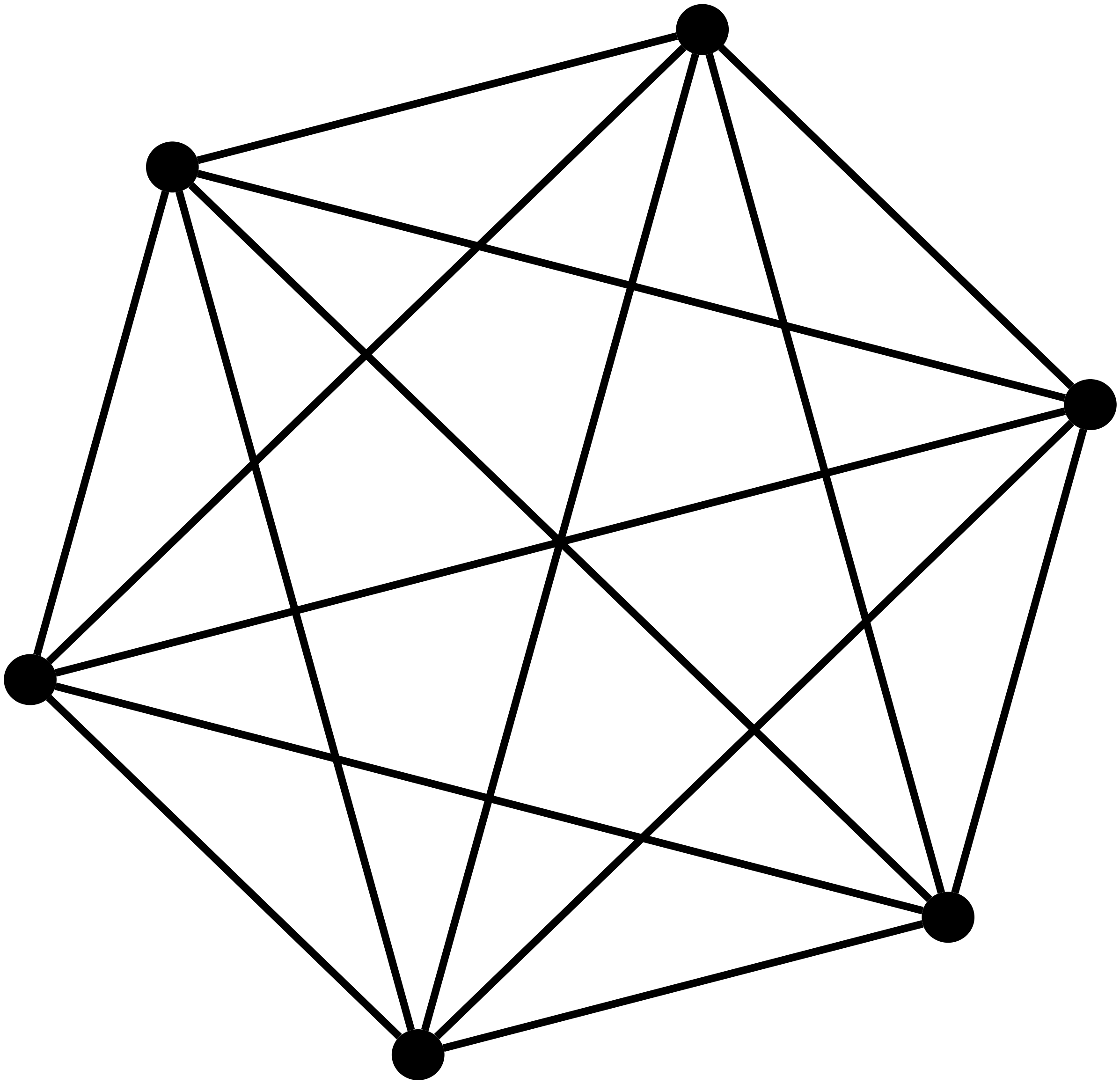
ELEMENTARY FOUNDATIONS

An introduction to topics in discrete mathematics



Jeremy Sylvestre

Elementary Foundations

An introduction to topics in discrete mathematics

Elementary Foundations

An introduction to topics in discrete mathematics

Jeremy Sylvestre
University of Alberta

August 27, 2023

Jeremy Sylvestre is Associate Professor of Mathematics at the University of Alberta's Augustana Campus.

Edition: Fall 2023

Website: Elementary Foundations[1]

[1] `sites.ualberta.ca/~jsylvest/books/EF/`

Preface

TODO

Jeremy Sylvestre
Camrose, Alberta 2018

Contents

I Logic

II Logic in Mathematics

III Set Theory

IV Graph Theory

V More Set Theory

VI Combinatorics

Appendices

Back Matter

Part I

Logic

Chapter 1

Symbolic language

1.1 Statements

statement
a sentence that is either true or false

Example 1.1.1 Logical statements.

1. All prime numbers are odd.
2. Some trees have leaves and some trees have needles.
3. If you pay attention in class and work through all the homework problems, then you will do well in this course.

□

substatement
part of a logical statement that could be considered a statement on its own

Example 1.1.2 Substatements. "Some trees have leaves" is a substatement of statement 2 in Example 1.1.1. □

simple statement
does not contain any proper substatements

compound statement
contains two or more substatements

connective
a connecting word between substatements in a compound statement

Example 1.1.3 Simple and compound statements. Reconsidering the statements in Example 1.1.1:

1. statement 1 is simple;
2. statement 2 is a compound statement made up of two (simple) substatements linked by the connective "and"; and
3. statement 3 is a compound statement made up of two substatements linked by the connective "if . . . then . . . ", where the substatement that constitutes the "if" part is itself a compound statement.

□

The substatements in a compound statement can be related to each other by **connectives** in various ways.

Definition 1.1.4 Five basic connectives.

negation “not”

conjuction
“and”

disjunction
“or”

conditional
“if . . . then . . . ”

biconditional
“if and only if”

Given statements A and B, we use these connectives to construct new statements:

negation of A
not A

conjuction of A **and** B
A and B

disjunction of A **and** B
A or B

conditional where A **implies** B
if A then B

biconditional involving A **and** B
A if and only if B

◊

Remark 1.1.5

1. All statements we will consider can be constructed starting from a finite number of simple statements and modifying/joining them using connectives as above.
2. *Always* take “A or B” to mean “A or B or both” (known as **inclusive or**). However, in everyday language it may be reasonable to take “*either* A or B” to mean “(A or B) and not (A and B)” (known as **exclusive or**).
3. The **conditional** and **biconditional** connectives are actually superfluous — they can be constructed from the first three. (See Worked Example 2.1.2 and Exercise 2.5.5.) But they occur frequently, and such constructions from other connectives obscure their meaning, so it is more convenient to include these two connectives in our list of of basic connectives.

Example 1.1.6 Translating everyday English into logical statements. A conversation.

Alice	It is raining.
Bob	No, it isn’t.
Alice	Either it’s raining or it isn’t.
Bob	How can we decide?
Alice	If we go outside and we get wet, then it’s raining.
Bob	We’d get wet outside if the sprinklers are on, too.

Alice Don't be silly!

Alice (continuing...)
We'll get wet if it's raining, and this is the only way we'll get wet.

Let us rewrite the above conversation to clearly identify the substatements and connectives.

Alice it is raining

Bob not (it is raining)

Alice (it is raining) or (not (it is raining))

Bob *[not a statement!]*

Alice if ((we are outside) and (we get wet)) then (it is raining)

Bob if ((we are outside) and (the sprinklers are on)) then (we get wet)

Alice *[not a statement!]*

Alice if (we are outside) then ((we get wet) if and only if (it is raining))

□

Test 1.1.7 Checking whether a sentence is a logical statement. *If S is an English language sentence and the phrase "It is true that S" makes sense as an English language sentence, then S is a logical statement.*

Strictly speaking, many mathematical statements are not logical statements, for a different reason then the one used in the test above.

Example 1.1.8 An ambiguous mathematical statement. The phrase "f is a differentiable function" is not a logical statement, since whether it is true or false depends on the **free variable** f. For example, if we substitute the function $f(x) = x$ into this statement, the statement becomes true. However, if we substitute the function $f(x) = |x|$, the statement becomes false. We will deal with this issue in Chapter 4. □

1.2 Converting language to symbols

As we have already begun to do, we will use letters to represent (possibly variable) logical statements and substatements. To complete the conversion from verbose language to compact symbolism, we will introduce symbols to represent the **Five basic connectives**.

negation of A
$\neg A$

conjuction of A **and** B
$A \wedge B$

disjunction of A **and** B
$A \vee B$

conditional where A **implies** B
$A \rightarrow B$

biconditional involving A **and** B
$A \leftrightarrow B$

Using variables to represent statements and the above symbols to represent connectives allows us to isolate the task of analyzing logical structure, without being distracted or influenced by the content of the statements.

Warning 1.2.1 In mathematics, the symbol → is also used in function notation; you will need to determine from the context which role this symbol is playing.

Example 1.2.2 Translating English language into symbolic language. Consider the statement “if we are outside and we get wet then it is raining.” Assign statement variables:

$$A = \text{“we are outside,”} \qquad B = \text{“we get wet,”} \qquad C = \text{“it is raining.”}$$

Then symbolically, the statement can be written

$$A \land B \to C.$$

□

Remark 1.2.3 Using *substatement* variables is not the same as using *free* variables. You should think of substatement variables as placeholders for specific logical statements which, by definition, can each be determined to be either true or false.

1.3 Logical analysis

We will now leave the English language behind and concentrate on logical statements consisting only of variables and connectives. Keep in mind that variables are not limited to representing simple statements; they can represent compound statements as well.

truth value
the property of being true or false

Given a logical statement, view the variables as *inputs* and the truth value of the entire statement as an *output*. We would like a systematic way to determine how the truth value of the output changes as we vary the truth values of the inputs.

logical analysis
the process of determining the truth value of a statement based on the truth values of its variable substatements

truth table
tabular method of carrying out logical analysis

If a statement involves a finite number of variables, then since each variable can have one of only two possible truth values, there are a finite number of different combinations of input truth values for the statement. So we can test each combination one after the other to determine all possible outputs. Arrange this analysis in a table with all possibilities for the input variables on the left and the resulting outputs on the right.

Note 1.3.1 In fact, if there are n variables, then there are exactly 2^n different combinations of truth values for the variables.

First, let’s establish the truth tables of the basic connectives (that is, of statements containing exactly one connective).

Negation

p	$\neg p$
T	F
F	T

Conjunction

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Disjunction

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

Conditional

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

Biconditional

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

Figure 1.3.2 The truth tables of the **Five basic connectives**.

Note 1.3.3

1. Conjunction $p \wedge q$ is true *only* when *both* p and q are true.

2. Disjunction $p \vee q$ is true when *at least one* of p and q is true.

   See. Statement 2 of Remark 1.1.5 for the difference between **inclusive or** and **exclusive or**.

3. The first two rows of the truth table for $p \to q$ are consistent with the reading "if p is true then q is also true." Really, this reading of the conditional says *nothing* in the case that p is actually false, but we cannot leave the "output" column of the truth table blank for the corresponding rows where $p = \text{F}$. Instead, the outputs in the last two rows of the truth table for $p \to q$ are "default" values chosen to avoid inconsistencies. (See Exercise 1.6.3.)

4. Looking at all four rows of the truth table for $p \to q$, we can succinctly say that $p \to q$ is true *except* when p is true but q is false.

5. Biconditional $p \leftrightarrow q$ reads "p is true when q is true, and *only* when q is true." But this means that when q is false, p cannot be true, hence must by false, which explains the fourth row of the truth table.

6. Looking at all four rows of the truth table for $p \leftrightarrow q$, we can succinctly say that $p \leftrightarrow q$ is true when p and q have the *same truth value*.

We can now use the truth tables of the basic connectives to analyze more complicated statements. Liberal use of extra "intermediate" columns to analyze substatements separately is highly recommended.

Worked Example 1.3.4 Analyze $\neg(p \leftrightarrow q)$.

Solution.

p	q	$p \leftrightarrow q$	$\neg(p \leftrightarrow q)$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	T	F

□

Note 1.3.5 The logical statement $\neg(p \leftrightarrow q)$ analyzed in Worked Example 1.3.4 is one way to realize *exclusive or*: p or q but not both.

Worked Example 1.3.6 Analyze $(p \wedge q) \to (p \leftrightarrow r)$.

Solution.

p	q	r	$p \wedge q$	$p \leftrightarrow r$	$(p \wedge q) \to (p \leftrightarrow r)$
T	T	T	T	T	T
T	T	F	T	F	F
T	F	T	F	T	T
T	F	F	F	F	T
F	T	T	F	F	T
F	T	F	F	T	T
F	F	T	F	F	T
F	F	F	F	T	T

□

Worked Example 1.3.7 Analyze $\big((p \to q) \to r\big) \leftrightarrow \big(p \to (q \to r)\big)$.

Solution.

p	q	r	$p \to q$	$q \to r$	A	B	$A \leftrightarrow B$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	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	T	F	F	T	F
F	F	T	T	T	T	T	T
F	F	F	T	T	F	T	F

□

1.4 Tautologies and contradictions

tautology a logical statement that is always true for all possible truth values of its variable substatements

logically true statement
synonym for **tautology**

Example 1.4.1 Basic tautologies.

1. $p \to p$.

2. $p \leftrightarrow p$.

3. **Law of the Excluded Middle**: $p \vee \neg p$.

   Verification:

p	$\neg p$	$p \vee \neg p$
T	F	T
F	T	T

   The table verifies that the statement is a tautology as the last column consists only of T values.

4. **Law of Contradiction**: $\neg\big(p \wedge \neg p\big)$.

   Verification:

p	$\neg p$	$p \wedge \neg p$	$\neg\big(p \wedge \neg p\big)$
T	F	F	T
F	T	F	T

The table verifies that the statement is a tautology as the last column consists only of T values.

□

Example 1.4.2 Not a tautology. Is $p \vee p$ a tautology? No, since it is false when p is false. □

contradiction
a statement that must always be false, regardless of the truth values of its variable substatements

logically false statement
synonym for **contradiction**

Example 1.4.3 Contradictions.

1. Negation of a tautology is always a contradiction (and negation of a contradiction is always a tautology).

2. Statement $(p \vee \neg p) \to (q \wedge \neg q)$ is a contradiction:

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

The table verifies that the statement is a contradiction as the last column consists only of F values.

□

Example 1.4.4 Conditional versus contradiction. Implication $A \to B$ can *only* be a contradiction if A is a tautology and B is a contradiction. □

Theorem 1.4.5 Substitution Rule. *Suppose A is a logical statement involving substatement variables $p_1, p_2, \ldots, p_m$. If A is logically true or logically false, then so is every statement obtained from A by replacing each statement variable p_i by some logical statement B_i, for every possible collection of logical statements $B_1, B_2, \ldots, B_m$.*

Example 1.4.6 Using the Substitution Rule.

1. We know $p \vee \neg p$ is a tautology, therefore so is

$$\big(q \to (r \wedge \neg s)\big) \vee \neg\big(q \to (r \wedge \neg s)\big)$$

using substitution $p = \big(q \to (r \wedge \neg s)\big)$.

2. We know $(p \vee \neg p) \to (q \wedge \neg q)$ is a contradiction, therefore so are

$$(p \vee \neg p) \to (p \wedge \neg p) \qquad \text{(by } p = p,\ q = p\text{)},$$

$$\big((r \vee s) \vee \neg(r \vee s)\big) \to \big(q \wedge \neg q\big) \qquad \text{(by } p = r \vee s,\ q = q\text{)},$$

$$\big(r \wedge (s \leftrightarrow t)\big) \vee \neg\big(r \wedge (s \leftrightarrow t)\big) \to \Big(t \wedge \neg t\Big) \qquad \text{(by } p = r \wedge (s \leftrightarrow t),\ q = t\text{)}.$$

□

In mathematics, we often wish to prove that a condition $A \to B$ is actually a tautology. (See Chapter 6.)

logically implies
if the conditional $A \to B$ is a tautology, we say that A **logically implies** B

$A \Rightarrow B$ notation for logical implication

Example 1.4.7 Logical implication.

1. If $A = p$ and $B = p \vee q$, then $A \Rightarrow B$.
2. If $A = p \wedge q$ and $B = p$, then $A \Rightarrow B$.

□

Remark 1.4.8 As we will see in Chapter 6, verifying logical implications in mathematical contexts is one of the main tasks of mathematical proof. And to verify a logical implication $A \Rightarrow B$, we want to focus on the idea of conditional as expressing "If A is true then B is true," and we really don't want to concern ourselves with what happens in the case that A is false. Here is where our "default" values in the rows of the truth table for the conditional $A \to B$ where A is false help out — as the conditional $A \to B$ is automatically true when A is false, regardless of the truth value of B, we really only need to consider what happens when A is true to verify $A \Rightarrow B$.

1.5 Activities

Activity 1.1 Consider the following statement.
If the game is on and the popcorn is ready, then Joe is happy.

(a) Assign statement variables and rewrite the statement in symbolic language.

(b) Write out the truth table for your symbolic statement.

(c) You visit Joe's residence room and found that Joe is unhappy even though the game is on. Assuming that the above conditional statement is a true statement about Joe, what can you conclude about the popcorn? Which rows in your truth table justify this conclusion?

Activity 1.2 Consider the logical statement

$$(p \to q) \to (\neg p \vee q).$$

(a) Make up an English language statement that has the same logical structure as this symbolic statement. (Do not just make a word-for-word translation of the logical connectives — make sure you have a reasonable-sounding English sentence when read out loud.)

(b) Argue convincingly that this symbolic statement is a tautology, not by writing out its truth table, but by arguing that it is not possible for the statement to be false.

Don't skip ahead. In Chapter 2, we will learn that the two substatements involved in this conditional are **logically equivalent**. If you have already read ahead into that chapter, do *not* just use this equivalence of these two statements to carry out this task.

Hint. Start with the assumption that this conditional statement *is* false, and then work backwards from the statement to the possible truth values of p and q based on that assumption to conclude that the statement being false is not actually possible.

Activity 1.3 Consider the logical statement

$$(p \wedge \neg r) \to \big[(p \to q) \to (p \wedge \neg r)\big].$$

(a) Make the statement simpler by assigning new variables to represent compound statements and rewriting the statement in terms of the new variables.

(b) Argue that your new statement is a tautology. What does this mean about the original statement?

Activity 1.4 First, re-familiarize yourself with what it means when one statement **logically implies** another.

Suppose that A logically implies B and B logically implies C. Must A logically imply C? Argue convincingly in support of your answer by arguing that the technical definition of **logically implies** is satisfied.

Activity 1.5 *If there is still time:* work through Exercise 1.6.3 from this chapter.

1.6 Exercises

1. Let p, q, and r represent the following statements.

p:	The game is on.
q:	The popcorn is ready.
r:	Joe is happy.

Suppose the following compound statement is true.

If the game is on and the popcorn is ready, then Joe is happy.

However, you just visited Joe's residence room and found that Joe is unhappy even though the game is on. What can you conclude about the popcorn? Use a truth table to justify your answer.

2. Consider the logical statement $(p \wedge q_1) \to (q_1 \vee q_2)$. Partially translated, this statement says:

if p and q_1 are both true, then at least one of q_1 and q_2 is true.

Would you expect this statement to be a tautology? ... a contradiction? ... neither?

Use a truth table to check.

3. The wizard Hatty Porrer is studying logic at Cowpimples School for Second-Rate Wizards. As an exercise, he is filling out the truth table for the conditional

$$(r \wedge s) \to (r \vee s).$$

But he forgets what to do for the lines where the "if" part of the conditional evaluates to false, so he only gets this far:

r	s	$r \wedge s$	$r \vee s$	$(r \wedge s) \to (r \vee s)$
T	T	T	T	T
T	F	F	T	?
F	T	F	T	?
F	F	F	F	?

(a) Help Hatty out by finishing his homework for him.

(b) While you were filling out the truth table, Hatty got bored and opened up a portal to a parallel universe. Parallel Hatty is also working on the same truth table, and is stuck at the same spot that normal Hatty was. However, you notice that parallel Hatty's textbook is open to the page with the truth table for the basic conditional $p \to q$, and it looks as follows.

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

Finish parallel Hatty's homework exercise. Make sure parallel Hatty's instructor will like the result!

(c) While you were finishing parallel Hatty's homework, Hatty got bored again and opened up a portal to *another* parallel universe! Parallel Hatty number two is also working on the same truth table, and is stuck at the same spot that the previous two Hattys were. This time, however, parallel Hatty number two's textbook says that the truth table for the basic conditional $p \to q$ is as follows.

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

Finish parallel Hatty number two's homework exercise. Make sure parallel Hatty number two's instructor will approve!

(d) You'll never believe what happened while you were finishing parallel Hatty number two's homework! Yep, Hatty got bored again and opened up a portal to *a third* parallel universe. Parallel Hatty number three is also working on the same truth table, and is stuck at the same spot that the previous three Hattys were. The truth table for the basic conditional $p \to q$ is different in parallel Hatty number three's universe, yet again.

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

Finish parallel Hatty number three's homework exercise. Make sure parallel Hatty number three's instructor will give him full marks!

(e) OK, so what the heck is the point of all this? The statement $(r \wedge s) \rightarrow (r \vee s)$ could be read as:

> If r and s are both true statements, then at least one of r and s is a true statement.

This conditional statement seems “obviously true”. Based on this example, what do you think of each parallel universe’s system of logic compared to our own?

4. Suppose A, E, U are logical statements such that U is a tautology and E is a contradiction.

(a) Show that $A \vee U$ is always a tautology.

(b) Show that $A \wedge E$ is always a contradiction.

5. Suppose that A, B, and C are logical statements such that $A \Rightarrow B$ and $B \Rightarrow C$. Must $A \Rightarrow C$?

CHAPTER 2

Logical equivalence

2.1 Equivalence

equivalent statements	
	statements A, B such that $A \leftrightarrow B$ is a tautology
$A \Leftrightarrow B$	statements A and B are equivalent

Test 2.1.1 Equivalence of logical statements. *Statements A and B are logically equivalent if A and B* always *have the same output truth value whenever the same input truth values are substituted for the substatement variables in each. That is, $A \Leftrightarrow B$ if A* and B have the same truth table.

Worked Example 2.1.2 Testing logical equivalence. Demonstrate that the following are equivalent statements.

A: If it's nice outside, I will ride my bike.
B: It's not nice outside, or I will ride my bike.

Solution. Let p represent the substatement "it's nice outside," and let q represent the substatement "I will ride my bike." Then the equivalence we want to establish is

$$p \to q \Leftrightarrow \neg p \vee q.$$

We can analyze the truth tables of both statements in the same table.

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

We see that the two statements always have the same truth value in all rows of the truth table, so they are equivalent. □

Note 2.1.3 Worked Example 2.1.2 shows that the basic conditional connective "if ... then ..." can be constructed out of the basic connectives "not" and "or".

Worked Example 2.1.4 Demonstrate the equivalence $p \leftrightarrow q \Leftrightarrow \neg p \leftrightarrow \neg q$.

Solution. Again we build a truth table, and see that the "output" columns for the two statements are identical.

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

□

Proposition 2.1.5 *Logical equivalence has the following properties.*

1. *It is* ***reflexive****. That is, $A \Leftrightarrow A$ is always true.*
2. *It is* ***symmetric****. That is, whenever $A \Leftrightarrow B$, then also $B \Leftrightarrow A$.*
3. *It is* ***transitive****. That is, whenever $A \Leftrightarrow B$ and $B \Leftrightarrow C$, then also $A \Leftrightarrow C$.*
4. *Every pair of tautologies is an equivalent pair of logical statements.*
5. *Every pair of contradictions is an equivalent pair of logical statements.*

Check your understanding. Thinking in terms of truth tables, consider why each of the statements of Proposition 2.1.5 holds.

2.2 Propositional calculus

Logical equivalence gives us something like an "equals sign" that we can use to perform logical "calculations" and manipulations, similar to algebraic calculations and manipulations. To enable us to do such calculations, we first need a "tool chest" of basic logical equivalences to use therein.

Proposition 2.2.1 Rules of Propositional Calculus. *Suppose A, B, C, E, U are logical statements, where E is a contradiction and U is a tautology. Then the following equivalences always hold.*

1. *Rules involving tautologies.*

   (a) $A \vee U \Leftrightarrow U$ (b) $A \wedge U \Leftrightarrow A$

2. *Rules involving contradictions.*

   (a) $A \vee E \Leftrightarrow A$ (b) $A \wedge E \Leftrightarrow E$

3. *Duality of tautologies and contradictions.*

   (a) $\neg U \Leftrightarrow E$ (b) $\neg E \Leftrightarrow U$

4. *Double negation.*

   $\neg\neg A \Leftrightarrow A$

5. *Idempotence.*

   (a) $A \vee A \Leftrightarrow A$ (b) $A \wedge A \Leftrightarrow A$

6. *Commutativity.*

   (a) $A \vee B \Leftrightarrow B \vee A$ (b) $A \wedge B \Leftrightarrow B \wedge A$

7. *Associativity.*

   (a) $(A \vee B) \vee C \Leftrightarrow A \vee (B \vee C)$ *(b)* $(A \wedge B) \wedge C \Leftrightarrow A \wedge (B \wedge C)$

8. *Distributivity.*

   (a) $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$

   (b) $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$

   (c) $(A \vee B) \wedge C \Leftrightarrow (A \wedge C) \vee (B \wedge C)$

   (d) $(A \wedge B) \vee C \Leftrightarrow (A \vee C) \wedge (B \vee C)$

9. *DeMorgan's Laws.*

   (a) $\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$ *(b)* $\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$

10. *Constructing the conditional and biconditional.*

    (a) $A \rightarrow B \Leftrightarrow \neg A \vee B$ *(b)* $A \leftrightarrow B \Leftrightarrow (A \rightarrow B) \wedge (B \rightarrow A)$

Remark 2.2.2 Each of these basic equivalences can be established with a truth table. See Exercise 2.5.4.

Example 2.2.3 DeMorgan. Using Rule 9.b of Proposition 2.2.1, the following are equivalent statements.

1. The triangle can't be both right and equilateral.

2. The triangle is not right or it is not equilateral.

To see how the rule applies, let p represent the statement "the triangle is right" and let q represent the statement "the triangle is equilateral." Then the first statement above is $\neg(p \wedge q)$, while the second statement above is $\neg p \vee \neg q$. □

Now we need some new substitution rules to enable us to use the rules of Proposition 2.2.1 in logical calculations.

Theorem 2.2.4 Substitution Rules.

1. *Replacing a substatement by an equivalent one.*

   Suppose A is a logical statement and X is a substatement of A. If statement Y is equivalent to X, then the new statement A' obtained from A by replacing substatement X by Y is equivalent to A. That is, if $Y \Leftrightarrow X$ then $A' \Leftrightarrow A$.

2. *Substituting into a known equivalence.*

   Suppose A and B are logical statements, each of which involves substatement variables $p_1, p_2, \ldots, p_m$. If A and B are equivalent, then so are new statements A' and B' obtained by applying substitution $p_i = C_i$ to both *A and B, for every collection of statements $C_1, C_2, \ldots, C_m$.*

Proof idea.

1. Think of X as an intermediate column in the calculation of the truth table of A. Replacing X by Y does not change this column, as the truth tables of X and Y are the same.

2. We leave this statement for you, the reader, to consider. (Again, think of the C_i as intermediate columns in the calculations of the truth tables of A'

and B'.)

■

Example 2.2.5 One of DeMorgan's Laws (Rule 9.a of Proposition 2.2.1) says that $\neg(p \vee q) \Leftrightarrow \neg p \wedge \neg q$. Therefore,

$$\neg\big((r \to t) \vee (t \to r)\big) \Leftrightarrow \neg(r \to t) \wedge \neg(t \to r),$$

using Rule 2 of our new Substitution Rules, with substitutions $p = r \to t$, $q = t \to r$.

□

Here is an example of a string of logical manipulations. It also demonstrates the use of Rule 10.a of Proposition 2.2.1 to manipulate an expression involving a conditional.

Example 2.2.6 DeMorgan with a conditional. Consider the statement $(p_1 \vee p_2) \to q$. We may read it as "if either p_1 or p_2 is true, then q will be true as well." So it seems that each of p_1 and p_2 must imply q on its own. Let's see what propositional calculus says about this:

$$\begin{aligned}
(p_1 \vee p_2) \to q &\Leftrightarrow \neg(p_1 \vee p_2) \vee q && \text{(i)}\\
&\Leftrightarrow (\neg p_1 \wedge \neg p_2) \vee q && \text{(ii)}\\
&\Leftrightarrow (\neg p_1 \vee q) \wedge (\neg p_2 \vee q) && \text{(iii)}\\
&\Leftrightarrow (p_1 \to q) \wedge (p_2 \to q) && \text{(iv)},
\end{aligned}$$

with justifications

(i) Rule 2 of our new Substitution Rules, where we substitute $A = p_1 \vee p_2$ into both sides of the construction of the conditional (Rule 10.a of Proposition 2.2.1);

(ii) Rule 1 of our new Substitution Rules, using DeMorgan (Rule 9.a of Proposition 2.2.1) on the substatement $\neg(p_1 \vee p_2)$;

(iii) distributivity (Rule 8.d of Proposition 2.2.1); and

(iv) Rule 1 of our new Substitution Rules, using the construction of the conditional (Rule 10.a of Proposition 2.2.1) on each of the two "factors" of the conjunction.

So our intuition about the logic of a disjunction in a conditional in this way was correct.

A look ahead. This observation will come in handy — see Section 6.4 and Section 6.5.

□

2.3 Converse, inverse, and contrapositive

Related to the conditional $p \to q$ are three important variations.

converse $q \to p$

inverse $\neg p \to \neg q$

contrapositive $\neg q \to \neg p$

Theorem 2.3.1 Modus tollens. *A conditional and its contrapositive are equivalent.*

Proof. We simply compare the truth tables.

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

As the two "output" columns are identical, we conclude that the statements are equivalent. ■

Corollary 2.3.2 Modus tollens for inverse and converse. *The inverse and converse of a conditional are equivalent.*

Proof. The inverse of the conditional $p \to q$ is $\neg p \to \neg q$. The contrapositive of this new conditional is $\neg\neg q \to \neg\neg p$, which is equivalent to $q \to p$ by double negation. ■

Warning 2.3.3 Common mistakes.

- Mixing up a conditional and its converse.
- Assuming that a conditional and its converse are equivalent.

Example 2.3.4 Related conditionals are not all equivalent.

1. Suppose m is a fixed but unspecified whole number that is greater than 2.

   conditional
   If m is a prime number, then it is an odd number.

   contrapositive
   If m is not an odd number, then it is not a prime number.

   converse If m is an odd number, then it is a prime number.

   inverse If m is not a prime number, then it is not an odd number.

   Only two of these four statements are true!

2. Suppose $f(x)$ is a fixed but unspecified function.

   conditional
   If f is continuous, then it is differentiable.

   contrapositive
   If f is not differentiable, then it is not continuous.

   converse If f is differentiable, then it is continuous.

   inverse If f is not continuous, then it is not differentiable.

   Only two of these four statements are true!

□

2.4 Activities

Activity 2.1 Write an English language statement that has the logical form $\neg(A \vee B)$. Then write one that has the form $\neg A \wedge \neg B$, where A and B are the same as in your first sentence. DeMorgan's Laws say your two sentences are logically equivalent. Do you agree?

Activity 2.2 What do you think DeMorgan's Laws would say about $\neg(A \wedge B \wedge C)$? Use propositional calculus to justify your answer.

Activity 2.3 Use Proposition 2.2.1 to simplify $\neg(p \to q)$. (Begin by applying Rule 10.a.)

Activity 2.4 Recall that a pair of coordinates (x, y) defines a point in the Cartesian plane.

Consider the following conditional statement.

> If Cartesian points (a, b) and (c, d) are actually the same point, then $a = c$.

(a) Write out the converse, inverse, and contrapositive of the above statement.

(b) You now have four conditional statements. For each of the four, decide whether it is true, and justify your answer.

(c) For each of the three new conditional statements from Task a in turn, take the view that that statement is the original conditional, and decide which of the others are *its* converse, inverse, and contrapositive.

Activity 2.5 In this activity, we will justify the equivalence

$$p \leftrightarrow q \Leftrightarrow (p \to q) \wedge (q \to p).$$

So consider the statements $A = p \leftrightarrow q$ and $B = (p \to q) \wedge (q \to p)$.

(a) Argue that if A is false, then so is B.

Do *not* use the proposed equivalence above as part of your argument.

(b) Argue that if B is false, then so is A.

Do *not* use the proposed equivalence above as part of your argument.

(c) Explain why the two arguments in Task a and Task b, taken together, justify the equivalence $A \Leftrightarrow B$. Do this *without* making any further arguments about the truth values of p and q.

Activity 2.6 Consider the statements $p \to (q_1 \vee q_2)$ and $(p \wedge \neg q_1) \to q_2$.

Use propositional calculus and substitution to show that these two statements are equivalent.

2.5 Exercises

1. Consider again the two collections of related conditional statements in Example 2.3.4.

(a) For each of these collections, determine which two of the four related statements are true and which two are false. For the two false statements in each collection, demonstrate it by providing examples where the statements are false.

(b) Give an example of a conditional statement involving mathematical objects for which all four of conditional, contrapositive, converse, and inverse are all true.

2. Suppose U is a tautology and E is a contradiction.

(a) Show that $P \wedge U \Leftrightarrow P$ for every statement P.

(b) Show that $P \vee E \Leftrightarrow P$ for every statement P.

3. Consider the equivalence of statements $p \rightarrow (q_1 \vee q_2) \Leftrightarrow (p \wedge \neg q_1) \rightarrow q_2$.

(a) Use a truth table to verify the equivalence.

(b) Use propositional calculus to demonstrate the equivalence.

4. Use truth tables to establish the double negation, idempotence, commutativity, associativity, distributivity, and DeMorgan's Law equivalences presented in Proposition 2.2.1.

5. This exercise asks you to demonstrate that the basic connective "if and only if" can be constructed out of the basic connectives "not", "and", and "or."

(a) Use a truth table to prove Rule 10.b from Proposition 2.2.1.

(b) Starting with Rule 10.b from Proposition 2.2.1, use propositional calculus to prove the equivalence

$$p \leftrightarrow q \Leftrightarrow (\neg p \vee q) \wedge (p \vee \neg q).$$

6. Use Exercise 5 to demonstrate that **exclusive or**

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

is equivalent to

$$p \leftrightarrow \neg q.$$

See. Statement 2 of Remark 1.1.5 for the difference between **inclusive or** and **exclusive or**.

CHAPTER 3

Boolean algebra

3.1 Boolean polynomials

We can proceed more algebraically by assigning value 0 to represent false and value 1 to represent true.

Example 3.1.1 Boolean multiplication. Comparing the two tables below, we see that Boolean multiplication is equivalent to logical conjunction.

x	y	xy
1	1	1
1	0	0
0	1	0
0	0	0

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

□

Example 3.1.2 Boolean addition. Comparing the following two tables, we see that Boolean addition is equivalent to **exclusive or**.

Boolean arithmetic. Notice that in the first row for Boolean addition, we use **mod** 2 **arithmetic** to define $1+1=0$.

x	y	$x+y$
1	1	0
1	0	1
0	1	1
0	0	0

p	q	$\neg(p \leftrightarrow q)$
T	T	F
T	F	T
F	T	T
F	F	F

□

Example 3.1.3 Boolean disjunction. In Boolean arithmetic we may realize disjunction by combining both addition and multiplication.

x	y	$x+y+xy$
1	1	1
1	0	1
0	1	1
0	0	0

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

□

Example 3.1.4 Boolean negation. In Boolean algebra, negation is just a matter of **shifting** one value to the next.

x	$x+1$
1	0
0	1

p	$\neg p$
T	F
F	T

□

For notation, we borrow symbols $\wedge$ and $\vee$ from logic, but add new negation notation.

x' Boolean negation

With this notation setup, we have

$$x \wedge y = xy, \qquad x \vee y = x + y + xy, \qquad x' = x + 1.$$

Boolean polynomial

an expression involving variables $x_1, x_2, \ldots, x_m$ representing Boolean values, and operations $\wedge, \vee, '$, often written in function notation

Note 3.1.5 Every Boolean polynomial can be interpreted as a logical statement.

Example 3.1.6 There are two special constant Boolean polynomials, the zero polynomial and the unit polynomial:

$$\mathbf{0}(x_1, x_2, \ldots, x_m) = 0, \qquad \mathbf{1}(x_1, x_2, \ldots, x_m) = 1.$$

□

Example 3.1.7 The Boolean polynomials $p(x,y) = x' \vee y$ and $q(x,y) = (x \wedge y')'$ have the same truth table.

x	y	x'	$p(x,y)$	y'	$x \wedge y'$	$q(x,y)$
1	1	0	1	0	0	1
1	0	0	0	1	1	0
0	1	1	1	0	0	1
0	0	1	1	1	0	1

Using our knowledge of logical equivalence, we see that the truth tables are the same because as logical statements, p and q are equivalent by DeMorgan. □

equivalent polynomials

Boolean polynomials which represent equivalent logical statements

Fact 3.1.8 Recognizing equivalent Boolean polynomials. *Polynomials p, q are equivalent if and only if they have the same truth table.*

3.2 Disjunctive normal form

It is often desired (e.g. in computer programming or logic circuit design) to reverse the process: starting with a desired truth table, can we construct a Boolean polynomial with the same outputs?

Worked Example 3.2.1 Determine a Boolean polynomial $p(x,y)$ that has the truth table below.

x	y	$p(x,y)$
1	1	1
1	0	0
0	1	0
0	0	1

Solution. We want a "true" output when the inputs match the first or fourth rows, and only then. The inputs match the first row precisely when both x and y are true (i.e. when the conjunction $x \wedge y$ is true), and they match the fourth row precisely when both x is not true and y is not true (i.e. when the conjunction $x' \wedge y'$ is true). So take the disjunction of these two conjunctions: $p(x,y) = (x \wedge y) \vee (x' \wedge y')$.
□

Remark 3.2.2 In the solution to the above worked example, it might seem like we should take a *conjunction* of the two conjunctions instead of a *disjunction*, since we see an output of 1 in the first row *and* in the fourth row. However, we cannot be in the input "state" described by those two rows simultaneously, since neither x nor y can be both 1 and 0 simultaneously. So you should think of it this way instead: if we see an output state of 1, then we know we must be either in the input state of the first row *or* of the fourth row.

disjunctive normal form

a Boolean polynomial in variables $x_1, x_2, \ldots, x_n$ which is the disjunction of *distinct* terms of the form $a_1 \wedge a_2 \wedge \cdots \wedge a_n$, where each a_i is either x_i or x_i'.

