$

Wouldn't it be nice if there were easy formula to compute such a sum? Hmm....

[1]Students have undoubtably seen this in the context of Riemann sums. Those who have taken Calc II have seen *infinite series* but we will only be dealing with *finite sums*.

(Only if time...)

Example. Recall our geometric progression $\{b_n\}$ with $b_n = \frac{1}{2}(-2)^n$. (Recall here $a = 1/2$, $r = (-2)^n$). Then the sum of the first 4 terms is

$$\sum_{j=0}^{3} = \frac{1}{2} + -1 + 2 + -4 = -5/2.$$

There is a nice formula for computing the sum of the first $(n+1)$ terms of a geometric progression. First note that if $r = 1$, then the sum $\sum_{j=0}^{n} ar^j$ is just $a + a + \cdots + a = (n+1)a$. Suppose $r \neq 1$ and set

$$S_n = \sum_{j=0}^{n} ar^j.$$

Then by reindexing we have

$$rS_n = \sum_{j=0}^{n} ar^{j+1} = \sum_{j=1}^{n+1} ar^j.$$

So these two sums are the same except for the first summand of S_n and the last summand of rS_n. This implies that

$$\begin{aligned} S_n - rS_n &= a - ar^{n+1} \\ (1-r)S_n &= a(1-r^{n+1}) \\ S_n &= \frac{a(1-r^{n+1})}{1-r}. \end{aligned}$$

Now check this with the previous example.

2.5 Cardinality of Sets

We now discuss how one compare the "sizes" of infinite set.

Definition: Cardinality

Nonempty sets A and B are said to have the same *cardinality* if there exists a bijection $f : A \to B$.

When two sets have the same cardinality, we write $|A| = |B|$.

Recall that when a set S is finite, say with n elements, then there is a bijection $f : \{1, \ldots, n\} \to S$ and we write $|S| = n$ (or $|S| = n$). So cardinality is really only interesting with regards to infinite sets. But, not all infinities are created equal.

If there exists an injective function $f : A \to B$, then we write $|A| \leq |B|$. If there exists an injection but no bijection, then we write $|A| < |B|$. Similarly, if there exists a surjective function $f : A \to B$, then $|B| \leq |A|$.

Definition: Countable, uncountable

A set that is either finite or has the same cardinality as the positive integers $\mathbb{N}$ is said to be *countable*. A set that is not countable is said to be *uncountable*.

We write $|\mathbb{N}| = \aleph_0$ ("aleph null").

Example. Here we demonstrate several countable sets.

(2) The positive integers $\mathbb{Z}^+$ (starting at 0) are countable. Define a map $f : \mathbb{N} \to \mathbb{Z}^+$ by $f(n) = n+1$. (Hilbert's Grand Hotel)

(1) The positive even integers, E^+, are countable. Define a map $f : \mathbb{Z}^+ \to E^+$ by $f(n) = 2n$. It is clear that this function is both injective and surjective. Hence it is bijective. A similar proof works for the positive odd integers.

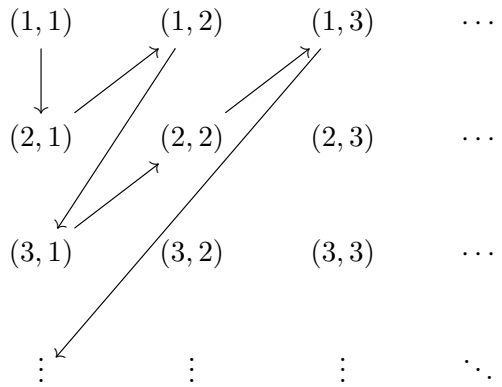
(3) The integers $\mathbb{Z}$ are countable. We could set up a formal correspondence, or we could simply note that we can list the integers in a specified order:

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

For those interested, we could define a bijection $f : \mathbb{N} \to \mathbb{Z}$ in the following way:

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

Example. We will show that $\mathbb{Q}^+$, the set of positive rationals, is countable. To do this, we set up an ordering



Example. On the other hand, $\mathbb{R}$ is not countably infinite. To do this, we use the decimal expansion of a real number.

Let $f : \mathbb{N} \to \mathbb{R}$ be a function. Choose an element $b = b_0.b_1b_2b_3b_4 \ldots \in \mathbb{R}$ in the following way:

- Choose b_0 so that it is different from the whole part of $f(0)$.
- Choose b_1 so that is is different from the first decimal place of $f(1)$.
- Choose b_2 so that is is different from the second decimal place of $f(2)$.
- And so on...
- Choose b_k so that is is different from the kth decimal place of $f(k)$.

Then for every n, b differs from $f(n)$ in at least one decimal place, so $f(n) \neq b$ for every n. That is, f is not surjective. Since f was chosen arbitrarily, this implies that there is no surjective function (and hence no bijective function) $\mathbb{N} \to \mathbb{R}$. However, there is an obvious injective map $\mathbb{N} \to \mathbb{R}$ (just send each element of $\mathbb{N}$ to itself in $\mathbb{R}$), so $|\mathbb{N}| < |\mathbb{R}|$.

Induction and Recursion

5.1 Mathematical Induction

Suppose that we can prove that it is possible to reach the first step of a staircase. Then we prove that from *any* step, we are able to get to the next step. By showing these two things we have now shown that it is possible to climb the entire staircase, no matter how many steps there are (possible infinite!). This illustrates the mathematical tool called *induction*. It is related in a very close way to the idea of recursion in programming.

Principle of Mathematical Induction

To prove that $P(n)$ is true for all positive integers n, where $P(n)$ is a propositional function, we complete two steps:

Basis step (base case): We verify that $P(1)$ is true.

Inductive step: We show that the conditional statement $P(k) \to P(k+1)$ is true for all positive integers k.

Depending on the situation, our basis step might be $k = 1$, but it might be $k = 0$ or some other value. The assumption (that $P(k)$ is true) is called the *inductive hypothesis*. It is important to indicate where we use the inductive hypothesis in our proof.

Example (Sum of the first n positive integers)**.** The sum of the first positive integer is 1. The sum of the first two is $1 + 2 = 3$. The sum of the first three is $1 + 2 + 3 = 6$. Finally, the sum of the first four is $1 + 2 + 3 + 4 = 10$. At this point, we might formulate a guess that the sum of the first n positive integers is $n(n+1)/2$. For n a positive integer, let $P(n)$ be the proposition that $1 + 2 + \cdots + n = n(n+1)/2$. (Base case) Since $1 = 1(1+1)/2$, then $P(1)$ is true.

Now suppose that for some positive integer k, $P(k)$ is true. That is, $1 + 2 + \cdots + k = k/(k+1)/2$ (inductive hypothesis). Now we must show that $P(k+1)$ is true. The key observation here is that $P(k+1)$ is the sum of the first $(k+1)$ integers, so it is the sum of the first k integers plus $(k+1)$. Using our inductive hypothesis and algebra we have

$$\begin{aligned} 1 + 2 + \cdots + (k+1) &= (1 + 2 + \cdots + k) + (k+1) \\ &= \left(\frac{k(k+1)}{2}\right) + (k+1) \quad \text{(by the inductive hypothesis)} \\ &= \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+1)(k+2)}{2}, \end{aligned}$$

which is the statement for $P(k+1)$. Since we have now verified $P(k+1)$, the given statement is true by the principle of mathematical induction.

Example (Sum of the first n positive odd integers)**.** We check the first few cases:

$$1 = 1, \qquad 1 + 3 = 4, \qquad 1 + 3 + 5 = 9, \qquad 1 + 3 + 5 + 7 = 16.$$

From this, we might conjecture that the sum of the first n positive odd integers is

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

(Base case) For n a positive integer, let $P(n)$ be the proposition that the sum of the first n positive integers is n^2. Since $1 = 1^2$, then $P(1)$ is true. (Inductive hypothesis) Suppose that $P(k)$ is true for some $k \geq 1$. That is, $1 + 2 + \cdots + (2k - 1) = k^2$. (Inductive step) Now we must show that $P(k + 1)$ is true. Using our inductive hypothesis and algebra we have

$$\begin{aligned} 1 + 2 + \cdots + (2k + 1) &= (1 + 2 + \cdots + (2k - 1)) + (2k + 1) \\ &= k^2 + 2k + 1 \quad \text{(by the inductive hypothesis)} \\ &= (k + 1)^2, \end{aligned}$$

which is the statement for $P(k + 1)$. Since we have now verified $P(k + 1)$, the given statement is true by the principle of mathematical induction.

In some cases, we may use summation notation to simplify an expression.

Example (Summation example)**.** Prove that for every positive integer n, $\sum_{j=1}^{n} j2^j = (n-1)2^{n+1} + 2$.

Let $P(n)$ be the proposition that $\sum_{j=1}^{n} j2^j = (n - 1)2^{n+1} + 2$. If $n = 1$ then we have

$$\sum_{j=1}^{1} j2^j = 1 \cdot 2^1 = 2 = (1 - 1)2^2 + 2.$$

Hence, $P(1)$ is true.

Now suppose that $P(k)$ is true for some positive integer k. Then we have

$$\begin{aligned} \sum_{j=1}^{k+1} j2^j &= (k + 1)2^{k+1} + \sum_{j=1}^{k} j2^j \\ &= (k + 1)2^{k+1} + \Big((k - 1)2^{k+1} + 2\Big) \quad \text{(by the inductive hypothesis)} \\ &= 2^{k+1}\left((k + 1) + (k - 1)\right) + 2 \\ &= 2^{k+1}(2k) + 2 \\ &= k2^{k+2} + 2, \end{aligned}$$

which is the statement for $P(k + 1)$. Since we have now verified $P(k + 1)$, the given statement is true by the principle of mathematical induction.

Induction can be used to verify inequalities, along with a host of other statements about the integers.

Example (Inequality example)**.** Prove that $3^n < n!$ if n is an integer greater than 6.

Let $P(n)$ be the proposition that $3^n < n!$. (Note that this is actually false for $n \geq 6$.) One checks (using a calculator, presumably) that $3^7 = 2187$ and $7! = 5040$, so $P(7)$ is true.

Now suppose that $P(k)$ is true for some $k \geq 7$ (that is, $3^k < k!$). Since $3 < k + 1$, then we have

$$\begin{aligned} 3^{k+1} &= 3^k \cdot 3 \\ &< k! \cdot 3 \quad \text{(by the inductive hypothesis)} \\ &< k! \cdot (k+1) \\ &= (k+1)! \end{aligned}$$

This is exactly the statement for $P(k+1)$. Since we have now verified $P(k+1)$, the given statement is true by the principle of mathematical induction.