Convention 3.2.3 The zero polynomial is also considered to be in disjunctive normal form.

Note 3.2.4 Disjunctive normal form is usually not the "nicest" or "simplest" Boolean polynomial with a desired truth table, but there is a relatively simple procedure to produce it.

Procedure 3.2.5 To produce the disjunctive normal form polynomial for a given Boolean truth table. *Given a truth table with nonzero output, we may obtain a Boolean polynomial in disjunctive normal form with that truth table as follows.*

1. *Identify rows the in truth table for which the desired output is* 1
2. *For each such row, form the conjunction of all variables, but negate those variables that have input value* 0 *for that row.*
3. *Form a polynomial by taking the disjunction of all those conjunctions.*

Worked Example 3.2.6 Determine a Boolean polynomial $p(x,y,z)$ that has the truth table below.

x	y	z	$p(x,y,z)$
1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	1

Solution. The fourth, fifth, seventh, and eigth rows have outcome 1. The corresponding conjunctions are

fourth row
$x \wedge y' \wedge z'$;

fifth row $x' \wedge y \wedge z$;

seventh row
$x' \wedge y' \wedge z$; and

eigth row $x' \wedge y' \wedge z'$.

Therefore, the Boolean polynomial

$$p(x,y,z) = (x \wedge y' \wedge z') \vee (x' \wedge y \wedge z) \vee (x' \wedge y' \wedge z) \vee (x' \wedge y' \wedge z')$$

is both in disjunctive normal form and will have the desired truth table. □

Worked Example 3.2.7 Determine a Boolean polynomial $q(x,y,z)$ that has the truth table below.

x	y	z	$q(x,y,z)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	1
0	0	1	1
0	0	0	1

Solution 1. Every row *except* the third has outcome 1, so we must form conjunctions for all rows except that one:

first row $x \wedge y \wedge z$;

second row
$x \wedge y \wedge z'$;

fourth row
$x \wedge y' \wedge z'$;

fifth row $x' \wedge y \wedge z$;

sixth row $x' \wedge y \wedge z'$;

seventh row
$x' \wedge y' \wedge z$; and

eigth row $x' \wedge y' \wedge z'$.

Therefore, the Boolean polynomial

$$\begin{aligned} q(x,y,z) = {} & (x \wedge y \wedge z) \vee (x \wedge y \wedge z') \vee (x \wedge y' \wedge z') \vee (x' \wedge y \wedge z) \\ & \vee (x' \wedge y \wedge z') \vee (x' \wedge y' \wedge z) \vee (x' \wedge y' \wedge z') \end{aligned}$$

is both in disjunctive normal form and will have the desired truth table.

Solution 2 (Alternative solution). We can get a much simpler expression for $q(x,y,z)$ by *not* using the procedure (though of course the result will not be in disjunctive normal form).

Notice that we want the third row to have output value 0. In logic terms, we want that combination (and only that combination) of input values to result in an output that is "*not* true". So the Boolean polynomial

$$q(x,y,z) = (x \wedge y' \wedge z)' = x' \vee y \vee z'$$

will produce the desired truth table. □

Note 3.2.8 The polynomials in the solutions to the preceding examples are in disjunctive normal form, but the alternative solution to the second example is not.

Fact 3.2.9 *From Procedure 3.2.5, it is easy to see that* any *Boolean polynomial can be expressed in disjunctive normal form.*

Worked Example 3.2.10 Converting a polynomial into disjunctive normal form. Rewrite the Boolean polynomial $p(x,y,z) = (x \wedge z)' \vee (x' \wedge y)$ in disjunctive normal form.

Solution. First, produce the truth table.

x	y	z	$x \wedge z$	$x' \wedge y$	$p(x,y,z)$
1	1	1	1	0	0
1	1	0	0	0	1
1	0	1	1	0	0
1	0	0	0	0	1
0	1	1	0	1	1
0	1	0	0	1	1
0	0	1	0	0	1
0	0	0	0	0	1

Then apply the disjunctive normal form procedure to obtain

$$\begin{aligned} p(x,y,z) = {} & (x \wedge y \wedge z') \vee (x \wedge y' \wedge z') \vee (x' \wedge y \wedge z) \\ & \vee (x' \wedge y \wedge z') \vee (x' \wedge y' \wedge z) \vee (x' \wedge y' \wedge z'). \end{aligned}$$

□

Check your understanding. What do you think **conjunctive normal form** should mean? Can you come up with a procedure which takes a truth table and determines a Boolean polynomial in conjunctive normal form with the desired truth table?

Hint. Extend the idea of the Alternative solution for Worked Example 3.2.7.

3.3 Exercises

Creating truth tables. In each of Exercises 1–2, write out the truth table for the given boolean polynomial.

1. $p(x,y) = (x \wedge y)' \wedge x'$.

2. $q(x,y,z) = (x \vee y)' \wedge (z \vee x) \wedge y$.

3. Explain why the boolean polynomial $p(x,y) = x \vee y \vee y'$ is *not* in disjunctive form.

Disjunctive normal form from a truth table. In each of Exercises 4–6, write out a boolean polynomial in disjunctive normal form that has the given truth table.

4.

x	y	$p(x,y)$
1	1	1
1	0	1
0	1	1
0	0	0

5.

x	y	$p(x,y)$
1	1	1
1	0	0
0	1	1
0	0	0

6.

x	y	z	$p(x,y,z)$
1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	0
0	0	0	0

Disjunctive normal form from a boolean polynomial. In each of Exercises 7–9, write out a boolean polynomial in disjunctive normal form that is equivalent to the given boolean polynomial.

7. $p(x,y,z) = (x \vee y) \wedge z$.

8. $q(x,y,z) = \left[(x \wedge y') \vee (x \wedge z)\right]' \vee x'$.

9. $r(x,y,z) = (x \wedge y') \vee (x \wedge z) \vee (x \wedge y)$.

CHAPTER 4

Predicate logic

4.1 Predicates and quantifiers

We often let variables represent arbitrary mathematical objects. However, as we have seen, object variables or **free variables** (as opposed to **statement variables**) lead to problems in logic. For example, the phrase "f is a differentiable function" can only be determined to be true or false when f represents a specific function.

In this section, we deal with these problems by **quantifying** such free variables: restricting ourselves to discussing whether "statements" involving one or more free variables are always/sometimes/never true for objects of the type represented by the free variables.

predicate	a statement whose truth value depends on one or more free variables
$A(x)$	a predicate statement A whose truth value depends on the free variable x
$A(x,y)$	a predicate statement A whose truth value depends on the free variables x and y

Example 4.1.1

1. Let $A(f)$ represent the phrase "f is differentiable", a predicate statement in one free variable f.

2. Let $B(m,n)$ represent the phrase "m is greater than n", a predicate statement in two free variables m and n.

□

Note 4.1.2 A predicate is not a logical statement unless all of its variables represent specific objects.

domain	the type of object that a variable in a predicate represents

Example 4.1.3 In Example 4.1.1, the domain of the variable f could be "functions of a single real variable", and the domains of the variables m and n could both be "natural numbers" (i.e. whole, nonnegative numbers). □

Note 4.1.4 Usually the domain of a variable in a predicate is implicit and can be determined from the context of the statement. However, if we want to make the domain explicit we can prefix it to the variable. For example,

$$A(f) = \text{“function } f \text{ is differentiable”},$$
$$B(m,n) = \text{“integer } m \text{ is greater than integer } n\text{”}.$$

We can turn a predicate into a logical statement by being more specific about which objects in their domains the variables represent. However, we often do not want to be *too* specific (or else we would usually not need variables).

Example 4.1.5 The following sentences are logical statements, because their truth value can be determined.

- “Every function f is differentiable.”
- “There exists an integer m that 2 divides.”

The first statement is false; for example, the function $f(x) = |x|$ is not differentiable at $x = 0$. The second statement is true; this statement basically says that even integers exist. □

quantifier
the sentence fragments “for every” and “there exists” **quantify** whether a predicate should apply to *all* or only *some* of the objects in the domain of one of its variables.

universal quantifier
the quantifier “for every”

$\forall$ symbol for the universal quantifier

existential quantifier
the quantifier “there exists”

$\exists$ symbol for the existential quantifier

Example 4.1.6 As before, let $A(f)$ represent the predicate “f is differentiable.” Then the statement $(\forall f)A(f)$ is false, because not every function is differentiable. However, the statement $(\exists f)A(f)$ is true — for example, polynomials are differentiable. □

Warning 4.1.7 For an existentially quantified statement to be true, it is not necessary for there to be *one and only one* object in the implied domain that satisfies the conditions of the predicate — there could be many such objects. So, just as you should always read a disjunction $p \vee q$ as “p or q *or both*,” you should always read an existentially quantified statement $(\exists x)A(x)$ as “there exists *at least one* x such that $A(x)$ is true.”

Mathematical statements often involve several quantified variables.

Worked Example 4.1.8 Working with quantified statements of several variables. Let $B(m,n)$ represent “m divides n”, where m and n are positive whole numbers. Which of the following statements are true?

1. $(\forall m)(\forall n)B(m,n)$
2. $(\exists m)(\exists n)B(m,n)$
3. $(\forall m)(\exists n)B(m,n)$
4. $(\exists m)(\forall n)B(m,n)$
5. $(\forall n)(\exists m)B(m,n)$
6. $(\exists n)(\forall m)B(m,n)$

Solution.

1. *False.*

   This statement says “for every m and for every n, m divides n.” One example of values for m and n with m *not* dividing n (such as $m = 3$ and $n = 2$) suffices to show that it is not *always* true that one number m divides another number n.

2. *True.*

   This statement says “there exists m such that there exists n such that m divides n.” To demonstrate that this statement is true, we have to explicitly show that at least one pair of values for m and n exists by giving an example (such as $m = n = 2$).

3. *True.*

   This statement says “for every m there exists an n such that m divides n.” To show that this statement is true, we have to provide, for every possible value of m, a value of n that works. When $m = 1$, we have example $n = 1$. When $m = 2$, we have example $n = 2$. When $m = 3$, we have example $n = 3$. Similarly, for each value of m, we can choose n to be the same value as m as an example.

   Note. If the domain for m and n included 0, then this third statement would actually be false, as demonstrated by the example $m = 0$.

4. *True.*

   This statement says “there exists m such that for every n we have m divides n.” This is true, as the example $m = 1$, which divides every number n, demonstrates existence of this special m (though it is the only example possible).

5. *True.*

   This statement says “for every n there exists an m so that m divides n.” Similarly to the justification for Statement 3, for every possible value of n we need to provide an example m so that m divides n, but this time it is the n that is arbitrary and the m that is to be the example. But again, every positive number divides itself, so we could always take m to be the same value as n as our example to demonstrate that such an m exists. (Or, actually we could always choose $m = 1$ as the example for each different value of n.)

6. *False.*

   This statement says “there exists n such that for every m we have m divides n.” However, there is no positive number that is divisible by every other positive number.

□

Warning 4.1.9 While the order of two quantifiers of the same type does not matter (which is why we didn’t consider the statements with quantifiers in the order $(\forall n)(\forall m)$ and $(\exists n)(\exists m)$ in Worked Example 4.1.8 above), ***the order of a “mixed” pair of quantifiers matters***! This is demonstrated by Statement 3 and Statement 6 in Worked Example 4.1.8 — both of these statements involve a $\forall m$ and a $\exists n$, but in opposite orders. Since one of these two statements is true and one is false, they obviously cannot be the same statement.

Even more, the order of a “mixed” pair of quantifiers implies a dependence of the second quantified variable on the first.

1. If the statement $(\forall x)(\exists y)A(x,y)$ is true, it means that, corresponding to *each and every* object x in the appropriate domain, there will exist at least one example of an object y in the appropriate domain so that $A(x,y)$ is true. But the corresponding example y could be different for different examples of the object x.

2. If the statement $(\exists y)(\forall x)A(x,y)$ is true, it means that there is at least one "special" example object y that enjoys the property that $A(x,y)$ will be true for *each and every* object x in the appropriate domain.

Example 4.1.10 Suppose $A(x,y)$ is a predicate statement, where x and y are variables that can only take on the values 0, 1, or 2. Further suppose that it is known that $A(x,y)$ is true in the specific instances

$$A(0,1), A(1,0), A(1,1), A(1,2), A(2,2).$$

1. Statement $(\forall x)(\exists y)A(x,y)$ is true because for each value of x we can exhibit at least one value y for which $A(x,y)$ is true:
   - when $x = 0$, we know $A(0,y)$ is true for at least one y (for example, $y = 1$);
   - when $x = 1$, we know $A(1,y)$ is true for at least one y (for example, $y = 1$); and
   - when $x = 2$, we know $A(2,y)$ is true for at least one y (for example, $y = 2$).

2. Statement $(\exists x)(\forall y)A(x,y)$ is true because we can exhibit at least one "special" value of x for which $A(x,y)$ is true for each and every value of y. In particular, we see that for $x = 1$ we have $A(1,y)$ true for each of $y = 0,1,2$.

□

Remark 4.1.11 Depending on grammar requirements or personal style, the quantifier "for every" might be expressed as "for all" or just "every" or "all". The quantifier "there exists" can also be expressed as "some" or "there is at least one", but remember that the reality of the situation *could* be "more than one."

Warning 4.1.12 Mathematicians are fond of using "any" or "for any" in place of "every" or "for every".

Worked Example 4.1.13 Prove that $n+1$ is an odd number for any even number n.

Solution 1 ((Incorrect)). It says to prove $n+1$ is odd for *any* even number n, so I will choose my favourite even number $n = 8$. Then $8+1 = 9$ is obviously odd.

Solution 2 ((Correct)). The problem statement is really asking to prove that *every* even number has the property that the subsequent number is odd. So let n represent an *arbitrary but unspecified* even number. Then n is divisible by 2, so there is some number m such that $n = 2m$. Now, $n+1 = 2m+1$ is not divisible by 2, since $(2m+1)/2 = m + \frac{1}{2}$ is not a whole number. Therefore, $n+1$ must be odd.

□

Note 4.1.14 We will practice proving statements involving universally quantified predicates in Chapter 6.

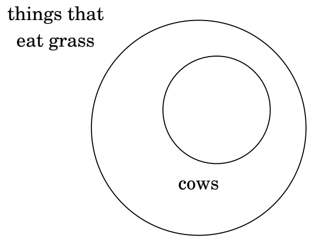
4.2 Manipulating quantified statements

4.2.1 Negation of quantified statements

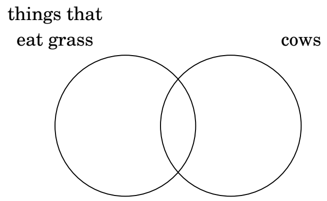
Negating quantified statements in English can be tricky, but we will establish rules that make it easy in symbolic logic.

Warning 4.2.1 The negation of the statement "all X are Y" is *not* "no X are Y" *nor* "all X are not Y".

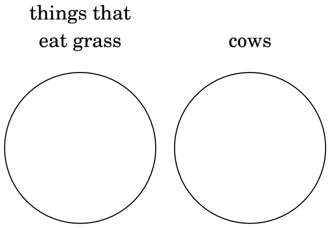
Example 4.2.2 What is the negation of the statement "all cows eat grass"? To avoid making the mistake in the preceding warning, consider the following question: what is the minimum number of cows that do not eat grass that can be used as evidence to demonstrate that the statement "all cows eat grass" is false?



In this case, both "all cows eat grass" and "some cows eat grass" are true.



In this case, both "some cows eat grass" and "some cows do not eat grass" are true.



In this case, each of "no cows eat grass", "all cows do not eat grass", and "some cows do not eat grass" are true.

It takes just *one* lasagna-eating cow to make "all cows eat grass" false, so the negation of "all cows eat grass" is "some cows do not eat grass" or "at least one cow does not eat grass". □

Warning 4.2.3 We have indicated that the statement "some cows eat grass" is technically true in the first diagram in Example 4.2.2, even though it is more precise to say "all cows eat grass" in that situation. Similarly, we have indicated that the statement "some cows do not eat grass" is technically true in the third diagram. Remember that truth and falsity are usually all that matter in logic, not necessarily making the most precise statement possible.

Proposition 4.2.4 Rules for negation of quantifiers. *Let $A(x)$ represent a predicate in the variable x.*

1. *Universal negation.*

   The negation of $(\forall x)A(x)$ is $(\exists x)(\neg A(x))$.

2. *Existential negation.*

   The negation of $(\exists x)A(x)$ is $(\forall x)(\neg A(x))$.

Check your understanding. Use the "cows eat grass" diagrams in Example 4.2.2 to convince yourself that these negation rules are correct.

Worked Example 4.2.5 Negating a quantified statement. Determine and "simplify" the negation of

$$(\forall x)\Big(A(x) \to \Big((\exists y)(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big)\Big).$$

Solution. Using the rules of quantifier negation and known logical equivalences, we can perform the following manipulations:

$$\begin{aligned}
&\neg(\forall x)\Big(A(x) \to (\exists y)(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big) && \\
&\Leftrightarrow (\exists x)\neg\Big(A(x) \to (\exists y)(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big) && \text{(i)}\\
&\Leftrightarrow (\exists x)\neg\Big(\neg A(x) \vee (\exists y)(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big) && \text{(ii)}\\
&\Leftrightarrow (\exists x)\Big(A(x) \wedge \neg(\exists y)(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big) && \text{(iii)}\\
&\Leftrightarrow (\exists x)\Big(A(x) \wedge (\forall y)\neg(\forall z)\big(B(y) \wedge \neg C(z)\big)\Big) && \text{(iv)}\\
&\Leftrightarrow (\exists x)\Big(A(x) \wedge (\forall y)(\exists z)\neg\big(B(y) \wedge \neg C(z)\big)\Big) && \text{(v)}\\
&\Leftrightarrow (\exists x)\Big(A(x) \wedge (\forall y)(\exists z)\big(\neg B(y) \vee C(z)\big)\Big) && \text{(vi)}
\end{aligned}$$

with justifications

(i) quantifier negation;

(ii) known equivalence $p \to q \Leftrightarrow \neg p \vee q$;

(iii) DeMorgan, double negation;

(iv) quantifier negation;

(v) quantifier negation; and

(vi) DeMorgan, double negation.

□

4.2.2 Distributing quantifiers

Proposition 4.2.6 Rules for distributing quantifiers. *Let $A(x), B(x)$ represent predicates in the variable x.*

1. *Universal distributes across conjunction.*

   $(\forall x)\big(A(x) \wedge B(x)\big) \Leftrightarrow (\forall x)A(x) \wedge (\forall x)B(x)$

2. *Existential distributes across disjunction.*

   $(\exists x)\big(A(x) \vee B(x)\big) \Leftrightarrow (\exists x)A(x) \vee (\exists x)B(x)$

Example 4.2.7

- The statement "every vegetable is delicious and nutritious" is the same as saying "every vegetable is delicious and every vegetable is nutritious".
- The statement "at least one vegetable in the garden is rotten or nibbled by squirrels" is the same as saying "at least one vegetable in the garden is rotten or at least one vegetable in the garden is nibbled by squirrels".

□

Warning 4.2.8

- The universal quantifier $\forall$ does *not* distribute over disjunction $\vee$.
- The existential quantifier $\exists$ does *not* distribute over conjunction $\wedge$.

Check your understanding.

1. Create an example of predicates $A(x)$ and $B(x)$ such that, of the statements
$$(\forall x)\big(A(x) \vee B(x)\big), \qquad (\forall x)A(x) \vee (\forall x)B(x),$$
the first is true but the second is false.
2. Create an example of predicates $A(x)$ and $B(x)$ such that, of the statements
$$(\exists x)\big(A(x) \wedge B(x)\big), \qquad (\exists x)A(x) \wedge (\exists x)B(x),$$
the first is false but the second is true.

4.3 Vacuously true statements

We have to be careful with quantified predicates because it is (seemingly) possible to violate the **Law of Contradiction** (see Basic Tautology 4 in Example 1.4.1).

Example 4.3.1 Let x be a variable in the domain of all living humans. Define predicates

$$\begin{aligned} A(x) &= \text{“}x\text{ is an Augustana student,”} \\ B(x) &= \text{“}x\text{ is three hundred years old,”} \\ C(x) &= \text{“}x\text{ is tall,”} \end{aligned}$$

and consider the statement

$$(\forall x)\Big(\big(A(x) \wedge B(x)\big) \to C(x)\Big),$$

which says “every three-hundred-year-old Augustana student is tall”. This statement is true, since a conditional $p \to q$ is true when p is false, and $A(x) \wedge B(x)$ is false for each and every x: there is no living human who is both three hundred years old and is an Augustana student (issues concerning the existence of vampires notwithstanding). But by the same reasoning, the statement “every three-hundred-year-old Augustana student is *not* tall” is true. This seems to be a contradiction: how can every three-hundred-year-old Augustana student be both tall and not tall? The answer is that you can say anything you like about things that *do not exist* and your statement will be true. So you should avoid altogether making claims about things that do not exist. □

vacuously true
a statement of the form $(\forall x)\big(P(x) \to Q(x)\big)$ where $P(x)$ is false for every x in its domain

Check your understanding. Determine the negation of $(\forall x)\big(P(x) \to Q(x)\big)$. Is the negation of a vacuously true statement true or false?

4.4 Activities

Activity 4.1

(a) Devise an example of predicates $A(x)$ and $B(x)$ such that, of the statements

- $(\forall x)\big(A(x) \lor B(x)\big)$, and
- $(\forall x)A(x) \lor (\forall x)B(x)$,

the first is true but the second is false.

(b) Devise an example of predicates $A(x)$ and $B(x)$ such that, of the statements

- $(\exists x)\big(A(x) \land B(x)\big)$, and
- $(\exists x)A(x) \land (\exists x)B(x)$,

the first is false but the second is true.

Activity 4.2 Let $P(f,g)$ represent the predicate $\frac{df}{dx} = g$, where f and g are free variables in the domain of *continuous* functions in the real variable x.

For each of the following, determine whether the statement is true or false. Explain your reasoning.

(a) $(\exists f)(\exists g)P(f,g)$

(b) $(\forall f)(\forall g)P(f,g)$

(c) $(\forall f)(\exists g)P(f,g)$

(d) $(\exists f)(\forall g)P(f,g)$

(e) $(\forall g)(\exists f)P(f,g)$

(f) $(\exists g)(\forall f)P(f,g)$

Activity 4.3 Let $P(f,g)$ represent the predicate $\frac{df}{dx} = g$, and let $E(f,g)$ represent the predicate $g = f$, where f and g are functions in the real variable x. Consider the statement

$$(\forall f)(\forall g)\Big((\exists h)\big(P(f,h) \land P(g,h)\big) \to E(f,g)\Big).$$

(a) Translate the statement into English.

(b) Determine whether the statement is true.

(c) Working with the originally provided symbolic version above, negate the statement. Simplify the negated version to so that any/all negation symbols appear directly to the left of one of the predicates P or E.

(d) Translate your simplified negated statement from Task c into English.

Activity 4.4 You've become an expert at predicate logic, and now make a (very meagre) living grading logic assignments for a large university. Here is the question you've been assigned to mark two thousand times.

> Let x represent a free variable from the domain of all living humans.
>
> Translate the following two statements into properly quantified predicate statements in the variable x.
>
> (i) All university students study diligently.
>
> (ii) Some university students study diligently.

You pick up the first assignment. Here is the student's answer.

> Let $U(x)$ mean "x is a university student". Let $S(x)$ mean "x studies diligently".
>
> (i) $(\forall x)\big[U(x) \to S(x)\big]$.
>
> (ii) $(\exists x)\big[U(x) \to S(x)\big]$.

Are the student's answers correct? Justify your assessment.

Hint. Try translating the student's symbolic language statements back into English, *explicitly using the stated domain of* x, and see what you get. Is it possible for the student's version of the statement to be true in a way that goes against the idea expressed by the original English version of the statement?

4.5 Exercises

Interpreting symbolic language. Let $A(x)$ represent the predicate "x is a wonderful learning experience", where x is a free variable in the domain of all university courses.

Translate each of the following into an English sentence that is grammatically correct.

1. $A(\text{AUMAT 250})$

2. $(\exists x)A(x)$

3. $(\forall x)A(x)$

4. $\neg(\forall x)A(x)$

5. $(\exists x)\neg A(x)$

Translating into symbolic language. Let $B(x)$ represent the predicate "x is excellent", where x is a free variable in the domain of all Augustana professors.

Translate each of the following into symbolic language.

6. The instructor for this course is an excellent professor.

7. Every professor at your university is excellent.

8. Some professor at your university is excellent.

9. Some professors at your university are excellent.

10. There is at least one professor at your university who is excellent.

11. Some professor at your university is not excellent.

12. Some professors at your university are not excellent.

13. Any professor at your university is excellent.

14. No professor at your university is excellent.

Analyzing predicate statements about integers. Let $P(m,n)$ represent the predicate $2m - 45n > 101$, where m and n are free variables in the domain of integers.

For each of the following, determine whether the statement is true or false. Explain your reasoning.

15. $P(25,-1)$

16. $P(30,-1)$

17. $P(100,2) \vee P(100,3)$

18. $P(100,2) \wedge P(100,3)$

19. $(\exists m)(\exists n)P(m,n)$

20. $(\forall m)(\forall n)P(m,n)$

21. $(\forall m)(\exists n)P(m,n)$

22. $(\exists m)(\forall n)P(m,n)$

23. $(\forall m)(\exists q)(\forall n)(P(q,n) \to P(m,n))$

Analyzing predicate statements about functions. *(Requires calculus.)* Let $P(f,g)$ represent the predicate $\frac{df}{dx} = g$, where f and g are free variables in the domain of *continuous* functions in the real variable x.

For each of the following, determine whether the statement is true or false. Explain your reasoning.

24. $(\exists f)(\exists g)P(f,g)$

25. $(\forall f)(\forall g)P(f,g)$

26. $(\forall f)(\exists g)P(f,g)$

27. $(\exists f)(\forall g)P(f,g)$

28. $(\forall g)(\exists f)P(f,g)$

29. $(\exists g)(\forall f)P(f,g)$

30. Consider the statement "every odd number is either 1 more or 3 more than a mulitple of 4."

(a) Assign appropriate predicates (with domains explictly stated), and then translate the statement into symbolic logic.

(b) Negate the statement, and simplify the logical expression so that any/all negation symbols appear directly to the left of a predicate.

(c) Translate your simplified negated statement from Task b into English.

31. Let $P(f,g)$ represent the predicate $\frac{df}{dx} = g$, and let $E(f,g)$ represent the predicate $g = f$, where f and g are free variables in the domain of functions in the real variable x. Consider the statement

$$(\forall f)(\forall g)\Big((\exists h)\big(P(f,h) \wedge P(g,h)\big) \to E(f,g)\Big).$$

(a) Translate the statement into English.

(b) Determine whether the statement is true.

(c) Working with the originally provided symbolic version above, negate the statement. Simplify the negated version to so that any/all negation symbols appear directly to the left of one of the predicates P or E.

(d) Translate your simplified negated statement from Part c into English.

32. You've become an expert at predicate logic, and now make a (very meagre) living grading logic assignments for a large university. Here is the question you've been assigned to mark two thousand times.

> Let x represent a free variable from the domain of all living humans.
>
> Translate the following two statements into properly quantified predicate statements in the variable x.
>
> (i) All university students study diligently.
>
> (ii) Some university students study diligently.

You pick up the first assignment. Here is the student's answer.

> Let $U(x)$ mean "x is a university student". Let $S(x)$ mean "x studies diligently".

(i) $(\forall x)[U(x) \to S(x)]$.

(ii) $(\exists x)[U(x) \to S(x)]$.

Are the student's answers correct? Justify your assessment.

Hint. Try translating the student's symbolic language statements back into English, *explicitly using the stated domain of* x, and see what you get.

Chapter 5

Arguments

5.1 Basics

Studying the logic of individual statements is an important first step, but ultimately we will need to analyze how statements can be combined into an **argument** (mathematical, philosophical, political, or otherwise) that tries to convince others that some particular **conclusion** is true.

argument a finite collection of statements, called **premises** or **hypotheses**, along with a final statement, called the **conclusion**

$A_1, A_2, \ldots, A_m \therefore C$
an argument with premises $A_1, A_2, \ldots, A_m$ and conclusion C

$$\begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_m \\ \hline C \end{array}$$

an argument with premises $A_1, A_2, \ldots, A_m$ and conclusion C

Example 5.1.1 Argument in English.

(premise)	If the world is flat, it has an edge.
(premise)	The world does not have an edge.
(conclusion)	Therefore, the world is not flat.

□

Example 5.1.2 Another argument in English.

(premise)	If the world is round, there exists a Titan named Atlas who holds it aloft in the heavens.
(premise)	The world is round.
(conclusion)	Therefore, the Titan Atlas exists.

□

Example 5.1.3 Yet another argument in English.

(premise)	Rectangles are geometric objects that have four sides.
(premise)	Parallelograms have four sides.
(premise)	Tetrahedrons have four sides.
(conclusion)	Therefore, parallelograms and tetrahedrons are rectangles.

□

When we analyse an argument, one component of the analysis should be to check whether or not its logical structure is **valid**, regardless of the content and truth/falsity of the individual statements making up the argument.

Question 5.1.4 Of the three provided English-language example arguments above, which are "true"? Which are "logically correct"? Is there a difference?

valid argument
whenever the premises are all true, the conclusion must be true as well

Warning 5.1.5 Whether the conclusion of an argument is *actually* true is irrelevant to the validity of the argument! It is the combination of possibilities of truth and falsity of the premises and conclusion *together* that determine whether an argument is valid.

Test 5.1.6 For validity of an argument in symbolic language. *If there is no choice of truth values for the statement variables that simultaneously make the premises all true but the conclusion false, then the argument is valid.*

Worked Example 5.1.7 Test the validity of the following argument.

$$\begin{array}{c} p \to q \\ q \to r \\ \hline p \to r \end{array}$$

Solution 1. Let's write out the truth tables for the statements in the argument. However, *we are only concerned with truth table rows where every premise is true*, so we won't bother completing any rows where a premise ends up being false.

			(pr)	(pr)	(c)	
p	q	r	$p \to q$	$q \to r$	$p \to r$	
T	T	T	T	T	T	✓
T	T	F	T	F	∗	
T	F	T	F	∗	∗	
T	F	F	F	∗	∗	
F	T	T	T	T	T	✓
F	T	F	T	F	∗	
F	F	T	T	T	T	✓
F	F	F	T	T	T	✓

As every row that resulted in both premises true also resulted in the conclusion true (as indicated by ✓), the argument is valid. (The ∗ symbol indicates a truth value that we don't care about, since it is in a row with at least one premise false.)

Solution 2 (Alternative solution). Rather than work out the whole truth table, let us consider the question: is there any possible way for the conclusion to be false but all the premises true? Start with the following partial truth table row.

			(pr)	(pr)	(c)
p	q	r	$p \to q$	$q \to r$	$p \to r$
			T	T	F

The conclusion is only false when $p = \mathrm{T}$ and $r = \mathrm{F}$; fill these into the row. Now since $p = \mathrm{T}$, the first premise can only be true if $q = \mathrm{T}$; fill this into the row.

			(pr)	(pr)	(c)	
p	q	r	$p \to q$	$q \to r$	$p \to r$	
T	T	F	T	T	F	×

We have marked this row as "incorrect," because $q = \text{T}$ and $r = \text{F}$ should make the second premise false! So the above truth table row is inconsistent, and therefore there is no way for conclusion to be false and all the premises true. Conclude that the argument is valid. □

Remark 5.1.8 The reasoning in the Alternative solution above is an example of a **proof by contradiction** — see Section 6.9.

Worked Example 5.1.9 Demonstrate that the following argument is *in*valid.

> If n is even, then n is divisible by 2.
> If n is odd, then $n+1$ is even.
> If n is not divisible by 2, then n is not divisible by 4.
> ―――――――――――――――――――――――――
> Therefore, if n is not divisible by 4, then $n+1$ is even.

Solution. Introduce statement variables and write the argument in symbolic form.

$p =$ "n is even"
$q =$ "$n+1$ is even"
$r =$ "n is divisble by 2"
$s =$ "n is divisble by 4"

$$\begin{array}{c} p \to r \\ \neg p \to q \\ \neg r \to \neg s \\ \hline \neg s \to q \end{array}$$

Try to construct a truth table row in which all the premises are true but the conclusion is false.

							(pr)	(pr)	(pr)	(c)
p	q	r	s	$\neg p$	$\neg r$	$\neg s$	$p \to r$	$\neg p \to q$	$\neg r \to \neg s$	$\neg s \to q$
							T	T	T	F

Start with the above partial truth table row. The conditional in the conclusion can only be false when $\neg s = \text{T}$ and $q = \text{F}$; fill these into the row, also entering $s = \text{F}$. Now since $q = \text{F}$, the second premise will only be true when $\neg p = \text{F}$; fill this and $p = \text{T}$ into the row. Now since $p = \text{T}$, the first premise will only be true when $r = \text{T}$; fill this and $\neg r = \text{F}$ into the row. Finally, check that our choices of truth values for p, q, r, s are consistent with the imposed truth value for the third premise.

							(pr)	(pr)	(pr)	(c)
p	q	r	s	$\neg p$	$\neg r$	$\neg s$	$p \to r$	$\neg p \to q$	$\neg r \to \neg s$	$\neg s \to q$
T	F	T	F	F	F	T	T	T	T	F

Since there exists a choice of truth values for the statement variables which makes all premises true but the conclusion false, the argument is invalid. □

Proposition 5.1.10 Some technicalities regarding argument validity.

1. *If the conclusion is a tautology, the argument is automatically valid.*

2. *If the premises are all contradictions (i.e. logically false), the argument is automatically valid.*

3. *If the argument is valid and the premises are all tautologies, then the conclusion must also be a tautology.*

4. *If the argument is valid and the conclusion is a contradiction, then the premises can't all be true at the same time. (That is, in this situation the conjunction of all the premises must be a contradiction.)*

Remark 5.1.11

1. In the logical analysis of an argument, we don't care if the premises are *actually* true. We only care that the conclusion follows from the premises.
2. *The order of the premises is irrelevant to the validity of the argument.* For arguments written in English language, there may be a preferred order that best illuminates validity or invalidity, but this is essentially only aesthetic from a logical-analysis point of view.

Check your understanding. Verify that an argument $A_1, A_2, \ldots, A_m \therefore C$ is valid if and only if $A_1 \wedge A_2 \wedge \cdots \wedge A_m \Rightarrow C$.

5.2 Standard arguments

5.2.1 Modus ponens

modus ponens

standard argument with form

$$\begin{array}{l} p \to q \\ p \\ \hline q \end{array}$$

Worked Example 5.2.1 Verify the validity of the modus ponens standard argument.

Solution. Verify the validity by ensuring that each row in the truth table with premises all true also has the conclusion true.

(pr)	(c)	(pr)	
p	q	$p \to q$	
T	T	T	✓ argument is valid
T	F	F	
F	T	*	
F	F	*	

□

Example 5.2.2 The argument in Example 5.1.2 has modus ponens form. So it is valid, even though the first premise and the conclusion are not actually true. □

5.2.2 Modus tollens

modus tollens

standard argument with form

$$\begin{array}{l} p \to q \\ \neg q \\ \hline \neg p \end{array}$$

Worked Example 5.2.3 Verify the validity of the modus tollens standard argument.

Solution. Verify the validity by ensuring that each row in the truth table with premises all true also has the conclusion true.

		(pr)	(pr)	(c)	
p	q	$p \to q$	$\neg q$	$\neg p$	
T	T	T	F	*	
T	F	F	*	*	
F	T	T	F	*	
F	F	T	T	T	✓ argument is valid

□

Example 5.2.4 The argument in Example 5.1.1 has modus tollens form. □

5.2.3 Law of Syllogism

Law of Syllogism
standard argument with form

$$\begin{array}{c} p \to q \\ q \to r \\ \hline p \to r \end{array}$$

Note 5.2.5 We already verified that the Law of Syllogism is valid in Worked Example 5.1.7.

The Law of Syllogism may be extended to chains of conditionals of arbitrary (finite) length.

Extended Law of Syllogism
standard argument with form

$$\begin{array}{c} p_1 \to p_2 \\ p_2 \to p_3 \\ \vdots \\ p_{n-1} \to p_n \\ \hline p_1 \to p_n \end{array}$$

Note 5.2.6 We will verify that the extended Law of Syllogism is a valid argument using mathematical induction in Section 7.2.

Example 5.2.7 A syllogistic argument in English.

If I don't study hard this term, I won't master the course material.
If I don't master the course material, I will fail the course.
If I fail the course, I will have to take it again next year.
If I take it again next year, I will have to study harder.

Therefore, if I don't study hard this term, I will have to study harder next year.

□

5.3 Substituting into an argument

Substituting into an argument does not change its validity.

Theorem 5.3.1 Substitution Rule. *Suppose $A_1, A_2, \ldots, A_m \therefore C$ is a valid argument involving statement variables $p_1, p_2, \ldots, p_\ell$. If we apply substitution $p_i \to B_i$ to each of $A_1, A_2, \ldots, A_m, C$, for some collections of statements $B_1, B_2, \ldots, B_\ell$, then the resulting argument is also valid.*

Example 5.3.2 Since modus tollens is a valid argument, using the substitution rule with the equivalences

$$r \wedge p \Leftrightarrow \neg(\neg r \vee \neg p) \Leftrightarrow \neg(r \to \neg p),$$

demonstrates that the following argument is also valid.

$$\begin{array}{l} (p \leftrightarrow q) \to (r \to \neg p) \\ r \wedge p \\ \hline \neg(p \leftrightarrow q) \end{array}$$

□

5.4 Activities

Activity 5.1 Write an argument in English that has modus ponens form where at least one premise is false, and the conclusion is true.

Does your argument contradict the fact that every modus ponens argument is valid?Write an argument in English that has modus tollens form where at least one premises is false *and* the conclusion is false. Does your argument contradict the fact that every modus tollens argument is valid?Write an argument in English that has syllogistic form where all the premises are true. Is your conclusion true or false?

Activity 5.2 Prove that modus tollens is valid without using a truth table. Instead, use the following facts:

- modus ponens is valid; and
- a conditional is equivalent to its contrapositive.

Activity 5.3 Discuss why an argument being valid is equivalent to its premises logically implying its conclusion.

Activity 5.4 The definition of valid argument is as follows.

> Whenever the premises are all true, the conclusion is true as well.

Create an equivalent definition that is the *contrapositive* of the definition above.