Our next example uses sets instead of numerical values.

Example (Set theory example)**.** Prove that if $A_1, A_2, \ldots, A_n$ and B are sets, then

$$\left(\bigcap_{j=1}^{n} A_j \right) \cup B = \bigcap_{j=1}^{n} (A_j \cup B).$$

Written out without summation notation, the above statement says

$$(A_1 \cap A_2 \cap \cdots \cap A_n) \cup B = (A_1 \cup B) \cap (A_2 \cup B) \cap \cdots \cap (A_n \cup B).$$

Let $P(n)$ be the statement that the above identity holds for the positive integer n. Note that $n = 1$ is trivial. We will use $n = 2$ as our basis step. In this case, the distributive law for sets gives

$$(A_1 \cap A_2) \cup B = (A_1 \cup B) \cap (A_2 \cup B).$$

That is, $P(2)$ is true.

Assume $P(k)$ is true for some integer $k \geq 2$. Then

$$\begin{aligned} (A_1 \cap A_2 \cap \cdots \cap A_{k+1}) \cup B &= ((A_1 \cap A_2 \cap \cdots \cap A_k) \cap A_{k+1}) \cup B \\ &= ((A_1 \cap A_2 \cap \cdots \cap A_k) \cup B) \cap (A_{k+1} \cup B) \quad \text{(distributive law)} \\ &= ((A_1 \cup B) \cap (A_2 \cup B) \cap \cdots \cap (A_k \cup B)) \cap (A_{k+1} \cup B) \quad \text{(inductive hypothesis)} \\ &= (A_1 \cup B) \cap (A_2 \cup B) \cap \cdots \cap (A_{k+1} \cup B) \quad \text{(associative law).} \end{aligned}$$

This is the statement for $P(k+1)$. Hence, the given statement is true by the principle of mathematical induction.

(Extra examples if time.)

Example. Let A be a finite set with $|A| = n$. Prove that $|\mathcal{P}(A)| = 2^n$.

Let $P(n)$ be the statement that $|\mathcal{P}(A)| = 2^n$ when $|A| = n$. If $|A| = 0$, then $A = \emptyset$. In this case $\mathcal{P}(A) = \{\{\emptyset\}\}$ so $|\mathcal{P}(A)| = 1$. Assume $P(k)$ is true and suppose that $|A| = k + 1$. Let $x \in A$. Set $B = A - \{x\}$, so that $|B| = k$. By the inductive hypothesis, $|\mathcal{P}(B)| = 2^k$. The subsets of A that contain x are obtained by adding x to each of the subsets of B. Hence, there are 2^k such subsets (again, by the inductive hypothesis). Thus, $|\mathcal{P}(A)| = 2^k + 2^k = 2(2^k) = 2^{k+1}$. Hence, $P(k + 1)$ is true and so the result holds by the principle of mathematical induction.

Finally, we consider an example involving divisibility.

Example. Prove that 6 divides $n^3 - n$ whenever n is a nonnegative integer.

(Here we take our base case to be $n = 0$.)

For n a nonnegative integer, let $P(n)$ be the statement that 6 divides $n^3 - n$. Clearly 6 divides $0^3 - 0 = 0$, so $P(0)$ is true.

Assume $P(k)$ is true for some integer $k \geq 0$. That is, 6 divides $k^3 - k$. Then

$$(k + 1)^3 - (k + 1) = (k^3 + 3k^2 + 3k + 1) - (k + 1) = (k^3 - k) + 3(k^2 + k).$$

Regardless of whether k is even or odd, $k^2 + k$ is even. That is, 2 divides $k^2 + k$. But then 6 divides $3(k^2 + k)$. By the inductive hypothesis, 6 divides $k^3 - k$. Hence, 6 divides $(k^3 - k) + 3(k^2 + k)$ so 6 divides $(k + 1)^3 - (k + 1)$. Thus, $P(k + 1)$ is true. Hence, the given statement is true by the principle of mathematical induction.

5.2 Strong induction

We now consider a variation on induction. In our staircase analogy, we now assume when we are on step k (inductive hypothesis) that we have climbed from step 1 to step k. **(Match game)**

Strong Induction

To prove that $P(n)$ is true for all positive integers n, where $P(n)$ is a propositional function, we complete two steps:

Basis step (base case): We verify that $P(1)$ is true.

Inductive step: We show that the conditional statement $[P(1) \wedge P(2) \wedge \cdots \wedge P(k)] \rightarrow P(k+1)$ is true for all positive integers k.

Example (Stamps)**.** Prove that we can make any postage of n cents using only 4-cent stamps and 7-cent stamps for $n \geq 18$.

Let $P(n)$ be the statement that a postage of n cents can be made using only these stamps. We claim $P(n)$ is true for $n \geq 18$. First we prove $P(n)$ for $n = 18, 19, 20, 21$.

$$18 = 7(2) + 4(1), \quad 19 = 7(1) + 4(3), \quad 20 = 7(0) + 4(5), \quad 21 = 7(3) + 4(0).$$

This proves our base cases. Now suppose that $P(k)$ is true for some $k \geq 21$. We claim that it is true for $k + 1$. Note that $18 \leq (k + 1) - 4 < k + 1$. Since $P(k - 3)$ is true by the inductive hypothesis, then we can write $k - 3$ using some combination of 4-cent and 7-cent stamps. Thus, we need only one more 4-cent stamp to produce $k + 1$. This proves $P(k + 1)$, so the result holds by strong induction.

Example (Products of primes)**.** Prove that if $n > 1$ is an integer, then n is a product of primes.

Recall that a positive integer p is *prime* if the only factors of p are 1 and p itself. An integer that is not prime is *composite*.

Let $P(n)$ be the proposition that n is a product of primes. Then $P(2)$ is true because 2 is itself prime (so it is a product of one prime). Now let k be some integer $k \geq 2$ and assume that $P(2), P(3), \ldots, P(k)$ are true. We must verify $P(k+1)$. There are two cases. The first is that $k+1$ is itself prime. Then there is nothing to prove and $P(k + 1)$ is true. Otherwise, $k + 1$ is composite. But then $k + 1 = ab$ where a, b are integers satisfying $1 < a, b < k + 1$. Since $P(a)$ and $P(b)$ are true by the (strong) inductive hypothesis, then a and b are products of primes. Thus $k + 1$ is a product of two products of primes, so it is itself a product of primes.

Though it is not proved above, this decomposition of an integer into a product of primes is unique up to rearranging the factors.

Example (Fibonacci)**.** Recall that the Fibonacci sequence is defined recursively by setting $f_0 = 0$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$. Let $\phi = (1 + \sqrt{5})/2$ (the golden ratio) and note that ϕ is a solution to the equation $x^2 - x - 1$. For $n \geq 3$, prove that $f_n > \phi^{n-2}$.

Let $P(n)$ be the statement that $f_n > \phi^{n-2}$. Since

$$f_3 = 2 > \phi \quad \text{and} \quad f_4 = 3 > (3 + \sqrt{5})/2 = \phi^2$$

then $P(3)$ and $P(4)$ are true.

Assume that $P(3), P(4), \ldots, P(k)$ are true for some $k \geq 4$. Then,

$$\begin{aligned}
\phi^{(k+1)-2} &= \phi^{k-1} = \phi^2 \phi^{k-3} \\
&= (\phi + 1)\phi^{k-3} \quad (\text{since } \phi^2 = \phi + 1) \\
&= \phi^{k-2} + \phi^{k-3} \\
&< f_k + f_{k-1} \quad (\text{by the inductive hypothesis}) \\
&= f_{k+1}.
\end{aligned}$$

Hence, $P(k+1)$ is true and so the result holds by strong induction.

5.3 Recursion

We have already discussed recursive sequences. We can now recursively define functions after specifying initial values. For example, given a sequence $\{a_n\}$, one can define the summation $\sum_{k=0}^{n} a_k$ recursively by setting $S_0 = a_0$ and $S_n = S_{n-1} + a_n$ for $n \geq 1$.

Recursive function

A *recursively defined function* is a function whose domain is the set of nonnegative integers and is defined recursively as follows:

Basis step: Specify the value of the function at zero.

Recursive step: Give a rule for finding its value at an integer from its values at smaller integers.

Example (Recursive function)**.** Let $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ be a function defined recursively by $f(1) = 1$, and $f(n) = 2f(n-1)$ for $n \geq 1$. This is well-defined because for every $n \in \mathbb{Z}^+$, we can determine the successive value of $f(n)$ and the output is an element of $\mathbb{N}$. The first few values of f are:

$$f(1) = 1, \quad f(2) = 2f(1) = 2, \quad f(3) = 2f(2) = 4, \quad f(4) = 2f(3) = 8.$$

So we see that for $n \geq 1$, f could be expressed by the closed formula $f(n) = 2^{n-1}$. We can prove that this formula is correct by (standard) induction.

Since $f(1) = 1 = 2^{1-1}$, then the base case is satisfied. Suppose $f(k) = 2^{k-1}$ holds for some $n = k \geq 1$. Then we have

$$f(k+1) = 2f(k) \overset{\star}{=} 2 \cdot 2^{k-1} = 2^k,$$

where the $\star$ step used the inductive hypothesis. Thus, the result holds at the step $n = k+1$. Thus, the formula holds for all $n \geq 1$.

Recursion can also be used to define sets.

Example (Multiples of 3)**.** Define a set S by the basis step $3 \in S$ and the recursive step is the rule: "If $x \in S$ and $y \in S$, then $x + y \in S$." We claim this defines the set of positive multiples of 3.

Let A be the set of positive multiples of 3. We claim $A = S$. We will prove this by showing that $A \subset S$ and $S \subset A$. Both directions will rely on induction.

Let $3n \in A$ and let $P(n)$ be the statement that $3n \in S$. Then $P(1)$ is true by hypothesis. Assume $P(k)$ is true for some $k \geq 1$. Since $3k \in S$ by the inductive hypothesis and $3 \in S$ by our given hypothesis, then $3(k+1) = 3k+3 \in S$ by the inductive rule for S. Thus, $P(k+1)$ is true so $A \subseteq S$.

Now we show the other inclusion. Our basis step in this case is simply to notice that $3 \in A$. Now assume that all elements $k \in S$ are also in A, for some $k \geq 3$. By the recursive definition, $k + 1 = x + y$ for $x, y \in S$. But then $x, y < k + 1$, so by the inductive hypothesis, 3 divides x and 3 divides y. Thus, 3 divides $x + y$ and so $k + 1 \in A$. This shows $S \subseteq A$.

As another example, let Σ be a set (called an *alphabet*), then Σ^* is the set of *strings* over the alphabet Σ. This is defined recursively. The basis step is the empty string $\lambda \in \Sigma^*$. The inductive step is to say that if $x \in \Sigma^*$ and $y \in \Sigma^*$, then $xy \in \Sigma^*$ (that is, the concatenation of x and y).

Example (Bit strings)**.** Let $\Sigma = \{0, 1\}$. Any string in Σ^* then is a combination of 0s and 1s. We have the empty string λ from the basis step. The first application of the recursive definition gives the strings 0 and 1. Using the recursive definition again we have the strings 00, 01, 10, and 11. Hence, Σ^* denotes the set of bit strings.

Now let Σ be an alphabet and let Σ^* be the set of strings on Σ. We can define the operation of *concatenation* on strings. For the basis step, we say that if $w \in \Sigma^*$, then $w \cdot \lambda = w$ (again, $\lambda \in \Sigma^*$ denotes the empty string). Recursively, if $w_1, w_2 \in \Sigma^*$ and $x \in \Sigma$, then $w_1 \cdot (w_2 x) = (w_1 \cdot w_2)x$. (We generally omit the $\cdot$ when writing the concatenation.)

Example (Length of a string)**.** We define the length of a string recursively. Our notation will be $\ell(w)$ for $w \in \Sigma^*$. First we set $\ell(\lambda) = 0$. Then we set $\ell(wx) = \ell(w) + 1$ for $w \in \Sigma^*$ and $x \in \Sigma$. In this way, the length of a string represents the number of elements of Σ that make it up.

We will show that this definition is consistent with concatenation. Let $y \in \Sigma^*$ and let $P(y)$ be the claim that $\ell(xy) = \ell(x) + \ell(y)$ for *any* string $x \in \Sigma^*$. For the basis step, we verify that $P(\lambda)$ is true. Note if $x \in \Sigma^*$ then

$$\ell(x\lambda) = \ell(x) = \ell(x) + 0 = \ell(x) + \ell(\lambda).$$

Now suppose that the result holds for *some* string $y \in \Sigma^*$ and let $a \in \Sigma$. Then

$$\begin{aligned} \ell(xya) &= \ell(xy) + 1 && \text{(by definition of length)} \\ &= \ell(x) + \ell(y) + 1 && \text{(by inductive hypothesis)} \\ &= \ell(x) + \ell(ya) && \text{(again by definition of length).} \end{aligned}$$

This proves the result for any string by induction. Formally, this is an example of *structural induction*.

Though we will study graphs in more detail later, this is a convenient opportunity to introduce a special type. A *graph* consists of vertices and edges which connect pairs of vertices.

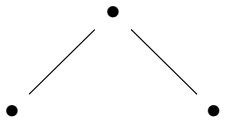
Full binary trees

The set of full binary trees can be defined recursively.

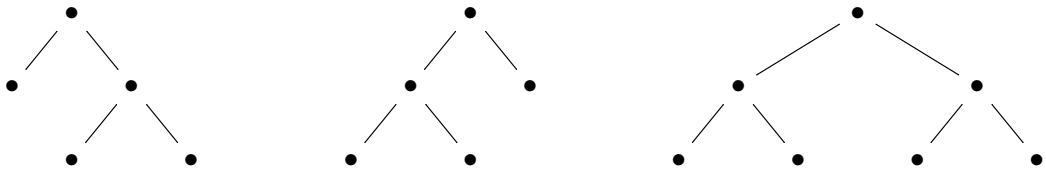
Basis step: There is a full binary tree consisting only of a single vertex r.

Recursive step: If T_1 and T_2 are disjoint full binary trees, there is a full binary tree, denoted $T_1 \cdot T_2$ consisting of a root r together with edges connecting the root to each of the roots of the left subtree T_1 and the right subtree T_2.

Example (Full binary trees)**.** At the basis step, we have only the vertex •. At step 1 we have



At step 2 we have the trees



The height $h(T)$ of a full binary tree T is defined recursively. For a full binary tree consisting of only one root, we set $h(T) = 0$. Now if T_1, T_2 are full binary trees and $T = T_1 \cdot T_2$, then $h(T) = 1 + \max(h(T_1), h(T_2))$. Similarly, we let $n(T)$ denoted the number of vertices of a full binary tree. When T has only one vertex, $n(T) = 1$ and otherwise $T = T_1 \cdot T_2$ and $n(T) = 1 + n(T_1) + n(T_2)$.

Example. Let T be a full binary tree. We will show that $n(T) \leq 2^{h(T)+1} - 1$.

If T is the full binary tree with only one root, then $n(T) = 1$ and $h(T) = 0$ (by definition), so $n(T) = 1 \leq 2^{0+1} - 1 = 1$.

For the inductive hypothesis, assume that $n(T_1) \leq 2^{h(T_1)+1} - 1$ and $n(T_2) \leq 2^{h(T_2)+1} - 1$ where T_1 and T_2 are full binary trees. Then

$$\begin{aligned} n(T) &= 1 + n(T_1) + n(T_2) \\ &\leq 1 + \left(2^{h(T_1)+1} - 1\right) + \left(2^{h(T_2)+1} - 1\right) \\ &\leq 2 \cdot \max\left(2^{h(T_1)}, 2^{h(T_2)}\right) - 1 \\ &= 2 \cdot 2^{\max(h(T_1), h(T_2))} - 1 \\ &= 2 \cdot 2^{h(T)} - 1 = 2^{h(T)+1} - 1. \end{aligned}$$

Note that we used the following algebraic fact $\max(2^x, 2^y) = 2^{\max(x,y)}$.

Counting

6.1 The basics of counting

In this chapter we introduce, formally, the study of Combinatorics. This area of math studies (complicated) counting problems, as well as connections between various types of counting problems. Some parts of this you are undoubtably aware of (e.g., binomial coefficients and basic probability) but other rules will be less familiar. We begin with two basic rules.

The Product Rule

Suppose a procedure can be broken down into two tasks. If there are n_1 ways to do the first task and there are n_2 ways to do the second, then there are $n_1 n_2$ ways to do the procedure.

The product rule scales inductively. If a procedure can be broken down into m tasks and for task k there are n_k ways to perform that task, then there are $n_1 \dot{n}_2 \dots n_m$ ways to do the procedure.

Example. (1) Suppose we have 8 boxes which can each hold one ball, and we have two balls. How many ways are there to assign a ball to a box? For the first ball, there are 8 choices. Then for the second ball there are only 7 choices. Hence there are $8 \cdot 7 = 56$ total ways to assign balls to boxes.

(2) Consider the last example but with four balls. Continuing our logic from before, there are $8 \cdot 7 \cdot 6 \cdot 5 = 1680$ ways to assign the four balls to different boxes.

(3) Suppose a license plate is made up of three letters followed by three digits (0-9). How many possible license plates combinations are there? There are 26 choices for each letter and there are 10 choices for each digit. Since we allow repeats, there are $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17,576,000$ possible combinations.

The Sum Rule

If a task can be done either one of n_1 ways or in one of n_2 ways, where none of the set of n_1 ways is the same as any of the of n_2 ways, then there are $n + 1 + n_2$ ways to do the task.

Example (Greek council)**.** Suppose a representative for the Greek council[1] from either the sorority $\Gamma\Gamma\Gamma$ or the fraternity $\Upsilon\Upsilon\Upsilon$. There are 37 members of $\Gamma\Gamma\Gamma$ and 52 members of $\Upsilon\Upsilon\Upsilon$. Hence, there are $52 + 37 = 89$ ways to choose a representative.

[1] Totally a real thing, right?

We can state these two rules in terms of sets. Let $A_1, A_2, \ldots, A_m$ be finite sets. Then the number of elements in the cartesian product of these sets is

$$|A_1 \times A_2 \times \cdots \times A_m| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_m|.$$

Choosing an element in the cartesian product involves choosing exactly one element from each set. One the other hand, if $A_1, A_2, \ldots, A_m$ are pairwise disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$) finite sets, then

$$|A_1 \cup A_2 \cup \cdots \cup A_m| = |A_1| + |A_2| + \cdots + |A_m|.$$

Example. A certain password is either eight or nine characters long. Each character is either an upper case letter or a digit (0-9). Each password must contain at least one letter and one digit. How many passwords are possible.

Let P_8 be the number of passwords that are eight characters long. Ignoring the extra rule, there are 36^8 possible passwords. On the other hand, there are 26^8 passwords that contain *only* letters. There are 10^8 passwords that contain *only* digits. These two sets have nothing in common. Thus,

$$P_8 = 36^8 - 26^8 - 10^8 = 2,612,182,842,880.$$

A similar computation shows that

$$P_9 = 36^9 - 26^9 - 10^9 = 96,129,452,989,440.$$

Thus, the total number of passwords is

$$P_8 + P_9 = 98,741,635,832,320.$$

The Subtraction Rule

If a task can be done either one of n_1 ways or in one of n_2 ways, then the number of ways to do the task is $n + 1 + n_2$ minus the number of ways to do the task that are common to both ways.

This is just the *principle of inclusion-exclusion*. If A_1 and A_2 are finite sets, then

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

Example. How many bit strings of length six start or end with 1?

A bit string consists of either a 0 or 1, and so there are $2^6 = 64$ total bit strings of length six. If a bit string starts with 1, then there are $2^5 = 32$ choices for the remaining bits. Similarly, there are 32 strings that end with 1. If a bit string starts *and* ends with 1, then there are $2^4 = 16$ choices for the remaining four bits. Consequently, there are $(32 + 32) - 16 = 48$ bit strings that start or end with 1.

The Division Rule

There are n/d ways to do a task if it can be done using a procedure carried out in n ways, and for every way w, exactly d of the n ways correspond to way w.