Activity 5.5 Show that the following argument is valid *without* using a truth table. Instead, argue that the argument fulfills the equivalent definition for **valid argument** that you created in Activity 5.4.

$$\begin{array}{l} p \to \neg q \\ r \to (p \wedge q) \\ \hline \neg r \end{array}$$

Part II

Logic in Mathematics

CHAPTER 6

Definitions and proof methods

6.1 Definitions

Definitions are used in mathematics to label objects that have special properties, and to group all such objects together. Be careful with definitions: as stated by mathematicians, they often contain implicit conditions.

Example 6.1.1 A number is called **prime** if its only divisors are 1 and itself.

This definition has some hidden parts: a more complete definition would be as follows.

A number is called **prime** if

(i) it is an integer,

(ii) it is strictly greater than 1, and

(iii) there does not exist any *other* number greater than 1 which divides it.

□

You should view a definition as a *technical test* or *collection of technical tests* that an object must pass before it can be given a specific label.

Worked Example 6.1.2 Demonstrate that, according to the technical definition of **prime**, 17 is prime but 21 is not.

Solution. Let us test 17.

(i) Yes, 17 is an integer.

(ii) Yes, $17 > 1$.

(iii) None of the numbers in the following list is an integer:

$$\frac{17}{2}, \frac{17}{3}, \frac{17}{4}, \ldots, \frac{17}{16}, \frac{17}{18}, \frac{17}{19}, \ldots.$$

So 17 is prime since it passes the technical tests that define the concept of **prime**.

Now let us test 21.

(i) Yes, 21 is an integer.

(ii) Yes, $21 > 1$.

(iii) However, clearly $21/3 = 7$ is an integer, so 3 divides 21.

So 21 is not prime, since it fails *at least one* of the technical tests that define the concept of **prime**. □

Often, the first thing we do in mathematics is to look for ways to make testing our definition easier.

Proposition 6.1.3 *Suppose n is an integer with $n \geq 2$. Then n is prime if and only if n/m is* not *an integer for every integer m with $2 \leq m < \frac{n}{2}$.*

The proof is left to you as Exercise 6.12.1.

See also. Exercise 6.12.2 for a refinement of the statement of Proposition 6.1.3.

Worked Example 6.1.4 Demonstrate that 17 is prime.

Solution ((Sketch)). By the proposition, to check that 17 is prime we now only need to note that none of the numbers in the following shorter list is an integer:

$$\frac{17}{2}, \frac{17}{3}, \frac{17}{4}, \dots \frac{17}{8}.$$

□

6.2 Common mathematical statements

In mathematics, we often want to prove that some statement P **logically implies** some other statement Q; i.e. we want to prove that $P \Rightarrow Q$ or $(\forall x)(P(x) \Rightarrow Q(x))$. Note that the universal form covers the common statement "all A are B", since this can be rephrased "for all x, if x is A then x is B".

Below are some common methods for proving $P \Rightarrow Q$. In the universal case $(\forall x)(P(x) \Rightarrow Q(x))$, the domain of x may be infinite, so we cannot prove $P(x) \Rightarrow Q(x)$ for each specific x, one-by-one. Instead, we treat x as a *fixed but arbitrary* object in the domain, and try to construct an argument proving $P(x) \Rightarrow Q(x)$ which does not depending on knowing the specific object x. So all of the methods below can also be used in the universal case.

Since a conditional $P \to Q$ is true automatically when P is false, it will be a tautology as long as we cannot have the case of P true and Q false at the same time. (See Figure 1.3.2.) Therefore, we (almost always) ***begin a proof by assuming that P is true***, and proceed to ***demonstrate that Q must then also be true***, based on that assumption.

6.3 Direct proof

Recall. The argument

$$A \to C_1, C_1 \to C_2, \ldots, C_{m-1} \to C_m, C_m \to B \therefore A \to B$$

is valid (Extended Law of Syllogism).

Procedure 6.3.1 Direct proof.

- *To prove $P \Rightarrow Q$, start by assuming that P is true. Then, through a sequence of (appropriately justified) intermediate conclusions, arrive at Q as a final conclusion.*

- *To prove $(\forall x)\big(P(x) \Rightarrow Q(x)\big)$, start by assuming that x is an* arbitrary but unspecified *element in the domain such that $P(x)$ is true. The first sentence in your argument should be: "Suppose x is a   such that $P(x)$", where the blank is filled in by the definition of the domain of x. Then, through a sequence of (appropriately justified) intermediate conclusions that do not depend on knowing the specific object x in the domain, arrive at $Q(x)$ as a conclusion.*

Worked Example 6.3.2 *Prove:* If n is even, then n^2 even.

Solution. Let $P(n)$ represent the predicate "n is even" and let $Q(n)$ represent the predicate "n^2 is even", with domain the integers.

Suppose that n is an arbitrary (but unspecified) integer such that n is even. Then there exists an integer m such that $n = 2m$, and so $n^2 = 4m^2 = 2(2m)$ is even. □

Check your understanding. Attempt Exercises 6.12.4–6.12.6.

6.4 Reduction to cases

Fact 6.4.1 *The following logical equivalence holds:*

$$(s_1 \vee s_2 \vee \cdots \vee s_m) \to t \Leftrightarrow (s_1 \to t) \wedge (s_2 \to t) \wedge \cdots \wedge (s_m \to t).$$

Proof idea. This is just an extended version of Example 2.2.6. ■

If

$$C_1 \vee C_2 \vee \cdots \vee C_m$$

is a tautology, then

$$P \Leftrightarrow P \wedge (C_1 \vee \cdots \vee C_m) \Leftrightarrow (P \wedge C_1) \vee \cdots \vee (P \wedge C_m).$$

By substitution and Fact 6.4.1,

$$P \to Q \Leftrightarrow (P \wedge C_1 \to Q) \wedge \cdots \wedge (P \wedge C_m \to Q).$$

A conjunction is only true if each "factor" in the conjunction is true, so the conjunction on the right above can only be a tautology if each conditional $P \wedge C_1 \to Q$ is a tautology. Therefore, when we have a collection of statements $C_1, \ldots, C_m$ so that

$$C_1 \vee C_2 \vee \cdots \vee C_m$$

is a tautology, we can prove $P \to Q$ by instead proving each of $P \wedge C_i \Rightarrow Q$ one at a time. This is also valid for universal statements, since $\forall$ distributes over $\wedge$ (Proposition 4.2.6).

Now, having to prove many slightly more complicated statements $P \wedge C_i \Rightarrow Q$ seems like a lot more work than just proving the single simple statement $P \to Q$ — why would we want to go to all this extra effort?

Idea 6.4.2 *Each case statement C_i provides extra information that can be combined with the assumption that P is true to arrive at the conclusion that Q must also be true.*

Procedure 6.4.3 Reduction to cases.

- *To prove $P \Rightarrow Q$, determine a set of cases $C_1, C_2, \ldots, C_m$ such that $C_1 \vee C_2 \vee \cdots \vee C_m$ is true, then provide a separate proof of each logical implication $P \wedge C_i \Rightarrow Q$.*

- *To prove $(\forall x)\big(P(x) \Rightarrow Q(x)\big)$, determine a set of cases $C_1(x), C_2(x), \ldots, C_m(x)$ such that*

  $$(\forall x)\big(C_1(x) \vee C_2(x) \vee \cdots \vee C_m(x)\big)$$

  is true, then provide a separate proof of each universally quantified logical implication $(\forall x)\big(P(x) \wedge C_i(x) \Rightarrow Q(x)\big)$.

Worked Example 6.4.4 Show $n^2 - n$ is always even.

Solution. Let $P(n)$ represent the predicate "n is an integer" and let $Q(n)$

represent the predicate “$n^2 - n$ is even”, each with domain the integers. Note that $P(n)$ is actually true for each n in the domain, since our original statement makes no extra premise on n besides its domain.

Suppose that n is an integer. Break into cases based on whether n is even or odd; in each case, proceed by direct proof.

Case n even. If n is even, then there exists an integer m such that $n = 2m$. Then,

$$n^2 - n = n(n-1) = 2m(2m-1)$$

is also even.

Case n odd. If n is odd, then $n-1$ is even, so there exists an integer m such that $n-1 = 2m$, or $n = 2m+1$. Then,

$$n^2 - n = n(n-1) = (2m+1)(2m)$$

is even. □

Warning 6.4.5 Make sure your cases cover all possibilities! (Though it is not necessary that your cases by non-overlapping.)

Check your understanding. Attempt Exercise 6.12.7.

6.5 Statements involving disjunction

First, let’s consider a conditional statement with a disjunction on the hypothesis side. To prove a statement of the form $(P_1 \vee P_2) \Rightarrow Q$, we can use Fact 6.4.1 to decompose into two conditionals:

$$(P_1 \vee P_2) \to Q \Leftrightarrow (P_1 \to Q) \wedge (P_2 \to Q).$$

Appealing to the properties of conjunction, as in our discussion of reduction to cases, we see that we can prove $P_1 \Rightarrow Q$ and $P_2 \Rightarrow Q$ by separate proofs.

What about a conditional with a disjunction on the conclusion side? To prove a statement of the form $P \Rightarrow (Q_1 \vee Q_2)$, we can again reduce to cases, but in a sort of tricky way. For any statement, there are only two possibilities — either the statement is true or it is false. (See Basic Tautology 3 in Example 1.4.1. Apply this fact to one of the statements we are trying to prove.

Procedure 6.5.1 Proof of conditional involving disjunction. *To prove a statement of the form $P \Rightarrow (Q_1 \vee Q_2)$, start by assuming that P is true and Q_1 is **false**. Try to show that these assumptions lead to Q_2 being true.*

Idea 6.5.2 *There are only two possibilities for Q_1: either it is true or it is false. If Q_1 is true, then $Q_1 \vee Q_2$ is already true, regardless of the truth values of P and Q_2, so there is nothing to prove in this case. On the other hand, if Q_1 is false, the only way $Q_1 \vee Q_2$ could be true is if Q_2 is true.*

Note 6.5.3

- Also see Exercise 2.5.3.
- Of course, you can swap the roles of Q_1 and Q_2 above: you could also start by assuming that P is true and Q_2 is false, then try to show that this leads to Q_1 being true.
- Another strategy is to attempt a **proof by contradiction** (discussed in Section 6.9 below). By DeMorgan, $\neg(Q_1 \vee Q_2) \Leftrightarrow \neg Q_1 \wedge \neg Q_2$, so for this

strategy, you should start by assuming that P is true and *both* Q_1 and Q_2 are false. Then, try to arrive at a contradiction.

Worked Example 6.5.4 *Prove:* Every odd number is either 1 more or 3 more than a multiple of 4.

Solution. Let $P(n)$ represent the predicate "n is odd", let $Q_1(n)$ represent the predicate "n is 1 more than a multiple of 4", and let $Q_2(n)$ represent the predicate "n is 3 more than a multiple of 4", each with domain the integers.

Start by assuming that n is an odd number that is *not* 1 more than a multiple of 4. We must now try to show that n *is* 3 more than a multiple of 4. We know that n is odd, so there exists a number m such that $n = 2m+1$. However, since n is *not* 1 more than a multiple of 4, $2m$ cannot be a multiple of 4, and so m cannot be a multiple of 2. Therefore, m is also odd, and so there exists another number ℓ such that $m = 2\ell+1$. Then

$$n = 2m+1 = 2(2\ell+1)+1 = 4\ell+3,$$

which says that n is 3 more than a multiple of 4, as desired. □

6.6 Proving the contrapositive

Recall. Modus tollens: $P \to Q \Leftrightarrow \neg Q \to \neg P$.

Procedure 6.6.1 Proof by proving the contrapositive. *To prove $P \Rightarrow Q$, you can instead prove $\neg Q \Rightarrow \neg P$.*

Example 6.6.2 In Worked Example 6.3.2, we proved that the square of an even number is also even. Therefore, this also constitutes a proof of the contrapositive statement: if the square of a number is odd, then that number is also odd. □

Worked Example 6.6.3 Prove that every prime number larger than 2 is odd.

Solution. We want to prove the following universally quantified conditional ("for all p" omitted, domain is positive integers).

conditional
if (p is prime and $p > 2$) then p is odd.

contrapositive
if p is not odd, then not (p is prime and $p > 2$)

DeMorgan substitution
if p is not odd, then (p is not prime or $p \leq 2$)

These are all equivalent.

Let's prove the last statement: as in the procedure for proving conditionals with a disjunction, start by assuming that p is not odd and $p > 2$. We must then show that p is not prime. Since p is not odd, it is divisible by 2. But since $p > 2$, p is divisible by a number *other than* 1 and p itself. Therefore, p is not prime. □

Check your understanding. Attempt Exercise 6.12.8.

6.7 Proof by counterexample

Sometimes we want to prove that $P \not\Rightarrow Q$; i.e. that $P \to Q$ is *not* a tautology.

Recall. The equivalence

$$P \to Q \Leftrightarrow (P \wedge C_1 \to Q) \wedge \cdots \wedge (P \wedge C_m \to Q)$$

holds for any set of cases $C_1, C_2, \ldots, C_m$ such that $C_1 \vee \cdots \vee C_m$ is a tautology. (See Section 6.4.)

So if $P \wedge C_i \to Q$ is *not* a tautology for at least one i, then $P \to Q$ also cannot be a tautology. Again, this also works in the universal case since $\forall$ distributes over $\wedge$ (Proposition 4.2.6).

counterexample
relative to the logical implication $P \Rightarrow Q$, a statement C such that $P \wedge C \to Q$ is false

Worked Example 6.7.1 In Exercise 6.12.8, you are asked to prove the following statement by proving the contrapositive.

If $2^n - 1$ prime, then n is prime.

Prove that the *converse* of this statement is *false*.

Solution. The converse statement is "If n is prime, then $2^n - 1$ is prime." But the case $n = 11$ is a **counterexample**:

$$2^{11} - 1 = 2047 = 23 \cdot 89$$

is not prime even though $n = 11$ is prime. □

Check your understanding. Attempt Exercise 6.12.9.

6.8 Proving biconditionals

We also often want to prove that two statements P, Q are equivalent; i.e. that $P \Leftrightarrow Q$.

Fact 6.8.1 *The equivalence*

$$P \leftrightarrow Q \;\Leftrightarrow\; (P \to Q) \wedge (Q \to P)$$

holds; i.e. a biconditional is equivalent to the conjunction of the corresponding conditional $P \to Q$ and *its converse.*

Proof. You are asked to prove this by truth table in Exercise 2.5.5. ■

Procedure 6.8.2 Proving a biconditional. *To prove $P \Leftrightarrow Q$, prove $P \Rightarrow Q$ and $Q \Rightarrow P$ separately.*

As usual, this also works in the universal case since $\forall$ distributes over $\wedge$ (Proposition 4.2.6).

Worked Example 6.8.3 *Prove:* A number is even if and only if its square is even.

Solution. We want to prove that the following quantified biconditional ("for all n" omitted, domain is nonnegative, whole numbers).

biconditional
n is even if and only if n^2 is even.

conditional and converse
(if n is even then n^2 is even) and (if n^2 is even then n is even)

contrapositive and converse
(if n^2 is odd then n is odd) and (if n^2 is even then n is even)

conditional and inverse
(if n is even then n^2 is even) and (if n is odd then n^2 is odd)

These are all equivalent, so we could prove any one pair.

Original conditional. This is proved as Worked Example 6.3.2.

Converse. If n^2 is even, then there exists an integer m such that $n^2 = 2m$, so that $n = \sqrt{2m}$ ... ? We seem to be stuck.

Inverse. If n is odd, then there exists an integer m such that $n = 2m+1$. Then, $n^2 = 4m^2 + 4m + 1$ is odd. □

Checkpoint 6.8.4 Attempt Exercise 6.12.10.

6.9 Proof by contradiction

Fact 6.9.1 *For any logically false statement e, we have*

$$s \to t \Leftrightarrow (s \wedge \neg t) \to e.$$

Proof. First, $s \to t$ is false precisely when s is true and t is false. On the other hand, $(s \wedge \neg t) \to e$ is false precisely when $s \wedge \neg t$ is true, and $s \wedge \neg t$ is true precisely when $s, \neg t$ are both true, i.e. when s is true and t is false. ■

Procedure 6.9.2 Proof by contradiction.

- *To prove $P \Rightarrow Q$, devise a false statement E such that $(P \wedge \neg Q) \Rightarrow E$.*
- *To prove $(\forall x)\big(P(x) \Rightarrow Q(x)\big)$, devise a predicate $E(x)$ such that $(\forall x)\big(\neg E(x)\big)$ is true (i.e. $E(x)$ is false for* all *x in the domain), but $(\forall x)\big[\big(P(x) \wedge \neg Q(x)\big) \Rightarrow E(x)\big]$.*

Note 6.9.3 Usually E is taken to be some variation of $C \wedge \neg C$, for some statement C. (See the **Law of Contradiction**, recorded as Basic Tautology 4 in Example 1.4.1.)

Worked Example 6.9.4 Prove that $\sqrt{2}$ is irrational.

Solution. We want to prove the quantified conditional with domain the real numbers: for all x, if $x^2 = 2$ and $x > 0$ then x is not rational.

Suppose that x is a real number such that $x^2 = 2$ and $x > 0$. By contradiction, also assume that x *is* rational. We want this extra assumption to lead to a false statement. Now, x rational means $x = a/b$ for some integers a, b. We may assume a, b are both positive, since $x > 0$. We may also assume a, b have no common factors (i.e. fraction a/b is in lowest terms). Then,

$$\begin{aligned} x^2 = 2 &\Rightarrow a^2 = 2b^2, \\ &\Rightarrow a^2 \text{ even}, \\ &\Rightarrow a \text{ even}, \\ &\Rightarrow a = 2m, \text{ some } m, \\ &\Rightarrow 2b^2 = a^2 = 4m^2, \\ &\Rightarrow b^2 = 2m^2, \\ &\Rightarrow b^2 \text{ even}, \\ &\Rightarrow b \text{ even}. \end{aligned}$$

But if both a, b are even, then a and b are both divisible by 2. We have arrived at our contradiction: a, b have no common factor but a, b have a common factor of 2. That is, we have shown the following.

> For all x, if $\big((x^2 = 2$ and $x > 0)$ and x not irrational$\big)$ then (there exist positive integers a, b such that $x = a/b$ and a, b have no common divisor and a, b have a common divisor).

□

Checkpoint 6.9.5 Attempt Exercises 6.12.11–6.12.13.

6.10 Existence and uniqueness

In mathematics we often want to know whether an object with specific desirable properties actually exists. In symbolic language, this is just $(\exists x)A(x)$. Conceptually, this is easy to do: just find an example! (In practice, this can often be quite difficult.)

Worked Example 6.10.1 Prove that 851 is not prime.

Solution. We want to prove the quantified statement

$$(\exists n)\big((n \neq 1) \wedge (n \neq 851) \wedge (n \text{ divides } 851)\big),$$

with domain the positive, whole numbers. Testing each number, one by one, starting at $n = 2$, we find that using $n = 23$ fits the bill. □

Once we have found an example for an existential statement, we also often want to know whether there are more examples, or whether the one we have found is **unique**. Suppose x_0 is our concrete example proving $(\exists x)A(x)$. To show that x_0 is unique, we should prove the universal statement: $(\forall y)\big(A(y) \to (y = x_0)\big)$. This translates as the following.

> For all y, if $A(y)$ is true, then $y = x_0$.

That is, the only way object y can satisfy $A(y)$ is if y is actually our original example x_0.

Procedure 6.10.2 Proving uniqueness. *To prove that $x = x_0$ is the unique instance of an object x such that $A(x)$ is true, assume that y is also an object such that $A(y)$ is true, and prove that $y = x_0$.*

Worked Example 6.10.3 Prove that 2 is the unique positive number that is both prime and even.

Solution. Suppose n is a positive number which is both prime and even. Since n is even, it is divisible by 2. But since n is prime, it is divisble by *only* 1 and itself. Therefore, 2 and "itself" must be the same, i.e. $n = 2$. □

6.11 Activities

Activity 6.1

(a) Write a technical definition for the word **car**.

(b) *Using only your technical definition* (i.e. ignoring your common sense notions of the word *car*), decide whether a transport truck should be called a car. Then do the same for a train.

Note. Do not go back and modify your definition of **car**; test the objects

transport truck and **train** against whatever definition you initially came up with in Task a.

(c) What is the point of this activity?

Activity 6.2 A **square number** is an integer which is equal to the square of some integer. An integer is **square free** if it is not divisible by any square number other than 1.

(a) Is 0 a square number? Is it square free?

(b) Does there exist a negative square number?

(c) Is every negative number square free?

(d) Is every prime number square free?

(e) Is every square free number prime?

(f) Does there exist an integer which is both a square number and square free?

Activity 6.3 The following statement is a basic (and very useful) fact about real numbers.

> **Triangle Inequality**: For every pair of real numbers x and y, $|x+y| \leq |x|+|y|$.

Use the above statement to directly prove the following extended version of the inequality, *without* resorting to considering cases of positive/negative for any of the variables.

> For every triple of real numbers x, y, and z, $|x+y+z| \leq |x|+|y|+|z|$.

Remark. Using the two-number version of the inequality to prove the three-number version is an example of *inductive reasoning*, something that we will soon investigate further.

Activity 6.4 Suppose you are analyzing the rules for a complicated table-top game, and you have come to the following realization.

> Given any trio of distinct wizards where the first is zapping the second, at least one of the following must also occur: the first is zapping the third or the third is zapping the second.

If you were to approach proving this statement using the advice you read on how to handle statements involving disjunction in Procedure 6.5.1, the first sentence of your proof would be

> Assume ______.

and the last sentence of your proof would be

> Therefore ______.

Activity 6.5 What is the difference between **proving the contrapositive** and **proof by contradiction** ?

Activity 6.6

(a) A positive integer that is greater than 1 and *not* prime is called **composite**.

Write a technical definition for the concept of **composite number** with a similar level of detail as in the "more complete" definition of **prime**

number given in Example 6.1.1.

Note. Don't just define it as "not prime." And make sure that the equality $7 = 1 \times 7$ can't be used to justify the statement "7 is composite" by your definition (because prime 7 is most definitely *not* composite).

(b) *Prove by proving the contrapositive:* If $2^n - 1$ is prime, then n is prime.

Hint. You may find the following "difference of powers" factorization formula useful:

$$a^m - b^m = (a-b)(a^{m-1} + a^{m-2}b + a^{m-3}b^2 + \cdots + a^2b^{m-3} + ab^{m-2} + b^{m-1}).$$

Activity 6.7

(a) Write down a technical definition of the term **rational number**.

(b) *Prove directly:* The sum of two rational numbers is a rational number.

(c) *Prove by contradiction:* The sum of a rational number and an irrational number is irrational.

(d) *Disprove by counterexample:* The sum of two irrational numbers is irrational.

Activity 6.8 Refer to Activity 6.2.

(a) Prove that a positive number n is square free if and only if for every factorization $n = ab$, the integers a and b do not share a common factor other than 1.

(b) Prove that a positive number is square free if and only if it is not divisible by the square of a prime number.

Activity 6.9 A pair of prime numbers p_1, p_2 is called a **twin prime pair** if $p_2 = p_1 + 2$. A prime number is called an **isolated prime** if it is not part of a twin prime pair.

(a) Determine the first (i.e. smallest) four twin prime pairs.

(b) Determine the first (i.e. smallest) two isolated primes.

(c) Prove that if $p, p+2$ is a twin prime pair with $p \geq 5$, then $p+1$ is divisible by 6.

(d) Prove that if $p, p+2$ is a twin prime pair, then $p-2, p$ and $p+2, p+4$ *cannot* be twin prime pairs.

6.12 Exercises

1. Let n represent an integer with $n \geq 2$. Prove that n is prime if and only if n/m is *not* an integer for every integer m with $2 \leq m < \frac{n}{2}$.

2. Let n represent an integer with $n \geq 2$. Suppose $p_1, p_2, \ldots, p_\ell$ is a complete list of prime numbers which are less than or equal to $n/2$. Prove that n is prime if and only if none of the p_i divide n. ***Careful:*** Is the statement actually true in the case $n = 2$? $n = 3$? (Why should these cases be given special consideration?)

3. Call two people **twins** if they share the same mother and the same birthdate. Consider the statement: "if two people are twins, then they share the same birthdate."

(a) Is the statement true?

(b) What is the converse of this statement? Is it true?

4. *Prove directly:* The sum of two rational numbers is a rational number.

5. *Prove directly:* If n is even, then n^2 is divisible by 4.

6. Recall that the **triangle inequality** states that $|x+y| \leq |x|+|y|$ for all numbers x and y.

Use the triangle inequality to prove directly: $|x+y+z| \leq |x|+|y|+|z|$ for all numbers x, y, z.

7. *Prove by reduction to cases:* $n^3 - n$ is always divisible by 3.

Hint. Use cases $n = 3m, 3m+1, 3m+2$.

8. *Prove by proving the contrapositive:* if $2^n - 1$ is prime, then n is prime.

Hint. You may find the following "difference of powers" factorization formula useful:

$$x^m - y^m = (x-y)(x^{m-1} + x^{m-2}y + x^{m-3}y^2 + \cdots + x^2y^{m-3} + xy^{m-2} + y^{m-1}).$$

9. Prove by counterexample that the following statement is false.

The sum of any two irrational numbers is irrational.

(See Exercise 6.12.4.)

10. *Prove the biconditional:* n is even if and only if n^2 is divisible by 4.

(See Exercise 6.12.5.)

11. *Prove by contradiction:* If m and n are integers such that $11m + 19n$ is odd, then either m or n (or both) must be odd.

12. *Prove by contradiction:* For $x, y > 0$, $\sqrt{x+y} \neq \sqrt{x} + \sqrt{y}$.

13. *Prove by contradiction:* The sum of a rational number and an irrational number is irrational.

(See Exercise 6.12.4 and Exercise 6.12.9.)

14. Prove that if ℓ, m, and n are integers such that ℓ divides m and ℓ divides n, then ℓ divides mn.

15. Prove that if ℓ, m, and n are integers such that mn divides ℓ, then both m and n divide ℓ.

16. Suppose that m and n are integers, and p is a prime number. Prove that if p does not divide the product mn, then p cannot divide either of m or n.

Working with a definition. Exercises 17–19 concern the following definitions.

A **square number** is an integer which is equal to the square of some integer. An integer is **square free** if it is not divisible by any square number other than 1.

17. For each of the following, provide a proof to justify your answer.

(a) Is 0 a square number? Is it square free?

(b) Does there exist a negative square number?

(c) Is every negative number square free?

(d) Is every prime number square free?

(e) Is every square free number prime?

(f) Does there exist an integer which is both a square number and square free?

18. Prove that a positive number n is square free if and only if for every factorization $n = ab$, the integers a and b do not share a common factor other than 1.

19. Prove that a positive number is square free if and only if it is not divisible by the square of a prime number.

20. A pair of prime numbers p_1, p_2 is called a **twin prime pair** if $p_2 = p_1 + 2$. A prime number is called an **isolated prime** if it is not part of a twin prime pair.

(a) Determine the first (i.e. smallest) four twin prime pairs.

(b) Determine the first (i.e. smallest) two isolated primes.

(c) Prove that if $p, p+2$ is a twin prime pair with $p \geq 5$, then $p+1$ is divisible by 6.

(d) Prove that if $p, p+2$ is a twin prime pair, then $p-2, p$ and $p+2, p+4$ *cannot* be twin prime pairs.

Chapter 7

Proof by mathematical induction

7.1 Principle of Mathematical Induction

Axiom 7.1.1 Principle of Mathematical Induction. *Suppose $P(n)$ is a predicate where the variable n has domain the positive, whole numbers. If*

(i) *$P(1)$ is true, and*

(ii) *$(\forall k)\big(P(k) \to P(k+1)\big)$ is true,*

then $(\forall n)P(n)$ is true.

It is usual to take the principle of mathematical induction as an **axiom**; that is, we *assume* that mathematical induction is valid without proving it.

A look ahead. We will discuss axioms a little more in Chapter 8.

Below is an outline of the idea behind why it is reasonable to assume that mathematical induction is valid. However, this outline does not constitute a proof since it technically uses mathematical induction implicitly.

Idea. Suppose n is fixed. We have a sequence of valid arguments:

$$\begin{array}{l} P(1) \to P(2) \\ P(1) \\ \hline P(2) \end{array} \qquad \begin{array}{l} P(2) \to P(3) \\ P(2) \\ \hline P(3) \end{array} \qquad \cdots \qquad \begin{array}{l} P(n-1) \to P(n) \\ P(n-1) \\ \hline P(n) \end{array}$$

Each is valid (modus ponens). So if we make the two assumptions stated in the principles (i.e. that $P(1)$ is true and that $P(k) \to P(k+1)$ is always true) we can trace the flow of truth from premises to conclusion in each argument in turn:

First argument

Premises true so conclusion is true.

Second argument

Premises true (using conclusion of the first argument) so conclusion is true.

Third argument

...

$(n-1)^{\text{th}}$ argument

Premises true (using conclusion of the $(n-2)^{\text{th}}$ argument) so conclusion is true.

The conclusion of $(n-1)^{\text{th}}$ argument is $P(n)$, so $P(n)$ is true.

Now, here is some specific terminology associated to proofs by induction.

base case the statement $P(1)$ in a proof by mathematical induction

induction step
the portion of a proof by mathematical induction that establishes the statement $(\forall k)\big(P(k) \to P(k+1)\big)$

induction hypothesis
the assumption $P(k)$ made as the first step in the induction step of a proof by mathematical induction

Procedure 7.1.2 Proof by induction. *To prove a universal statement indexed by whole numbers:*

Base case. Start by proving the statement obtained from the universally quantified predicate for the base case $n = 1$.

Induction step. Next, assume that k is a fixed number such that $k \geq 1$, and that the statement obtained from the universally quantified predicate is true for $n = k$. Based on this assumption, try to prove that the next case, $n = k + 1$, is also true.

Worked Example 7.1.3 Prove that the sum of the first n positive integers is

$$\frac{n(n+1)}{2}.$$

Solution. We want to prove $(\forall n)P(n)$, where $P(n)$ is as follows.

$$P(1)\colon\ 1 = \frac{1 \cdot 2}{2}, \qquad P(2)\colon\ 1+2 = \frac{2 \cdot 3}{2}, \qquad P(3)\colon\ 1+2+3 = \frac{3 \cdot 4}{2},$$
$$\ldots, \qquad P(n)\colon\ 1+2+\cdots+n = \frac{n(n+1)}{2}, \qquad \ldots$$

We will prove this by induction.

Base case. $1 = (1 \cdot 2)/2$ is obviously true.

Induction step. Assume the statement is true for $n = k$; i.e. assume that

$$1+2+\cdots+k = k(k+1)/2.$$

We want to show that this implies the statement is true for $n = k+1$; i.e. show

$$1+2+\cdots+k+(k+1) = (k+1)(k+2)/2.$$

We have

$$\begin{aligned} 1+2+\cdots+k+(k+1) &= (1+2+\cdots+k)+(k+1) \\ &= \frac{k(k+1)}{2}+(k+1) \\ &= \frac{k^2+3k+2}{2} \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned}$$

□

Worked Example 7.1.4 Prove that $n^3+(n+1)^3+(n+2)^3$ is always divisible by 9 for every $n \geq 1$.

Solution. We want to prove $(\forall n)P(n)$, where $P(n)$ is as follows.

$$\begin{aligned}
P(1)&: \quad 1^3+2^3+3^3 \text{ is divisible by } 9\\
P(2)&: \quad 2^3+3^3+4^3 \text{ is divisible by } 9\\
P(3)&: \quad 3^3+4^3+5^3 \text{ is divisible by } 9\\
&\vdots\\
P(n)&: \quad n^3+(n+1)^3+(n+2)^3 \text{ is divisible by } 9\\
&\vdots
\end{aligned}$$

We will prove this by induction.

Base case. For $n=1$, $n^3+(n+1)^3+(n+2)^3=36=9\cdot 4$.

Induction step. Assume $k^3+(k+1)^3+(k+2)^3$ is divisible by 9. This means that there exists some whole number m so that

$$k^3+(k+1)^3+(k+2)^3=9m.$$

We want to show that $(k+1)^3+(k+2)^3+(k+3)^3$ is also divisible by 9. To make the connection between this sum of cubes and the "previous case" sum of cubes above, we can add in (and simultaneously subtract out) a k^3 term:

$$\begin{aligned}
(k+1)^3+(k+2)^3+(k+3)^3 &= \left(k^3+(k+1)^3+(k+2)^3\right)+(k+3)^3-k^3\\
&=9m+(k+3)^3-k^3\\
&=9m+(k^3+9k^2+27k+27)-k^3\\
&=9(m+k^2+3k+3).
\end{aligned}$$

Since we have factored our sum of cubes into a product involving 9, that sum of cubes is divisible by 9. □

Worked Example 7.1.5 Prove that $3^{3n}+1$ is divisible by 7 whenever n is odd.

Solution. We do not want to use n as our induction index, since it jumps by twos. But n odd means that $n=2m-1$ for some $m\geq 1$, so we want to prove $(\forall m)P(m)$, where $P(m)$ is as follows.

$$\begin{aligned}
P(1)&: \quad 3^3+1 \text{ is divisible by } 7\\
P(2)&: \quad 3^9+1 \text{ is divisible by } 7\\
P(3)&: \quad 3^{15}+1 \text{ is divisible by } 7\\
&\vdots\\
P(m)&: \quad 3^{6m-3}+1 \text{ is divisible by } 7\\
&\vdots
\end{aligned}$$

We will prove this by induction.

Base case. For $m=1$, $3^3+1=28=7\cdot 4$.

Induction step. Assume $3^{6k-3}+1$ is divisible by 7. This means that there exists some whole number ℓ so that

$$3^{6k-3}+1=7\ell.$$

We want to show $3^{6(k+1)-3}+1$ is divisible by 7. We have

$$\begin{aligned} 3^{6(k+1)-3}+1 &= (3^{6k-3})(3^6)+1 \\ &= (3^{6k-3}+1-1)(3^6)+1 \\ &= (3^{6k-3}+1)(3^6)-3^6+1 \\ &= (7\ell)(3^6)-728 \\ &= 7(3^6\ell-104). \end{aligned}$$

Since we have our expression factored into a product involving 7, our expression is divisible by 7 as desired. □

Remark 7.1.6 Indexing of statements does not have to start at 1.

Worked Example 7.1.7 Prove $2^n < n!$ whenever $n \geq 4$.

Note. This statement is actually false for $n = 1, 2, 3$.

Solution.

Base case. For $n = 4$, $2^4 = 16$ and $4! = 24$.

Induction step. Assume $2^k < k!$ for some $k \geq 4$. We want to show $2^{k+1} < (k+1)!$. We have

$$2^{k+1} = 2(2^k) < 2(k!) < (k+1)(k!) = (k+1)!.$$

□

7.2 An application to logic

Theorem 7.2.1 Validity of the Extended Law of Syllogism. *The Extended Law of Syllogism is a valid argument.*

Proof. By mathematical induction.

Base case $n = 3$. This is just the ordinary Law of Syllogism.

Induction step. Let $k \geq 3$. Consider the $n = k$ version (below left) and the $n = k+1$ version (below right) of the Extended Law of Syllogism.

$$\begin{array}{c} p_1 \to p_2 \\ p_2 \to p_3 \\ \vdots \\ p_{k-1} \to p_k \\ \hline p_1 \to p_k \end{array} \qquad\qquad \begin{array}{c} p_1 \to p_2 \\ p_2 \to p_3 \\ \vdots \\ p_{k-1} \to p_k \\ p_k \to p_{k+1} \\ \hline p_1 \to p_{k+1} \end{array}$$

Assume the $n = k$ version of the argument is valid. We want to show that the $n = k+1$ version is also valid. So suppose that premises of that latter version are all true. We need to show that the conclusion $p_1 \to p_{k+1}$ must then also be true.

But each premise of the $n = k$ version is also a premise of the $n = k+1$ version, so we can say that we have assumed that every premise of the $n = k$ version is true. But we have also assumed that version to be valid, so we may take its conclusion $p_1 \to p_k$ to be true.

Consider the following syllogism.

$$\begin{array}{c} p_1 \to p_k \\ p_k \to p_{k+1} \\ \hline p_1 \to p_{k+1} \end{array}$$

Since this is valid (base case $n = 2$) and its premises are all true, the conclusion is true. ■

7.3 Strong form of Mathematical Induction

Axiom 7.3.1 Principle of Mathematical Induction (Strong Form). *Suppose $P(n)$ is a predicate where the variable n has domain the positive, whole numbers. If*

(i) $P(1)$ *is true, and*

(ii) $(\forall k)\Big(\big(P(1) \wedge P(2) \wedge \cdots \wedge P(k)\big) \to P(k+1)\Big)$ *is true,*

then $(\forall n)P(n)$ *is true.*

Idea. The idea here is the same as for regular mathematical induction. However, in the strong form, we allow ourselves more than just the immediately preceding case to justify the current case.

If the first case $P(1)$ is true, and $P(1) \to P(2)$, then $P(2)$ must be true as well. Now, if $P(1) \wedge P(2) \to P(3)$, and we already have $P(1)$ and $P(2)$ both true, then $P(3)$ must be true as well. Furthermore, if $P(1) \wedge P(2) \wedge P(3) \to P(4)$, and we already have $P(1)$, $P(2)$, and $P(3)$ all true, then $P(4)$ must be true as well. And so on, until we have reached $P(n)$, for n whatever value we wish.

Procedure 7.3.2 Proof by strong induction.

Base case. Start by proving the statement for the base case $n = 1$.

Induction step. Next, assume that k is a fixed number such that $k \geq 1$, and that the statement is true for all $n \leq k$. Based on this assumption, try to prove that the next case, $n = k+1$, is also true.

Worked Example 7.3.3 Prove that each whole number greater than 1 can be factored into a product of (one or more) primes.

Solution.

Base case. The first number greater than 1 is $n = 2$, and it is prime. So it can be considered a product of *one* prime.

Note. Our base case is at $n = 2$ because our original statement only concerns numbers *greater* than 1.

Induction step. Let k represent a whole number that is greater than 1. Assume that $2, 3, 4, \ldots, k$ can each be factored into primes. We want to show $k+1$ can also be factored into primes.

Break into cases.

Case $k+1$ prime. In this case $k+1$ is already a product of a single prime, itself.

Case $k+1$ not prime. If $k+1$ is not prime, then it has nontrivial divisors. So there exist integers m_1, m_2, with $2 \leq m_1, m_2 \leq k$, such that $k+1 = m_1 m_2$. By our induction hypothesis, each of m_1, m_2 can be factored into a product of primes, so their product $k+1$ can as well. □

Warning 7.3.4 If your proof of the induction step requires knowing a very specific number of previous cases are true, you may need to use a variant of the strong form of mathematical induction where *several* base cases are first proved. For example, if, in the induction step, proving that $P(k+1)$ is true relies specifically on knowing that both $P(k-1)$ and $P(k)$ are true, then this argument does *not* prove that $P(1) \to P(2)$, and so you must prove both base cases of $P(1)$ *and* $P(2)$ explicitly.

7.4 Activities

Below is a more detailed version of Procedure 7.1.2. Follow the steps of Procedure 7.4.1 to create a proof by induction for each of the requested proofs in this activity set.

Procedure 7.4.1 Mathematical induction, step-by-step.

(a) Write the statement with n replaced by k.

(b) Write the statement with n replaced by $k+1$.

(c) Identify the connection between the k^{th} statement and the $(k+1)^{\text{th}}$ statement.

(d) Complete the ***induction step*** *by* assuming *that the $n = k$ version of the statement is* true*, and using this assumption to* prove *that the $n = k+1$ version of the statement is* true*.*

(e) Complete the induction proof by proving the ***base case****.*

Activity 7.1 A **binary string** is a “word” in which each “letter” can only be 0 or 1.

Prove that there are 2^n different binary strings of length n.

Activity 7.2 Prove that for every positive integer n, the binomial $1-x^n$ can be factored as $(1-x)(1+x+x^2+\cdots+x^{n-1})$.

Activity 7.3 Prove that the following argument is valid for all positive integers n.

$$\begin{array}{l}
(p_1 \wedge q_1) \to r_1 \\
(p_2 \wedge q_2) \to r_2 \\
\qquad\vdots \\
(p_n \wedge q_n) \to r_n \\
p_1 \wedge p_2 \wedge \cdots \wedge p_n \\
\hline
(q_1 \to r_1) \wedge (q_2 \to r_2) \wedge \cdots \wedge (q_n \to r_n)
\end{array}$$

Careful. Recall that in this context, the words **valid** and **true** do not have the same meaning.

Activity 7.4 Prove that a truth table involving n statement variables requires 2^n rows.

Activity 7.5 Prove that a knight can be moved from any square to any other square on an $n \times n$ chess board by some sequence of allowed moves, for every $n \geq 4$.

CHAPTER 8

Axiomatic systems

8.1 Basics and examples

Any mathematical system must have a starting point; we cannot create something out of nothing. The starting point of a mathematical system (or any logical system, for that matter) is a collection of basic terminology accompanied by a collection of assumed facts about the things the terminology describes.

primitive term
a label for an object or action that is left *undefined*

axiom a statement (usually involving primitive terms or terms defined in terms of primitive terms) that is held to be true *without proof*

axiomatic system
a collection of primitive terms and axioms

Example 8.1.1 An axiomatic system.

Primitive terms.

- **woozle** (noun),
- **dorple** (noun),
- **snarf** (verb).

Axioms.

1. There exist at least three distinct woozles.
2. A woozle snarfs a dorple if and only if the dorple snarfs the woozle.
3. Each pair of distinct woozles snarfs exactly one dorple in common.
4. There is at least one trio of distinct woozles that snarf no dorple in common.
5. Each dorple is snarfed by at least two distinct woozles.

□

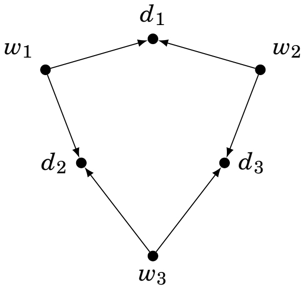
Remark 8.1.2 In the axiomatic system of Axiom 8.1.1, Axiom 1 is redundant as we may infer from Axiom 4 that there exist three distinct woozles. But there is no harm in including this axiom for clarity. As well, we will later investigate the effect of altering it.

The axiomatic system of Example 8.1.1 seems like nonsense, but we can actually prove things from it.

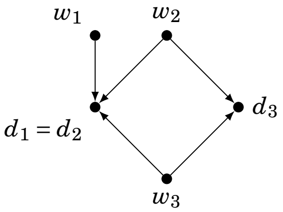
Theorem 8.1.3 *There exist at least three distinct dorples.*

Proof. (In this proof, all references to axioms refer to the axioms of Example 8.1.1.)

By Axiom 4, there exists a trio w_1, w_2, w_3 of distinct woozles that snarf no dorple in common. Breaking this trio into various pairs and applying Axiom 3, we see that there exists a dorple d_1 that w_1 and w_2 both snarf in common, there also exists a dorple d_2 that w_1 and w_3 both snarf in common, and there also exists a dorple d_3 that w_2 and w_3 both snarf in common. These snarfing relationships are illustrated in the diagram below.



Now, suppose d_1 and d_2 were actually the same dorple — then all three woozles would snarf it in common.



As this would contradict our initial assumption, it must be the case that d_1 and d_2 are distinct. Similar arguments allow us to also conclude that $d_1 \neq d_3$ and $d_2 \neq d_3$. ■

It is often useful to give names to important properties of objects.

defined term
a label for an object or action that is defined in terms of primitive terms, axioms, and/or other defined terms

definition an formal explanation of the meaning of a **defined term**

Example 8.1.4 Making a definition. Here is a definition relative to the axiomatic system of Example 8.1.1.

snarf buddies
two distinct dorples that snarf a common woozle

□

A definition allows us to more succinctly communicate ideas and facts about the objects of an axiomatic system.

Theorem 8.1.5 *A pair of* ***snarf buddies*** *snarf a* unique *woozle in common.*

Proof. Suppose d_1, d_2 are snarf buddies. By contradiction, suppose they snarf *more than one* woozle in common: let w_1, w_2 be distinct woozles both snarfed by

d_1 and d_2. By Axiom 2, each of w_1, w_2 snarfs each of d_1, d_2. But this contradicts Axiom 3, as two distinct woozles cannot snarf more than one dorple in common. ■

Suppose we replace Axiom 1 in the system of Example 8.1.1 with the following.

1. There exist *exactly* three distinct woozles.

In the new, modified axiomatic system, our previous two theorems (Theorem 8.1.3 and Theorem 8.1.5) remain true, because it is still true that there exist at least three distinct woozles. But we can now also prove the following.

Theorem 8.1.6 *In the axiomatic system of Example 8.1.1 with the above modified version of Axiom 1, there exist exactly three distinct dorples.*

Proof. You are asked to prove this in the exercises. ■

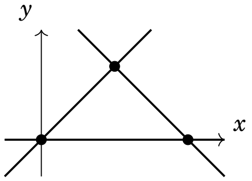
A nonsense system like the one in Example 8.1.1 is just that — nonsense — and not much use unless there are actual examples to which the developed theory can be applied.

model a system obtained by replacing the primitive terms in an axiomatic system with more "concrete" terms in such a way that all the axioms are true statements about the new terms

If we agree that the axiom statements are still all true with the new terms, then any theorems proved under the abstract system are still valid in the new model system.

Example 8.1.7 A model for the woozel-dorple system. Again consider the axiomatic system of Example 8.1.1, still using the modified version of Axiom 1. Let the three distinct woozles be the points $(0,0)$, $(1,1)$, and $(2,0)$ in the Cartesian plane. Let **dorple** now mean **line in the plane**, and let **snarf** now mean **lies on**. Convince yourself that the axioms of the system are all true with this interpretation of the primitive terms.

Theorem 8.1.6 now says that there exist exactly three distinct lines in the plane which fit into our axiomatic system; can you find their equations?



□

Remark 8.1.8 Using nonsense terms like **woozle**, **dorple**, and **snarf** for the primitive terms in an axiomatic system is usually not a good idea, as it takes all intuition out of the process of discovering statements that can be deduced from the axioms. It would have been much better if we had used the words **point** instead of **woozle**, **line** instead of **dorple**, and **lies on** instead of **snarfs** as our primitive terms, to be able to use our intuition about how such objects interact. In such a case, the axioms we choose should be a reflection of our idea of the simplest possible properties about the primitive terms, properties that everyone could reasonably agree are "true" without proof. However, for the theorems deduced from such an axiomatic system to have the widest possible applicability, we should leave the words **point** and **line** as truly **primitive**, undefined terms — that is, **point** and **line** should not be taken to mean **point in the plane** and **line in the plane**, as in the example above, but rather just left as some abstract, intuitive idea of **point** and **line**.

8.2 Incompleteness of axiomatic systems

It turns out that if we want to create an axiomatic system on which to base mathematics, we will always run into problems, and some things will remain out of our reach.

Theorem 8.2.1 Gödel's First Incompleteness Theorem. *In any axiomatic system that is sufficiently complex for it to be possible to prove certain basic facts about the nonnegative whole numbers, it is possible to devise a statement that is true but unprovable.*

May you never attempt to prove a statement that is true but unprovable!

8.3 Exercises

Reasoning in an abstract axiomatic system. Exercises 1–5 concern the axiomatic system described in Example 8.1.1.

1. Rewrite each axiom of the system and each subsequent theorem proved in Section 8.1, replacing the words **woozle** by **point**, **dorple** by **line**, and **snarfs** by **lies on**. Come up with a replacement for the terminology **snarf buddies** that is consistent with these replacement primitive terms. Do the statements make more sense now?

2. Rewrite Theorem 8.1.5 as an "if … then … " statement. Then form the converse of this conditional. Now prove the converse.

3. Prove each of the following statements. In your proofs, you may use as justification any combination of the five axioms in the system, Theorem 8.1.3 and Theorem 8.1.5 already proved in this chapter, and/or any of the statements of this exercise that you have already proved.

- **(a)** There is no dorple who snarfs all woozles.
- **(b)** Each woozle snarfs at least two distinct dorples.
- **(c)** Each dorple belongs to at least two distinct snarf buddy pairs.
- **(d)** There is no woolze who snarfs all dorples.
- **(e)** There is at least one trio of distinct dorples that snarf no woozle in common.
- **(f)** Each woozle belongs to a trio of woozles that snarf no dorple in common.
- **(g)** Every pair of woozles can be increased to a trio of woozles that snarf no dorple in common.

Note. Statement f and Statement g in Exercise 8.3.3 are indeed different statements and require separate proofs (and each of these statements is different from Axiom 4).

4. Rewrite each statement in Exercise 8.3.3 using the replacement primitive terms **point** for **woozle**, **line** for **dorple**, and **lies on** for **snarfs**. Also replace **snarf buddies** by whatever terminology you came up with in Exercise 8.3.1.

5. Now consider the system with the revised version of Axiom 1. Prove that there exist exactly three dorples.

Hint. Start with the first diagram in the proof of Theorem 8.1.3. Now argue by contradiction: what do the axioms say would happen if you added a fourth dorple d_4?

6. Consider the following axiomatic system.

Primitive terms.

- **wizard** (noun),
- **zaps** (verb).

Axioms.

1. There are at least three distinct wizards.
2. If W_1, W_2 are distinct wizards, then W_1 zaps W_2 or W_2 zaps W_1.
3. No wizard zaps itself.
4. If W_1, W_2, W_3 are wizards such that W_1 zaps W_2 and W_2 zaps W_3, then W_1 zaps W_3.

Notes.

- Recall that in mathematics and logic, we always interpret “or” as *inclusive or*: one or the other or possibly both.
- In Axiom 2 and Axiom 4, you should treat W_1, W_2, W_3 as variables or placeholders that can be “substituted into”. These axioms are not stating facts about *specific* wizards; rather, they are stating facts about *all* wizards, and their relationships to each other through zapping. In particular, Axiom 4 could (in principle) be applied to a collection W_1, W_2, W_3 of wizards where W_1 and W_3 are in fact the *same* wizard.

Prove the following statements based on this axiomatic system.

(a) **Principle of Non-Retaliation**. If wizard A zaps wizard B, then B does not zap A.

(b) **Friends and Enemies Theorem**. If A, B, and C are *distinct* wizards such that A zaps B, then A zaps C or C zaps B.

Hint. You may wish to refer back to Activity 6.4.

(c) **Bully Theorem**. Given four distinct wizards, exactly one of the four zaps all of the others.

Hint. First argue there cannot be more than one of the four that zaps the other three. Then show there is at least one. You may need to consider several cases — draw diagrams to help.)

Part III

Set Theory

CHAPTER 9

Sets

9.1 Basics

object any distinct entity

Example 9.1.1 Some objects.

- The number 2.
- The real number line.
- A monkey.
- A basket of tennis balls.

□

set a collection of objects

Example 9.1.2 Some sets. From our list of example objects above, we would intuitively consider

- the number 2 to not be a set;
- the real number line to be a set as it is a collection of points, each representing a different real number;
- a monkey to not be a set; and
- a basket of tennis balls to be a set as it is a collection of tennis balls (though the basket itself is not part of this set, just the container for the objects making up the set).

However, the answers above may depend on your point of view. For example, a monkey could be considered a collection of cells. Even the number 2 is sometimes *defined* to be a set! (See Example 11.4.2.) □

Remark 9.1.3 Formally, we leave **object** and **set** as primitive terms in the axiomatic system of set theory. The reason for leaving these terms undefined is because any attempt to define them would lead us down a never-ending rabbit-hole of definitions: what is an “entity”? what is a “collection”?

We will not discuss any axiomatic basis for set theory, but instead rely on

naive set theory.

naive set theory
whatever axioms for set theory the experts decide upon, we are safe (usually, see Warning 9.7.7) to assume that all the mathematical objects that we would like to be **sets**, will be

We need one more primitive term to make set theory workable.

membership
a property of sets relative to other objects: given object x and set S, exactly *one* of the statements "x is a member of S" and "x is not a member of S" is true

element an object that is a member of a set

$x \in S$ object x is an element of set S

9.2 Defining sets

Remember that mathematical notation is about *communicating mathematical information*. Since a set is defined by its member objects, to communicate the details of a set of objects one needs to provide a means to decide whether any given object is or is not an element of the set.

9.2.1 Listing elements

One way to communicate the details of a set definition is to *explicitly* list or describe *all* elements of the set. Such a list should be enclosed in **braces** to indicate that the objects in the list are being collected into a set.

Example 9.2.1 Listing the elements of a set. If we write

$$A = \{\text{monkey, tennis ball, the number 2}\},$$

then we intend for the letter A to become a label representing the set consisting of some specific monkey, some specific tennis ball, and the number 2. □

Here are some sets containing familiar collections of numbers. Notice how in the first two examples we "list" the elements by providing a pattern and then using ... to imply that the pattern continues as expected, and in the second two examples we merely describe what the elements are in words.

$\mathbb{N}$ the set $\{0,1,2,\ldots\}$ of **natural numbers**

$\mathbb{Z}$ the set $\{\ldots,-2,-1,0,1,2,\ldots\}$ of **integers**

$\mathbb{Q}$ the set of all fractions, call the set of **rational numbers**

$\mathbb{R}$ the set of all decimal numbers, called the set of **real numbers**

Note 9.2.2 Keep the following in mind for a set defined by listing elements.

- *Order does not matter*. For example, $\{a,b\}$ and $\{b,a\}$ are the same set because they consist of precisely the same member elements.

- *Repetition does not matter*. For example, $\{a,a,b\}$ and $\{a,b\}$ are the same set because they consist of precisely the same member elements.

9.2.2 Candidate-condition notation

Another way to define a set is **candidate-condition notation**:

$$\text{set} = \{\,\text{candidate domain} \mid \text{condition(s) on candidates}\,\}.$$

This notation provides a means to decide whether an object is a member of the set by first using an already-defined set as a pool of "member candidates" as well a condition or a list of conditions each candidate must satisfy in order to actually be a member.

If we write S for the set being defined, C for the set of candidates, and T for the test those candidates must satisfy to be included in S (that is, T is a predicate with domain C), then the candidate-condition notation takes the form

$$S = \{\, x \in C \mid T(x) \,\},$$

and can be read as

> S is the set of those elements x in C for which $T(x)$ is true.

Example 9.2.3 Using candidate-condition notation to define a set. Consider the set

$$A = \{0, 3, 6, 9, 12, \ldots\}.$$

We could define this set in a more precise manner (i.e. without resorting to using dots) as follows.

$$A = \{\, n \in \mathbb{N} \mid n \text{ divisible by } 3 \,\}$$

The "$n \in \mathbb{N}$" part to the left of the divider tells us that the pool of "member candidates" for A is the set of natural numbers, and the test to the right of the divider tells us how to decide when a given candidate natural number n is actually a member of A. In words, you should think of the above definition as saying the following.

> Set A consists of those elements of $\mathbb{N}$ which are divisible by 3.

□

9.2.3 Form-parameter notation

Finally, sets can be defined by **form-parameter notation**:

$$\text{set} = \{\,\text{form involving parameter} \mid \text{parameter domain}\,\}.$$

This notation describes the members of a set by providing a "form" to which the members must conform. Usually the "form" is based on parameter variables that can range over a set of possibilities.

Example 9.2.4 Using form-parameter notation to define a set. Again consider the set

$$A = \{0, 3, 6, 9, 12, \ldots\}.$$

We could also define this set as

$$A = \{\, 3n \mid n \in \mathbb{N} \,\}.$$

Here, the **form** of the elements of A is given to the left of the divider as "3 times a number", where the number is represented by the **parameter** n. Then the allowed range of the number parameter n is given to the right of the divider. In words, you should think of the above definition as saying the following.

The elements of set A are precisely those objects that are 3 times a natural number.

□

Example 9.2.5 Defining the set of fractions. We could define the set $\mathbb{Q}$ of **rational numbers** in this way:

$$\mathbb{Q} = \left\{ \frac{m}{n} \;\middle|\; m,n \in \mathbb{Z},\, n \neq 0 \right\}.$$

This says that the set $\mathbb{Q}$ consists of all symbols of the form "number over number", where the numbers can be any integers, as long as the bottom number is not zero. However, we need to be a little bit careful here, since we allow different symbols of this form to represent the same element. For example,

$$\frac{3}{6} = \frac{1}{2}, \qquad \frac{2}{-9} = \frac{-2}{9}, \qquad \frac{0}{n} = \frac{0}{1} \quad (\text{any } n \neq 0).$$

We really should make this element form duplication explicit in the definition of the set, but to do this would be really cumbersome and would be expressing something that is learned in grade school, so it is usually omitted. □

9.2.4 Empty set

There is one special set, the elements of which are very easy to list.

empty set the set which has *no* elements

$\varnothing$ the empty set

Remark 9.2.6 The empty set is defined by requiring that the statement "x is an element of $\varnothing$" is *always* false, for every object x.

Warning 9.2.7 Be careful not to inadvertently try to prove some property of members of the empty set! You will be proving a vacuously true statement. (See Section 4.3.)

9.3 Subsets and equality of sets

Often we want to distinguish a collection of certain "special" elements within a larger set of elements.

subset a set whose elements are all members of another set

$A \subseteq B$ set A is a **subset of** or is **contained in** set B

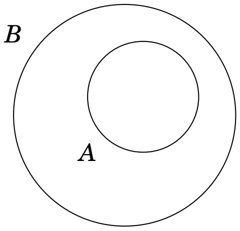


Figure 9.3.1 A Venn diagram demonstrating a subset relationship.

Warning 9.3.2 We also sometimes use the phrase "contained in" to mean an object is an element of a set.

Test 9.3.3 Subset. *To demonstrate $A \subseteq B$, prove $(\forall x)(x \in A \Rightarrow x \in B)$.*

Example 9.3.4 Basic examples involving familiar sets of numbers.

- Every natural number is an integer, so $\mathbb{N} \subseteq \mathbb{Z}$. To emphasize this, we could write $\mathbb{N} = \{\, m \in \mathbb{Z} \mid m \geq 0 \,\}$.
- Every integer can be considered to be a rational number, since for every $m \in \mathbb{Z}$ we can write $m = \frac{a}{b}$ with $a = m$ and $b = 1$. Thus $\mathbb{Z} \subseteq \mathbb{Q}$.
- Every rational number can be considered to be a real number if we identify fractions with their decimal expansions via long division. Thus $\mathbb{Q} \subseteq \mathbb{R}$.

□

Example 9.3.5 Candidate-condition notation always defines subsets. When we define a set by **Candidate-condition notation**, we first specify a pool of candidate elements, and then a condition or collection of conditions that those candidates must satisfy in order to actually be included in the set. But then every element in the set we are defining must first be from the set of candidate elements, so our defined set must be a subset of the candidate set.

For example, in Example 9.3.4, we provided a definition for the set $\mathbb{N}$ in candidate-condition form where the pool of candidates is the set $\mathbb{Z}$. This definition makes it explicit that $\mathbb{N} \subseteq \mathbb{Z}$. □

Worked Example 9.3.6 Prove that $A \subseteq B$ for

$$A = \{\, 3m + 1 \mid m \in \mathbb{Z} \,\}, \qquad B = \left\{ x \in \mathbb{R} \,\middle|\, \sin\left(\frac{\pi(x-1)}{3}\right) = 0 \right\}.$$

Solution. There are an infinite number of elements of A, so we cannot check that all elements of A are also elements of B one-by-one. Instead, we let a variable x represent an *arbitrary* but unspecified element of A. Since all elements of A have the form $3m + 1$ for some $m \in \mathbb{Z}$, we have $x = 3m + 1$ for some $m \in \mathbb{Z}$. Check the condition for being an element of B by calculating

$$\sin\left(\frac{\pi(x-1)}{3}\right) = \sin\left(\frac{\pi\big((3m+1)-1\big)}{3}\right) = \sin(m\pi) = 0.$$

Therefore, $x \in B$. Since the above calculation works for every $m \in \mathbb{Z}$, all elements of A are elements of B. □

Proposition 9.3.7 Basic properties of the subset relationship.

1. *Every set has an empty set as a subset. That is, $\varnothing \subseteq S$ is always true for a set S.*
2. *Every set is a subset of itself. That is, $S \subseteq S$ is always true for a set S.*
3. *The subset relation is **transitive**. That is, whenever $A \subseteq B$ and $B \subseteq C$ are true, then $A \subseteq C$ is true as well.*

A look ahead. We will study abstract notions of **relation** and the **transitive property** in Chapter 17.

set equality

write $A = B$ if both sets consist of precisely the same elements

Test 9.3.8 For set equality. *To demonstrate $A = B$, check that both $A \subseteq B$ and $B \subseteq A$.*

This requires two applications of the Subset Test:

(i) begin with the assumption $x \in A$ and proceed to the conclusion $x \in B$;

(ii) begin with the assumption $x \in B$ and proceed to the conclusion $x \in A$.

Remark 9.3.9 One could combine both applications of the Subset Test described in the Test for Set Equality above into one biconditional: $A = B$ is true if

$$(\forall x)(x \in A \Leftrightarrow x \in B)$$

is true. If the logic of $x \in A \Rightarrow x \in B$ is easily reversed, then it makes sense to argue $x \in A \Leftrightarrow x \in B$ instead of separately arguing $x \in A \Rightarrow x \in B$ and $x \in B \Rightarrow x \in A$. However, in most cases separate arguments of these logical implications is preferred.

Worked Example 9.3.10 Let A and B be as in Worked Example 9.3.6. Prove that $A = B$.

Solution.

Show $A \subseteq B$. See Worked Example 9.3.6.

Show $B \subseteq A$. Let x represent an arbitrary element of B. This means that

$$\sin\left(\frac{\pi(x-1)}{3}\right) = 0.$$

However, we know from trigonometry that $\sin\theta = 0$ if and only if θ is an integer multiple of π; i.e. $\theta = m\pi$ for some $m \in \mathbb{Z}$. If we set

$$\frac{\pi(x-1)}{3} = m\pi$$

and solve for x, we get $x = 3m + 1 \in A$. □

proper subset

a set contained in *but not equal to* another set

$A \subsetneqq B$	set A is a proper subset of set B

Note 9.3.11 Some people exclude $\varnothing$ from the definition of proper subset.

Test 9.3.12 For a proper subset. *To demonstrate $A \subsetneqq B$, first test $A \subseteq B$ as usual (Test 9.3.3), but also demonstrate that there exists some $x \in B$ such that $x \notin A$.*

Example 9.3.13 Proper subsets of number sets. We already know that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$, but we have

1. $\mathbb{N} \subsetneqq \mathbb{Z}$, since, for example, $-1 \in \mathbb{Z}$ but $-1 \notin \mathbb{N}$;

2. $\mathbb{Z} \subsetneqq \mathbb{Q}$, since, for example, $\frac{1}{2} \in \mathbb{Q}$ but $\frac{1}{2} \notin \mathbb{Z}$; and

3. $\mathbb{Q} \subsetneqq \mathbb{R}$, since, for example, $\sqrt{2} \in \mathbb{R}$ but $\sqrt{2} \notin \mathbb{Q}$.

□

Remark 9.3.14 To show the $A \neq B$ part of $A \subsetneqq B$, you only need to exhibit *one* example element of B which is not in A; i.e. you need to find a **counterexample** for the logical implication $x \in B \Rightarrow x \in A$.

9.4 Complement, union, and intersection

First, it is often convenient to restrict the scope of the discussion.

universal set
a set that contains all objects currently under consideration

We will consider all of the following set operations to be performed within a universal set U. In particular, suppose $A, B \subseteq U$.

9.4.1 Universal and relative complement

complement
the set of elements of U which are *not* in A

A^{c} the complement of A (in U), so that

$$A^{\mathrm{c}} = \{\, x \in U \mid x \notin A \,\}$$

relative complement
if $A, B \subseteq U$, the **complement of** A **in** B is the set of elements of B which are not in A

$B \smallsetminus A$ the complement of A in B, so that

$$B \smallsetminus A = \{\, x \in B \mid x \notin A \,\}$$

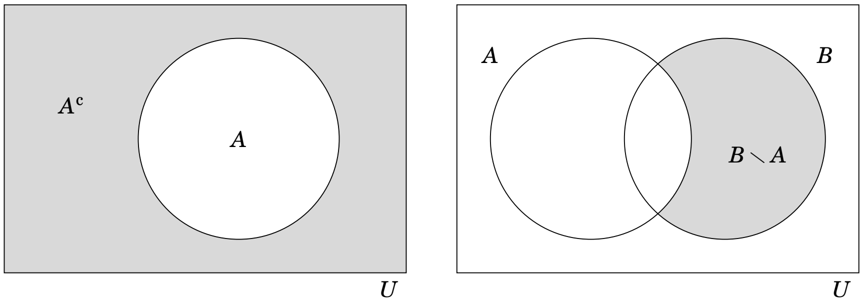


Figure 9.4.1 Venn diagrams of universal and relative set complements.

Note 9.4.2 Another common notation for relative complement is $B - A$. However, this conflicts with the notation for the algebraic operation of subtraction in certain contexts, so we will prefer the notation $B \smallsetminus A$.

Example 9.4.3 Some examples of relative complement involving number sets.

- Suppose $B = \{1,2,3,4,5,6\}$ and $A = \{1,3,5\}$. Then $B \smallsetminus A = \{2,4,6\}$.
- The complement of the set of rational numbers $\mathbb{Q}$ inside the set of real numbers $\mathbb{R}$ is called the **set of irrational numbers**, and we write $\mathbb{I} = \mathbb{R} \smallsetminus \mathbb{Q}$ for this set. If you are thinking of real numbers in terms of their decimal expanions, the irrational numbers are precisely those that have nonterminating, nonrepeating decimal expansions.

□

9.4.2 Union, intersection, and disjoint sets

union the combined collection of all elements in a pair of sets

$A \cup B$ the union of sets A and B, so that

$$A \cup B = \{\, x \in U \mid x \in A \text{ or } x \in B \text{ (or both)} \,\}$$

intersection

the collection of only those elements common to a pair of sets

$A \cap B$ the intersection of A and B, so that

$$A \cap B = \{\, x \in U \mid x \in A \text{ and } x \in B \,\}$$

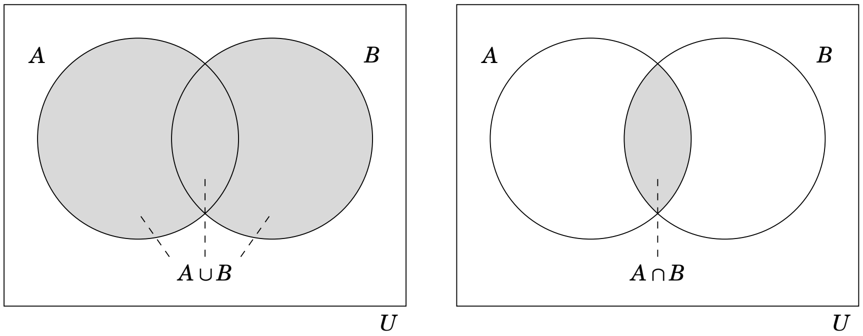


Figure 9.4.4 Venn diagrams of set union and intersection.

Note 9.4.5 A union contains every element from both sets, so it contains both sets as subsets:

$$A, B \subseteq A \cup B.$$

On the other hand, every element in an intersection is in both sets, so the intersection is a subset of both sets:

$$A \cap B \subseteq A, B.$$

Example 9.4.6 For subsets $A = \{1,2,3,4\}$ and $B = \{3,4,5,6\}$ of $\mathbb{N}$, we have

$$A \cup B = \{1,2,3,4,5,6\}, \qquad A \cap B = \{3,4\}.$$

□

Example 9.4.7 Consider the following subsets of $\mathbb{N}$.

$$\begin{array}{ll} \mathcal{E} = \{\, n \in \mathbb{N} \mid n \text{ even} \,\} & \mathcal{P} = \{\, n \in \mathbb{N} \mid n \text{ prime, } n \neq 0 \,\} \\ \mathcal{O} = \{\, n \in \mathbb{N} \mid n \text{ odd} \,\} & \mathcal{T} = \{\, 3n \mid n \in \mathbb{N} \,\} = \{0, 3, 6, 9, \ldots\} \end{array}$$

Then,

$$\begin{array}{lll} \mathcal{E} \cup \mathcal{O} = \mathbb{N}, & \mathcal{E} \cap \mathcal{P} = \{2\}, & \mathcal{E} \cap \mathcal{T} = \{\, 6n \mid n \in \mathbb{N} \,\}, \\ \mathcal{E} \cap \mathcal{O} = \varnothing, & \mathcal{O} \cap \mathcal{P} = \mathcal{P} \setminus \{2\}, & \mathcal{O} \cap \mathcal{T} = \{\, 6n + 3 \mid n \in \mathbb{N} \,\}. \end{array}$$

□

disjoint sets

sets that have no elements in common, i.e. sets A, B such that $A \cap B = \varnothing$

disjoint union
a union $A \cup B$ where A and B are disjoint

$A \sqcup B$ the disjoint union of sets A and B

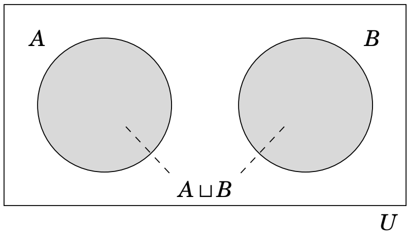


Figure 9.4.8 A Venn diagram of a disjoint set union.

Example 9.4.9 Sets $\mathcal{E}, \mathcal{O}$ from Example 9.4.7 are disjoint, and $\mathbb{N} = \mathcal{E} \sqcup \mathcal{O}$. □

Remark 9.4.10 If $A \subseteq U$, then we can express U as a disjoint union $U = A \sqcup A^{c}$. Similarly, if $U = A \sqcup B$, then we must have $B = A^{c}$.

9.4.3 Rules for set operations

Proposition 9.4.11 Rules for Operations on Sets. *Suppose A, B, C are subsets of a universal set U. Then the following set equalities hold.*

1. *Properties of the universal set.*

   (a) $A \cup U = U$ *(b)* $A \cap U = A$

2. *Properties of the empty set.*

   (a) $A \cup \varnothing = A$ *(b)* $A \cap \varnothing = \varnothing$

3. *Duality of universal and empty sets.*

   (a) $U^{c} = \varnothing$ *(b)* $\varnothing^{c} = U$

4. *Double complement.*
   $(A^{c})^{c} = A$

5. *Idempotence.*

   (a) $A \cup A = A$ *(b)* $A \cap A = A$

6. *Commutativity.*

   (a) $A \cup B = B \cup A$ *(b)* $A \cap B = B \cap A$

7. *Associativity.*

   (a) $(A \cup B) \cup C = A \cup (B \cup C)$ *(b)* $(A \cap B) \cap C = A \cap (B \cap C)$

8. *Distributivity.*

   (a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(c) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

(d) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

9. *DeMorgan's Laws.*

(a) $(A \cup B)^c = A^c \cap B^c$ (b) $(A \cap B)^c = A^c \cup B^c$

Proof of Rule 9.a. Recall that to prove this set equality, we need to show both

$$(A \cup B)^c \subseteq A^c \cap B^c, \qquad A^c \cap B^c \subseteq (A \cup B)^c.$$

Show $(A \cup B)^c \subseteq A^c \cap B^c$. We need to show

$$x \in (A \cup B)^c \Rightarrow x \in A^c \cap B^c.$$

If $x \in (A \cup B)^c$ then by definition of complement, $x \in U$ but $x \notin A \cup B$. Then $x \notin A$ must be true, since if x were in A then it would also be in $A \cup B$. Similarly, $x \notin B$ must also be true. So $x \in A^c$ and $x \in B^c$; i.e. $x \in A^c \cap B^c$.

Show $A^c \cap B^c \subseteq (A \cup B)^c$. We need to show

$$x \in A^c \cap B^c \Rightarrow x \in (A \cup B)^c.$$

If $x \in A^c \cap B^c$ then by definition of intersection, both $x \in A^c$ and $x \in B^c$ are true.; i.e. $x \notin A$ and $x \notin B$. Since $A \cup B$ is all elements of U which are in one (or both) of A, B, we must have $x \notin A \cup B$. Thus $x \in (A \cup B)^c$. ■

Proofs of the other rules. These are left to you, the reader, in the Exercise 9.9.1. ■

Remark 9.4.12 Compare the set operation rules of the proposition above with the Rules of Propositional Calculus.

9.5 Cartesian Product

9.5.1 Definition and examples

Cartesian product
the set of all possible ordered pairs of elements from two given sets A and B, where the first element in a pair is from A and the second is from B

$A \times B$ the Cartesian product of A and B: $A \times B = \{\, (a,b) \mid a \in A,\ b \in B \,\}$

For "small" sets, we can list the elements of the Cartesian product by listing all ways of combining an element from the first with an element from the second.

Example 9.5.1 A Cartesian product of "small" sets. Suppose $A = \{1,2\}$ and $B = \{a,b,c\}$. Then

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

□

Example 9.5.2 A special subset of a certain Cartesian product. Let $\mathbb{N}^+$ represent the positive natural numbers: $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. Then we can describe the

Cartesian product $\mathbb{Z} \times \mathbb{N}^+$ as

$$\mathbb{Z} \times \mathbb{N}^+ = \{\, (m,n) \mid m,n \in \mathbb{Z},\ n > 0 \,\} \subseteq \mathbb{Z} \times \mathbb{Z}.$$

Consider the subset

$$A = \left\{\, (m,n) \in \mathbb{Z} \times \mathbb{N}^+ \;\middle|\; n \text{ has no divisors in common with } |m| \,\right\} \subseteq \mathbb{Z} \times \mathbb{N}^+.$$

Does A resemble some more familiar set . . . ? □

Extend. Define $A \times B \times C = \{\, (a,b,c) \mid a \in A,\ b \in B,\ c \in C \,\}$.

Example 9.5.3 Suppose $A = \{1,2\}$, $B = \{a,b,c\}$, $C = \{\alpha,\beta\}$. Then,

$$\begin{aligned} A \times B \times C = \{\; &(1,a,\alpha), (1,a,\beta), (1,b,\alpha), (1,b,\beta), (1,c,\alpha), (1,c,\beta), \\ &(2,a,\alpha), (2,a,\beta), (2,b,\alpha), (2,b,\beta), (2,c,\alpha), (2,c,\beta) \;\} \end{aligned}$$

□

Remark 9.5.4 Technically, there is a difference between the elements of each of the sets

$$\begin{aligned} (A \times B) \times C &= \left\{\, \big((a,b),c\big) \;\middle|\; a \in A,\ b \in B,\ c \in C \,\right\}, \\ A \times (B \times C) &= \left\{\, \big(a,(b,c)\big) \;\middle|\; a \in A,\ b \in B,\ c \in C \,\right\}, \\ A \times B \times C &= \{\, (a,b,c) \mid a \in A,\ b \in B,\ c \in C \,\}, \end{aligned}$$

but it is rare that anyone actually observes this technicality. Usually, we consider these three sets to be the same set.

We use special notation for Cartesian products of a set with itself.

A^2	notation to mean $A \times A$
A^3	notation to mean $A \times A \times A$
A^n	notation to mean $A \times A \times \cdots \times A$ involving n “factors” of A

And so on.

Example 9.5.5 Cartesian products in linear algebra. You have probably already encountered the notation

$$\begin{aligned} \mathbb{R}^2 &= \{\, (x,y) \mid x,y \in \mathbb{R} \,\}, \\ \mathbb{R}^3 &= \{\, (x,y,z) \mid x,y,z \in \mathbb{R} \,\}, \\ &\ \vdots \\ \mathbb{R}^n &= \left\{\, (x_1,x_2,\ldots,x_n) \;\middle|\; x_j \in \mathbb{R} \,\right\}, \\ &\ \vdots \end{aligned}$$

used to represent 2-, 3-, and higher-dimensional (real) vector spaces. □

9.5.2 Visualizing Cartesian products

Cartesian products do not really lend themselves to visualization with Venn diagrams. So how should we visualize them?

The example we are probably most familiar with is that of the **Cartesian plane**, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, where each element (x,y) is visualized as a point in a two-dimensional diagram, plotted according to the x- and y-coordinates of the element relative to a central set of xy-axes.

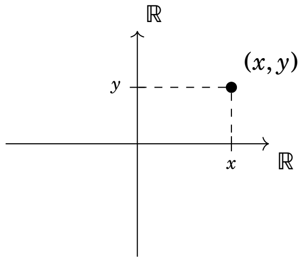


Figure 9.5.6 Visualizing elements of $\mathbb{R} \times \mathbb{R}$ as points in a plane relative to a set of perpendicular axes.

To produce this visualization, we imagine the elements of the "first" $\mathbb{R}$ arrayed along the x-axis, and the elements of the "second" $\mathbb{R}$ as arrayed along the y-axis, and then imagine each element of the Cartesian product $\mathbb{R} \times \mathbb{R}$ as a point located at a position that "lines up" with the corresponding positions on the x- and y-axes of the element's coordinates. This is possible because the concepts of "less than" and "greater than" allow us to think of elements of $\mathbb{R}$ as "progressing" from left to right along the x-axis and from bottom to top along the y-axis.