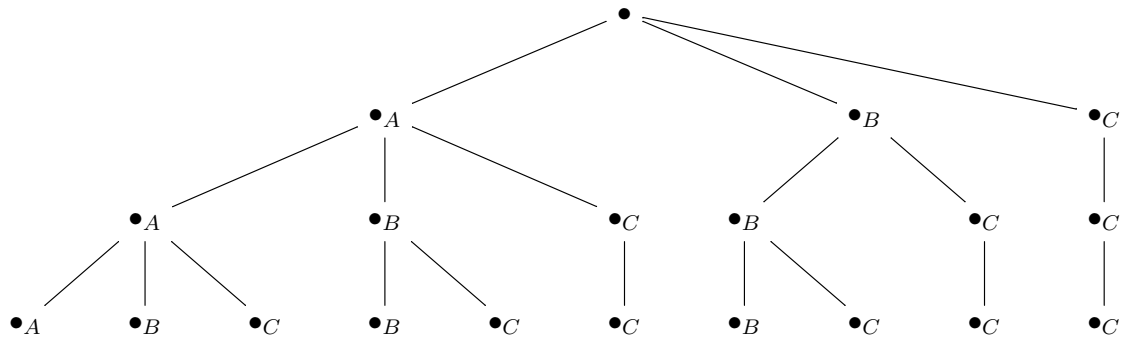
If a finite set A is the union of n pairwise disjoint subsets each with d elements, then $n = |A|/d$.

Example. Four people are to be seated around a circular table (with four seats). Two seatings are considered the same when each person has the same left neighbor and the same right neighbor. How many distinct seatings are possible.

Label one seat 1 and proceed labeling 2,3,4 clockwise around the table. There are four ways to seat one of the people in seat 1, three ways to seat someone in seat 2, two ways to seat someone in seat 1, and then only one choice for the last seat. Thus, there are $4! = 24$ possible seatings, ignoring the rule. However, each of the four choices for seat 1 leads to the same arrangements, since we only distinguish arrangements when the neighbors are both different. There were four choices for seat 1, and so there are $24/4 = 6$ possible seatings.

A final strategy, that we will discuss, for solving counting problems is to use a tree diagram. In this diagram, the leaves represent the possible outcomes.

Example. A passcode is a string from the alphabet $\{A, B, C\}$. Repeated letters are allowed but letters must appear alphabetically (e.g., a B may not appear before an A). How many passcodes of length 3 are possible?



We can read off from the graph that there are 10 possibilities. These are

$$\{AAA, AAB, AAC, ABB, ABC, ACC, BBB, BBC, BCC, CCC\}.$$

6.2 Pigeonhole Principle

Imagine we have k pigeonholes and $k+1$ pigeons. If those $k+1$ pigeons fly into the pigeonholes, then at least one of the pigeonholes contains more than one pigeon. This may seem exceptionally trivial, but there are significant applications of this fact.

The Pigeonhole Principle

If k is a positive integer and $k+1$ or more objects are put into k boxes, then there is at least one box containing two or more of the objects.

Proof. We use proof by contraposition. Suppose no box contains more than one object. Then there are at most k objects. □

Example. (1) In any group of 367 people, there must be at least two with the same birthday. In any group of 27 English words, there must be at least two that begin with the same letter. In a class of 102 people, at least two students will receive the same grade on an exam.

(2) If $f : A \to B$ is a function with $|A| = k+1$ and $|B| = k$, then f is not 1-1. The contrapositive of this statement is also useful. If $f : A \to B$ a one-to-one function between finite sets, then $|A| \leq |B|$.

The Generalized Pigeonhole Principle

If N objects are placed into k boxes, then at least one box contains at least $\lceil N/k \rceil$ objects.

Proof. Suppose no box contains more than $\lceil N/k \rceil$ objects. Note that $\lceil N/k \rceil < (N/k) + 1$. Then the total number of objects is at most

$$k\left(\lceil N/k \rceil - 1\right) < k\left(\left(\frac{N}{k} + 1\right) - 1\right) = N.$$ □

Example. (1) Suppose we have 100 people together in a room. Since there are only 12 months, then at least 9 people were born in the same month.

(2) Suppose we have a standard deck of 52 cards. If we select cards at random, how many must we select to guarantee we have at least 3 of the same suit. Let N be the number of cards we need to select. The four "boxes" in this case are the four suits. So what we want is $\lceil N/4 \rceil \geq 3$. The smallest positive integer that makes this true is $N = 9$.

A sequence $\{a_n\}$ (indexed by $\mathbb{Z}^+$) is said to be *increasing* if $a_i \geq a_{i-1}$ for all i and *strictly increasing* if $a_i > a_{i-1}$ for all i. Similarly, we say $\{a_n\}$ is *decreasing* if $a_i \leq a_{i-1}$ for all i and *strictly decreasing* if $a_i < a_{i-1}$ for all i. A *subsequence* of $\{a_n\}$ is a sequence of the form $a_{i_1}, a_{i_2}, a_{i_3}$ where $1 \leq i_1 < i_2 < i_3 < \dots$. That is, a subsequence is a collection of elements in $\{a_n\}$ in the same order that they appear in the original sequence.

Example. Suppose 101 people are standing in a line, all of whom have different heights. Can we find 11 people in the order they are standing with heights that are increasing or decreasing.

We let a_i denote the height of the ith person in line, so that our sequence $a_1, a_2, \dots, a_{101}$ represents the line of people. We are looking for a subsequence of $\{a_n\}$ that is either (strictly) increasing or (strictly) decreasing. (The parentheses are there because, as everyone has different heights, then in this context they mean the same thing.)

To each a_k, we associate the pair (i_k, d_k) where i_k is the length of the longest increasing sequence starting at a_k, and d_k is the length of the longest decreasing sequence starting at a_k.

Suppose (by way of contradiction) that there exists no increasing or decreasing sequences of length 11. Then $1 \leq i_k, d_k \leq 10$ for all k. By the product rule, this implies that there are $10 \cdot 10$ possible ordered pairs for the (i_k, d_k). But there are actually 101 such ordered pairs, so by the pigeonhole principle, two of these are equal. That is, there are some terms a_s and a_t with $s < t$ where $i_s = i_t$ and $d_s = d_t$. We claim this is impossible.

Because the terms are distinct, we have either $a_s < a_t$ or $a_s > a_t$. If $a_s < a_t$, then, because $i_s = i_t$, we can form an increasing sequence by taking a_s followed by the increasing sequence of length i_t starting at a_t. But then this increasing sequence starts at a_s has length $i_t + 1 = i_s + 1$. But this contradicts the definition of i_s. A similar contradiction arrives by assuming $a_s > a_t$ but using decreasing sequences.

We conclude then that there must be a subsequence of $\{a_n\}$ of length (at least) 11 which is either increasing or decreasing.

A similar idea as in the example justifies the following theorem.

Theorem 1

Every sequence of $n^2 + 1$ distinct real numbers contains a subsequence of length $n + 1$ that is either strictly increasing or decreasing.

Example. Suppose in a group of 10 people, an executive committee is to be formed consisting of a President, Vice President, and a Secretary. How many different committees can be formed?

Order matters! There are 10 possibilities for President, but once this person has been chosen, there are only 9 possibilities for Vice President, then 8 for Secretary. By the product rule, there are $10 \cdot 9 \cdot 8 = 720$ possible committees.

Definition: Permutation, r-permutation

A *permutation* of a set of distinct objects is an ordered arrangement of those objects. An ordered arrangement of r elements of a set is called an *r-permutation.*

Denote the number of r-permutations of a set with n elements by $P(n, r)$.

Example. Given a set $\{a, b, c, d\}$ of 4 elements, we have

$$P(4, 4) = 4! = 24, \quad P(4, 3) = 4 \cdot 3 \cdot 2 = 24, \quad P(4, 2) = 4 \cdot 3 = 12, \quad P(4, 1) = 4.$$

Theorem 2

If n is a positive integer and r is an integer with $1 \leq r \leq n$, then

$$P(n, r) = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}.$$

The last formula actually works if $r = 0$ as well. Note $P(n, 0) = 1$ for any $n \geq 0$.

Example. A first prize, second prize, and third prize are to be awarded for best cow at a state fair. If 20 cows are entered, how many different ways can these be awarded.

We are counting $P(20, 3)$. By the formula, this is

$$P(20, 3) = \frac{20!}{(20-3)!} = \frac{20!}{17!} = 20 \cdot 19 \cdot 18 = 6840.$$

Example. How many permutations of $ABCDEFGH$ contain the string ABC?

We treat the string ABC as a single character, so we must determine the permutations of 6 objects, which is $6! = 720$.

Definition: r-combination

An *r-combination* of elements of a set is an unordered selection of r elements from the set.

We denote the number of r-combinations of a set with n distinct elements by $C(n, r)$. This is also denoted by the *binomial coefficient*, $\binom{n}{r}$.

Example. Suppose the math department has a faculty of 10 professors. Three need to be chosen for the awards committee. How many different groups of three professors can we form?

If order mattered, then there would be $P(10, 3) = 720$ such committees. But order *does not* matter so we have overcounted. For a particular committee, there are $3! = 6$ permutations of that committee which are the same. By the division rule, the actual number is $720/6 = 120$.

Example. Given a set $\{a, b, c, d\}$ of 4 elements, we have

$$C(4, 4) = 1, \quad C(4, 3) = 4, \quad C(4, 2) = 6, \quad C(4, 1) = 4, \quad C(4, 0) = 1.$$

Theorem 3

The number of r-combinations of a set with n elements, where $0 \leq r \leq n$, is

$$C(n, r) = \frac{n!}{r!(n-r)!}.$$

Proof. By the division rule, $C(n, r) = \dfrac{P(n, r)}{P(r, r)} = \dfrac{n!/(n-r)!}{r!(r-r)!} = \dfrac{n!}{r!(n-r)!}$. □

Example. (1) In the faculty example, we have

$$C(10, 3) = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8}{6} = 120.$$

(2) How many five-card poker hands can be dealt from a standard deck of of 52 cards?

We have $C(52, 47) = C(52, 5) = \dfrac{52!}{5!47!} = \dfrac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2,598,960.$

Corollary 4

Let n and r be integers with $0 \leq r \leq n$. Then $C(n, r) = C(n, n - r)$.

We can see this by switching terms in the denominator of $C(n, r)$ and note that $n - (n - r) = r$.

Another method is by a *bijective proofs*, where we show that there is a bijection between the sets of objects counted on the two sides of the identity. Let S be a set with n elements. Define a map f on the power set of S by $f(A) = \overline{A}$. This map is a bijection on $\mathcal{P}(S)$ since $f(f(A)) = A$. If A has r elements, then $\overline{A}$ has $n - r$ elements.

A *combinatorial proof*, or *double counting proof*, uses counting arguments to prove that both sides of the identity count the same objects but in different ways. Choose a subset A of S by specifying the elements in A is the same as specifying the elements *not* in A. Thus, counting the number of subsets with r elements is the same counting the number of subsets with $n - r$ elements.

Example. A school has 10 math professors and 8 computer science professors. How many ways are there to choose two professors from each to be on a committee.

There are $C(10, 2)$ ways to choose math faculty and $C(8, 2)$ ways to choose computer science faculty. Hence, by the product rule, the total is,

$$C(10, 2) \cdot C(8, 2) = \frac{10!}{2!8!} \cdot \frac{8!}{2!6!} = 45 \cdot 28 = 1260.$$

6.4 Binomial Theorem

A binomial is an expression of the form $x + y$. The binomial theorem gives a mechanism for representing powers of this expression. Recall that $\binom{n}{j}$ is the same as $C(n, j)$.

Theorem 5: The Binomial Theorem

Let x and y be variables and let n be a nonnegative integer. Then

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j.$$

Writing out the first few terms, we have

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j = \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n.$$

The proof of this theorem is surprisingly simple. We count how many times $x^{n-j}y^j$ shows up in the product

$$(x+y)^n = (x+y)(x+y)\ldots(x+y) \quad \text{(n times)}.$$

Notice that a term in the product is obtained by choosing either an x or a y from each binomial factor. Choosing $n - j$ x's from n possible choices is $\binom{n}{n-j}$ and this is equal to $\binom{n}{j}$.

Example. When $n = 3$ we have

$$\begin{aligned}(x+y)^3 &= \binom{3}{0} x^3 + \binom{3}{1} x^2 y + \binom{3}{2} x y^2 + \binom{3}{3} y^3 \\ &= x^3 + 3x^2 y + 3xy^2 + y^3.\end{aligned}$$

We can use this to find a specific coefficient even when writing out the entire expansion is unrealistic.

Example. The coefficient of $x^4 y^{16}$ in $(x+y)^{20}$ is

$$\binom{20}{16} = \frac{20!}{16!4!} = 4845.$$

Now consider the coefficient of $x^4 y^{16}$ in $(2x - 3y)^{20}$. Observe that $(2x - 3y) = (2x + (-3y))$ so the binomial theorem here gives

$$(x+y)^{20} = \sum_{j=0}^{20} \binom{20}{j} (2x)^{n-j} (-3y)^j.$$

Hence, the coefficient of $x^4 y^{16}$ is

$$\binom{20}{16} (2)^4 (-3)^{16} = 3,336,981,811,920.$$

The binomial theorem leads to some other very interesting identities. The first one gives a way to count the number of elements in a power set of n elements. This could be proved by induction, but it is much simpler to prove using the Binomial Theorem.

Example (Cardinality of the power set)**.** Let A be a finite set with n elements. Use the binomial theorem to show that $|\mathcal{P}(A)| = 2^n$.

The number of subsets of A with j elements, $0 \leq j \leq n$, is $\binom{n}{j}$. Then by the binomial theorem,

$$|\mathcal{P}(A)| = \sum_{j=0}^{n} \binom{n}{j} = \sum_{j=0}^{n} \binom{n}{j} 1^{n-j} 1^j = (1+1)^n = 2^n.$$

Example (Binomial identity)**.** Let n be a positive integer. Then

$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} = 0.$$

By the binomial theorem with $x = 1$ and $y = -1$ we have

$$0 = 0^n = (1 + (-1))^n = \sum_{j=0}^{n} \binom{n}{j} 1^{n-j} (-1)^j = \sum_{j=0}^{n} \binom{n}{j} (-1)^j.$$

Example (Binomial coefficient identity)**.** Let n be a positive integer and $0 \leq k \leq n$. Then

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

We use the definition of binomial coefficients and algebra:

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-(k-1))!} \\
&= \frac{n!}{k!(n-k)!} \cdot \frac{((n+1)-k))}{((n+1)-k))} + \frac{n!}{(k-1)!((n+1)-k))!} \cdot \frac{k}{k} \\
&= \frac{n!((n+1)-k)) + n!k}{k!((n+1)-k))!} \\
&= \frac{n!\left(((n+1)-k)) + k\right)}{k!((n+1)-k))!} \\
&= \frac{(n+1)!}{k!((n+1)-k))!} \\
&= \binom{n+1}{k}.
\end{aligned}$$

6.5 Generalized Permutations and Combinations

In this section we generalize our counting methods to include repetition.

The next theorem we have already seen. Suppose we want a bit string of length r. Then for each spot we can choose a 0 or a 1. So the total number is 2^r. (Order matters here.)

Theorem 6

The number of r-permutations of a set of n objects with repetition allowed is n^r.

Example. Suppose at a school there are at least 3 faculty each in Math, CSE, and Stat. How many committees of three faculty can be formed if multiple are allowed from the same department.

Order does not matter here, so we are considering a combination as opposed to a permutation. One way to do this would be to list all of the options, we use M,C,S to refer to the respective departments. By listing we find that there are 10 possibilities:

MMM, MMC, MMS, MCC, MCS, MSS, CCC, CCS, CSS, SSS

Theorem 7

The number of r-combinations from a set with n elements when repetition is allowed is

$$C(n+r-1, r) = C(n+r-1, n-1).$$

Sketch. We are choosing r elements from a set of n elements where order does not matter and repetition is allowed. We represent the r elements by stars $*$. Order does not matter so we may assume the stars are in order from the "first" element to the "last". We put a bar every time there is a switch from one element to another.

(For example, if we have 6 elements from a set with four consisting of two of the first element, one of the second, none of the third, and three of the fourth, we would represent as $**|*||***$.)

Thus, we actually have $n-1+r$ choices total of where to place stars and bars, but out of these we are choosing r positions for the stars (or, equivalently, choosing $n-1$ positions for the bars). □

Example. Suppose a market has 5 types of fruit (bananas, apples, oranges, kiwi, and plums). How many ways are there to choose 3 pieces of fruit. (Assume there are at least 3 of each type.)

The number of ways to choose 3 elements (with repetition from a set of 5 (unordered) elements is

$$C(5+3-1, 3) = C(7, 3) = \frac{7!}{3!4!} = 35.$$

Example. A passcode is a string from the alphabet $\{A, B, C\}$. Repeated letters are allowed but letters must appear alphabetically (e.g., a B may not appear before an A). How many passcodes of length 3 are possible?

Previously we solved this problem using a graph. Now we can use it by thinking of the entries in the passcode as the stars, so 3 stars, and the bars as indicating when we switch letters (from A to B, or B to C), so 3 bars. Hence, the number of passcodes is

$$C(3+3-1,3) = C(5,3) = \frac{5!}{3!2!} = 10.$$

Example. How many solutions does the equation

$$x_1 + x_2 + x_3 + x_4 = 14$$

have, where the x_i are nonnegative integers? How many if the x_i are required to be positive.

One such solution would be $1 + 1 + 1 + 11$. It should be clear that counting these directly would be quite exhausting! We should think about this a different way.

Using our "stars-and-bars" idea above, we can think about the stars representing 14 elements. Each placement of $4 - 1 = 3$ bars then gives a way of adding four numbers to get 14. Obviously order does not matter and repetition is allowed. So the total would be

$$C(4+14-1,4-1) = C(17,3) = \frac{17!}{14!3!} = \frac{17 \cdot 16 \cdot 15}{6} = 680.$$

If the x_i are required to be positive (i.e., $x_i \geq 1$), then we make a replacement. Set $y_i = x_i - 1$ for each i, so that $y_i \geq 0$ for all i. Then

$$y_1 + y_2 + y_3 + y_4 = (x_1 - 1) + (x_2 - 1) + (x_3 - 1) + (x_4 - 1) = (x_1 + x_2 + x_3 + x_4) - 4 = 14 - 4 = 10.$$

So, we have reduced this to a problem very similar to the above. Then we have

$$C(4+10-1,4-1) = \frac{13!}{10!3!} = \frac{13 \cdot 12 \cdot 11}{3 \cdot 2 \cdot 1} = 286.$$

Example. Suppose we have the word SAMPLE. How many strings can we make by reordering the letters? What if the word is SUCCESS?

This is a simple problem because all of the letters are distinct. So we are just asking for the number of permutations from the set $\{S, A, M, P, L, E\}$. This is just $P(6,6) = 6! = 720$.

Now suppose we have the word SUCCESS. This has repetition. Any order of the three S's would be the same string. We can put the three S's in the seven positions in $C(7,3)$ ways. Subsequently, we can choose the positions for the two $C's$ in $C(4,2)$ ways. Then we have $C(2,1)$ ways to place the U leaving only $C(1,1) = 1$ way to place the E. This gives a total of

$$C(7,3)C(4,2)C(2,1)C(1,1) = 420.$$

Theorem 8

The number of different permutations of n objects, where there are n_i indistinguishable objects of type k, is

$$\frac{n!}{n_1!n_2!\dots n_k!}.$$

This is also the number of ways to distribute n distinguishable objects into k distinguishable boxes so that n_i objects are placed into box i.

Example. How many ways can we deal hands of 5 playing cards to four people.

We already know that for one person there are $C(52,5)$ ways to deal out a poker hand (because the order of the cards does not matter). After we have dealt the hand to the first person, then there are only 47 cards left, so there are $C(47,5)$ ways to deal the cards to the second person. Continuing in this way we find that

$$C(52,5)C(47,5)C(42,5)C(37,5) = \frac{52!}{5!5!5!5!32!}$$

ways to deal cards to all of the players.

If the objects themselves are indistinguishable, then this is no different from how we counted above. A set of n-indistinguishable objects is no different than a set of n elements. On the other hand, if the objects are distinguishable but the boxes are not, then the solution is different.

Example. How many ways can four professors by put into three offices leaving no office empty?

Each office needs at least one professor, and then exactly one office will have two professors. There are six ways which can list explicitly. Let $\{A,B,C,D\}$ denote the four professors. Then we have

$$\{\{A,B\},\{C\},\{D\}\},\quad \{\{A,C\},\{B\},\{D\}\},\quad \{\{A,D\},\{B\},\{C\}\},$$
$$\{\{B,C\},\{A\},\{D\}\},\quad \{\{B,D\},\{A\},\{C\}\},\quad \{\{C,D\},\{A\},\{B\}\}.$$

The next formula gives a way of counting in the previous example.

Stirling numbers

The number of ways to put n distinguishable objects into j indistinguishable boxes is

$$S(n,j) = \frac{1}{j!}\sum_{i=0}^{j-1}(-1)^i\binom{j}{i}(j-i)^n.$$

The numbers $S(n,j)$ are called *Stirling numbers* (of the second kind).