Even though other types of sets may not have readily available notions of "less than" and "greater than", we may still visualize a Cartesian product $A \times B$ as "points" in a "plane" plotted relative to a central set of AB-axes. In the particular case that A and B are finite, it doesn't really matter in what "order" we place the elements of A along the A-axis or the elements of B along the B-axis.

A look ahead. We will study abstract notions of "less than" and "greater than" in Chapter 19.

Example 9.5.7 Visualizing a Cartesian product between finite sets. Consider the sets $A = \{a, \alpha, \phi, z\}$ and $B = \{1,2,3\}$. Instead of listing the elements of $A \times B$, we will place them in a grid. In Figure 9.5.8, you might imagine that the elements of A listed along the bottom row make up a horizontal axis and the elements of B along the leftmost column make up a vertical axis. But take care that these "axes" are not a continuum of values — there are no other "values" between these "coordinate values" along the "axes". For example, there is no 1.5 along the "horizontal axis." Similarly, there are no other "points" between the elements of $A \times B$.

B					
3	$(a,3)$	$(\alpha,3)$	$(\phi,3)$	$(z,3)$	
2	$(a,2)$	$(\alpha,2)$	$(\phi,2)$	$(z,2)$	
1	$(a,1)$	$(\alpha,1)$	$(\phi,1)$	$(z,1)$	
	a	α	ϕ	z	A

Figure 9.5.8 Visualizing elements of $A \times B$ as "points" in a grid relative to a set of perpendicular "axes."

□

9.6 Alphabets and words

alphabet	any set can be considered an **alphabet**
letters	the elements of an alphabet set
word	a finite-length, ordered list of letters
Σ^*	the set of words using **alphabet** set Σ

Remark 9.6.1 Even if the alphabet set Σ is the usual English-language alphabet, we do not restrict ourselves to *actual* English-language words — nonsense words are allowed.

Example 9.6.2 English is not a full set of words. Using $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$, words

$$\text{math, qwerty, aabbccddijzuuu}$$

are examples of elements in Σ^*. So, ignoring punctuation, hyphenation, and capitalization, the English language is a *proper* subset of Σ^*. □

Example 9.6.3 If digits are letters then numbers are words. Using alphabet $\Sigma = \{0, 1, 2, \ldots, 9\}$, then $\mathbb{N} \subsetneq \Sigma^*$. □

Checkpoint 9.6.4 Why is $\mathbb{N} \neq \Sigma^*$ in Example 9.6.3?

In computing science, a certain set of words is of particular importance.

binary word
a word using alphabet $\{0, 1\}$

binary string
synonym for **binary word**

Warning 9.6.5 Order matters! For example, using the alphabet

$$\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\},$$

the words ab and ba are *different* words in Σ^*.

length (of a word)
given $w \in \Sigma^*$, the **length of** w is the number of elements from Σ used to form w, counting repetition

$|w|$ length of the word $w \in \Sigma^*$

Example 9.6.6 Using alphabet $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$, we have

$$|\mathrm{qwerty}| = 6, \qquad |\mathrm{aabab}| = 5.$$

□

The concept of **length** allows us to identify some special subsets and a special element of Σ^*.

Σ_n^* for $n \in \mathbb{N}$, the subset of Σ^* consisting of all words of length n

empty word
given an alphabet Σ, we always consider Σ^* to contain a unique word of length 0

$\emptyset$ the empty word

9.7 Sets of sets

Note 9.7.1 Sets can be made up of *any* kind of objects, even other sets! (But now we must be careful of the use of the phrase “contained in”.)

Example 9.7.2 Consider

$$\mathcal{T} = \{\, 3n \mid n \in \mathbb{N} \,\}, \qquad X = \{\, A \subset \mathbb{N} \mid A \cap \mathcal{T} = \varnothing \,\}, \qquad Y = X \cup \mathcal{T}.$$

Elements of $\mathcal{T}$ are numbers. Elements of X are *subsets* of $\mathbb{N}$ — that is, X is a set of subsets of $\mathbb{N}$, but *is not* itself a subset of $\mathbb{N}$. Elements of Y are either from X or from $\mathcal{T}$, so some elements of Y are numbers, and some elements of Y are sets of numbers. □

power set given a set A, the **power set of** A is the set $\{B \subseteq A\}$ of all subsets of A

$\mathscr{P}(A)$ the power set of the set A

Warning 9.7.3 The **elements** of a power set are **subsets** of the set in question.

Fact 9.7.4 A power set is never empty. *For every set A,* $\mathscr{P}(A) \neq \varnothing$.

Proof. Both $\varnothing$ and A are *subsets* of A, so both are *elements* of $\mathscr{P}(A)$. Even if $A = \varnothing$, we still have

$$\mathscr{P}(\varnothing) = \{\varnothing\} \neq \varnothing.$$

■

Example 9.7.5 Power set of a “small” set. For $A = \{a, b, c\}$, we have

$$\mathscr{P}(A) = \{\, \varnothing,\ \{a\},\ \{b\},\ \{c\},\ \{a,b\},\ \{a,c\},\ \{b,c\},\ \{a,b,c\} \,\}.$$

Note the use of curly braces here. In particular, note that $\varnothing$ has *not* been placed in its own set of curly braces because it is already a set itself. □

Example 9.7.6 A set of sets as a subset of a power set. For X as in Example 9.7.2, we have $X \subseteq \mathscr{P}(\mathbb{N})$. □

Warning 9.7.7 We are not completely free to define sets any way we want.

Example 9.7.8 Let

$$R = \{\, \text{any set } X \mid X \text{ is not an element of itself.} \,\}.$$

First note that there exist sets which satisfy the condition for membership in R; for example, the empty set. So R should not be not empty. If R is a set, then it is a “candidate” for membership in itself! Break into cases.

Case $R \in R$. Then $R \notin R$, which contradicts the case assumption.

Case $R \notin R$. Then $R \in R$, which contradicts the case assumption. Since all cases lead to a contradiction, R is *cannot be a set*! This is called **Russell’s Paradox**, and is one of the reasons we rely upon “naive set theory” in this course. □

Remark 9.7.9 One of the ways to avoid Russell’s Paradox is by requiring every object, including sets, to have a **type**, similar to how variables in a computer language can be declared to have a type. In such a scheme, a **set** is never just a set — it is always a **set of a certain kind of object**. Then an operation such as $\mathbb{N} \cup \mathscr{P}(\mathbb{N})$ would not be allowed, as $\mathbb{N}$ is a **set of numbers** while $\mathscr{P}(\mathbb{N})$ is a **set of sets of numbers**, and we have a type mismatch. And, more importantly, asking a question like “Is $R \in R$?” becomes nonsensical, as on the left of the $\in$ symbol R

is required to be some type of object while on the right R is required to be a *set* of that type of object, and again we have a type mismatch.

9.8 Activities

Activity 9.1 For each member of your group, consider the set of all math and computing science courses you have taken so far at university. What is the intersection of these sets for your group?

Activity 9.2 Is it possible to have two sets A and B with $A \cup B = A \cap B$?

Activity 9.3 Cancellation is not always valid.

(a) Demonstrate using an example that $(A \cup B) \setminus B = A$ is *not* a valid simplification in set theory.

(b) Demonstrate using an example that $A \cup B = A \cup C \Rightarrow A = B$ is *not* a valid simplification in set theory.

Activity 9.4 Fill in the blank with a concept from the reading.

Breaking the students in a class into groups is an example of ______.

Activity 9.5

(a) Write a definition in Candidate-condition notation for the set of all points on the graph of the parabola $f(x) = x^2$.

(b) Write a definition in Form-parameter notation for the set of all numbers that are one less than a power of two.

Activity 9.6 Recall that $\mathrm{M}_n(\mathbb{R})$ is the set of all $n \times n$ matrices. Let V be the subset of **invertible** $n \times n$ matrices, and S the set of **scalar** $n \times n$ matrices. Write $\mathbf{0}$ for the $n \times n$ zero matrix.

Recall. Scalar matrix means a scalar multiple of the identity matrix.
Singular matrix means not invertible.

Express each of the following statements using the symbols of set theory:

$$\in, \ \subseteq, \ \cup, \ \cap, \ \varnothing, \ \text{etc.}$$

(a) $\mathbf{0}$ is a scalar matrix.

(b) $\mathbf{0}$ is scalar and singular.

(c) $\mathbf{0}$ is the *only* scalar, singular matrix.

(d) Every scalar matrix besides $\mathbf{0}$ is invertible.

(e) Every matrix is either invertible or singular.

Activity 9.7 Pick another group in the class and list the elements of the Cartesian product of your group with that other group. If that group happened to also choose your group for this task, would their answer be the same as yours?

Activity 9.8 List the elements of the power set of your group. Make sure you have all the { }-pairs you need in all the right places.

Activity 9.9 For alphabet $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$, describe the elements of Σ^* and $(\Sigma^*)^*$:
Elements of Σ^* are ______.
Elements of $(\Sigma^*)^*$ are ______.

Is the equality of sets $(\Sigma^*)^* = \Sigma^*$ true?

Activity 9.10 The equality of sets

$$A \times (B \setminus C) = (A \times B) \setminus (A \times C)$$

is true in general.

Write a ***formal*** proof of this equality, using the Test for Set Equality.

Activity 9.11 The equality of sets $(A \times B) \cup (C \times D) = (A \cup C) \times (B \cup D)$ is false in general.

(a) Write down definitions for example sets A, B, C, D that form a counterexample.

(b) Can you come up with some conditions on A, B, C, D that make this equality true?

Activity 9.12 Write a ***formal*** proof of the equality

$$\mathscr{P}(A \cap B) = \mathscr{P}(A) \cap \mathscr{P}(B)$$

using the Test for Set Equality.

Keep Warning 9.7.3 in mind as you do this!

Activity 9.13 Informally explain why the set equality $\mathscr{P}(A \cup B) = \mathscr{P}(A) \cup \mathscr{P}(B)$ is not true in general.

9.9 Exercises

1. Prove each of the set operation rules in Proposition 9.4.11. Use the provided proof of the first of DeMorgan's Laws as a model for your proofs.

Expressing relationships using the symbols of set theory. In each of Exercises 2–4, you are given a collection of sets (and possibly some elements of those sets), a collection of symbols, and a collection of statements about those sets and their elements. Use the given symbols to express the given statements in symbolic language.

Note that there may be more than one correct answer for each statement.

2. Sets:

$A =$ the set of all Augustana students,
$R =$ the set of Augustana students who attend class regularly,
$S =$ the set of Augustana students who study diligently,
$P =$ the set of Augustana students who will pass all their courses.

Symbols:

$$A, \quad R, \quad S, \quad P, \quad R^{c}, \quad S^{c}, \quad P^{c}, \quad \cap, \quad \cup, \quad =, \quad \neq, \quad \subseteq, \quad \subsetneq, \quad \emptyset.$$

Statements:

(a) All Augustana students who attend class regularly and study diligently will pass all their courses.

(b) Some Augustana students attend class regularly but do not study diligently.

(c) Some Augustana students who study diligently will still fail a course.

3. Recall that a square number is an integer which is equal to the square of some integer. (See the introduction preceding Exercise 6.12.17 in Exercises 6.12.)

Sets:

$$\begin{aligned} P &= \text{the set of prime numbers,} \\ E &= \text{the set of even numbers,} \\ S &= \text{the set of square numbers.} \end{aligned}$$

Symbols:

$2,\ \mathbb{N},\ P,\ E,\ S,\ \mathbb{N}^{c},\ P^{c},\ E^{c},\ S^{c},\ \in,\ \cap,\ \cup,\ =,\ \neq,\ \subseteq,\ \subsetneq,\ \varnothing,\ \{\ \}.$

Statements:

(a) 2 is the only even, prime number.

(b) There exist odd square numbers.

(c) No prime number is square.

(d) No square number is prime.

(e) It is not true that every natural number is either even or prime.

4. Sets:

$$\begin{aligned} \mathcal{F} &= \text{the set of all functions in a single real variable,} \\ \mathcal{C} &= \text{the set of continuous functions,} \\ \mathcal{D} &= \text{the set of differentiable functions,} \\ \mathcal{P} &= \text{nonnegative functions} \\ &= \{\, f(x) \mid f(x) \geq 0 \text{ for all } x \text{ in the domain of } f \,\}. \end{aligned}$$

Elements:

$$f_1(x) = x^2 \qquad f_2(x) = |x| \qquad f_3(x) = \tan x$$

Symbols:

$\mathcal{F},\ \mathcal{C},\ \mathcal{D},\ \mathcal{P},\ \mathcal{F}^{c},\ \mathcal{C}^{c},\ \mathcal{D}^{c},\ \mathcal{P}^{c},\ f_1(x),\ f_2(x),\ f_3(x),\ \in,\ \cap,\ \cup,\ =,\ \neq,\ \subseteq,\ \subsetneq,\ \varnothing.$

Statements:

(a) The function $f_1(x)$ is differentiable and nonnegative.

(b) The function $f_2(x)$ is continuous and nonnegative, but not differentiable.

(c) The function $f_3(x)$ is neither continuous nor nonnegative.

(d) Every differentiable function is continuous.

(e) Some continuous functions are not differentiable.

(f) Not every function is continuous.

Testing set equalitye. For each of Exercises 5–8, either formally prove the given equivalence of sets (using the Test for Set Equality) or demonstrate that it is false by providing a *specific* counterexample.

5. $A = (A \setminus B) \sqcup (A \cap B)$
6. $A \setminus (A \setminus B) = B$
7. $(A \times B) \cup (C \times D) = (A \cup C) \times (B \cup D)$
8. $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$

9. Suppose Σ is an alphabet. Prove that Σ^* is the disjoint union of the subsets

$$\Sigma_0^*, \Sigma_1^*, \Sigma_2^*, \ldots, \Sigma_n^*, \ldots.$$

10. Write out the elements of each of the sets

$$\mathscr{P}(\varnothing), \qquad \mathscr{P}(\mathscr{P}(\varnothing)), \qquad \mathscr{P}(\mathscr{P}(\mathscr{P}(\varnothing))), \qquad \mathscr{P}(\mathscr{P}(\mathscr{P}(\mathscr{P}(\varnothing)))).$$

Make sure you have all the pairs of braces { } you should have.

Without computing it, make a conjecture about the number of elements in the set

$$\mathscr{P}(\mathscr{P}(\mathscr{P}(\mathscr{P}(\mathscr{P}(\varnothing))))).$$

Properties of power sets. For each of Exercises 11–14, either formally prove the given statement about power sets or demonstrate that it is false by providing a *specific* counterexample.

11. $\mathscr{P}(A \cup B) = \mathscr{P}(A) \cup \mathscr{P}(B)$
12. $\mathscr{P}(A \cap B) = \mathscr{P}(A) \cap \mathscr{P}(B)$
13. If $A \subseteq B$, then $\mathscr{P}(A) \subseteq \mathscr{P}(B)$.
14. If $A \subseteq B$, then $\mathscr{P}(B \setminus A) = \mathscr{P}(B) \setminus \mathscr{P}(A)$.

Chapter 10

Functions

10.1 Basics

10.1.1 Terminology and basic concepts

function (working definition)
a rule which assigns to each input element from a set A a single output element from a set B

domain the set of all possible input elements for a function

codomain a set *containing* all possible output elements for a function

$f : A \to B$ f is a function with domain A and codomain B

input-output rule
the process/algorithm/rule/formula that describes how each input element from the domain will be transformed into an output element in the codomain

$f(a) = b$ function $f : A \to B$ associates the codomain element $b \in B$ to the domain element $a \in A$

$a \mapsto b$ alternative notation for $f(a) = b$

image (of an domain element)
when $f(a) = b$ we say that b is the **image** of a under f, or that f **maps** a to b

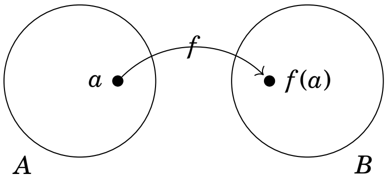


Figure 10.1.1 A Venn diagram of a function transforming a domain element into a codomain element.

Warning 10.1.2 Domain elements are necessarily inputs, but codomain elements are not necessarily outputs. When we define a function, the domain should either be implicitly clear from the input-output rule, or explicitly stated so that the precise collection of allowable input elements is known.

However, it would be too onerous to do the same for the precise collection of output elements — often when we create a function we won't initially know exactly what outputs it will produce. The purpose of stating a **codomain** is so that it is at least clear what *type* of output element is produced.

10.1.2 Defining functions

Defining a function is a ***two-step process***, in which we need to specify ***three pieces of information***:

(i) the **domain**,

(ii) the **codomain**, and

(iii) the **input-output rule**.

The first two pieces of information are specified in one step, when we write

$$f : A \to B.$$

This notation indicates that A will be the domain and B will be the codomain for the function named f. Of course, the **name** of the function is an additional piece of information being specified with this notation, but naming a function is optional (though highly recommended!).

Specifying the input-output rule may be done in many different ways, e.g. by a formula, table of values, a description of a step-by-step process or algorithm to determine or compute an output given an arbitrary input, etc.

Example 10.1.3 Defining a function by an input-output formula. An input-output formula like $f(x) = \sqrt{x}$ defines a function, but we here need to be careful about the domain. The domain and codomain for this function could be specified as $f : \mathbb{R}_{\ge 0} \to \mathbb{R}$, where $\mathbb{R}_{\ge 0}$ represents the set of nonnegative real numbers. □

Example 10.1.4 Correctly stating a domain and codomain. In the function definition

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto \frac{1}{x^2}, \end{aligned}$$

the first line of the definition tells us the **domain** ($\mathbb{R}$), **codomain** (again $\mathbb{R}$), and a name for the function (f). The second line tells us the **input-output rule**, so that

$$f(x) = \frac{1}{x^2}.$$

However, on closer inspection we discover that the domain has been ***incorrectly*** specified, as $x = 0$ is not a permissible input for the input-output rule. Instead, we should write

$$f : \mathbb{R} \setminus \{0\} \to \mathbb{R}.$$

Even though this function will only ever produce *positive* real numbers as outputs, the codomain is acceptable as stated. It *would* be more precise to write

$$f : \mathbb{R} \setminus \{0\} \to \mathbb{R}_{>0},$$

where $\mathbb{R}_{>0}$ represents the set of positive real numbers, but it is not necessary to do so. □

Example 10.1.5 Defining a function by an input-output algorithm. Consider the function $\mathscr{P}(\mathbb{Z}) \to \mathbb{N}$ where outputs are computed according to the following algorithm.

Given an input element $X \in \mathscr{P}(\mathbb{Z})$ (which, by definition, is a *set* of integers), carry out the following.

1. Compute the absolute value of each element in X. (If X is empty, skip this

step.)

2. Determine the minimum result of the absolute value computations in the previous step. (If X is empty, there will not be any absolute value computation results to compare, so take 0 as the "mininum" instead.)

3. Multiply the minimum value found in the previous step by 2 and add 1. Output this final result.

However, with the right notation, an algorithm like the above can often be converted into an input-output formula — see Example 10.4.7. □

Example 10.1.6 Defining a function by listing input-ouput pairs. For $\mathcal{N} = \{1,2,3\}$ and $\mathcal{A} = \{a,b,c,d\}$, one way to define a function $f\colon \mathcal{N} \to \mathcal{A}$ is

$$f(1) = d, \qquad f(2) = a, \qquad f(3) = d.$$

□

Example 10.1.7 Multi-variable functions. In a first course in calculus a student typically studies only **single-variable functions**, i.e. functions with a single input variable and a single output variable. In subsequent calculus courses a student may study **multi-variable functions** with multiple input variables, such as

$$f(x,y) = x^2 + y^2.$$

Technically, we should write

$$f\big((x,y)\big) = x^2 + y^2,$$

as the proper definition of f is $f\colon \mathbb{R}^2 \to \mathbb{R}$, but the extra brackets convey no additional information and only clutter things up.

Functions with multiple real output variables are often called **vector functions**. For example, $g\colon \mathbb{R} \to \mathbb{R}^2$ defined by

$$g(t) = (t, t^2)$$

can be considered as a vector parametrization of a parabola in the plane.

And of course we could consider **multi-variable vector functions** as well. A function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ like

$$\varphi(s,t) = (s-t, s+t)$$

could be considered as a change of variables

$$x = s - t, \qquad y = s + t.$$

□

Example 10.1.8 Logical statements as functions. A logical statement S involving statement variables $p_1, p_2, \ldots, p_m$ is essentially a multi-variable function

$$S\colon \Lambda^m \to \Lambda,$$

where $\Lambda = \{\mathrm{T}, \mathrm{F}\}$. For example, the statement

$$S(p_1,p_2) = (p_1 \to p_2)$$

is a function $S\colon \Lambda \times \Lambda \to \Lambda$, where

$$S(\mathrm{T},\mathrm{T}) = \mathrm{T}, \qquad S(\mathrm{T},\mathrm{F}) = \mathrm{F}, \qquad S(\mathrm{F},\mathrm{T}) = \mathrm{T}, \qquad S(\mathrm{F},\mathrm{F}) = \mathrm{T}.$$

□

10.1.3 Graph of a function

graph (of a function)
the set of all input-output pairs for the function

$\Delta(f)$ the graph of function $f\colon A \to B$, so that

$$\Delta(f) = \{\, (a, f(a)) \mid a \in A \,\} \subseteq A \times B$$

Example 10.1.9 Graph of a single-variable, real-valued function. The graph of a function $f\colon \mathbb{R} \to \mathbb{R}$ is a subset of $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. We usually represent $\mathbb{R}^2$ visually as the xy-plane and the graph $\Delta(f) \subseteq \mathbb{R}^2$ as a curve in the plane.

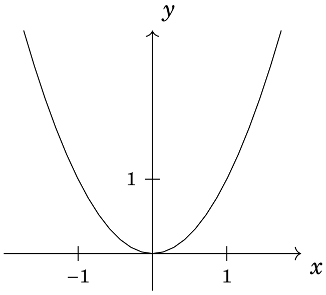


Figure 10.1.10 The graph $\Delta(x \mapsto x^2)$ represented as a curve in the Cartesian plane.

In the graph of $f(x) = x^2$ above, each point on the curve represents an element of $\mathbb{R}^2$ which is in the subset $\Delta(f)$. For example, $(-1,1) \in \Delta(f)$ but $(-1,\pi) \notin \Delta(f)$. □

Example 10.1.11 Graph of a multi-variable, real-valued function. The graph of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is technically a subset of $\mathbb{R}^2 \times \mathbb{R}$, but usually we just think of this as $\mathbb{R}^3$, or 3-space. Instead of a curve, such a graph defines a *surface* in $\mathbb{R}^3$. For example, the graph of the function $f(x,y) = x^2 + y^2$ from Example 10.1.7 is a parabolic cone, i.e. a (non-solid) cone-like surface with parabolic sides. □

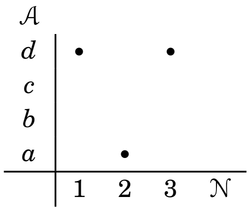
Example 10.1.12 Graph of a function defined by a list. To describe the graph of the function $f\colon \mathcal{N} \to \mathcal{A}$ defined in Example 10.1.6, we just need to collect the defined input-output pairs into Cartesian product elements:

$$\Delta(f) = \{(1,d),(2,a),(3,d)\}.$$

This graph can most simply be represented by a table:

x	1	2	3
$f(x)$	d	a	d

Similarly to Example 9.5.7, we can also visualize this graph as a set of "points" relative to a set of perpendicular "axes," where the horizontal axis represents the domain set and the vertical axis represents the codomain set.



Notice that we have not joined the points in the above visualization with lines or a curve, since the three points pictured are the *only* points on the graph. □

Example 10.1.13 Graph of a logical statement. We've already encountered the graph of a logical statement: it is usually represented as a truth table. For example, the graph $\Delta(S)$ of the logical statement

$$S\colon \Lambda \times \Lambda \to \Lambda, \qquad S(p_1,p_2) = p_1 \to p_2,$$

where $\Lambda = \{\mathrm{T},\mathrm{F}\}$ as usual, can be represented as below.

p_1	p_2	$S(p_1,p_2)$
T	T	T
T	F	F
F	T	T
F	F	T

Figure 10.1.14 The graph $\Delta(S)$ of the logical statement $S(p_1,p_2) = p_1 \to p_2$.

□

Unfortunately, our working definition for **function** is lacking: what is a "rule"? Rather than chasing some circle of definitions, we can come up with a better definition by noticing that the graph of a function contains all the necessary information about the function.

function (formal definition)
a subset $F \subseteq A \times B$ such that for every $x \in A$ there is *exactly one* element $(a,b) \in F$ with $a = x$

In this formal definition, we are defining a function to *be* what we previously would have called its **graph**.

Example 10.1.15 Formal definition for a single-variable, real-valued function. We are now defining a function $f\colon \mathbb{R} \to \mathbb{R}$ to be the subset of the Cartesian plane $\mathbb{R}^2$ consisting of the graph of the function. In this case, you can think of the "exactly one" requirement as equivalent to the **vertical line test**: an input value may not produce more than one output value. (Though the "one" part of "exactly one" captures our requirement that a function be defined on every domain element.) □

10.1.4 Undefined and well-defined

We have to be careful defining functions; sometimes what we think is a function turns out to not be a function.

Example 10.1.16 A function must be defined on the whole domain. Again write $\mathcal{N} = \{1,2,3\}$ and $\mathcal{A} = \{a,b,c,d\}$, and consider

$$F = \{(1,a),(3,d)\} \subseteq \mathcal{N} \times \mathcal{A}.$$

Does this subset define a function with domain $\mathcal{N}$ and codomain $\mathcal{A}$? That is, does there exist a function $f\colon \mathcal{N} \to \mathcal{A}$ such that $F = \Delta(f)$? The answer is ***no***, because there is no input-output pair in F with domain element 2. If we attempt to consider a function f with graph $\Delta(f) = F$, we have no way to tell what result $f(2)$ should return. In other words, such an f will have been left **undefined** on element 2, which is supposed to be part of the domain.

The set F *does* define a function, just not one with domain $\mathcal{N}$. If we consider the smaller set $\mathcal{N}' = \{1,3\}$, then there *is* a function $f\colon \mathcal{N}' \to \mathcal{A}$ with $F = \Delta(f)$. □

Example 10.1.17 A function must be well-defined. Again write $\mathcal{N} = \{1,2,3\}$ and $\mathcal{A} = \{a,b,c,d\}$, and consider

$$F = \{(1,a),(3,a),(3,d)\} \subseteq \mathcal{N} \times \mathcal{A}.$$

Does this subset define a function with domain $\mathcal{N}$ and codomain $\mathcal{A}$? That is, does there exist a function $f\colon \mathcal{N} \to \mathcal{A}$ such that $F = \Delta(f)$? The answer is ***no***, because there are *more than one* input-output pairs with domain element 3. In other words, a function f with graph $\Delta(f) = F$ is not **well-defined**, because we have no way to tell whether $f(3)$ should be a or d. □

Example 10.1.18 An input-output rule does not necessarily define a function. Recall that

$$\mathbb{Q} = \left\{ \frac{m}{n} \,\middle|\, m,n \in \mathbb{Z},\, n \neq 0 \right\}.$$

Suppose we attempt to define $f\colon \mathbb{Q} \to \mathbb{Z}$ by $f(\frac{m}{n}) = m+n$. This seems like a valid way to define a function, until we realize that, for example,

$$f\left(\frac{1}{2}\right) = 1+2 = 3, \qquad f\left(\frac{2}{4}\right) = 2+4 = 6.$$

This is nonsense, because $\frac{1}{2}$ and $\frac{2}{4}$ represent the *same element* of $\mathbb{Q}$. Thus, rule f is not well-defined as a function, since to each element of the domain $\mathbb{Q}$ it associates more than one element of the codomain $\mathbb{Z}$. □

10.1.5 Equality of functions

equality of functions

for $f\colon A \to B$ and $g\colon A \to B$, write $f = g$ if $f(a) = g(a)$ for all $a \in A$

Example 10.1.19 Seemingly different input-output rules can define the same function. The functions $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$, and $g\colon \mathbb{R} \to \mathbb{R}$, $g(x) = \sqrt{x^2}$, are equal. □

10.1.6 Image of a function

image of a function

the set of all possible outputs of the function

$f(A)$ the image of function $f\colon A \to B$, so that

$$f(A) = \{\, f(a) \mid a \in A \,\} \subseteq B$$

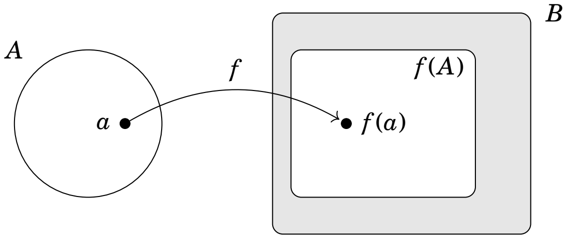


Figure 10.1.20 A Venn diagram of the image of a function.

Warning 10.1.21 Codomain elements are not necessarily image elements. We have stated before that a codomain in a function definition may be “larger” than necessary because we do not always know precisely what output elements a given input-output rule will produce. Our idea of codomain is that it should at least tell us what “type” of outputs will be produced, but not necessarily *exactly* what outputs will be produced.

With our new concept of **function image**, we can now repeat this more technically: ***a function image is always a subset of the codomain, but it might be a proper subset***.

How do we know if a codomain element is an image element? For function $f : A \to B$ and codomain element $b \in B$, we have $b \in f(A)$ if and only if there exists $a \in A$ such that $b = f(a)$.

Worked Example 10.1.22 Verifying a function image description. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Prove $f(\mathbb{R}) = \mathbb{R}_{\geq 0}$, where $\mathbb{R}_{\geq 0}$ is the set of nonnegative real numbers.

Solution. Following the Test for Set Equality, we need to show both

$$f(\mathbb{R}) \subseteq \mathbb{R}_{\geq 0}, \qquad\qquad f(\mathbb{R}) \supseteq \mathbb{R}_{\geq 0}.$$

To be more explicit about the second set $\mathbb{R}_{\geq 0}$, we can write

$$\mathbb{R}_{\geq 0} = \{\, x \in \mathbb{R} \mid x \geq 0 \,\}.$$

Show $f(\mathbb{R}) \subseteq \mathbb{R}_{\geq 0}$. Let y represent an *arbitrary* element of $f(\mathbb{R})$. As an element of the image of f, y is an output corresponding to some input. That is, there exists some $x \in \mathbb{R}$ such that

$$y = f(x) = x^2.$$

Therefore, since square numbers are always positive, we have $y \geq 0$, and hence $y \in \mathbb{R}_{\geq 0}$.

Show $f(\mathbb{R}) \supseteq \mathbb{R}_{\geq 0}$. Let y represent an *arbitrary* element of $\mathbb{R}_{\geq 0}$. To show $y \in f(\mathbb{R})$, we need to find $x \in \mathbb{R}$ such that $f(x) = y$. Let $x = \sqrt{y}$, which is defined since $y \in \mathbb{R}_{\geq 0}$ implies $y \geq 0$. Then

$$f(x) = x^2 = (\sqrt{y})^2 = y,$$

as desired. □

image of a function on a subset

the set of all outputs of a function when only fed inputs from a given subset

$f(A')$ the image of the subset $A' \subseteq A$ under a function $f : A \to B$, so that

$$f(A') = \left\{\, f(a) \,\middle|\, a \in A' \,\right\} \subseteq B$$

Example 10.1.23 We saw in Worked Example 10.1.22 that for $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, we have $f(\mathbb{R}) = \mathbb{R}_{\geq 0}$. Now, the set of integers $\mathbb{Z}$ is a subset of the domain $\mathbb{R}$, so we can compute

$$f(\mathbb{Z}) = \{0, 1, 4, 9, 16, \ldots, n^2, \ldots\} \subseteq \mathbb{R}_{\geq 0}.$$

□

10.2 Properties of functions

surjective function
a function whose image is all of its codomain — that is, every element of the codomain is an output for the function;

surjection
a surjective function

onto synonym for surjective

$f\colon A \twoheadrightarrow B$ function f is surjective

A function $f\colon A \to B$ is surjective if $f(A) = B$. Since we have $f(A) \subseteq B$ by definition of image, to show that a function is surjective we only need to show $f(A) \supseteq B$.

Test 10.2.1 Surjective function.

- *Function $f\colon A \to B$ is surjective if $B \subseteq f(A)$. That is, f is surjective if* for every *element $b \in B$, there exists* at least one *element $a \in A$ such that $f(a) = b$.*
- *Function $f\colon A \to B$ is* not *surjective if there exists* at least one *element $b \in B$ for which there is* no *element $a \in A$ satisfying $f(a) = b$. (Equivalently, there exists $b \in B$ for which every $a \in A$ satisifes $f(a) \neq b$.)*

Worked Example 10.2.2 Show that, of the following functions, f is surjective and g is not.

$$\begin{aligned} f\colon \mathbb{Z} &\to \mathbb{N} & \qquad g\colon \mathbb{Z} &\to \mathbb{Q} \\ m &\mapsto |m| & m &\mapsto m/2 \end{aligned}$$

Solution.

Show that f is surjective. Consider an *arbitrary* element n of the codomain $\mathbb{N}$. Since $\mathbb{N} \subseteq \mathbb{Z}$, n is also an element of the domain. In particular, $f(n) = n$, since $n \geq 0$. Therefore, as an element of the codomain, we have $n \in f(\mathbb{Z})$.

Show that g is not surjective. We need to find a *specific* example of a rational number that is not an output for g. For this, we could use 1/3, since there is no integer such that $m/2 = 1/3$. □

injective function
a function for which two different inputs *never* produce the same output

injection an injective function

embedding
synonym for injection

one-to-one
synonym for injective

$f\colon A \hookrightarrow B$ function f is injective

Test 10.2.3 Injective function.

- *Function $f\colon A \to B$ is injective if the following conditional always holds for elements $a_1, a_2 \in A$:*

if $a_1 \neq a_2$ *then* $f(a_1) \neq f(a_2)$.

Alternatively, one can establish that the contrapositive of the above conditional always holds for elements $a_1, a_2 \in A$*:*

if $f(a_1) = f(a_2)$ *then* $a_1 = a_2$.

- *Function* $f : A \to B$ *is* not *injective if there exists* at least one *pair of elements* $a_1, a_2 \in A$ *with* $a_1 \neq a_2$ *but* $f(a_1) = f(a_2)$.

Example 10.2.4 Demonstrating that a function is not injective. The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, is *not* injective, since f has repeated outputs. For example, $f(-1) = f(1)$. And in fact, $f(-x) = f(x)$ for every $x \in \mathbb{R}$. □

Worked Example 10.2.5 Demonstrating that a function is injective. Verify that the function $f : \mathbb{N} \to \mathbb{N}$, $f(n) = 2n + 1$, is injective.

Solution. Using the contrapositive version of the Injective Function Test, suppose domain elements $n_1, n_2 \in \mathbb{N}$ satisfy $f(n_1) = f(n_2)$. Then using the formula defining the input-output rule for f, we have

$$2n_1 + 1 = 2n_2 + 1,$$

which reduces to $n_1 = n_2$. □

An injection $f : A \hookrightarrow B$ gives us a way of thinking of A as a *subset* of B, by considering $f(A) \subseteq B$.

Example 10.2.6 Turning letters into numbers. Let Σ = {a, b, ..., z}, and define $\varphi : \Sigma \to \mathbb{N}$ by the following table.

σ	a	b	c	$\cdots$	z
$\varphi(\sigma)$	1	2	3	$\cdots$	26

Then f **embeds** Σ into $\mathbb{N}$ in a familiar way, and lets us think of letters as numbers. □

bijective function
a function that is both injective and surjective

bijection an bijective function

one-to-one correspondence
synonym for bijection

Example 10.2.7 Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$ is bijective. □

A bijection $f : A \to B$ allows us to think of A and B as essentially the same sets.

Example 10.2.8 Identifying letters with numbers. Consider again $f : \Sigma \to \mathbb{N}$ from Example 10.2.6. If we write $B = f(\Sigma) = \{1, 2, 3, \ldots, 26\}$, then really we could think of the function as being defined $f : \Sigma \to B$. This version of f is bijective, and allows us to identify each letter with a corresponding number:

$$a \leftrightarrow 1, \qquad b \leftrightarrow 2, \qquad c \leftrightarrow 3, \qquad \ldots, \qquad z \leftrightarrow 26.$$

In this way, we can think of Σ and B as essentially the same set. □

Worked Example 10.2.9 Recognizing bijections. Which of the following functions are bijections?

$$f : \mathbb{Z} \to \mathbb{Z}, \qquad g : \mathbb{Z} \to \mathbb{N}, \qquad h : \mathbb{Z} \to \mathbb{Z},$$

$$m \mapsto 2m, \qquad m \mapsto |m|, \qquad m \mapsto -m.$$

Solution.

Is f bijective?. No, f is not bijective because it is not surjective. For example, there is no $m \in \mathbb{Z}$ such that $f(m) = 1$.

Is g bijective?. No, g is not bijective because it is not injective. For example, $g(-1) = g(1)$.

Is h bijective?. Yes, h is bijective. It is injective because if $m_1 \neq m_2$ then $-m_1 \neq -m_2$. And it is surjective because for $n \in \mathbb{Z}$, we can realize n as an output $n = h(m)$ by setting $m = -n$. □

Checkpoint 10.2.10 Bijections of counting sets. For $m \in \mathbb{N}$ write

$$\mathbb{N}_{<m} = \{\, n \in \mathbb{N} \mid n < m \,\} = \{0, 1, \dots, m-1\}.$$

Prove that there exists a bijection $\mathbb{N}_{<\ell} \to \mathbb{N}_{<m}$ if and only if $\ell = m$.

10.3 Important examples

identity function (on a set A)
the function $A \to A$ defined by $a \mapsto a$

$\mathrm{id}_A : A \to A$
the identity function on on set A

Note 10.3.1 An identity function is always a bijection.

inclusion function (on subset $A \subseteq X$)
the function $A \to X$ defined by $a \mapsto a$

Check your understanding. Do you understand the difference between the definitions of **identity function** and **inclusion function**?