Example. In the previous example, we found $S(4,3)$. We now use the formula to compute this:

$$\begin{aligned} S(4,3) &= \frac{1}{3!}\sum_{i=0}^{2}(-1)^i\binom{j}{i}(j-i)^n \\ &= \frac{1}{6}\left(\binom{3}{0}3^4 - \binom{3}{1}2^4 + \binom{3}{2}1^4\right) \\ &= \frac{1}{6}(81-48+3) = 6. \end{aligned}$$

Example. How many ways can we put five copies of the same book into three identical boxes?

This is counting *partitions*:

$$5, \quad 4+1, \quad 3+2, \quad 3+1+1, \quad 2+2+1, \quad 2+1+1+1, \quad 1+1+1+1+1$$

Relations

9.1 Relations and their properties

Definition: Binary relation

Let A and B be sets. A *binary relation from A to B* is a subset of $A \times B$.

If $R \subset A \times B$ is a relation, then we use the notation aRb to mean $(a, b) \in R$ and we say "a is *related* to b". We write $a \not R b$ to mean $(a, b) \notin R$.

For example, we might let A be the set of US capitals and B US states. Then let R to be pairs (a, b) if a is the capital of state b. So (Columbus, Ohio) is in R as is (Frankfort, Kentucky). However, (Indianapolis, Wisconsin) would not be in R.

A function is a special kind of relation. A function is a relation $R \subset A \times B$ if there is some element $(a, b) \in R$ for every $a \in A$, and $(a, b), (a, b') \in R$ implies $b = b'$. Using function notation we define f by $f(a) = b$.

Definition: Relation on a set

A *relation on a set A* is a relation from A to A.

Another way to state the above is that a relation on a set A is a subset of $A \times A$.

Example. Define a relation on the set $A = \{1, 2, 3, 4, 5, 6\}$ by $R = \{(a, b) \mid a \text{ divides } b\}$. Then

$$R = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,2), (2,4), (2,6), (3,3), (3,6), (4,4), (5,5), (6,6)\}$$

We could represent this graphically by making two columns 1-6 and drawing an arrow if elements are related.

Now we can consider special properties of relations.

Definition: Reflexive, symmetric, antisymmetric, transitive

Let R be a relation on a set A

(1) R is called *reflexive* if $(a, a) \in R$ for every element $a \in A$.

(2) R is called *symmetric* if $(b, a) \in R$ whenever $(a, b) \in R$ for all $a, b \in A$.

(3) R is called *antisymmetric* if $(a, b) \in R$ and $(b, a) \in R$ implies $a = b$ for all $a, b \in A$.

(4) R is called *transitive* if whenever $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$ for all $a, b, c \in A$.

Example. Consider each relation R on the set $\mathbb{R}$.

(1) $\{(x, y) \mid x + y = 0\}$

This is not reflexive since $(1, 1) \notin R$. It is symmetric. It is not antisymmetric since $(1, -1), (-1, 1) \in R$ but $1 \neq -1$. It is not transitive.

(2) $\{(x, y) \mid x = \pm y\}$

This is reflexive since $x = x$, so $(x, x) \in R$.

Since $(x, y) \in R$, then $x = \pm y$, which implies that $y = \pm x$. Hence, $(y, x) \in R$.

It is not antisymmetric because $(1, -1), (-1, 1) \in R$ but $1 \neq -1$.

It is transitive. If $(x, y), (y, z) \in R$, then $x = \pm y$ and $y = \pm z$. Considering the four cases, this implies that $x = \pm z$. Thus, $(x, z) \in R$.

(3) $\{(x, y) \mid x - y \text{ is rational}\}$

(4) $\{(x, y) \mid xy \geq 0\}$

(5) $\{(x, y) \mid xy = 0\}$

This is not reflexive. Let $x \neq 0$, then $xx = x^2 \neq 0$, so $(x, x) \notin R$.

(6) $\{(x, y) \mid x = 1\}$

(7) $\{(x, y) \mid x = 1 \text{ or } y = 1\}$

We can combine relations using our usual set operations .

Example. Consider the following relations from $\{1, 2, 3\}$ to $\{1, 2, 3, 4\}$.

$$R_1 = \{(1, 2), (2, 3), (3, 4)\}$$
$$R_2 = \{(1, 1), (1, 2), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}.$$

Determine $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 \oplus R_2$ (exclusive or), $R_1 - R_2$, and $R_2 - R_1$.

Definition: Composite relation

Let R be a relation from a set A to a set B and S a relation from B to a set C. The *composite* of R and S is the relation consisting of ordered pairs (a, c) where $a \in A$ and $c \in C$, and for which there exists an element $b \in B$ such that $(a, b) \in R$ and $(b, c) \in S$. We denote this composition by $S \circ R$.

Example. Consider the following relations from $\{1, 2, 3, 4\}$ to $\{1, 2, 3\}$.

$$R = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 1)\}$$
$$S = \{(2, 1), (3, 1), (3, 2), (4, 2)\}.$$

Find $S \circ R$.

9.4 Closure of relations

If R is a relation on a set A, then it may or may not have some property P, such as reflexivity, symmetry, or transitivity. If R does not have a property P, we would like to find the smallest relation containing R that has property P.

Definition: Closure

If R is a relation on a set A, then the *closure* of R with respect to P, if it exists, is the relation S on A with property P that contains R and is a subset of every subset of $A \times A$ containing R with property P.

We will consider closures related to our three properties: reflexive, symmetric, and transitive.

Let R be a relation on a set A.

(1) The *diagonal relation* on A is $\Delta = \{(a, a) \mid a \in A\}$. The *reflexive closure* of R is $R \cup \Delta$.
(2) The *inverse relation* of R is $R^{-1} = \{(b, a) \mid (a, b) \in R\}$. The *symmetric closure* of R is $R \cup R^{-1}$.

Example (Reflexive and symmetric closures)**.** Let $A = \{1, 2, 3\}$. Determine the reflexive and symmetric closure of the relation $R = \{(1, 1), (1, 2), (2, 1), (3, 2)\}$.

The relation R is not reflexive because it is missing $(2, 2)$ and $(3, 3)$. The diagonal relation of A is $\Delta = \{(1, 1), (2, 2), (3, 3)\}$. Then the reflexive closure is

$$R \cup \Delta = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 2), (3, 3)\}$$

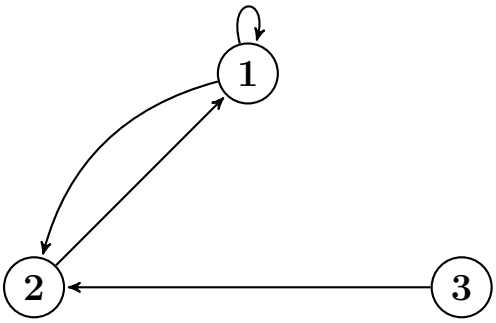
The relation R is not symmetric either because it is missing $(2, 3)$. The inverse relation is $R^{-1} = \{(1, 1), (2, 1), (1, 2), (2, 3)\}$. Then the symmetric closure is

$$R \cup R^{-1} = \{(1, 1), (1, 2), (2, 1), (2, 3), (3, 2)\}$$

We can form a directed graph from R by marking all elements of A vertices and drawing an arrow from a to b if $(a, b) \in R$. We denote by R^n the pairs (a, b) such that there is a path of length n from a to b. The *connectivity relation* R^* consists of the pairs (a, b) such that there is a path of at least one from a to b in R. The *transitivity closure* of R equals the connectivity relation R^*.

Example (Transitive closures)**.** Let $A = \{1, 2, 3\}$. Determine the transitive closure of the relation $R = \{(1, 1), (1, 2), (2, 1), (3, 2)\}$.

We note that R is not transitive because it is missing $(2, 2)$ and $(3, 1)$. We construct the graph described above to determine the connectivity relation of R.



Then $R^1 = R$ (that is, the paths of length 1). The paths of length 2 are

$$R^2 = \{(1, 1), (2, 1), (2, 2), (3, 1)\}$$

We can also compute R^3, but this does not give us any new paths. Since the graph only has three vertices, then it is not necessary to compute beyond R^3.

Hence, the transitive closure of R is

$$R^* = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)\}$$

Definition: Equivalence relation, equivalent

A relation R on a set A is called an *equivalence relation* if it is reflexive, symmetric, and transitive. If R is an equivalence relation and $(a, b) \in R$, then we say a and b are *equivalent*.

We use the notation $a \sim b$ to indicate that $(a, b) \in R$ when R is understood. If it is not clear which relation we are referring to, we may write $a \sim_R b$.

Example (Basic examples)**.**

(1) Triangle congruence is an equivalence relation, as is triangle similarity.
(2) Those who have taken linear algebra have seen matrix similarity, which is also an equivalence relation.
(3) (We have seen this example previously) Consider the relation R on $\mathbb{Z}$ defined by aRb if $a = \pm b$. Then R is an equivalence relation. In other notation, we write $a \sim a$ and $a \sim -a$.
(4) Recall that we say that an integer a divides an integer b if b is a multiple of a. We write $a \mid b$ in this situation.

    The relation $R = \{(a, b) \mid a \textit{ divides } b\}$ is not an equivalence relation since it is not symmetric. For example, $2 \mid 4$ but $4 \nmid 2$.
(5) Let A be a set and let $f : A \to A$ be a function. Define a relation R on A by $(x, y) \in R$ if and only if $f(x) = f(y)$. Then R is an equivalence relation. Two element $x, y \in A$ are equivalent $(x \sim y)$ if they have the same image under f.

Example (Congruence mod 5)**.** Define a relation R on $\mathbb{Z}$ by aRb if $a - b$ is divisible by 5. We claim R in an equivalence relation. (Called "congruence mod 5").

Let $a \in \mathbb{Z}$. Since $a - a = 0$ and 5 divides 0, then aRa. Thus, R is reflexive.

Suppose aRb. Then 5 divides $a - b$ so $a - b = 5k$ for some integer k. But them $b - a = -(a - b) = (-k)5$, so 5 divides $b - a$. That is, bRa so R is symmetric.

Finally, suppose aRb and bRc. Then $a - b = 5k$ and $b - c = 5\ell$ for some integers k and ℓ. Then $a - c = (a - b) + (b - c) = 5(k + \ell)$, so m divides $a - c$. That is, aRc so R is transitive. Thus, R is an equivalence relation.

Definition: Equivalence class

Let R be an equivalence relation of a set A. The *equivalence class of* $a \in A$ is the set

$$[a]_R = \{b \in A \mid (a, b) \in R\} = \{b \in A \mid a \sim b\}.$$

We drop the subscript R if the relation is understood. Colloquially, we can define the equivalence class of a as the set of all elements that are related to a. Note that $[a]_R$ is never empty since the relation is reflexive and so $a \sim a$. Also, if $b \in [a]_R$, then by symmetry and transitivity we see that $[a]_R = [b]_R$. We call any element $b \in [a]_R$ a *representative* of $[a]_R$.

Example (Congruence mod 5, again)**.** Let R be congruence mod 5. The equivalence classes are

$$[0] = \{\dots, -10, -5, 0, 5, 10, \dots\} \qquad [1] = \{\dots, -9, -4, 1, 6, 11, \dots\}$$
$$[2] = \{\dots, -8, -3, 2, 7, 12, \dots\} \qquad [3] = \{\dots, -7, -2, 3, 8, 13, \dots\}$$
$$[4] = \{\dots, -6, -1, 4, 8, 14, \dots\}.$$

Theorem 1: Properties of equivalence classes

Let R be an equivalence relation on a set A and let $a, b \in A$. The following are equivalent:

(i) $(a, b) \in R$
(ii) $[a] = [b]$
(iii) $[a] \cap [b] \neq \emptyset$.

Now let us summarize what this theorem says. Since R is reflexive, then every element belongs to *at least one* equivalence class. Since (iii) $\Rightarrow$ (ii), then if $[a]$ and $[b]$ have *anything* in common, then they are the same. Thus, equivalence classes are *either equal or disjoint*. Thus, the equivalence classes split the set A into pairwise disjoint subsets. This is known as *partitioning* sets. While one may define partitions independently, it turns out that every partition defines an equivalence class, so in a sense, partitions and equivalence relations are the same thing.

We observe that every integer belongs to exactly one of these equivalence classes. Another way to express this phenomenon is that *equivalence classes partition a set*.

9.6 Partial orderings

Now we study another special kind of relation.

Definition: Partial ordering, poset

A relation R on a set S is call a *partial ordering* (or *partial order*) if it is reflexive, antisymmetric, and transitive. A set S together with a partial order R is called a *partially ordered set* (or *poset*), and is denoted (S, R). Members of S are called *elements* of the poset.

Example. (1) One of the most recognizable partial orders is $\leq$ on $\mathbb{Z}$. Since $a \leq a$, then $\leq$ is reflexive. If $a \leq b$ and $b \leq a$, then $a = b$, so $\geq$ is antisymmetric. Finally, if $a \leq b$ and $b \leq c$, then $a \leq c$, so $\leq$ is transitive. Hence, $(\mathbb{Z}, \leq)$ is a poset.

(2) One can similarly see that "divides" is partial order on $\mathbb{Z}^+$.

(3) The inclusion relation $\subseteq$ is a partial ordering on the power set of a set S.

Definition: Comparable, incomparable

The elements a and b of a poset $(S, \preccurlyeq)$ are called *comparable* if either $a \preccurlyeq b$ or $b \preccurlyeq a$. When a and b are elements such that neither $a \preccurlyeq b$ nor $b \preccurlyeq a$, then a and b are called *incomparable*.

Example. (1) In $(\mathbb{Z}, \leq)$, all elements are comparable.

(2) In $(\mathbb{Z}^+, |)$, the numbers 3 and 9 are comparable because $3 \mid 9$ but 5 and 7 are incomparable because $5 \nmid 7$ and $7 \nmid 5$.

(3) In $(P(\mathbb{Z}), \subseteq)$, not all elements are comparable, even though $\{1, 2\}$ and $\{1, 3\}$ are in $P(\mathbb{Z})$, it does not make sense to write $\{1, 2\} \subseteq \{1, 3\}$ or $\{1, 3\} \subseteq \{1, 2\}$. We use the notation $a \preccurlyeq b$ to indicate that $(a, b) \in R$.

Definition: Totally ordered set, total order

If $(S, \preccurlyeq)$ is a poset and every two elements of S are comparable, S is called a *totally ordered set* (or *linearly ordered set*), and $\preccurlyeq$ is called a *total order* (or *linear order*). A totally ordered set is also called a chain.

Example. The poset $(\mathbb{Z}, \leq)$ is a totally ordered set. However, $(\mathbb{Z}^+, |)$ is not totally ordered.

Definition: Well-ordered set

A poset $(S, \preccurlyeq)$ is a *well-ordered set* if $\preccurlyeq$ is a total ordering and every nonempty subset of S has a least element.

Example. The set $(\mathbb{Z}, \leq)$ is not well-ordered. However, the set $(\mathbb{Z}^+, \leq)$ is well-ordered.

Example. Define a relation $\preccurlyeq$ on $\mathbb{Z}^+ \times \mathbb{Z}^+$ by $(a_1, a_2) \preccurlyeq (b_1, b_2)$ if $a_1 < b_1$ or if $a_1 = b_1$ and $a_2 \leq b_2$. This is known as *lexicographic ordering*. The relation $\preccurlyeq$ makes $\mathbb{Z}^+ \times \mathbb{Z}^+$ into a well-ordered set.

(Show lattice of $\mathbb{Z}^+ \times \mathbb{Z}^+$ and how to compare points.)

This extends easily to $\mathbb{Z}^+ \times \mathbb{Z}^+ = (\mathbb{Z}^+)^n$. We can also extend the ideal of a lexicographic ordering to any pair of posets. Let $(A_1, \preccurlyeq_1)$ and $(A_2, \preccurlyeq_2)$ be posets. Define a partial order $\preccurlyeq$ on $A_1 \times A_2$ by $(a_1, a_2) \preccurlyeq (b_1, b_2)$ if either $a_1 \preccurlyeq_1 b_1$ or if $a_1 = b_1$ and $a_2 \preccurlyeq_2 b_2$.

A *Hasse diagram* for a poset is the directed graph for the relation R associated to the poset. We drew this previously when considering the transitive closure of a relation. However, we perform some reductions on the diagram to simplify.

(1) Start by drawing arrows $a \to b$ if $(a, b) \in R$.
(2) Remove all loops at a vertex and all arrows that are implied by the transitivity condition.
(3) If we assume that arrows are pointed upwards, then we can remove the direction on the arrows.

Example (Hasse diagram)**.** Draw the Hasse diagram for $\{(a, b) \mid a \text{ divides } b\}$ on $\{1, 2, 3, 4, 6, 8, 12\}$.

Finite-state machines

13.2 Finite-state machines with output

Definition: Finite state machine

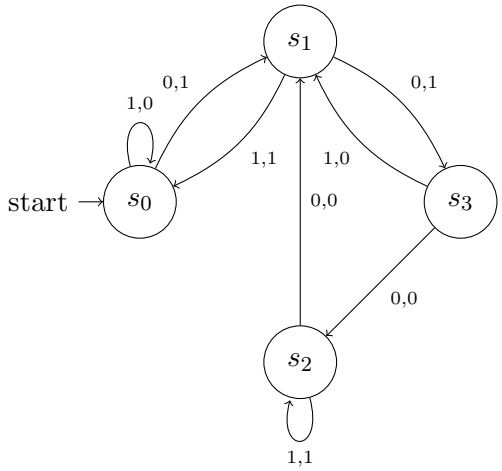
A finite-state machine $M = (S, I, O, f, g, s_0)$ consists of a finite set S of *states*, a finite *input alphabet* I, a finite *output alphabet* O, a *transition function* f that assigns to each state and input a new state, an *output function* g that assigns to each state and input pair an output, and an *initial state* s_0.

In a finite state machine, two things happen on any input. First, the state changes. Secondly, the machine decides what the output should be. Sometimes there may be no output, only a change in the state. For example, given a vending machine, putting in money changes the state (how much money is in the sale) but only produces an output when the state reaches a certain level.

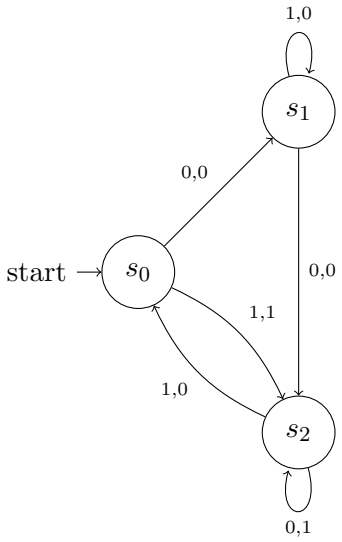
We can represent a finite-state machine using a *state table*. This table shows how f and g are evaluated on each input for each state. We can also draw a *state diagram*. The edges for this graph are the various stages. The arrows are labeled with the input and output for that transition.

Example. Consider a finite-state machine with three states $S = \{s_0, s_1, s_2, s_3\}$, $I = \{0, 1\}$, and $O = \{0, 1\}$. The state table is given:

State	f		g	
	Input		Input	
	0	1	0	1
s_0	s_1	s_0	1	0
s_1	s_3	s_0	1	0
s_2	s_1	s_2	0	1
s_3	s_2	s_1	0	0



Example. Given the following state diagram we can reconstruct the state table.



State	f		g	
	Input		Input	
	0	1	0	1
s_0	s_1	s_2	0	1
s_1	s_2	s_1	0	0
s_2	s_2	s_0	1	0

Given a bit string $x = x_1x_2 \dots x_k$, we can form a new string from any finite-state machine. Let $q_0 = s_0$. Then $q_1 = f(q_0, x_1)$, $q_2 = f(q_1, x_2)$, and so on so $q_k = f(q_{k-1}, x_k)$. This produces an output string $y = y_1y_2 \dots y_k$ where $y_1 = g(q_0, x_1)$, $y_2 = g(q_1, x_2)$, and so on so $y_k = g(q_{k-1}, x_k)$.

Example. Determine the output string from the finite state machine in the previous example with input string 1010.