$\iota_A^X : A \to X$ the inclusion function on subset $A \subseteq X$

Note 10.3.2 An inclusion function is always an injection.

projection functions (on a Cartesian product $A \times B$)
the functions $A \times B \to A$ and $A \times B \to B$ defined by $(a, b) \mapsto a$ and $(a, b) \mapsto b$

$\rho_A : A \times B \to A$
the projection function onto the first factor A in the Cartesian product $A \times B$

$\rho_B : A \times B \to B$
the projection function onto the second factor B in the Cartesian product $A \times B$

Example 10.3.3 Projection images. Consider $(\frac{1}{2}, \pi) \in \mathbb{Q} \times \mathbb{R}$. Then

$$p_{\mathbb{Q}}\left(\frac{1}{2},\pi\right)=\frac{1}{2}, \qquad p_{\mathbb{R}}\left(\frac{1}{2},\pi\right)=\pi.$$

□

Extend. We may of course similarly define a projection function on a Cartesian product with any number of factors. Write

$$\rho_i : A_1 \times A_2 \times \cdots \times A_n \to A_i$$

to mean the projection function onto the i^{th} factor A_i in the Cartesian product

$$A_1 \times A_2 \times \cdots \times A_n.$$

Alternatively, we may write

$$\operatorname{proj}_i : A_1 \times A_2 \times \cdots \times A_n \to A_i$$

for this function.

Note 10.3.4 A projection is always surjective (except possibly when one or more of the factors in the Cartesian product is the empty set).

restricting the domain

	the “induced” function $A \to Y$ created from function $f : X \to Y$ and subset $A \subseteq X$ by “forgetting” about all elements of X that do not lie in A
$f\vert_A$	restriction of function $f : X \to Y$ to subset $A \subseteq X$
$f\vert A$	alternative domain restriction notation
$\operatorname{res}^X_A f$	alternative domain restriction notation

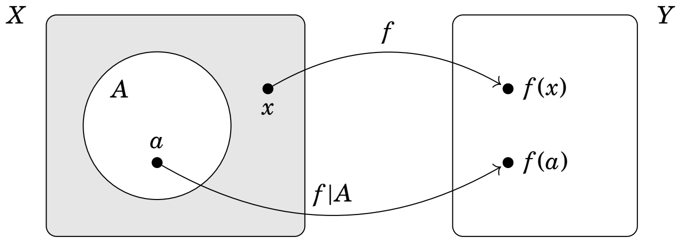


Figure 10.3.5 A Venn diagram of restricting the domain of a function.

Example 10.3.6 Domain restriction. For $f : \mathbb{Z} \to \mathbb{N}$, $f(m) = |m|$, we have $f|_{\mathbb{N}} = \operatorname{id}_{\mathbb{N}}$. □

Checkpoint 10.3.7 Properties of restrictions. Consider function $f : X \to Y$ and subset $A \subseteq X$.

1. If f is injective, is $f|_A$ injective?

2. If $f|_A$ is injective, must f be injective?

3. Answer the previous two questions replacing “injective” with “surjective”.

Remark 10.3.8 The concept of **restricting the domain** makes our previously defined concept **image of a function on a subset** unnecessary: for function $f\colon X \to Y$ and subset $A \subseteq X$, the image of f on A is the same as the image of the restriction $f|_A$.

restricting the codomain

the "induced" function $X \to B$ created from function $f\colon X \to Y$ and subset $B \subseteq Y$ by "forgetting" about all elements of Y that do not lie in B, *where B must contain the image of f*

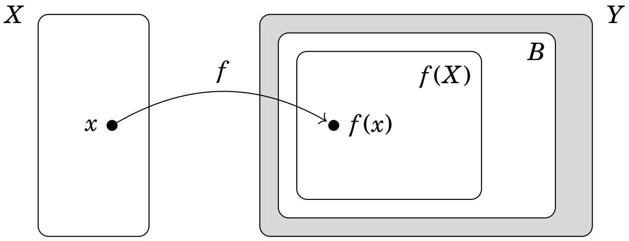


Figure 10.3.9 A Venn diagram of restricting the codomain of a function.

Example 10.3.10 Codomain restriction. Consider $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. It would be more precise to write $f\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$, since $x^2 \geq 0$ for all $x \in \mathbb{R}$. □

Note 10.3.11 If we restrict the codomain all the way down to the image set $f(X)$, the resulting map $f\colon X \to f(X)$ is always surjective. In particular, if $f\colon X \hookrightarrow Y$ is injective, then by restricting the codomain we can obtain a *bijection* $f\colon X \to f(X)$.

extension of a function

relative to function $f\colon A \to B$ and *super*set $X \supseteq A$, a function $g\colon X \to B$ so that $g(a) = f(a)$ for all $a \in A$

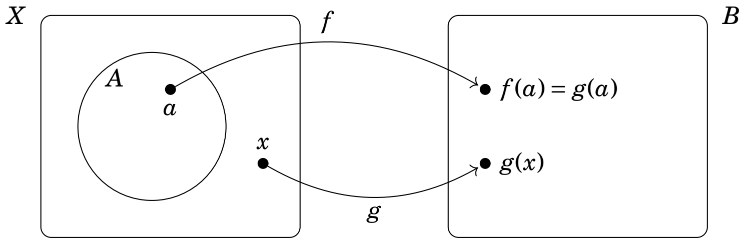


Figure 10.3.12 A Venn diagram of a function extension.

Note 10.3.13 The condition defining the concept **extension function** can be more succinctly stated as requiring function $g\colon X \to B$ with $A \subseteq X$ satisfy $g|_A = f$.

Example 10.3.14 Floor function. Write $\operatorname{flr}\colon \mathbb{R} \to \mathbb{Z}$ to mean the **floor function**: for real input x, the output $\operatorname{flr}(x)$ is defined to be the greatest integer that is less than or equal to x. Usually we write

$$\operatorname{flr}(x) = \lfloor x \rfloor.$$

As every integer is less than or equal to itself, we have $\operatorname{flr}(z) = z$ for every $z \in \mathbb{Z}$. This says that the floor function is an extension of the identity function $\mathrm{id}_{\mathbb{Z}}$. □

One of the most common ways to extend a function to a larger domain is to pick an appropriate constant value in the codomain to assign to all "new" inputs in the enlarged domain.

extenstion by zero

relative to function $f\colon A \to Z$ and superset $X \supseteq A$, where Z is a set of "numbers" containing a zero element, the extension function $g\colon X \to Z$ defined by

$$g(x) = \begin{cases} f(x), & x \in X, \\ 0, & \text{otherwise.} \end{cases}$$

Example 10.3.15 Extending the identity function by zero. Define $\widetilde{\mathrm{id}}_{\mathbb{Z}}\colon \mathbb{R} \to \mathbb{Z}$ by

$$\widetilde{\mathrm{id}}_{\mathbb{Z}}(x) = \begin{cases} x, & x \in \mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\widetilde{\mathrm{id}}_{\mathbb{Z}}$ is the extension by zero of the identity function $\mathrm{id}_{\mathbb{Z}}$.

Compare. Example 10.3.14 also involved an extension of the identity function $\mathrm{id}_{\mathbb{Z}}$ — was it an **extension by zero**?

□

10.4 Composition of functions

composition function

a function $A \to C$ created from given functions $f\colon A \to B$ and $g\colon B \to C$ by $a \mapsto g\big(f(a)\big)$

$g \circ f$

the composition of functions $f\colon A \to B$ and $g\colon B \to C$, so that $g \circ f\colon A \to C$ by $(g \circ f)(a) = g\big(f(a)\big)$

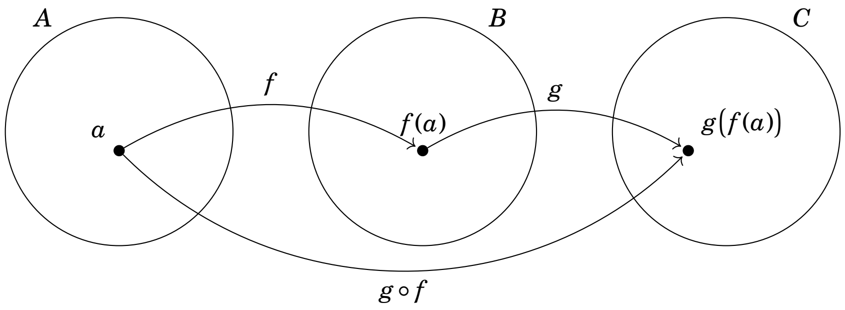


Figure 10.4.1 A Venn diagram of a function composition.

Example 10.4.2 A composition of two functions. Consider the functions

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R}_{\geq 0}, & g\colon \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0}, \\ x &\mapsto x^2, & x &\mapsto \sqrt{x}, \end{aligned}$$

Then we have

$$\begin{aligned} g \circ f\colon \mathbb{R} &\to \mathbb{R}_{\geq 0}, \\ x &\mapsto \sqrt{x^2} = |x|. \end{aligned}$$

□

Warning 10.4.3 Composition order matters. The notation for the composition of functions f and g involves a reversal of order, so that we write $g \circ f$. This is so that when we use this notation with input-output notation $(g \circ f)(a)$, the notation reminds us that f must first be applied to the input a, and then g is applied to the result $f(a)$.

In general, $f \circ g \neq g \circ f$. Usually, one of the two is not even defined, because domains and codomains of f and g will not necessarily match up in both orders. And when both *are* defined, the two different orders of composition usually have different domains and codomains.

Example 10.4.4 Comparing composition order. Consider functions

$$f \colon \mathbb{N} \to \mathbb{N}, \qquad\qquad g \colon \mathbb{N} \to \mathbb{N},$$
$$n \mapsto n^2, \qquad\qquad n \mapsto n+1.$$

Then, both $f \circ g \colon \mathbb{N} \to \mathbb{N}$ and $g \circ f \colon \mathbb{N} \to \mathbb{N}$ are defined. But they are not equal, as

$$(f \circ g)(n) = (n+1)^2 = n^2 + 2n + 1, \qquad\qquad (g \circ f)(n) = n^2 + 1.$$

□

Example 10.4.5 An undefined composition. Consider functions

$$\mathrm{sqrt} \colon \mathbb{N} \to \mathbb{R}, \qquad\qquad \mathrm{flr} \colon \mathbb{R} \to \mathbb{Z},$$
$$n \mapsto \sqrt{n}, \qquad\qquad x \mapsto \lfloor x \rfloor.$$

(See Example 10.3.14 for a description of the flr function.)

Then, $\mathrm{flr} \circ \mathrm{sqrt} \colon \mathbb{N} \to \mathbb{Z}$ is defined, with

$$(\mathrm{flr} \circ \mathrm{sqrt})(n) = \lfloor \sqrt{n} \rfloor.$$

But $\mathrm{sqrt} \circ \mathrm{flr}$ is not defined, as the codomain of flr does not match the domain of sqrt. In particular, flr will sometimes return a negative output, and we cannot use such an output as an input in sqrt. □

Checkpoint 10.4.6 Properties of compositions. Consider functions $f \colon A \to B$ and $g \colon B \to C$.

1. If $g \circ f$ is injective, are either or both of f, g necessarily injective?
2. Answer the same question as above with “injective” replaced by “surjective”.
3. Demonstrate that if both f and g are bijective, then the composition $g \circ f$ is also bijective.

Of course, we can compose any number of functions.

Example 10.4.7 A composition of three functions. Let us reconsider the function defined by algorithm in Example 10.1.5. As the function description involved a multi-step algorithm, we should be able to break the steps involved into their own functions, then recreate the original functions as a composition.

First, define $\mathrm{abs} \colon \mathscr{P}(\mathbb{Z}) \to \mathscr{P}(\mathbb{N})$ by

$$\mathrm{abs}(X) = \{\, |x| \mid x \in X \,\}.$$

Next, define $\min \colon \mathscr{P}(\mathbb{N}) \to \mathbb{N}$ so that $\min(X)$ outputs the minimum number in input set X, and outputs 0 in case $X = \varnothing$.

Finally, define $f \colon \mathbb{N} \to \mathbb{N}$ by $f(n) = 2n + 1$.

Each of these functions represents one step in the algorithm defining the function in Example 10.1.5, but to recreate that function we need to compose the functions in the correct order: write $\varphi = f \circ \min \circ \operatorname{abs}$, so that

$$\varphi(X) = 2\min\bigl(\operatorname{abs}(X)\bigr) + 1$$

computes the same result for an input set X as the algorithm described in Example 10.1.5. □

10.5 Inverses

Suppose $f : A \to B$ is a function. By definition, f associates an element of B to each element of A. Sometimes we want to reverse this process: given an element $b \in B$, can we determine an element $a \in A$ such that $f(a) = b$? We'll begin to answer this question by first finding all possible "reverse results" from elements in subsets of B.

inverse image (of a subset C of the codomain B)
the set of all domain elements $a \in A$ for function $f : A \to B$ for which the corresponding output element $f(a)$ lies in the subset C of the codomain

$f^{-1}(C)$ the inverse image of the subset $C \subseteq B$ under the function $f : A \to B$, so that

$$f^{-1}(C) = \{\, a \in A \mid f(a) \in C \,\}$$

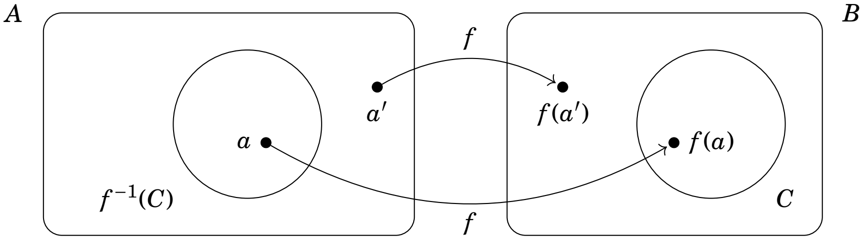


Figure 10.5.1 A Venn diagram of a function inverse image.

Idea 10.5.2 *As in the figure above, $f^{-1}(C)$ collects together all those elements of A whose images under f land inside C.*

Example 10.5.3 Some inverse images under sine. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sin x$.

Then

$$f^{-1}\bigl(\{-1, 0, 1\}\bigr) = \left\{\, \frac{m\pi}{2} \,\middle|\, m \in \mathbb{Z} \,\right\}$$

because

$$\sin\left(\frac{m\pi}{2}\right)$$

will equal 0 when m is even and will equal 1 or -1 when m is odd, and no other input values will produce outputs of 0, 1, or -1.

However,

$$f^{-1}\bigl(\{\, y \in \mathbb{R} \mid y > 1 \,\}\bigr) = \varnothing$$

because there are *no* input values for sine that will produce an output value greater than 1. □

Now let's return to the question of trying to reverse an input-output relationship $f(a) = b$: the set $f^{-1}\bigl(\{b\}\bigr)$ collects together all possible candidates for the inverse image of b.

inverse image (of an element b of the codomain B)
the inverse image $f^{-1}(\{b\})$, which consists of all domain elements $a \in A$ for which $f(a) = b$

$f^{-1}(b)$ simplified notation to mean the inverse image of element b

This gives us a way to associate to an element $b \in B$ a set $f^{-1}(b)$ of elements of A.

Question 10.5.4 When does this association $b \mapsto f^{-1}(b)$ give us a function $f^{-1}\colon B \to A$?

There are two possible ways that this will fail to give us a function.

1. Suppose there is an element $b \in B$ such that the set $f^{-1}(b)$ contains (at least) two distinct elements a_1, a_2. Then in general there is no way to choose between $f^{-1}(b) = a_1$ and $f^{-1}(b) = a_2$. Therefore, if f is *not injective*, the function $f^{-1}\colon B \to A$ is *not well-defined*.

2. Suppose there is an element $b \in B$ such that $f^{-1}(b) = \varnothing$. Then there is *no* element of A which we can assign to $f^{-1}(b)$. Therefore, if f is *not surjective*, the function $f^{-1}\colon B \to A$ is *undefined* on some elements of B.

So it seems we will need a function to be *bijective* in order to be able to reverse the input-output rule to obtain an inverse function.

inverse function
for a bijective function f, the **inverse function** associates to each codomain element of f the corresponding unique domain element that produces it through f

f^{-1} the inverse function $f^{-1}\colon B \to A$ for bijective function $f\colon A \to B$, so that for $b \in B$ we have $f^{-1}(()b)$ defined to be the unique element $a \in A$ such that $f(a) = b$

Example 10.5.5 An invertible single-variable, real-valued function. The function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, is bijective and has inverse $f^{-1}(x) = x^{\frac{1}{3}}$. □

Example 10.5.6 Inverting a numerical encoding of the alphabet. Returning again to the bijection $\varphi\colon \Sigma \to B$ encountered in Example 10.2.6 and Example 10.2.8, where

$$\Sigma = \{\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}, \qquad B = \{1, 2, \ldots, 26\},$$

the inverse function $\varphi^{-1}\colon B \to \Sigma$ associates to each number $1 \le b \le 26$ the corresponding letter at that position of the alphabet. For example, $\varphi^{-1}(11) = \mathrm{k}$. □

Example 10.5.7 A non-invertible function. The function $g\colon \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$, does not have an inverse since it is not bijective. However, the function $h\colon \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$, $h(x) = x^2$, so that $h = g|_{\mathbb{R}_{\ge 0}}$ but with codomain also restricted down to the image of g, has inverse $h^{-1}(x) = \sqrt{x}$. □

Note 10.5.8 If f is bijective, then so is f^{-1}, and f^{-1} is the *unique* function $B \to A$ such that both

$$f^{-1} \circ f = \mathrm{id}_A, \qquad f \circ f^{-1} = \mathrm{id}_B.$$

Checkpoint 10.5.9 Prove that if f is bijective then so is f^{-1}, and $(f^{-1})^{-1} = f$.

10.6 Activities

Activity 10.1 Suppose n is a fixed but unknown positive integer, and let $D: \mathbb{R} \to \mathbb{R}^n$ represent the function defined by $D(x) = (x, x, \ldots, x)$.

Write a set definition in Candidate-condition notation for the image set $D(\mathbb{R})$. Then do the same for the graph $\Delta(D)$.

Activity 10.2

(a) Devise an example of a function $\mathbb{N} \to \mathbb{N}$ that is bijective.

(b) Devise an example of a function $\mathbb{N} \to \mathbb{N}$ that is injective but not surjective.

(c) Devise an example of a function $\mathbb{N} \to \mathbb{N}$ that is surjective but not injective.

Note that when you define a function, you don't necessarily have to give an input-output *formula* — you can also use a table of input-output values or just a description in words (i.e. an **algorithm**) of how an output is to be produced from an input.

Activity 10.3 For each function $f: A \to B$ defined below, carry out the following.

(i) Decide whether the function is injective. Use the Injective Function Test to verify your answer.

(ii) Determine some pattern that all elements of the image $f(B)$ have in common. That is, if you were handed an arbitrary element of the codomain B, describe what property or properties you would use to determine whether it was in the subset $f(A) \subseteq B$.

(iii) Decide whether the function is surjective. Use the Surjective Function Test to verify your answer.

(a) $\Sigma = \{0,1\}$, $c: \Sigma^* \to \Sigma^*$ is the **bitwise complement** function: for input word w, the output word $c(w)$ is the word of the same length as w but with a 0 at every position that w has a 1, and a 1 at every position that w has a 0.

(b) $f: \mathbb{R} \to \mathbb{R} \times \mathbb{R}$, $f(x) = (x+1, x-1)$.

(c) $A = \mathscr{P}(\mathbb{N}) \setminus \{\varnothing\}$, $m: A \to \mathbb{N}$, $m(X) =$ the smallest number in X.

Activity 10.4 Consider $\Sigma = \{0,1\}$, and recall that for $n \in \mathbb{N}$, Σ_n^* is the subset of Σ^* consisting of all words from the alphabet Σ with length n. Suppose $A = \{a_1, a_2, \ldots, a_n\}$ is a set with n distinct elements. Construct a bijection $\mathscr{P}(A) \to \Sigma_n^*$.

When attempting this activity, remember that when you define a function you don't necessarily have to give an input-output *formula* — you can also use a description in words (i.e. an **algorithm**) of how an output is to be produced from an input.

Activity 10.5 Suppose A is a set that definitely does not contain any cats, and let

$$f: \mathscr{P}(A) \to \mathscr{P}(A \cup \{\text{Grumpy Cat}\})$$

represent the function defined by

$$f(X) = X \cup \{\text{Grumpy Cat}\}.$$

(a) Verify that f is injective.

(b) Verify that f is *not* surjective.

(c) Describe specifically how to make f bijective by **restricting the codomain**.

(d) As all bijective functions are invertible, the bijective version of f from Task c has an inverse f^{-1}. Describe this inverse by specifying its

(i) domain,

(ii) codomain, and

(iii) input-output rule.

Activity 10.6 Let $\ell\colon \Sigma^* \to \mathbb{N}$ represent the length function, using alphabet is $\Sigma = \{\alpha, \omega\}$.

(a) Compute $\ell\big(\ell^{-1}(B)\big)$ for $B = \{1, 10, 100\}$.

(b) How many elements are there in $\ell^{-1}\big(\ell(A)\big)$ for $A = \{\alpha\alpha, \alpha\omega, \omega\omega\alpha\omega\}$?

Activity 10.7 Suppose $f\colon A \to B$ is a function, and B_1,B_2 are subsets of B.

(a) Draw a Venn diagram illustrating that

$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

Include all of the sets

$$A,\ B,\ B_1,\ B_2,\ B_1 \cap B_2,\ f^{-1}(B_1),\ f^{-1}(B_2),\\ f^{-1}(B_1) \cap f^{-1}(B_2),\ \text{ and }\ f^{-1}(B_1 \cap B_2)$$

in your diagram.

(b) Formally prove that $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}B_2$, using the Test for Set Equality.

Activity 10.8 (*Note:* The parts of this question are independent of one another.) Suppose $f\colon A \to B$ and $g\colon B \to C$ are functions.

(a) Argue that if f and g are both surjective, then so is $g \circ f$.

(b) If $g \circ f$ is surjective, must g be? Must f be?

(c) Argue that if f and g are both injective, then so is $g \circ f$.

(d) If $g \circ f$ is injective, must g be? Must f be?

10.7 Exercises

1. Use predicate logic to write formal definitions of **surjective function**, **injective function**, and **bijective function**. Be sure to state the domains of your free variables.

2. Let A represent the set of all university students and let C be the set of all university courses. Does the rule $f\colon A \to C$ given by

$$f(a) = c \text{ if student is registered in course}$$

define a function? Justify your answer.

Testing bijectivity and determining inverses. In each of Exercises 3–7, determine whether or not the described function is a bijection. For those functions that are bijective, describe the inverse function; that is, specify the inverse function's

(i) domain,

(ii) codomain, and

(iii) input-output rule.

3. $\Lambda = \{\mathrm{T}, \mathrm{F}\}$, $n \colon \Lambda \to \Lambda$ is the logical negation function $n(p) = \neg p$.

4. $\mathcal{L}$ represents the set of all possible logical statements, $N \colon \mathcal{L} \to \mathcal{L}$ is the logical negation function $N(A) = \neg A$ for A a logical statement.

(*Note:* You may treat equivalent statements as being the same statement.)

5. $\mathcal{N} \colon \mathbb{Z} \to \mathbb{Z}$ is the numerical negation function $\mathcal{N}(n) = -n$.

6. $\Sigma = \{0,1\}$, Σ^* represents the set of all binary words, $c \colon \Sigma^* \to \Sigma^*$ is the *bitwise complement* function defined by: if w is a binary word, let $c(w)$ be a binary word of the same length but with a 0 at every position that w has a 1, and a 1 at every position that w has a 0. For example, $c(010) = 101$ and $c(0000) = 1111$.

7. U represents a universal set, $C \colon \mathscr{P}(U) \to \mathscr{P}(U)$ is the complement function $C(A) = A^{\mathrm{c}}$, for $A \subseteq U$.

8. Let $E \subseteq \mathbb{Z}$ represent the set of even integers, and consider the function $f \colon \mathbb{Z} \to E$, $f(n) = 2n$.

(a) Prove that f is a bijection.

(b) Describe the inverse function $f^{-1} \colon E \to \mathbb{Z}$. That is, describe the rule to determine $f^{-1}(n)$, given even number n.

9. As usual, $\mathbb{R}^m = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ represents the Cartesian product of m copies of $\mathbb{R}$, where m is a positive integer. Consider the function $D \colon \mathbb{R} \to \mathbb{R}^m$ defined by $D(x) = (x, x, \ldots, x)$.

Terminology. The function D in this exercise is called a **diagonal embedding**. We can define a similar diagonal embedding $D \colon A \hookrightarrow A^m$ for every nonempty set A.

(a) Prove that D embeds $\mathbb{R}$ into $\mathbb{R}^m$.

(b) Fill in the right-hand side of the set definition in Candidate-condition notation for the image of D below.

$$D(\mathbb{R}) = \left\{ (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \;\middle|\; \qquad\qquad\qquad \right\}$$

(c) Provide a set definition for the graph $\Delta(D)$ in Form-parameter notation. Of what set is $\Delta(D)$ a subset?

(d) Can you come up with other "natural" embeddings $\mathbb{R} \hookrightarrow \mathbb{R}^m$?

10. Let $A = \{0,1,2,3,4,5,6,7,8,9\}$ and let $P \subseteq \mathscr{P}(A)$ represent the set of all subsets of A which contain an odd number of elements. Define $v \colon P \to A$ by setting $v(X)$ to be the "middle" element of X when the elements of X are listed in order by size. For example, $v(\{0,8,9\}) = 8$.

Is v injective? Surjective? Bijective?

11. Let $\Sigma = \{0,1\}$. Recall that for $n \in \mathbb{N}$, Σ_n^* is the subset of Σ^* consisting of all binary words of length n.

Suppose $A = \{a_1, a_2, \ldots, a_n\}$ is a set with n (distinct) elements. Construct a bijection $\mathscr{P}(A) \to \Sigma_n^*$.

12. Call a function with domain $\varnothing$ an **empty function**.

(a) Verify that every empty function is injective.

Hint. Use your formal expression of **injective** from Exercise 10.7.1, along with what you learned in Section 4.3.

(b) Verify that an empty function with empty codomain is bijective.

Hint. You have already verified injectivity of an empty function more generally in Task a. For surjectivity in this more specific setting, use your formal expression of **surjective** from Exercise 10.7.1, along with what you learned in Section 4.3.

13.

(a) Prove that if f and g are both surjective, then $g \circ f$ is surjective.

(b) If $g \circ f$ is surjective, must either or both of f, g necessarily be surjective? Justify your answers.

(c) Prove that if f and g are both injective, then $g \circ f$ is injective.

(d) If $g \circ f$ is injective, must either or both of f, g necessarily be injective? Justify your answers.

(e) Task a and Task c together prove that if f and g are both bijective, then $g \circ f$ is bijective.

Prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. (See the definition of **equality of functions**.)

14.

(a) Prove that f is a bijection.

(b) Prove that $g = f^{-1}$.

Function image sets and inverse image sets. In each of Exercises 15–18, consider abstract function $f : A \to B$ and subsets $A_1, A_2 \subseteq A$, $B_1, B_2 \subseteq B$.

15.

(a) Draw a Venn diagram illustrating that $A_1 \subseteq f^{-1}\big(f(A_1)\big)$.

Include all of the sets

$$A,\ B,\ A_1,\ f(A_1),\ \text{and}\ \ f^{-1}\big(f(A_1)\big)$$

in your diagram.

(b) Formally prove that $A_1 \subseteq f^{-1}\big(f(A_1)\big)$, using the Subset Test.

(c) Devise an explicit example where $A_1 \subsetneq f^{-1}\big(f(A_1)\big)$.

16.

(a) Draw a diagram illustrating that $f\big(f^{-1}(B_1)\big) \subseteq B_1$.

Include all of the sets

$$A,\ B,\ B_1,\ f^{-1}(B_1),\ \text{and}\ \ f\big(f^{-1}(B_1)\big)$$

in your diagram.

(b) Formally prove that $f\big(f^{-1}(B_1)\big) \subseteq B_1$, using the Subset Test.

(c) Devise an explicit example where $f\big(f^{-1}(B_1)\big) \subsetneq B_1$.

17.

(a) Draw a diagram illustrating that $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$. Include all of the sets

$$A,\ B,\ A_1,\ A_2,\ A_1 \cap A_2,\ f(A_1),\ f(A_2),$$
$$f(A_1) \cap f(A_2),\ \text{ and }\ f(A_1 \cap A_2)$$

in your diagram.

(b) Formally prove that $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$, using the Subset Test.

(c) Devise an explicit example where $f(A_1 \cap A_2) \subsetneq f(A_1) \cap f(A_2)$.

18.

(a) Draw a diagram illustrating that

$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

Include all of the sets

$$A,\ B,\ B_1,\ B_2,\ B_1 \cap B_2,\ f^{-1}(B_1),\ f^{-1}(B_2),$$
$$f^{-1}(B_1) \cap f^{-1}(B_2),\ \text{ and }\ f^{-1}(B_1 \cap B_2)$$

in your diagram.

(b) Formally prove that $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$, using the Test for Set Equality.

19. Suppose $f : A \to B$ is an injection. Use f to devise an injection $F : \mathscr{P}(A) \hookrightarrow \mathscr{P}(B)$. Be sure to verify that your proposed function F is injective. If f is bijective, will F also be bijective?

CHAPTER 11

Recurrence and induction

11.1 Sequences

counting set
the set

$$\mathbb{N}_{<m} = \{\, n \in \mathbb{N} \mid n < m \,\} = \{0, 1, \ldots, m-1\}$$

The set $\mathbb{N}_{<m}$ has exactly m elements in it. In Chapter 12 we will use these counting sets to, well, *count* the elements in other sets. For now, we will use them to index the objects in an ordered list.

finite sequence (from a set A)
a function $\mathbb{N}_{<m} \to A$

infinite sequence (from a set A)
a function $\mathbb{N} \to A$

term in a sequence
one of the image elements of the function defining the sequence

a_k the k^{th} term in a sequence, so that if $f : \mathbb{N}_{<m} \to A$ or $f : \mathbb{N} \to A$ is a sequence then $a_k = f(k)$

$\{a_k\}$ the collection of all terms in a sequence

$\{a_k\}_0^m$ the collection of the terms in a sequence up to (and including) the m^{th} term (if the sequence is finite, this could represent all terms in the sequence for the appropriate m value)

$\{a_k\}_0^\infty$ the collection of all terms in a sequence, where we are explicit that it is an infinite sequence

Remark 11.1.1

- Of course, we do not restrict ourselves to the letter a to represent the terms of a sequence. We might write b_k, or s_k, etc..
- While we use set-like notation {} to represent the collection of all terms in a sequence, this collection is *not* a set, since order and repetition matter.

Example 11.1.2 Sequence of squares. The sequence $\{k^2\}$ has terms $0, 1, 4, 9, 16, 25, \ldots, k^2, \ldots$. □

Example 11.1.3 Sequence of definite integral values. The sequence

$$\left\{ (-1)^k \int_1^{k+1} \frac{dx}{x} \right\}$$

has terms $0, -\ln 2, \ln 3, -\ln 4, \ldots, (-1)^k \ln(k+1), \ldots$. □

11.2 Recurrence relations

recursively-defined sequence
a sequence $\{a_k\}$ from a set A, where $a_0, a_1, \ldots, a_{K-1}$ are defined explicitly, and for $k \geq K$, the term a_k is defined in terms of some (or all) of the previous terms in the sequence $a_0, a_1, \ldots, a_{k-1}$

recurrence relation
for a recursively-defined sequence, the formula that defines the general term a_k recursively in the previous terms $a_0, a_1, \ldots, a_{k-1}$

Example 11.2.1 A bouncing ball. A ball is dropped from a height of 100 cm. On each bounce, it returns to 75% of its previous height.

Let h_k be the height in centimetres after the k^{th} bounce. Then $h_0 = 100$ and the recurrence relation is

$$h_k = \frac{3h_{k-1}}{4}, \qquad k \geq 1.$$

The terms of the sequence are

$$100, 75, 56.25, 42.1875, \ldots.$$

□

Example 11.2.2 Factorial. Set $a_0 = 1$, and let $a_k = ka_{k-1}$ for $k \geq 1$. Then the terms of the sequence are

$$1, 1, 2, 6, 24, 120, \ldots.$$

□

Example 11.2.3 Fibonacci sequence. The sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

can be defined recursively by $a_0 = 0$, $a_1 = 1$, and

$$a_k = a_{k-1} + a_{k-2}, \qquad k \geq 2.$$

□

Example 11.2.4 A sequence of sets. Define a sequence $\{A_k\}$ from $\mathscr{P}(\mathbb{N})$ recursively as follows. Let $A_0 = \varnothing$, and take the recurrence relation to be

$$A_k = A_{k-1} \cup \{k\}, \qquad k \geq 1.$$

Then the terms of the sequence are

$$\varnothing, \{1\}, \{1,2\}, \{1,2,3\}, \ldots, \{1,2,\ldots,k\}, \ldots.$$

□

11.3 Solving through iteration

Given a recursively defined sequence $\{a_k\}$, we can "unravel" the recursive definition to determine an explicit formula for the general term a_k which involves only the index k.

Worked Example 11.3.1 Solve the recurrence relation from Example 11.2.1.

Solution. The sequence in the example was defined recursively by $h_0 = 100$ and

$$h_k = \frac{3}{4} h_{k-1}, \qquad\qquad k \geq 1.$$

We can apply this formula to every term in the sequence, except for the first, using the pattern "each term is three-quarters of the previous term." That is,

$$h_k = \frac{3}{4} h_{k-1}, \qquad h_{k-1} = \frac{3}{4} h_{k-2}, \qquad h_{k-2} = \frac{3}{4} h_{k-3}, \qquad \ldots.$$

Therefore, for $k \geq 1$, we can calculate

$$\begin{aligned} h_k &= \frac{3}{4} h_{k-1} \\ &= \frac{3}{4}\left(\frac{3}{4} h_{k-2}\right) = \left(\frac{3}{4}\right)^2 h_{k-2} \\ &= \left(\frac{3}{4}\right)^2 \left(\frac{3}{4} h_{k-3}\right) = \left(\frac{3}{4}\right)^3 h_{k-3} \\ &\vdots \\ &= \left(\frac{3}{4}\right)^k h_0 = \left(\frac{3}{4}\right)^k (100). \end{aligned}$$

(Note that this formula is also valid for $k = 0$.)

We can verify our formula by substituting it into the original recurrence relation:

$$\text{RHS} = \frac{3}{4} h_{k-1} = \frac{3}{4}\left(\frac{3}{4}\right)^{k-1} h_0 = \left(\frac{3}{4}\right)^k h_0 = h_k = \text{LHS}.$$

We could also prove our formula is correct by induction. □

Worked Example 11.3.2 Solve the recurrence relation $a_k = r a_{k-1}$ for $k \geq 1$, where r is a constant and the first term a_0 is arbitrarily chosen.

Compare. This example generalizes the previous example.

Solution. Through iteration, we obtain

$$a_k = r a_{k-1} = r(r a_{k-2}) = r\big(r(r a_{n-3})\big) = \cdots = r^k a_0.$$

□

Worked Example 11.3.3 Solve the recurrence relation from Example 11.2.2.

Solution. The sequence in the example was defined recursively by $a_0 = 1$ and

$$a_k = k a_{k-1}, \qquad\qquad k \geq 1.$$

Therefore, for $k \geq 1$, we have

$$\begin{aligned} a_k &= k a_{k-1} \\ &= k\big((k-1) a_{k-2}\big) \end{aligned}$$

$$\begin{aligned}
&= k\Big((k-1)\big((k-2)a_{k-3}\big)\Big)\\
&\vdots\\
&= k(k-1)(k-2)\cdots\big(k-(k-1)\big)a_{k-k}.
\end{aligned}$$

Simplifying this last expression leads to

$$a_k = k(k-1)(k-2)\cdots 1\cdot a_0 = k!.$$

Note that the formula $a_k = k!$ is also valid for $k = 0$ when we adopt the convention $0! = 1$. □

Checkpoint 11.3.4 Verify that the formula in the solution to the above worked example satisfies the recurrence relation.

Worked Example 11.3.5 Solve the recurrence relation $a_0 = 1$, $a_1 = \frac{1}{2}$, and

$$a_k = \frac{k}{2(k-1)}a_{k-1}, \qquad\qquad k \geq 2.$$

Solution. Iterating, we obtain

$$\begin{aligned}
a_k &= \frac{k}{2(k-1)}a_{k-1}\\
&= \frac{k}{2(k-1)}\left(\frac{k-1}{2(k-2)}a_{k-2}\right) = \frac{k}{2^2(k-2)}a_{k-2}\\
&= \frac{k}{2^2(k-2)}\left(\frac{k-2}{2(k-3)}a_{k-3}\right) = \frac{k}{2^3(k-3)}a_{k-3}\\
&\vdots\\
&= \frac{k}{2^{k-1}\big(k-(k-1)\big)}a_{k-(k-1)}.
\end{aligned}$$

Simplifying this last expression, we obtain

$$a_k = \frac{k}{2^{k-1}}a_1 = \frac{k}{2^k}.$$

□

11.4 Inductive definitions

We can use the idea of recursive definitions in a more general manner.

inductive definition
a method of defining a collection of objects, where each object in the collection can be constructed from objects assumed or already known to exist in the collection

base clause
a statement specifying some specific initial objects that belong to the inductively-defined set

inductive clause
a statement describing a means to determine new objects in the inductively-defined set from those already known to belong

limiting clause
: a declaration that no objects belong to the inductively-defined set unless obtained from a finite number of applications of the base and inductive clauses

Example 11.4.1 Set of all possible logical statements. Let us define a set $\mathcal{L}$ with the following inductive definition.

Base clause. For every $m \in \mathbb{N}$, the statement variable p_m belongs to $\mathcal{L}$.

Inductive clause. Given statements $A, B \in \mathcal{L}$, the statements

$$\neg A, \quad A \wedge B, \quad A \vee B, \quad A \rightarrow B, \quad A \leftrightarrow B$$

are also elements of $\mathcal{L}$.

Limiting clause. The set $\mathcal{L}$ does not contain any elements except those that can be obtained from a finite number of applications of the base and inductive clauses. For example, the logical statement $\neg p_2 \wedge (p_1 \rightarrow p_2)$ is in $\mathcal{L}$ by the following construction.

$$p_1, p_2 \in \mathcal{L} \quad \Rightarrow \quad \neg p_2 \in \mathcal{L},\ p_1 \rightarrow p_2 \in \mathcal{L} \quad \Rightarrow \quad \neg p_2 \wedge (p_1 \rightarrow p_2) \in \mathcal{L}$$

□

Example 11.4.2 Set-theoretic construction of the natural numbers. We assume that an empty set $\varnothing$ exists. Let us define a set N inductively.

Base clause. The empty set $\varnothing$ is an *element* of N.

Inductive clause. If X is an element of N and X itself is a set, then the set $X^+ = X \cup \{X\}$ is also an element of N.

Limiting clause. The set N does not contain any other elements except those that can be obtained from a finite number of applications of the base and inductive clauses. Note that the three clauses together imply that every element of N must be a set, so the "and X itself is a set" part of the inductive clause is superfluous.

Since the base clause involves a single initial element of N and the inductive clause produces one new element of N from a single old element of N, we can explicitly carry out the construction step-by-step. We now *define* the natural numbers to be the elements in this construction:

$$\begin{aligned}
&0 = \varnothing, \\
&1 = 0^+ = 0 \cup \{0\} = \varnothing \cup \{0\} = \{0\} \neq 0, \\
&2 = 1^+ = 1 \cup \{1\} = \{0\} \cup \{1\} = \{0,1\} \neq 0,1, \\
&3 = 2^+ = 2 \cup \{2\} = \{0,1\} \cup \{2\} = \{0,1,2\} \neq 0,1,2 \\
&\quad\vdots
\end{aligned}$$

We usually write $\mathbb{N}$ for this set instead of N.

Note that the number of elements in each natural number (as a set) is equal to the number defined by that set, and that each natural number m is defined to be the set that we have previously called $\mathbb{N}_{<m}$. □

Bonus. In Example 11.4.2 above, we constructed the set $\mathbb{N}$ inductively using only the axioms of set theory. But how do we do arithmetic with this definition? We can define addition as an infinite collection of inductively-defined functions: for each $m \in \mathbb{N}$, define a "sum with m" function $s_m : \mathbb{N} \rightarrow \mathbb{N}$ as follows.

Base clause. Set $s_m(0) = m$.

Inductive clause. For $n \in \mathbb{N}$ such that $s_m(n)$ is already defined, set

$$s_m(n^+) = s_m(n)^+ = s_m(n) \cup \{s_m(n)\}.$$

That is, if $s_m(n)$ is defined and n^+ is the next natural number after n in the inductive definition of $\mathbb{N}$, then define $s_m(n^+)$ to be the next natural number after $s_m(n)$. We then use the symbols $m+n$ to mean $s_m(n)$. In this notation, you can think of the inductive clause above as saying that once $m+n$ is defined, we can define $m+(n+1)$ as $(m+n)+1$.

If you are bored on a Friday or Saturday night, you can try the following using the above definition of addition in $\mathbb{N}$.

Checkpoint 11.4.3

1. Prove that addition in $\mathbb{N}$ is:

   (a) **commutative**: $m+n = n+m$, i.e. $s_m(n) = s_n(m)$ for all $m, n \in \mathbb{N}$; and

   (b) **associative**: $(m+n)+\ell = m+(n+\ell)$, i.e. $s_{s_m(n)}(\ell) = s_m\big(s_n(\ell)\big)$, for all $m, n, \ell \in \mathbb{N}$.

2. Use the idea that every positive integer should have a negative to define $\mathbb{Z}$ as a subset of the Cartesian product $\mathbb{N} \times \mathbb{N}$. Then define addition and subtraction in $\mathbb{Z}$.

Hint. To define $\mathbb{Z}$, first choose an appropriate one-to-one function embedding $\mathbb{N}$ into $\mathbb{N} \times \mathbb{N}$ in such a way that will then allow you to attach an additional second piece of information to each natural number (namely, a designator of the sign of the number).

11.5 Activities

Activity 11.1 Develop an inductive definition of the set of words Σ^* from the alphabet $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$.

Then verify that the word ccababb is in the set by tracing it back to the base clause.

Hint. Steps:

(i) Think of a simple way to form new words from old (inductive clause).

(ii) Then think about the basic words you need to get the process started (base clause).

(iii) Finally, decide whether you are certain you can form *every possible* word in a *finite* number steps starting at some base word.

Activity 11.2 Let $\Sigma = \{\mathrm{a}, \mathrm{z}\}$. Write an inductive definition for the set of words in Σ^* that have the same number of a letters as z letters.

11.6 Exercises

1. Compute each of the terms s_2, s_3, s_4, s_5, s_6 for the sequence defined recursively by

$$s_n = \sqrt{s_{n-2}^2 + s_{n-1}^2}, \quad n \geq 2,$$

with initial terms $s_0 = 3$ and $s_1 = 4$.

Solving by iteration. In each of Exercises 2–8, use iteration to determine an expression for the n^{th} term of the sequence as a formula in n (and the initial term(s) of the sequence, if necessary).

In some of these, you may find the following formulas useful.

$$1+2+3+\cdots+m=\frac{m(m+1)}{2}$$
$$1^2+2^2+3^2+\cdots+m^2=\frac{m(m+1)(2m+1)}{6}$$
$$r^0+r^1+r^2+\cdots+r^{m-1}=\frac{r^m-1}{r-1}, \quad r\neq 0,1$$

2. $a_n=2na_{n-1}$, $a_0=1$.

3. $a_n=(2n-1)a_{n-1}$, $a_0=1$.

4. $a_n=a_{n-1}+3^{n-1}$, $a_0=1$.

5. $a_n=a_{n-1}+n-1$, $a_0=1$.

6. $a_n=a_{n-1}+n+n^2$, $a_0=1$.

7. $a_n=f(a_{n-1})$, where $f(x)$ is the linear function $f(x)=mx+b$ for some fixed constants m,b, and with arbitrary initial term a_0.

8. $a_n=4a_{n-2}$, $n\geq 2$, $a_0=1$, $a_1=2$.

Hint. Treat the cases n even and n odd separately.

9. **Fibonacci numbers** are those that appear in the sequence defined recursively by

$$a_n=a_{n-1}+a_{n-2}, \qquad n\geq 2,$$

for some choice of initial terms a_0,a_1.

See. Example 11.2.3.

Using initial terms $a_0=a_1=1$, use mathematical induction to prove that every Fibonacci number a_n satisfies $a_n<2^n$ (except, of course, for a_0.

10. You are attempting to predict population dynamics on a yearly basis.

Suppose a population increases by a factor of i each year. That is, if we set $p=100i$, then the population increases by p percent. (***Careful:*** This is a description of the *increase* in population, not the total population. For example, $i=1$ means that the population doubles.)

(a) Write down a recurrence relation that expresses the population P_n in the n^{th} year relative to the previous year.

(b) Use iteration to determine an expression for the population in the n^{th} year as a formula in n, i, and the initial population P_0.

(c) Suppose that on top of the natural population increase of i percent per year, immigration increases the population by fixed amount A people annually. Design a new recurrence relation for P_n, and use iteration to determine an expression for the population in the n^{th} year as a formula in n, i, A, and the initial population P_0.

11. Explicitly describe how to construct the following logical statement in a finite number of steps using the inductive definition for $\mathcal{L}$, the set of all possible logical statements, given in Example 11.4.1.

$$(p_1\wedge p_2)\to\big((\neg p_3\vee p_1)\leftrightarrow(p_3\wedge\neg p_2)\big)$$

12. The set $\mathcal{C}$ of *constructible numbers* can be defined inductively as follows.

Base clause. Assume $1 \in \mathcal{C}$.

Inductive clauses.

(i) Whenever $a, b \in \mathcal{C}$, then so are

$$a+b, \quad ab, \quad a/b, \quad \sqrt{a}.$$

(ii) Whenever $a, b \in \mathcal{C}$ with $a > b$, then $a-b$ is also in $\mathcal{C}$.

Limiting clause. The set $\mathcal{C}$ contains no elements other than those that can be obtained through a finite number of applications of the base and/or inductive clauses. Explicitly verify, by listing each application of the relevant clauses, that the roots of the polynomial $2x^2 - 3x + \frac{7}{8}$ are both constructible numbers.

13. Consider the following inductively defined set $A \subseteq \mathbb{N}$.

Base clause. Assume $32879 \in A$.

Inductive clauses.

(i) When a is an element of A, then each of the prime factors of a is also an element of A.

(ii) Whenever prime p is an element of A, then $p+1$ is also an element of A.

Limiting clause. The set A contains no elements other than those that can be obtained through a finite number of applications of the base and/or inductive clauses. Determine all elements of A.

Hint. To help with this question, you may wish to search for “list of small primes” on the internet.

14. Devise an algorithm that will produce an answer to the following question in a finite number of applications of the inductive clause that we used to define the natural numbers in Example 11.4.2.

Given $m, n \in \mathbb{N}$ with $m \neq n$, is $m > n$ or is $n > m$?

CHAPTER 12

Cardinality

In this chapter we will discuss how to measure and compare "sizes" of sets using bijections.

12.1 Finite sets

Recall. For $m \in \mathbb{N}$ we have defined the **counting set**

$$\mathbb{N}_{<m} = \{\, n \in \mathbb{N} \mid n < m \,\} = \{0, 1, \dots, m-1\}.$$

Clearly, $\mathbb{N}_{<m}$ contains exactly m elements. In fact, we have *defined* the number m to be the set $\mathbb{N}_{<m}$. (See Example 11.4.2.)

As the terminology implies, we will use these sets to count the elements of other sets. In particular, given a set A, if we can match up the elements of A with the elements of $\mathbb{N}_{<m}$, one for one, then A must also contain exactly m elements.

finite set a set A for which there exists a bijection $\mathbb{N}_{<m} \to A$ for some $m \in \mathbb{N}$, $m > 0$

Fact 12.1.1 Uniqueness of counting number. *For finite set A there exists* one unique *natural number m for which a bijection $\mathbb{N}_{<m} \to A$ exists.*

Remark 12.1.2 Suppose A is finite. While there is only one number m for which a bijection $\mathbb{N}_{<m} \to A$ exists, there can be many such bijections, and the number of bijections increases as m increases.

Checkpoint 12.1.3 Prove Fact 12.1.1.

cardinality (of a finite set A)
the unique natural number m for which a bijection $\mathbb{N}_{<m} \to A$ exists

$|A|$ the cardinality of the finite set A

$\operatorname{card} A$ alternative notation for the cardinality of the finite set A

$\#\{\dots\}$ alternative notation for the cardinality of the set defined by $\{\dots\}$

Example 12.1.4 For $\Sigma = \{\mathrm{a}, \mathrm{b}, \dots, \mathrm{z}\}$, we have $|\Sigma| = 26$. Below are two example bijections $\varphi, \psi \colon \mathbb{N}_{<26} \to \Sigma$ that verify this cardinality number.

σ	0	1	2	3	$\cdots$	24	25
$\varphi(\sigma)$	a	b	c	d	$\cdots$	y	z
$\psi(\sigma)$	a	z	b	y	$\cdots$	m	n

Figure 12.1.5 Bijections $\varphi, \psi \colon \mathbb{N}_{<26} \to \Sigma$ defined by a table of values.

□

Cardinality of an empty set. What about the empty set? Clearly we should have $|\varnothing| = 0$. But is this consistent with our definition of cardinality?

empty function
a function with domain $\varnothing$

If we accept the existence of an **empty function** $\varnothing \to X$ for every set X, then the properties of such functions that we need in order to establish $|\varnothing| = 0$ will be vacuously true.

Proposition 12.1.6 Properties of empty functions.

1. *For every set X, an empty function $\varnothing \to X$ is injective.*
2. *An empty function $\varnothing \to \varnothing$ is a bijection.*

Proof. You were asked to verify these statements in Exercise 10.7.12. ■

Corollary 12.1.7 *The cardinality of the empty set is* 0.

Proof. We are required to demonstrate an example of a bijection $\mathbb{N}_{<0} \to \varnothing$. But

$$\mathbb{N}_{<0} = \{\, n \in \mathbb{N} \mid n < 0 \,\} = \varnothing,$$

so Statement 2 of Proposition 12.1.6 tells that the empty function $\mathbb{N}_{<0} \to \varnothing$ is indeed a bijection. ■

12.2 Properties of finite sets and their cardinality

12.2.1 Finite sets versus finite sequences

Recall that a function $f \colon \mathbb{N}_{<n} \to A$ defines a **finite sequence** of elements from the set A, by setting

$$a_0 = f(0), \qquad a_1 = f(1), \qquad a_2 = f(2), \qquad \ldots, \qquad a_{n-1} = f(n-1).$$

If A is finite, then there exists such a function f that is **bijective**, which leads to the following.

Fact 12.2.1 Characterization of finiteness using sequences. *A set A is finite if and only if there exists a finite sequence from A in which each element of A appears* exactly *once.*

Remark 12.2.2

1. The above fact makes an important connection between functions and counting. If A is a finite set, $f \colon \mathbb{N}_{<n} \to A$ is a corresponding bijection, and we create sequence $\{a_k\}_{k=0}^{m}$ with $a_k = f(k)$ as above, then we are able to list the elements of A in sequence:

   $$A = \{a_0, a_1, \ldots, a_{n-1}\}.$$

   In turn, writing the elements of A in sequence is really just a way of *counting* them, in a manner that is roughly equivalent to counting on your fingers (if you had a lot of fingers). In fact, counting the elements of A **totally orders** the elements of A (a concept we will meet in a future chapter). In Chapter 13, we will adapt this connection between functions and counting to determine whether it is possible to "count" infinite sets.

2. We should note, however, that the above fact is essentially trivial once we "unravel" the definitions of **finite set** and **finite sequence**, as both involve a function with domain $\mathbb{N}_{<m}$ for some m and codomain A.

Corollary 12.2.3 *A set A is finite if and only if there exists a finite sequence from A which contains each element of A* at least *once.*

Proof idea. If we have a sequence that contains each element of A at least once, we could turn it into a sequence that contains each element of A *exactly* once by removing repeated terms. ■

12.2.2 Finite sets versus bijections, subsets, and unions

Bijections compose to create bijections (see Exercise 10.7.13). This fact lets us relate finite sets to each other.

Fact 12.2.4 Bijection implies same cardinality. *If one of A,B is finite and there exists a bijection $f: A \to B$, then* both *are finite and $|A| = |B|$.*

Proof. Reconsider "one of A,B finite" as a disjunction: "A is finite or B is finite". Then break into cases.

Assume A is finite. Suppose $|A| = n$, so that there exists a bijection $g: \mathbb{N}_{<n} \to A$. Then $f \circ g: \mathbb{N}_{<n} \to B$ is also a bijection, so $|B| = n$.

Assume B is finite. Suppose $|B| = n$. Repeat the argument from the previous case, swapping roles of A and B and using the bijection $f^{-1}: B \to A$ in place of f. ■

Fact 12.2.5 Subset of finite is finite. *Assume B is a finite set.*

1. *Every subset $A \subseteq B$ is finite, with $|A| \le |B|$.*

2. *If $f: C \to B$ is an injection, then C is finite with $|C| \le |B|$.*

Proof idea.

1. This is left to you as Exercise 12.6.1.

2. Let A represent the image $f(C)$. Then $A \subseteq B$, so we can apply Statement 1 and Fact 12.2.4.

■

We may also relate cardinality of finite sets to the union operation.

Fact 12.2.6 Cardinality of unions. *Suppose A and B are finite subsets of a universal set U.*

1. *If A and B are disjoint, then $|A \sqcup B| = |A| + |B|$.*

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

Proof idea. The idea behind these formulas should be obvious once you draw appropriate Venn diagrams, but formal proofs are left to you in Exercise 12.6.2. ■

12.2.3 Cardinality of power sets of finite sets

Worked Example 12.2.7 What is the cardinality of $\mathscr{P}(\{1,2,3,\ldots,k\})$?

Solution. We can solve this using recursion! In Example 11.2.4, we defined the following sequence of subsets of $\mathbb{N}$,

$$\begin{aligned} A_0 = \varnothing, \quad A_1 = \{1\}, \quad A_2 = \{1,2\}, \quad A_3 = \{1,2,3\}, \\ \ldots, \quad A_k = \{1,2,\ldots,k\}, \quad \ldots, \end{aligned}$$

recursively. We can also express the sequence $N_k = |\mathscr{P}(A_k)|$ recursively. First, $N_0 = 1$. Then, since

$$A_{k-1} = \{1,2,\ldots,k-1\} \subsetneq \{1,2,\ldots,k-1,k\} = A_k,$$

we can consider $\mathscr{P}(A_{k-1}) \subseteq \mathscr{P}(A_k)$. (See Exercise 9.9.13.) In doing so, we can break $\mathscr{P}(A_k)$ into the disjoint union

$$\mathscr{P}(A_k) = \mathscr{P}(A_{k-1}) \sqcup \mathscr{P}(A_{k-1})^{\mathrm{c}}.$$

Notice that the elements of $\mathscr{P}(A_{k-1})$ are precisely those subsets of A_k that *do not* contain the element k, and therefore the elements of $\mathscr{P}(A_{k-1})^{\mathrm{c}}$ are precisely those subsets of A_k that *do* contain the element k. So

$$B \in \mathscr{P}(A_{k-1}) \quad \Rightarrow \quad B \cup \{k\} \in \mathscr{P}(A_{k-1})^{\mathrm{c}}.$$

This correspondence actually gives us a bijection

$$\begin{aligned} \mathscr{P}(A_{k-1}) &\to \mathscr{P}(A_{k-1})^{\mathrm{c}}, \\ B &\mapsto B \cup \{k\}. \end{aligned}$$

(Check!)

Now we have

$$N_{k-1} = |\mathscr{P}(A_{k-1})| = \left|\mathscr{P}(A_{k-1})^{\mathrm{c}}\right|.$$

Since

$$\mathscr{P}(A_k) = \mathscr{P}(A_{k-1}) \sqcup \mathscr{P}(A_{k-1})^{\mathrm{c}},$$

we then have

$$N_k = N_{k-1} + N_{k-1} = 2N_{k-1},$$

a recursively defined sequence. Solving this recurrence relation by iteration yields

$$N_k = 2^k.$$

□

Checkpoint 12.2.8 Verify that the map

$$\begin{aligned} \mathscr{P}(A_{k-1}) &\to \mathscr{P}(A_{k-1})^{\mathrm{c}}, \\ B &\mapsto B \cup \{k\}. \end{aligned}$$

in the solution to the preceding Worked Example is a bijection.

We can use the idea of Worked Example 12.2.7 to prove a similar but more general fact.

Theorem 12.2.9 Cardinality of a power set. *If $|A| = n$, then $|\mathscr{P}(A)| = 2^n$.*

Proof idea. Since A has the same cardinality as the set $\{1,2,3,\ldots,n\}$, there exists a bijection between the two sets. In Exercise 12.6.8, you are asked to prove that two sets of the same cardinality also have power sets of the same cardinality. Using this fact and the result of Worked Example 12.2.7, we have

$$|\mathscr{P}(A)| = |\mathscr{P}(\{1,2,3,\ldots,n\})| = 2^n.$$

■

12.2.4 Infinite sets

infinite set

	a set that is not finite
$\lvert A\rvert = \infty$	set A is infinite
$\lvert A\rvert < \infty$	set A is finite

To demonstrate that a set A is infinite using the technical definition, we must demonstrate that no bijection $\mathbb{N}_{<m} \to A$ can exist, for every cardinality value m. But if A *is* infinite, there will be many *injective* functions $\mathbb{N}_{<m} \hookrightarrow A$ for each m. Therefore, one must demonstrate that no *surjection* $\mathbb{N}_{<m} \twoheadrightarrow A$ can exist, for every m.

Example 12.2.10 Demonstrating that a set is infinite from the definition. Suppose we have an alphabet consisting of a single letter $\Sigma = \{\text{x}\}$. Then the set of words

$$\Sigma^* = \{x, xx, xxx, \ldots\}$$

is infinite.

To verify this, we will argue that no function $\mathbb{N}_{<m} \to \Sigma^*$ can be surjective, no matter the cardinality value m. So suppose $m \in \mathbb{N}$ is fixed but arbitrary, and $f : \mathbb{N}_{<m} \to \Sigma^*$ is an arbitrary function. Following the Test for Surjectivity (which also describes how to demonstrate that a function is *not* surjective), we must exhibit an example element in Σ^* that is *not* the image under f of any domain element in $\mathbb{N}_{<m}$.

Function f defines a finite sequence of words from Σ^*:

$$w_0, w_1, w_2, \ldots, w_{m-1},$$

where each w_j is the image $f(j)$. We have two cases.

Each w_j is the empty word. In this case, clearly f cannot be surjective since the word consisting of the single letter x is not in the sequence of outputs for f.

Otherwise. In this case, consider the word we get by concatenating the words $w_0, w_1, \ldots, w_{m-1}$ together *twice over*:

$$w = w_0 w_1 w_2 \cdots w_{m-1} w_0 w_1 w_2 \cdots w_{m-1}.$$

(Note that this is not multiplication, we are just writing the words one after the other to create one big word.) Then this word is certainly longer than any of the individual words w_j, and so cannot be equal to one of those words. (The reason we have concatenated all the w_j *twice over* is so that we don't have to separately consider the case that *all but one* of the w_j is empty, since in that case concatenating all the w_j just once over wouldn't actually produce a result that is longer than that one non-empty w_j.) Since this long word is not in the sequence of image elements for f, the function f cannot be surjective. □

Remark 12.2.11 While we have no hope of demonstrating that a set A is infinite by demonstrating that functions $\mathbb{N}_{<m} \to A$ cannot be injective, if we wish we *can* argue using injectivity by just turning things around. If a bijection $\mathbb{N}_{<m} \to A$

were possible, its inverse $A \to \mathbb{N}_{<m}$ would also be a bijection. So another way to demonstrate that a set A is infinite is to demonstrate that no injection $A \to \mathbb{N}_{<m}$ is possible, for every cardinality number m.

We can also demonstrate that a set is infinite by relating it to known infinite sets.

Fact 12.2.12 Contains infinite subset implies infinite. *Assume A is an infinite set.*

1. *Every set B that contains A as a subset (i.e. $B \supseteq A$) is infinite.*

2. *If $f: A \to B$ is an injection, then B is also infinite.*

Proof.

1. This is left to you as Exercise 12.6.4.

2. This is just the contrapositive of Statement 2 of Fact 12.2.5.

■

Worked Example 12.2.13 Show that if $\Sigma \neq \varnothing$, then $|\Sigma^*| = \infty$, regardless of whether $|\Sigma| < \infty$ or $|\Sigma| = \infty$.

Solution. If $\Sigma \neq \varnothing$, then there exists some $x \in \Sigma$. Consider the restricted alphabet $\Xi = \{x\}$. In Example 12.2.10, we demonstrated that Ξ^* was infinite. Clearly $\Xi \subseteq \Sigma$ implies $\Xi^* \subseteq \Sigma^*$, so applying Statement 1 of Fact 12.2.12 we may conclude that Σ^* is also infinite. □

12.3 Relative sizes of sets

We have defined a set A to be finite when we can count its elements by matching them bijectively with the elements of some counting set $\mathbb{N}_{<m}$. And in this case, by defining $|A| = m$, we are declaring that A has the same "size" as $\mathbb{N}_{<m}$.

Expanding on this idea, we can think of *every* bijection as using the elements of one set to "count" the elements of another.

same size sets A and B for which there exists a bijection $A \to B$

Fact 12.3.1 Symmetry of size. *If B has the same size as A, then A has the same size as B.*

Proof. If $f: A \to B$ is a bijection, then so is $f^{-1}: B \to A$. ■

Fact 12.3.2 Transitivity of size. *If A has the same size as B and B has the same size as C, then A has the same size as C.*

Proof. This is left to you as Exercise 12.6.5. ■

We expect our general notion of **same size** to match with just counting elements of finite sets and getting the same result.

Fact 12.3.3 Finite sets with equal cardinality have the same size. *Assume A and B are finite sets. Then $|A| = |B|$ if and only if A and B have the same size.*

Proof.

Assume equal cardinality, show same size. Assume $|A| = |B| = m$. Then by definition there exist bijections $f: \mathbb{N}_{<m} \to A$ and $g: \mathbb{N}_{<m} \to B$. Now $g \circ f^{-1}$ is a bijection $A \to B$, so A and B have the same size according to the technical definition.

Assume same size, show equal cardinality. Assume A and B have the same size. Then by definition there exists a bijection $f\colon A \to B$. Now, we have also assumed that A is finite, so there exists a bijection $g\colon \mathbb{N}_{<m} \to A$, where $m = |A|$. Then $f \circ g\colon \mathbb{N}_{<m} \to B$ is a bijection that demonstrates $|B| = m$ as well. ■

Warning 12.3.4 Your intuition may fail you when considering "sizes" of infinite sets. In particular, it is possible to have $|A| = |B| = \infty$, where A and B *do not* have the same size.

Example 12.3.5 Sets of integers and natural numbers have the same size. Even though $\mathbb{N} \subsetneqq \mathbb{Z}$, $\mathbb{N}$ and $\mathbb{Z}$ have the *same size*! The following defines a bijection $f\colon \mathbb{N} \to \mathbb{Z}$.

n	0	1	2	3	4	$\cdots$
$f(n)$	0	-1	1	-2	2	$\cdots$

This bijection can be expressed by the formula

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

□

Example 12.3.6 Sets of real numbers and natural numbers do not have the same size. In Chapter 13, we will see that even though $|\mathbb{N}| = |\mathbb{R}| = \infty$, the sets $\mathbb{N}$ and $\mathbb{R}$ *do not* have the same size! □

Example 12.3.7 Intervals of real numbers of different lengths have the same size. Recall from first-year calculus that for $a, b \in \mathbb{R}$ with $a < b$, we define the **open interval** from a to b to be the set of all real numbers *strictly* between a and b:

$$(a,b) = \{\, x \in \mathbb{R} \mid a < x < b \,\}.$$

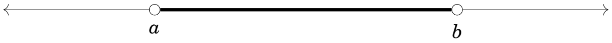


Figure 12.3.8 An interval on the real number line.

It turns out that, even though they may have different **lengths**, the interval (a,b) and the unit interval $(0,1)$ have the *same size*! (That is, they somehow contain the same "number" of numbers.)

Construct a bijection $(0,1) \to (a,b)$ in two steps.

1. The map
$$\begin{aligned} f\colon (0,1) &\to (0, b-a), \\ x &\mapsto (b-a)x, \end{aligned}$$
is a bijection. (Check!)
2. The map
$$\begin{aligned} g\colon (0,b-a) &\to (a,b), \\ x &\mapsto x+a, \end{aligned}$$
is a bijection. (Check!)

Then $g \circ f\colon (0,1) \to (a,b)$ is a bijection.

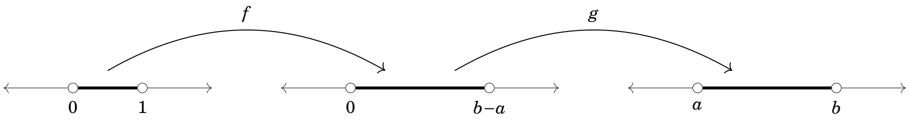


Figure 12.3.9 Scaling and translating the unit interval onto another interval.

□

Example 12.3.10 A punctured circle has the same size as $\mathbb{R}$. Define

$$S = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; x^2 + (y - \frac{1}{2})^2 = \frac{1}{4} \right\}, \qquad \hat{S} = S \setminus \{(0,1)\}.$$

Here, S is a circle in the plane with radius $\frac{1}{2}$ and centre $(0, \frac{1}{2})$, and $\hat{S}$ is the circle S “punctured” at the “north pole”.

We claim that $\hat{S}$ has the same size as $\mathbb{R}$. Construct a bijection $\hat{S} \to \mathbb{R}$ in two steps.

1. Let X represent the x-axis in the plane, i.e.

   $$X = \{ (x,0) \mid x \in \mathbb{R} \} \subseteq \mathbb{R}^2.$$

   Let $f \colon \hat{S} \to X$ be defined as follows: for $(x,y) \in \hat{S}$, let $f(x,y)$ be the x-intercept of the line through points $(0,1), (x,y)$.

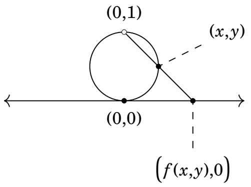
   

   Figure 12.3.11 Projecting the punctured circle onto the real number line.

   Then f is a bijection. (Check!)

2. We also have a bijection $g \colon X \to \mathbb{R}$ by $g(x,0) = x$.

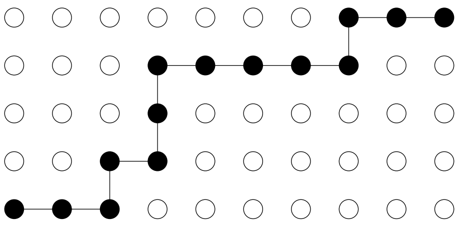
Therefore, the composition $g \circ f \colon \hat{S} \to \mathbb{R}$ is a bijection. □

Example 12.3.12 Every interval of real numbers has the same size as the entire set of real numbers. Example 12.3.7 and Example 12.3.10 can be combined to demonstrate that every finite-length interval (a,b) of real numbers has the *same size* as the entire set $\mathbb{R}$ of real numbers. See Exercise 12.6.6. □

12.4 Counting elements of finite sets with bijections

In a future chapter, we will begin learning how to count complicated collections by counting the “choices” needed to determine an arbitrary element in the collection. In this section, we look at how to count collections by finding a collection of the *same size* which is easier to count.

Example 12.4.1 Counting paths with words. Consider paths in the 5×10 grid below that start at the bottom left and end at the top right, and only move up or right at each step. (One such path is drawn in.) How many such paths are there?



Let P represent the set of all such paths. We can distinguish each element of P by the sequence of directions it takes at each step. Let $\Sigma = \{R, U\}$, where R and U stand for the directions "right" and "up", respectively. Then for each path $p \in P$ we can build a word $w_p \in \Sigma^*$ by setting the letters of w_p to correspond to the steps in the path. For example, the path in the diagram above would correspond to the word $RRURUURRRRURR$.

This assignment of words in Σ^* to paths in P is a function! Let's call it $f : P \to \Sigma^*$, and set $W = f(P)$, the image of f in Σ^*. This function is clearly one-to-one, as different paths must produce different words of direction indicators. And since every function maps its domain surjectively onto its image, we obtain a bijection $f : P \to W$ by restricting the codomain. Therefore, we can count the paths in P by instead counting the words in W!

Since each path in P takes exactly 13 steps, exactly 4 of which must be upwards and exactly 9 of which must be to the right, we see that W consists precisely of those words in Σ^* that have length 13 and contain exactly 4 Us and 9 Rs. Once we learn some basic counting techniques later in the course, you will be able to come back to this example to verify that

$$|P| = |W| = 715.$$

□

12.5 Activities

Activity 12.1 Use the definition of **cardinality** to verify that $|A| = 8$ for

$$A = \{1, 2, 4, 8, 16, 32, 64, 128\}.$$

Activity 12.2 The steps below will guide you through a proof of the following statement.

> If B is finite and $A \subseteq B$, then A is also finite.

(a) Start by assuming that B is finite. Write out what this means. (You may do this using the technical definition of **finite set**, or you may do this using the sequence characterization of finiteness in Fact 12.2.1.)

(b) Now add the assumption that $A \subseteq B$. Try to use your technical expression of the assumption "B is finite" from Task a to determine a similar technical expression of the desired conclusion "A is finite."

Activity 12.3 Use the sequence characterization of finiteness in Fact 12.2.1 to prove the following statement.

> If A and B are finite and do not intersect, then $|A \sqcup B| = |A| + |B|$.

Hint. Use separate finite sequences to "count" the elements of A and B. Then use these two sequences to build a sequence that "counts" the elements of $A \sqcup B$.

Activity 12.4 In each of the following, demonstrate that the two sets satisfy the technical definition of **same size** by explicitly describing a bijection between them.

(a) The set of natural numbers and the set of positive natural numbers.

(b) The set of natural numbers and the set of natural numbers that are greater than 9,999,999.

(c) The set of even natural numbers and the set of odd natural numbers.

(d) The set of even natural numbers and the set of natural numbers.

(e) The set of natural numbers and the set of integer powers of 2.

Activity 12.5 Set

$$A = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}\}, \qquad \Sigma = \{\mathrm{Y}, \mathrm{N}\}.$$

(a) Demonstrate that $\mathscr{P}(A)$ and Σ_5^* have the same size. (Recall that Σ_5^* means the set of words in Σ^* of length exactly 5.) Do this not by determining the cardinality of each of the two sets, but by showing that the sets satisfy the technical definition of **same size**. As in Activity 12.4, this will require that you explicitly describe a bijection between the two sets.

Hint. Think of a 5-letter word in the alphabet Σ as the answers to five yes-or-no questions. How does such a string of answers correspond to some subset of A?

(b) Describe how you could use the bijection you set up in Task a to turn the problem of counting the number of subsets of A that have exactly 3 elements into a problem of counting a related collection of words in the alphabet Σ.

(*Note:* You are not asked to actually determine the number of such subsets. You are only asked to describe *how* the result of Task a can be adapted to this counting problem.)

12.6 Exercises

1. *Prove:* If B is finite and $A \subseteq B$, then A is finite and $|A| \leq |B|$.

2. Suppose that A, B, and C are finite subsets of a universal set U.

(a) *Prove:* If A and B are disjoint, then $|A \sqcup B| = |A| + |B|$.

(b) *Prove:* $|A \cup B| = |A| + |B| - |A \cap B|$.

Hint. See Exercise 9.9.5, and use the equality from Task a.

(c) Determine a similar formula for $|A \cup B \cup C|$.

Hint. Draw a Venn diagram first.

3. Use induction to prove directly that if $|A| = n$ then $|\mathscr{P}(A)| = 2^n$. Use Worked Example 12.2.7 as a model for your proof of the induction step.

4. *Prove:* If $|A| = \infty$ and $A \subseteq B$, then $|B| = \infty$.

5. Prove Fact 12.3.2.

6. Combine Example 12.3.7 and Example 12.3.10 to verify that the unit interval $(0,1)$ and $\mathbb{R}$ have the same size.

Hint. First map the punctured circle $\hat{S}$ onto some open interval in the x-axis by "unrolling" $\hat{S}$.

7. Use Example 12.3.7 and the function $f(x) = \tan x$ to prove that the interval $(-\pi/2, \pi/2)$ and $\mathbb{R}$ have the same size.

Hint. The function $f(x) = \tan x$ is not one-to-one, but it becomes one-to-one if you restrict its domain to an appropriate interval

8. Prove that if A and B have the same size, then so do $\mathscr{P}(A)$ and $\mathscr{P}(B)$.

Hint. See Exercise 10.7.19.

9. Suppose A is a set with $|A| = n$. Then we can enumerate its elements as $A = \{a_1, a_2, \ldots, a_n\}$.

(a) Construct a bijection from the power set of A to the set of words in the alphabet $\Sigma = \{T, F\}$ of length n.

Note that there are two tasks required here.

(i) Explicitly describe a function $f : \mathscr{P}(A) \to \Sigma_n^*$ by describing the input-output rule: give a detailed description of how, given a subset $B \subseteq A$, the word $f(B)$ should be produced.

(ii) Prove that your function f is a bijection.

Hint. When determining the input-output rule for your function $f : \mathscr{P}(A) \to \Sigma_n^*$, think of how one might construct an arbitrary subset of A, and then relate that process to a sequence of answers to n true/false questions.

(b) Use Task a to determine the cardinality of $\mathscr{P}(A)$. Explain.

Hint. See Note 1.3.1.

(c) Suppose k is some fixed (but unknown) integer, with $0 \leq k \leq n$. Let $\mathscr{P}(A)_k$ represent the subset of $\mathscr{P}(A)$ consisting of all subsets of A that have exactly k elements. Describe how your bijection from Task a, could be used to count the elements of $\mathscr{P}(A)_k$.

Hint. Consider how restricting the domain might help.

Chapter 13

Countable and uncountable sets

13.1 Basics and examples

If A is a set that has the same size as $\mathbb{N}$, then we can think of a bijection $\mathbb{N} \to A$ as "counting" the elements of A (even though there are an infinite number of elements to count), in exactly the same way that we use our counting sets $\mathbb{N}_{<m}$ to count finite sets.

countable
a set that is finite or has the same size as $\mathbb{N}$

countably infinite
a countable set which has the same size as $\mathbb{N}$

uncountable
a set that is not countable

Note 13.1.1

1. An uncountable set is necessarily infinite.
2. Two sets which have the same size (i.e. there exists a bijection between them) are either both countable or both uncountable.

Fact 13.1.2 Characterization of countable sets using sequences. *A nonempty set A is countable if and only if there exists a sequence of elements from A in which each element of A appears exactly once.*

Proof idea. In case A is finite, the statement is precisely that of Fact 12.2.1. So assume A is infinite, in which case a sequence of the type described in the statement must also be infinite. Technically, an infinite sequence from A is a function $\mathbb{N} \to A$. The "each element of A" property is the same as saying the function is surjective, and the "exactly once" property is the same as saying the function is injective. So a sequence of the described kind is exactly the same as *bijection* $\mathbb{N} \to A$, which is what is required for A to be the same size as $\mathbb{N}$ (i.e. countably infinite). ■

Remark 13.1.3 Compare this fact with Fact 12.2.1.

Theorem 13.1.4 Countability of integers and rationals. *Sets $\mathbb{Z}$ and $\mathbb{Q}$ are both countable.*

Proof idea. We have already constructed a bijection $\mathbb{N} \to \mathbb{Z}$ in Example 12.3.5, which shows that $\mathbb{Z}$ is countable.

To show that $\mathbb{Q}$ is countable, we will use Fact 13.1.2 and construct an infinite sequence which contains each element of $\mathbb{Q}$ exactly once. First, construct an infinite grid which contains *all positive* rational numbers. By zig-zagging through the grid, we obtain an infinite sequence which contains each positive element of $\mathbb{Q}$ *at least* once, though there are duplicates because an element of $\mathbb{Q}$ can have many different representations as a fraction.

1/1	2/1	3/1	4/1	5/1	…
1/2	2/2	3/2	4/2	5/2	
1/3	2/3	3/3	4/3	⋱	
1/4	2/4	3/4	⋱		
1/5	2/5	⋱			
⋮					

Figure 13.1.5 A grid containing all positive rational numbers.

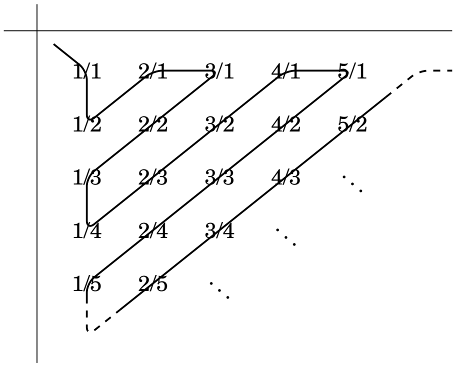


Figure 13.1.6 A path through the positive rational numbers.