We have $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, and $x_4 = 0$. Then

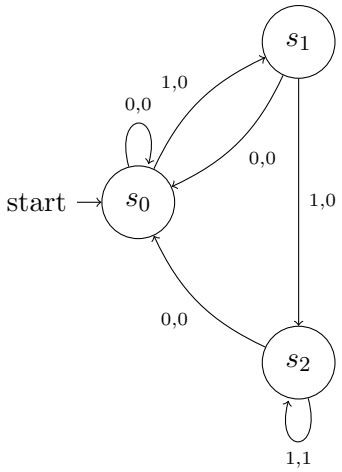
$$q_0 = s_0, \quad q_1 = f(q_0, x_1) = f(s_0, 1) = s_2, \quad q_2 = f(q_1, x_2) = f(s_2, 0) = s_2,$$
$$q_3 = f(q_2, x_3) = f(s_2, 1) = s_0, \quad q_4 = f(q_3, x_4) = f(s_0, 0) = s_1.$$

Now we have

$$y_1 = g(q_0, x_1) = g(s_0, 1) = 1, \quad y_2 = g(q_1, x_2) = g(s_2, 0) = 1,$$
$$y_3 = g(q_2, x_3) = g(s_2, 1) = 0, \quad y_4 = g(q_3, x_4) = g(s_0, 0) = 0.$$

This gives the output string 1100.

Example. Suppose in a given coding scheme, three consecutive 1's indicate that there has been a transmission error. We can represent this using a finite-state machine. The state function f will move between states s_0, s_1, s_2 every time a 1 is given as an input. The function g will output 0 if no error has been produced but a 1 if an error is produced. That is, $g(s_2, 1) = 1$. The state diagram for this is given as follows.



State	f		g	
	Input		Input	
	0	1	0	1
s_0	s_0	s_1	0	1
s_1	s_0	s_2	0	0
s_2	s_0	s_2	1	0

Recall that for a set I, in this case an alphabet, I^* denotes the strings in the alphabet I.

Definition: Recognizes

Let $M = (S, I, O, f, g, s_0)$ be a finite-state machine and $L \subset I^*$. We say that M *recognizes* (or *accepts*) L if a bit string belongs to L if and only if the last output bit produced by M when given x as input is a 1.

Example. In the previous example, determine which bit strings are recognized.

The bit string 111 is recognized.

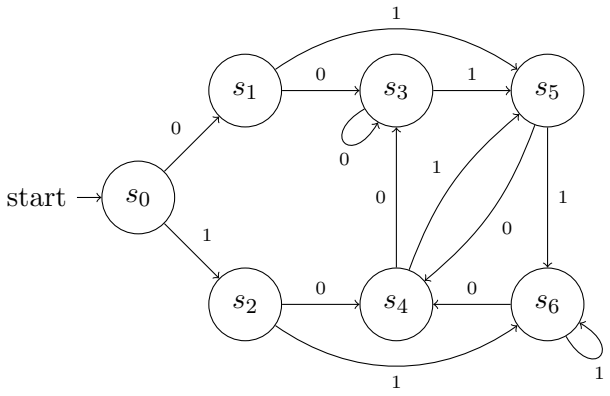
Example. Construct a finite-state machine that recognizes a bit string with 1 in the last position and 0 in the third-to-last position read so far.

Let s_0 be our starting state. Initially we may have a 0 or a 1, so let s_1 and s_2 correspond to these two options. The remaining states, s_3 through s_6, correspond to the 4 options for the next pair of bits. So

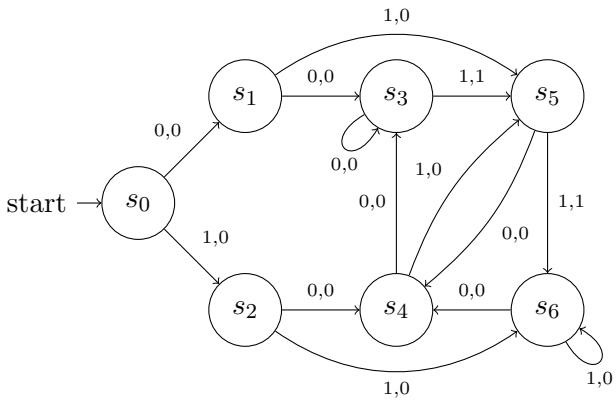
s_3 = the string 00 $\qquad$ s_4 = the string 10

s_5 = the string 01 $\qquad$ s_6 = the string 11

Now it is easy to read the transitions from this. This gives the diagram



Now we just need to follow along and indicate when an error has been produced. That is, when a 0 is followed two steps later by a 1. This happens either at the transition from s_3 to s_5, or from s_5 to s_6. Otherwise there is no error.



Recall that if V is an alphabet (also called a *vocabulary*), then V^* denotes the set of strings in V.

Definition: Concatenation

Suppose that V is a vocabulary and $A, B \subseteq V^*$. The *concatenation* of A and B, denoted AB, is the set of all strings of the form xy with $x \in A$ and $y \in B$.

Given a single subset A of V^*, we can concatenate A with itself. We set $A^0 = \{\lambda\}$ where λ is the empty word, and recursively define A^n by $A^{n+1} = A^n A$.

Example. Let $A = \{0, 11\}$ and $B = \{1, 10, 110\}$. Determine AB, BA, and A^n for $n = 0, 1, 2, 3$.

We have

$$AB = \{01, 010, 0110, 111, 1110, 11110\}$$
$$BA = \{10, 111, 100, 1011, 1100, 11011\}.$$

Note that $AB \neq BA$.

Now $A^0 = \{\lambda\}$, $A^1 = A = \{0, 11\}$ and

$$A^2 = AA = \{00, 011, 110, 1111\}$$
$$A^3 = A^2A = \{000, 0110, 1100, 11110, 0011, 01111, 11011, 111111\}.$$

Definition: Kleene closure

Suppose that V is a vocabulary and $A \subseteq V^*$. The *Kleene closure* of A, denote A^*, is the set consisting of concatenations of arbitrarily many strings from A. That is, $A^* = \bigcup_{k=0}^{\infty} A^k$.

Example (Kleene closures)**.** Let $A = \{0\}$ (note that this is different from the set containing the empty string). Then for each $k \geq 1$, $A^k = \{0^k\}$ (here $0^k = 000\cdots0$, k times). Hence, the Kleene closure is

$$A^n = \{0^n \mid n = 0, 1, 2, \ldots\}.$$

On the other hand, if $B = \{0, 1\}$, then B^* is the set of all bit strings.

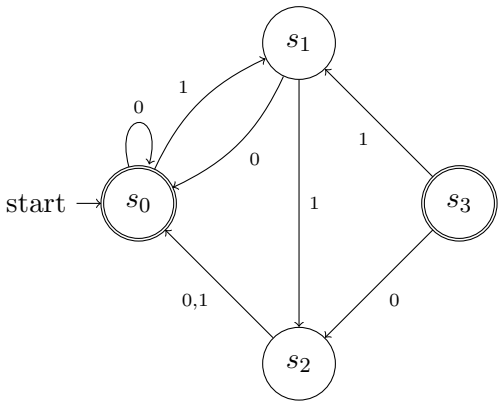
Definition: Finite state automaton

A *finite-state automaton* $M = (S, I, f, s_0, F)$ consists of a finite set of states S, a finite input alphabet I, a transition function $f : S \times I \to S$, an initial (or start) state s_0, and a subset F of S consisting of *final states* (or *accepting states*).

State diagrams and tables are similar to before, but we do not need to consider an output function. Hence, we need only decorate our arrows with a single number corresponding to the transition function. Final/accepting states are indicated by double circles.

Example. Consider the state diagram for the finite-state automaton $M = (S, I, f, s_0, F)$ where $S = \{s_0, s_1, s_2, s_3\}$, $I = \{0, 1\}$, $F = \{s_0, s_3\}$ and f is as given below.

State	f	
	Input	
	0	1
s_0	s_0	s_1
s_1	s_0	s_2
s_2	s_0	s_0
s_3	s_2	s_1



The transition function above is only defined for bits, but we can extend it to strings. Let $M = (S, I, f, s_0, F)$ be a finite-state machine. We extend f to a function $f : S \times I^* \to S$ in the following way. Let $x = x_1 x_2 \cdots x_k$ be a string in I^*. For any state s, $f(s, \lambda) = s$ (where λ is the empty string). Inductively we then define

$$f(s, xa) = f(f(s, x), a) \quad \text{for all } s \in S,\ x \in I^*,\ a \in I$$

Example. Let M be as in the previous example and let $x = 110$. We will compute $f(s_3, x)$.

We have $x = x_1 x_2 x_3$ where $x_1 = 1$, $x_2 = 1$, and $x_3 = 0$. Then

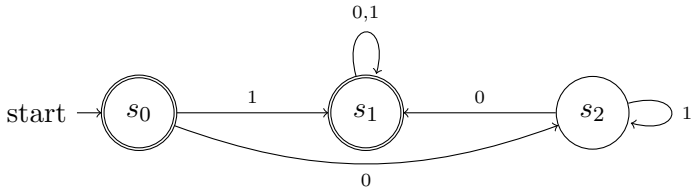
$$f(s_3, x) = f(s_3, x_1 x_2 x_3) = f(f(s_3, x_1 x_2), x_3) = f(f(f(s_3, x_1), x_2), x_3)$$
$$= f(f(f(s_3, 1), 1), 0) = f(f(s_1, 1), 0) = f(s_2, 0) = s_0.$$

Definition: Recognized

A string x is said to be *recognized* (or *accepted*) by the machine $M = (S, I, f, s_0, F)$ if it takes the initial state s_0 to a final state, that is, $f(s_0, x)$ is a state in F. The *language recognized* (or *recognized*) by the machine M, denoted $L(M)$, is the set of all strings that are recognized by M. Two finite-state automata are called *equivalent* if they recognize the same language.

Example. In the previous example, the machine recognizes any string ending in 00 or 10.

Example. Consider the following deterministic finite-state automaton.

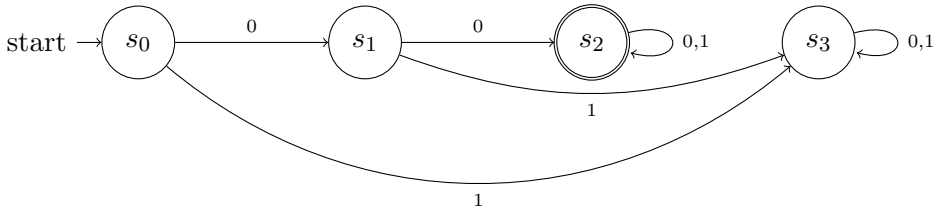


There are no input arrows for s_0, so we only consider paths that begin at s_0 and end at s_1. Thus,

$$L(M) = \{1x, 01^n0x \mid n \in \mathbb{N}, x \text{ is any string}\}$$

Going the other way, we can construct deterministic FSAs that recognize certain languages.

Example. Construct a deterministic FSA that recognizes the set of bit strings that begin with two 0s.



Example. Construct a deterministic FSA that recognizes the set of bit strings that contain two consecutive 0s.