The path through the grid creates the following sequence of positive rational numbers. By crossing out duplicates, we obtain an infinite sequence which contains each positive rational number exactly once.

$$1,\ \frac{1}{2},\ 2,\ 3,\ \cancel{1},\ \frac{1}{3},\ \frac{1}{4},\ \frac{2}{3},\ \frac{3}{2},\ 4,\ 5,$$
$$\cancel{2},\ \cancel{1},\ \cancel{\frac{1}{2}},\ \frac{1}{5},\ \frac{1}{6},\ \frac{2}{5},\ \frac{3}{4},\ \frac{4}{3},\ \frac{5}{2},\ \cdots$$

Finally, interleave the negative rational numbers into the above sequence, and insert 0 at the beginning.

$$0,\ 1,\ -1,\ \frac{1}{2},\ -\frac{1}{2},\ 2,\ -2,\ 3,\ -3,\ \frac{1}{3},\ -\frac{1}{3},$$
$$\frac{1}{4},\ -\frac{1}{4},\ \frac{2}{3},\ -\frac{2}{3},\ \frac{3}{2},\ -\frac{3}{2},\ 4,\ -4,\ 5,\ -5,\ \ldots$$

■

Here is an example of an uncountable set. The argument to prove the set is uncountable is a famous one, so we encapsulate it as the proof of a Lemma, rather than just a plain Example.

Lemma 13.1.7 An uncountable set of real numbers. *Let $\mathcal{C}$ represent the set of all real numbers between* 0 *and* 0.2 *(including* 0*) whose decimal expansions involve only the digits* 0 *and* 1.

Set $\mathcal{C}$ is uncountable.

Proof.

The argument in this proof is called **Cantor's diagonal argument**.

We will show that no sequence of numbers from $\mathcal{C}$ can contain *every* element of $\mathcal{C}$.

Suppose $\{a_k\}$ is an infinite sequence of elements of $\mathcal{C}$. We can create an element $r \in \mathcal{C}$ which is not in the sequence as follows. Set

$$r = 0.r_1r_2r_3r_4\cdots,$$

where r_k is the digit in the k^{th} decimal place of r, according to the following rules.

- If the k^{th} decimal place of a_{k-1} is 0, set $r_k = 1$.

- If the k^{th} decimal place of a_{k-1} is 1, set $r_k = 0$.

Clearly every digit of r will be either a 0 or 1, and $0 \le r < 0.2$, so $r \in \mathcal{C}$. (The reason we use a_{k-1} instead of a_k in the rules to create r is to make sure we consider sequence element a_0 somewhere in there.)

Now we have $a_k \neq r$ for every $k \in \mathbb{N}$, since r and a_k differ in the $(k+1)^{\text{th}}$ decimal place. Furthermore, $r \in \mathcal{C}$ because it is between 0 and 0.2 and its decimal expansion involves only digits 0 and 1. Therefore, sequence $\{a_k\}$ does not contain every element of $\mathcal{C}$ because it does not contain r. ■

Remark 13.1.8

1. The "diagonal" part of the name **Cantor's diagonal argument** refers to the following. If the decimal expansions of the real numbers in the sequence $\{a_k\}$ are written out in a grid so that each row is one of the numbers a_k and each column represents a specific decimal place in the sequence numbers, then the rules to create r can be thought of as "flipping" the digits that occur in the diagonal positions of this grid. (Draw the grid for yourself to see the pattern!)

2. Later in this chapter we will use Lemma 13.1.7 to prove that $\mathbb{R}$ itself is uncountable. (See Theorem 13.2.5.)

13.2 Properties

The following facts outline some relationships countability and the set operations. They can be used to more easily prove that a set is countable or uncountable using the already-known countability or uncountability of a related set.

Proposition 13.2.1 Countability properties.

1. *Every subset of* $\mathbb{N}$ *is countable.*

2. *If there exists an injection* $A \hookrightarrow \mathbb{N}$*, then the set* A *is countable.*

3. *Suppose* $A \subseteq B$*. If* B *is countable, then so is* A*.*

4. *Suppose* $A \subseteq B$*. If* A *is* uncountable*, then so is* B*.*

5. *If* A *and* B *are countable, then* $A \cup B$ *and* $A \cap B$ *are both countable.*

Proof outline.

1. Assume $A \subseteq \mathbb{N}$. If A is finite, then it is countable by definition. So assume that $|A| = \infty$. We can construct a sequence $\{a_k\}$ that contains each element of A exactly once as follows.

$$\begin{aligned} a_0 &= \text{ smallest number in } A, \\ a_1 &= \text{ next smallest number in } A, \\ a_2 &= \text{ next smallest number in } A, \\ &\vdots \end{aligned}$$

Therefore, A is countable.

2. If $f : A \hookrightarrow \mathbb{N}$ is injective, then $f : A \to f(A)$ is a bijection, so that A and its image $f(A)$ have the same size. But $f(A)$ is countable by Statement 1, so using the definition of countable along with Fact 12.3.2, conclude that A is countable.

3. If B is countable, then by definition there exists a bijection $f : B \to \mathbb{N}$. Then $f|_A : A \to \mathbb{N}$ is an injection. Apply Statement 2.

4. This is the contrapositive of Statement 3, under the common assumption $A \subseteq B$.

5. For $A \cap B$, consider $A \cap B \subseteq A$ and apply Statement 3.

   Now consider $A \cup B$. For simplicity, we will assume $A \cap B = \varnothing$, so that $A \cup B = A \sqcup B$. Since A and B are countable, we can write their elements as sequences:

   $$A = \{a_0, a_1, a_2, \ldots\}, \qquad B = \{b_0, b_1, b_2, \ldots\}.$$

   We can then write the elements of $A \sqcup B$ in a sequence by interleaving these two sequences:

   $$A \sqcup B = \{a_0, b_0, a_1, b_1, a_2, b_2, \ldots\}.$$

■

Checkpoint 13.2.2 Prove $A \cup B$ is countable even in the case $A \cap B \neq \varnothing$.

Hint. Consider the sets

$$A' = A \setminus (A \cap B), \qquad B' = B \setminus (A \cap B), \qquad C = A' \sqcup B'.$$

Then $A \cup B$ is the disjoint union of C and $A \cap B$.

Example 13.2.3 Primes are countable. The set of prime numbers is countable, since it is a subset of $\mathbb{N}$. □

Example 13.2.4 Unit interval is uncountable. The unit interval $(0,1)$ on the real number line is uncountable because it contains the uncountable subset $\mathcal{C}$ from Lemma 13.1.7. □

Theorem 13.2.5 *Set $\mathbb{R}$ is uncountable.*

Proof. This follows from Lemma 13.1.7 and Statement 4 of Proposition 13.2.1. ■

Example 13.2.6 The Cartesian product set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is uncountable because it has an uncountable subset: the x-axis has the same size as $\mathbb{R}$. □

13.3 More about relative sizes of sets

Question 13.3.1 What is the “size” of $\mathbb{R}$? □

We know $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ all have the same size (countably infinite). But $\mathbb{R}$ is so large that it is *un*countable, so it seems like $\mathbb{R}$ should be “larger” than $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$.

larger set set B is **larger** than set A if

(i) B has a subset the same size as A, and

(ii) every subset of B which is the same size as A is proper

smallet set

if B is larger than A, then A is **smaller** than B

Test 13.3.2 To show set B is larger than set A.

(i) Show there exists an injection $A \hookrightarrow B$.

(ii) Show that every *injection $A \hookrightarrow B$ cannot also be a surjection.*

However, if one or both of A,B are finite, one can instead just verify that $|B| > |A|$.

Remark 13.3.3

1. Matching up the parts of the Test with the parts of the definition of **larger set**:

   (i) The existence of an injection $f\colon A \hookrightarrow B$ demonstrates that B has a subset that is the same size as A, as restricting the codomain to $f\colon A \to f(A)$ creates a bijection.

   (ii) By definition, a subset $C \subseteq B$ has the same size as A when there exists a bijection $A \to C$. Enlarging the codomain, such a bijection can be thought of as an injection $A \hookrightarrow B$ whose image is $C \subseteq B$. If no such injection can also be surjective, then $C \subsetneq B$, i.e. C is a proper subset of B.

2. In the second part of the test, one can simply show that every *function* $A \to B$ cannot be a surjection, in which case surely every injection $A \hookrightarrow B$ cannot be a surjection. It may seem like it should be more difficult to prove this more general statement, but if you will find that your argument that every injection cannot be a surjection does not actually rely on the injective assumption, then there is no need for that assumption.

Example 13.3.4 Set $\mathbb{R}$ is larger than each of the sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. □

Here follows an important comparison of set sizes.

Theorem 13.3.5 (Cantor) *Every set is smaller than its own power set.*

Proof. Let A represent an arbitrary set. We will apply the Larger Set Test to demonstrate that $\mathscr{P}(A)$ is larger than A.

(i) There exists a natural injection

$$\begin{aligned} A &\hookrightarrow \mathscr{P}(A), \\ x &\mapsto \{x\}. \end{aligned}$$

(If A is empty, then this is just the **empty function**, which is always injective by Statement 1 of Proposition 12.1.6.)

(ii) Suppose $f\colon A \to \mathscr{P}(A)$ is an arbitrary function. Using the Surjective Function Test, we will demonstrate that it cannot be surjective by exhibiting an element $X \in \mathscr{P}(A)$ for which *no* element $a \in A$ satisfies $f(a) = X$. (We will not need to make the assumption that f is injective — see Statement 2 of Remark 13.3.3.)

Note that for each $a \in A$, the image element $f(a) \in \mathscr{P}(A)$, being a power set element, is a *subset* of A. So for each $a \in A$ we can ask whether a is contained in the subset $f(a)$ or not. Collecting together the "or not" answers, set

$$X = \{\, a \in A \mid a \notin f(a) \,\}.$$

Note that X is a subset of A, so again this means that it is also an *element* of $\mathscr{P}(A)$.

Could $f(a) = X$ be possible for some $a \in A$? Since $X \subseteq A$, for each $a \in A$ we have either $a \in X$ or $a \notin X$.

Case $a \in X$. Then by definition of X above, we have $a \notin f(a)$. Since X contains element a but $f(a)$ does not, sets X and $f(a)$ cannot be equal.

> *Case $a \notin X$.* Then by definition of X above, we must have $a \in f(a)$, since otherwise a would satisfy the condition to be in X. But now $f(a)$ contains element a but X does not, so again sets X and $f(a)$ cannot be equal. Since $f(a) = X$ is *not* possible in all cases, we have found an element in $\mathscr{P}(A)$ that is not in the image $f(A)$, as required to demonstrate that f is not surjective.

■

Remark 13.3.6 Cantor's Theorem implies that there are an infinite number of "levels" of infinity! For, if A is an infinite set, then $\mathscr{P}(A)$ is a *larger* infinite set, and $\mathscr{P}(\mathscr{P}(A))$ is a *still larger* infinite set, and $\mathscr{P}(\mathscr{P}(\mathscr{P}(A)))$ is a *still larger* infinite set, ...

The size of the set of natural numbers $\mathbb{N}$ seems like the lowest possible level of infinity, since every subset of $\mathbb{N}$ is either finite or has the same size as $\mathbb{N}$. (See Statement 1 of Proposition 13.2.1.) The set of real numbers $\mathbb{R}$ is larger than $\mathbb{N}$, since it contains $\mathbb{N}$ as a proper subset but is not itself the same size as $\mathbb{N}$. So writing $|\mathbb{N}| = \infty$ is not the same as writing $|\mathbb{R}| = \infty$, as they are evidently different levels of infinity. Is there any level of infinity between these two?

Conjecture 13.3.7 Continuum Hypothesis. *There does* not *exist a set larger than $\mathbb{N}$ but smaller than $\mathbb{R}$.*

Remark 13.3.8 It is not known whether the Continuum Hypothesis is true! In fact, it has been proved that the Continuum Hypothesis can be neither proved nor disproved in certain common axiomatic systems for set theory!

We have seen that funny things can happen with sizes of infinite sets — for example, $\mathbb{N}$ is an proper subset of $\mathbb{Z}$, but the two have the same size! This is not a defect in our definitions, it just demonstrates that for infinite sets, the subset relation is not a good measure of size. But it also demonstrates that we should be vigilant about other possible unintuitive consequences of our definitions, because they *might* reveal defects in our definitions. For example, from our definitions of **smaller** and **larger** sets, there is no obvious reason why there could not be some weird example of a pair of sets A and B with both B larger than A *and* A larger than B. Luckily, that cannot happen thanks to the following theorem.

Theorem 13.3.9 (Cantor, Bernstein) *Suppose A and B are infinite sets. If there exists both an injection $A \hookrightarrow B$ and an injection $B \hookrightarrow A$, then A and B have the same size.*

Proof idea. Suppose $f : A \hookrightarrow B$ and $g : B \hookrightarrow A$ are injections. We need to exhibit a bijection from A to B (or vice versa, but we will construct one from A to B).

For every element $a \in A$, we can construct a chain of alternating elements from A and B as follows. Working forwards from a, the injection f maps a to some element of B, and the injection g maps that element of B to some element of A, which is mapped by f to some element of B, and so on.

$$\begin{aligned}
a_0 &= a, \\
b_0 &= f(a_0), \\
a_1 &= g(b_0), \\
b_1 &= f(a_1), \\
a_2 &= g(b_1), \\
&\vdots
\end{aligned}$$

The chain will go on infinitely because the functions f and g always provide a next element.

We can also try to trace the chain backwards: starting at our original element

$a \in A$, we can look for some element of B that g maps to a, though at first consideration it's possible that none exists. If we do find such an element of B, we can then look for some element of A that f maps to it, and so on. While the chain extends infinitely in the forward direction, we cannot be sure at this point that it will extend infinitely in the backward direction.

Now, every element of A can be placed into such a chain, and because f and g are injective the chain in which we find an element a is always the same: the next element in the chain is always $f(a)$, and the element before a is always the unique b in B so that $g(b) = a$ (if such an element exists). And the elements after $f(a)$ and before b are also uniquely determined by the injectiveness of f and g. And so on.

So we end up finding every element of A in a unique alternating chain, and each chain has one of four patterns:

- a chain with some element repeated, in which case we could force the chain to loop back on itself at the repetition to form a finite, circular chain;

- a chain with no repetition and no end or start;

- a chain with no repetition and no end, but the process to trace it backward failed at some point, and the last element in the "backward" direction (which we could view as the first element in the whole chain) is an element of A; and

- a chain with no repetition and no end, but starts with an element of B.

Now that we have cut out possible repetition by creating circular chains, every element of A appears *exactly once* in a unique chain, and by symmetry the same can be said about B. We can then create a bijection from A to B by mapping every element of A to the element of B that follows it in the chain it appears in. *Except* for elements of A that appear in a chain that has a beginning with a starting element from B — instead each of those elements of A should be mapped to the element of B that *precedes* it in its chain. This will create a bijective correspondence between the elements of A and B. ■

13.4 Activities

Activity 13.1 In each of the following, prove that the given set is countable by exhibiting an explicitly defined bijective correspondence between it and $\mathbb{N}$.

(a) The set of natural numbers excluding 0.

(b) The set of natural numbers that are greater than 9,999,999.

(c) The set of odd natural numbers.

(d) The set of integer powers of 2 (including both positive and negative exponents).

Activity 13.2 Without cheating and looking at the proofs in this chapter, prove each of the following statements. You may wish to make use of the characterization of countability in Fact 13.1.2 instead of the technical definition of **countable set**.

Note: Each statement except the first two can be proved directly from the preceding statements.

(a) Every subset of $\mathbb{N}$ is countable.

(b) If two sets have the same size and one of them is countable, then so is the other.

(c) Every set that is the same size as a subset of $\mathbb{N}$ is countable.

(d) Every subset of a countable set is countable.

(e) Every set that is the same size as a subset of a countable set is countable.

(f) A set that contains an uncountable subset is uncountable.

Activity 13.3

(a) Prove that $\mathbb{N} \times \mathbb{N}$ is countable.

Hint. Use a zig-zag-through-a-grid method similar to the proof of the countability of the rational numbers. (See Theorem 13.1.4 and its proof.)

(b) Prove that if A and B are both countable, then so is $A \times B$.

Hint. You could do more zig-zagging, or you could use the statement of Task a.

(c) Prove that if X, Y, and Z are each countable, then so is $X \times Y \times Z$.

Hint. Use the statement of Task b twice.

(d) What proof method do you think you would use to prove the following statement?

If $A_1, A_2, \ldots, A_n$ are all countable, then so is

$$A_1 \times A_2 \times \cdots \times A_n.$$

Activity 13.4 The Infinite Orchard Problem. You own a magical apple orchard that contains an infinite number of trees, each of which bears an infinite number of apples. Describe a method to pick all of the apples in the orchard, ***one apple at a time***. (No shaking the trees, please! However, you may assume an infinite amount of time.)

Activity 13.5 Prove that if $A_0, A_1, A_2, \ldots$ is an infinite collection of sets, each of which is countably infinite, then the union

$$\bigcup_{n=0}^{\infty} A_n = A_0 \cup A_1 \cup A_2 \cup \cdots$$

is also countably infinite.

Hint. What if each set was an apple tree?

Activity 13.6 Let $\mathcal{F}$ represent the set of all functions with domain $\{0,1\}$ and codomain $\mathbb{N}$.

(a) Determine a bijective correspondence between $\mathcal{F}$ and $\mathbb{N} \times \mathbb{N}$.

(b) Explain why Task a proves that $\mathcal{F}$ is countable.

Hint. See Activity 13.2 and Activity 13.3.

Activity 13.7 Let $\mathcal{F}'$ represent the set of all functions with domain $\mathbb{N}$ and codomain $\{0,1\}$.

Note that each element of $\mathcal{F}'$ defines an infinite sequence of 0s and 1s.

(a) Suppose A is a countable subset of $\mathcal{F}'$. (So A is an infinite list of infinite sequences of 0s and 1s.)

Describe how to construct an element of $\mathcal{F}'$ that is definitely not in A. That is, build an infinite sequence of 0s and 1s that is definitely not the same as any of the infinite sequences in the infinite list of A.

Hint. Use Cantor's diagonal argument from the proof of Lemma 13.1.7.

(b) Explain why Task a proves that $\mathcal{F}'$ is uncountable.

Hint. $\mathcal{F}' \subseteq \mathcal{F}'$.

Part IV

Graph Theory

CHAPTER 14

Graphs

14.1 Basics and examples

graph (working definition)
a diagram consisting of a finite collection of points linked by line segments or arcs

graph (formal definition)
an ordered pair (V,E), where V is a finite set and E is a finite, unordered list of *subsets* of V, each of which has exactly 1 or 2 elements

vertex a point in a graph (i.e. an element of V)

node synonym for vertex

edge a line or arc linking two vertices (i.e. an element of E)

empty graph
the graph $(\varnothing,\varnothing)$, with no vertices and no edges

Warning 14.1.1 The list E is not a set, since duplicate entries in this list have a graphical meaning; see below.

An element $e \in E$ represents an edge as follows. If e consists of exactly two elements of V, draw a line between these two vertices. If e consists of exactly one element of V, draw a line from this vertex to itself.

Example 14.1.2 A very basic graph. The graph $G=(V,E)$, where

$$V=\{v_1,v_2,v_3\}, \qquad E=\big\{\{v_1,v_2\},\{v_1,v_3\},\{v_2,v_3\}\big\},$$

has three vertices and three edges.

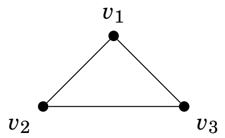


Figure 14.1.3 Diagram of the graph $G=(V,E)$.

□

Example 14.1.4 A slightly more complicated example graph. The graph $G=(V,E)$, where

$$V=\{v_1,v_2,v_3,v_4\}, \qquad E=\big\{\{v_1,v_2\},\{v_1,v_2\},\{v_1,v_2\},\{v_3\},\{v_3\}\big\},$$

has four vertices and five edges.

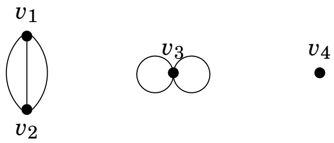


Figure 14.1.5 Diagram of the graph $G = (V, E)$.

□

Worked Example 14.1.6 Using a graph to solve a problem. Suppose we have jugs A, B, C with capacities $8, 5, 3$ litres, respectively. Jug A is full of water and jugs B, C are empty. Can we divide the water into two *exactly* equal parts? If so, find the most efficient pouring sequence.

Solution. Construct a graph with points labelled by elements $(a, b, c) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}$, where a, b, c are the volumes of water in jugs A, B, C, respectively. Join two points with a line segment if we can obtain one set of volumes from the other in a single pour. We will ignore pours that return us to a configuration previously achievable by fewer pours.

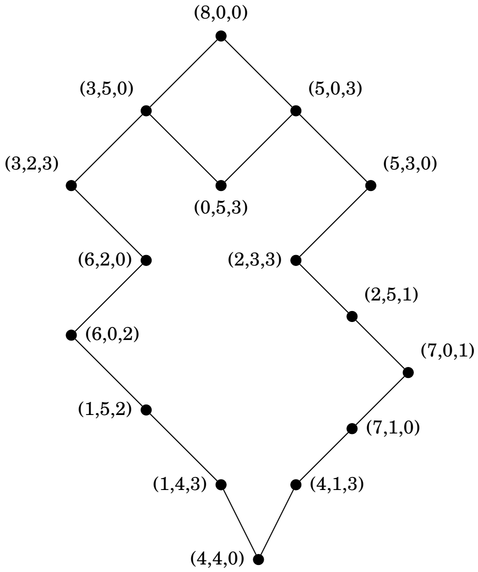


Figure 14.1.7 Graph to track possible jug fill states.

Following the left leg of the graph gets us to the desired configuration in 7 pours. □

14.2 Properties of graphs

14.2.1 Properties of vertices and edges

adjacent vertices
linked by an edge

adjacent edges
share a common vertex

incident a pair of a vertex and edge where the edge links that vertex either to itself or another vertex of the graph

loop an edge which links a vertex to itself

parallel edges
edges which link the same vertices

simple graph
no loops or parallel edges

isolated vertex
a vertex that is incident with *no* edges

degree (of a vertex)
the number of times that the vertex is incident with an edge of the graph

$\deg v$ degree of vertex v

Note 14.2.1 In a *simple* graph, the degree of each vertex is equal to the number of incident edges. However, in a *non*-simple graph, a loop is incident to its vertex *twice*, and we count that in the degree:

$$\deg v = \#\{\text{ edges that are incident to } v \text{ but not loops at } v\} + 2 \cdot \#\{\text{ loops at } v\}.$$

Example 14.2.2 Properties of our very basic example graph. The graph of Example 14.1.2 has the following properties.

- It is a simple graph.
- Each pair of vertices is adjacent.
- Each pair of edges is adjacent but not parallel.
- There are no loops.
- Each vertex is incident to two non-loop edges, so each vertex has degree 2.

□

Example 14.2.3 Properties of our slightly more complicated example graph. The graph of Example 14.1.4 has the following properties.

- It is not a simple graph.
- There are three parallel edges linking v_1 and v_2.
- There are two loops at vertex v_3 (and these are also parallel edges).
- The parallel edges between v_1 and v_2 are adjacent, as are the two loops at v_3.
- Vertices v_1 and v_2 are adjacent, and vertex v_3 is adjacent to itself.
- Vertices v_1 and v_2 are incident to the three edges running between them, and vertex v_3 is incident to its two loops.
- Vertex v_4 is isolated
- For degrees we have

$$\deg v_1 = \deg v_2 = 3, \qquad \deg v_3 = 4, \qquad \deg v_4 = 0.$$

□

The number of edges in a graph is an important measure both of how “connected” the graph is, as well as how much “redundancy” the graph contains.

$\lvert E\rvert$	the number of edges in the graph $G=(V,E)$

Theorem 14.2.4 Sum of degrees is twice the number of edges. *Suppose $G=(V,E)$ is a graph with vertex set $V=\{v_1,v_2,\ldots,v_n\}$. Then,*

$$\deg v_1+\deg v_2+\ldots\deg v_n=2\lvert E\rvert .$$

Proof idea. If an edge e is a loop at vertex v_i, then it contributes 2 to $\deg v_i$. Otherwise, if edge e links vertices v_i and v_j ($v_i\neq v_j$), then it contributes 1 to each of $\deg v_i$ and $\deg v_j$. In every case, each edge contributes exactly 2 to the sum of the vertex degrees. ■

Corollary 14.2.5 Odd degrees are even. *In every graph, the number of vertices of odd degree is even.*

Proof idea. Otherwise, the sum of the degrees of all vertices would be odd, which contradicts the theorem above. ■

Worked Example 14.2.6 An odd fellow throws an odd party and invites an even number of other equally-odd people. Each odd person at the party is friends with an odd number of other odd people at the party. Is this odd party even possible?

Solution. Create a simple graph with the people at the party as vertices, where two vertices are linked by a single edge if and only if the two people are friends. As each person has an odd number of friends at the party, the degree of each vertex is odd. But the number of party attendees is also odd, since there are an even number of invitees, plus the host himself. So we have an odd number of vertices each with odd degree, which the corollary above says is not possible. □

14.2.2 Subgraphs

subgraph	a graph that is a part of a larger graph
$G'\preceq G$	graph G' is a subgraph of graph G

Remark 14.2.7 Without a diagram, how can we tell if a graph $G'=(V',E')$ is a subgraph of another graph $G=(V,E)$? First, each vertex of G' should also be a vertex of G, so that $V'\subseteq V$. And also, each edge of G' should also appear as an edge in G. (Though we shouldn't just write $E'\subseteq E$, and not only because E' and E are actually sets — see Activity 14.2.)

Note 14.2.8 Just as $\varnothing\subseteq A$ and $A\subseteq A$ for every set A, we consider $(\varnothing,\varnothing)\preceq G$ and $G\preceq G$ for every graph G.

Worked Example 14.2.9 Determine all possible subgraphs of the graph in Example 14.1.2.

Solution. First, every graph contains the empty graph as a subgraph. Next, a nonempty subgraph of this particular graph can contain one, two, or all three vertices. We can write out all nonempty possibilities in a general way based on the number of vertices in the subgraph. In each graph below we require the vertex indices i,j,k to all be distinct and to satisfy $1\leq i,j,k\leq 3$.

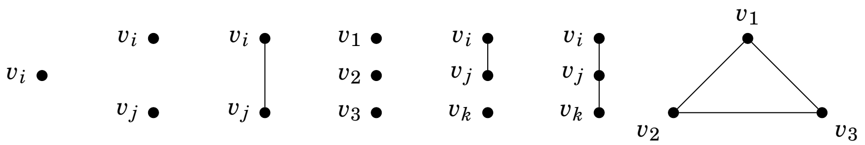


Figure 14.2.10 All possible subgraphs of our basic example graph from Example 14.1.2.

There are 3 subgraphs of each of the first three types in Figure 14.2.10. There are also 3 subgraphs of the fourth and fifth types. Therefore, including the empty graph, there are 18 subgraphs of this example graph. □

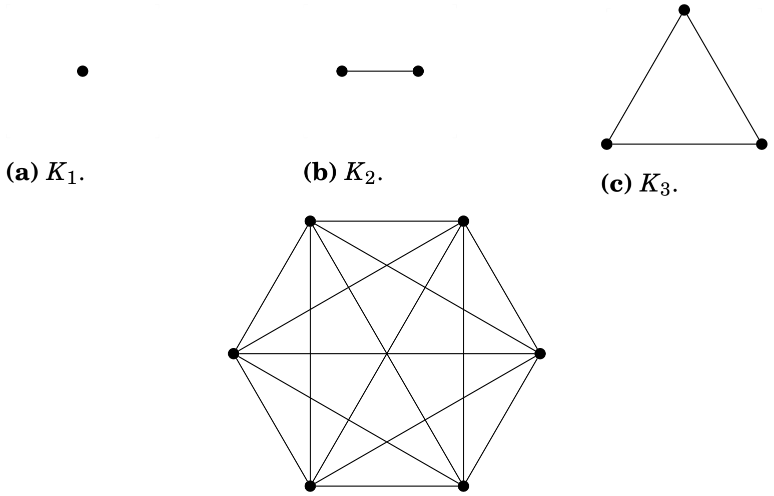
14.2.3 Complete graphs

complete graph

a simple graph in which every pair of distinct vertices is adjacent

Proposition 14.2.11 Properties of complete graphs.

1. *For each $n \geq 0$, there is a* unique *complete graph $K_n = (V, E)$ with $|V| = n$.*
2. *If $n \geq 1$, then every vertex in K_n has degree $n - 1$.*
3. *Every simple graph with n or fewer vertices is a subgraph of K_n.*



(a) K_1.

(b) K_2.

(c) K_3.

(d) K_6.

Figure 14.2.12 Complete graphs with 1, 2, 3, and 6 vertices.

Checkpoint 14.2.13 Draw the complete graphs K_4 and K_5.

Checkpoint 14.2.14 Is there a complete graph K_0?

14.3 Adding information to graphs

weighted graph (working definition)

a graph in which each edge is assigned a *weight* or *cost*, usually a numerical value

weighted graph (formal definition)
an ordered triple (V,E,w), where (V,E) is an ordinary graph and $w\colon E \to W$ is a function with some set W as codomain

weights elements of the image $w(E) \subseteq W$

We usually label each edge with its weight on diagrams for the graph.

Example 14.3.1 A road map weighted by distances. A road map with distances as weights is a weighted graph. Below is a simplified road map of the area around Camrose, Alberta. The vertex set is the set of cities, and the edge set is the set of highways. For example, the two-city set {Camrose, Edmonton} represents the edge on the graph between Camrose and Edmonton, and corresponds to Highway 21.

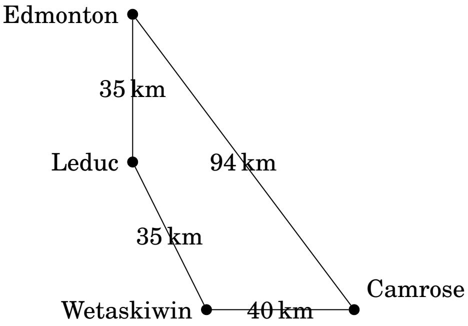


Figure 14.3.2 Road map of the area around Camrose, Alberta.

e	$w(e)$
{Camrose, Edmonton}	94
{Edmonton, Leduc}	35
{Leduc, Wetaskiwin}	35
{Wetaskiwin, Camrose}	40

Figure 14.3.3 Table of values for distance weight function.

The edges in the graph are weighted by the (rounded) highway distances between cities. Formally, this is a function w from the edge set to the natural numbers. The input-output relationship defining this function is tabulated above right. □

Example 14.3.4 Variations on Example 14.3.1 include any kind of transportation or communication network with transportation/communication lines as edges. Possible weights assigned to an edge include: length of the line; amount of time it takes a vehicle/message to travel along the line from one node to the next; capacity of the line in vehicles/passengers/messages/data per unit time; etc.. □

directed graph (working definition)
a graph in which each edge can be given a direction, “pointing” to one of its incident vertices

directed graph (formal definition)
an ordered pair (V,E), where V is a finite set and E is an unordered, possibly empty list of elements of $V \times V$

Again, elements of V are the vertices and elements of E are the edges of the graph. For an ordinary graph, edges were represented by subsets of V because when specifying an edge, the order of the vertices which are to be incident to the edge is irrelevant. For a directed graph, the order of the vertices incident to an edge now matters, so we use ordered pairs of vertices to specify an edge. If $e = (v,v') \in E$ for some $v,v' \in V$, consider the direction of e to be $v \to v'$.

Example 14.3.5 A basic directed graph. Consider

$$V = \{v_1, v_2, v_3, v_4\},$$
$$E = \{(v_1,v_2),(v_1,v_2),(v_2,v_3),(v_3,v_2),(v_4,v_3),(v_4,v_4)\}.$$

We draw the graph $G = (V,E)$ with arrows to indicate the direction of edges.

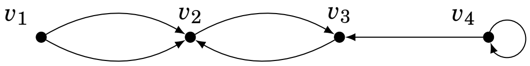


Figure 14.3.6 Diagram of the directed graph $G = (V,E)$.

□

Checkpoint 14.3.7 Invent a formal definition for **directed, weighted graph**.

14.4 Important examples

Example 14.4.1 A power set graph. We can use a graph to visualize the power set of a finite set A: let $(\mathscr{P}(A),E)$ be the directed graph where, for vertices $B,C \in \mathscr{P}(A)$ (that is, subsets $B,C \subseteq A$), the ordered pair (B,C) is an edge in E if the following two conditions are satisfied:

(i) $B \subsetneqq C$; and

(ii) there *does not* exist a subset $D \subseteq A$ such that $B \subsetneqq D \subsetneqq C$.

Note: The second condition is to avoid cluttering the graph with extra edges that do not add any extra information.

For example, consider $A = \{a,b,c\}$ and

$$\mathscr{P}(A) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}.$$

So we can draw a graph to represent the subset relationships of $\mathscr{P}(A)$, with arrows to point from subset to superset.

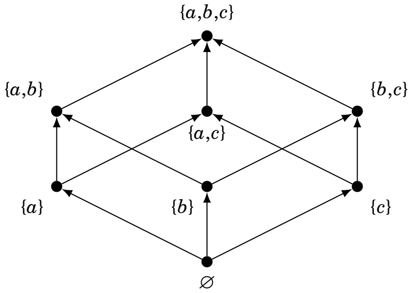


Figure 14.4.2 Directed graph to represent the subset relationships between elements of $\mathscr{P}(A)$.

It is somewhat natural to draw the graph with the largest subsets of A at the top (or at the bottom). If we decide that arrows will always point *upwards*, we can unclutter our graph by just drawing lines instead of arrows for edges. □

We can use graphs to visualize other kinds of mathematical relationships.

Example 14.4.3 A division graph. For integers m,n, write $m \mid n$ if m divides n; that is, if n/m is also an integer. In this case, we say that m **divides** n.

Set

$$V = \{2,3,4,5,6,7,8,9,10,11,12,13,14\}.$$

Let $G = (V,E)$ be the directed graph where, for integers $m,n \in V$ with $m < n$, the ordered pair (m,n) is an edge in E if the following two conditions are satisfied:

(i) $m \mid n$;

(ii) there *does not* exist an integer $\ell \in V$ such that $m \mid \ell$ and $\ell \mid n$, but $\ell \neq m,n$.

We can use this directed graph to visualize the set V with respect to the **divides** relationship. Again, we agree that arrows will always point *up*, but do not actually draw them. Note that the prime numbers all appear at the bottom.

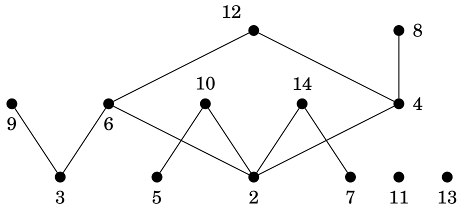


Figure 14.4.4 Directed graph to represent the relationship of division between elements of V.

□

Example 14.4.5 Travelling salesman problem. You get a summer job with the Prospective Student Office at the University of Alberta's Augustana Campus in Camrose, Alberta. In May and June, your job is to visit Alberta high schools and meet with students who are thinking of applying to Augustana. Below is a map of the cities and towns you must visit, given as a weighted graph with distances in kilometres as weights. To save on gas, you would like to visit each city and town *exactly* once, and do so while travelling the shortest distance possible.

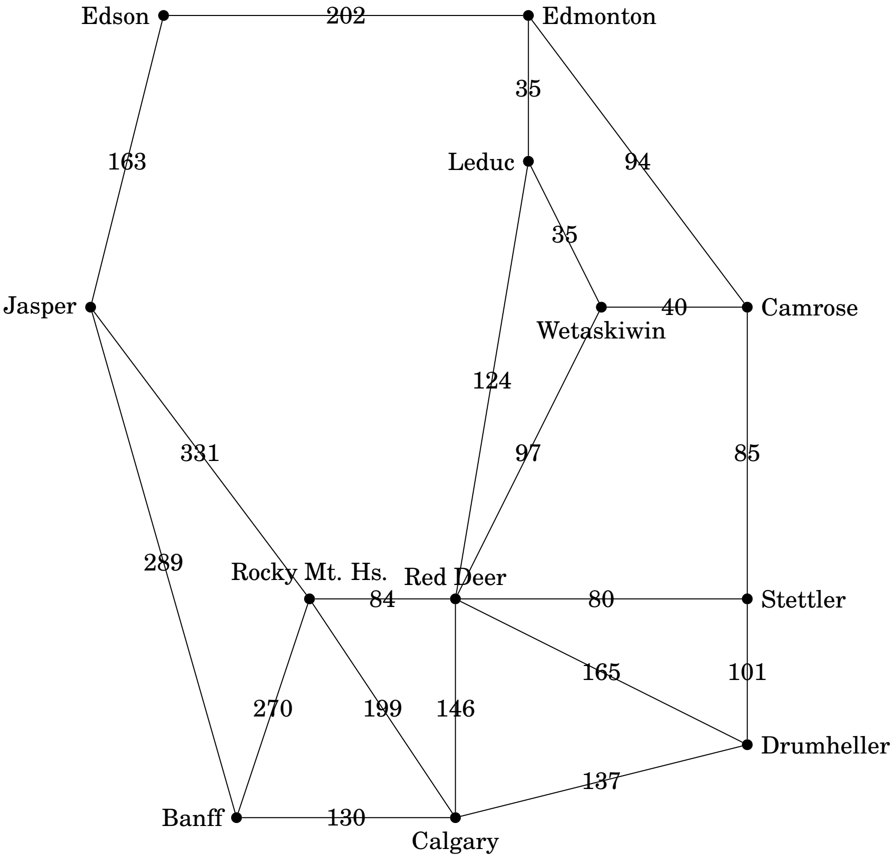


Figure 14.4.6 Road map of major centres in Alberta as a weighted graph.

This is a difficult problem, and gets more difficult as the number of cities and roads increases. □

14.5 Activities

Activity 14.1 Draw all possible simple graphs with 4 vertices.

Hint. See Statement 3 of Proposition 14.2.11.

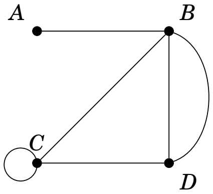
Activity 14.2 Suppose $G = (V, E)$ is a graph. Decide the truth of the following statement.

> Every pair of a subset $V' \subseteq V$ and a subcollection $E' \subseteq E$ defines a subgraph $G' = (V', E')$ of G.

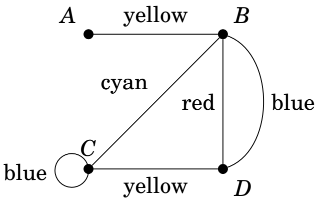
Activity 14.3 Draw a graph where the nodes are students present in today's class. Draw edges between pairs of students that are in the same group today. Additionally, draw an edge between a member of your group and another student if that pair was in a group together last class.

Activity 14.4 For each of the following graphs, write out its formal definition as either a (regular) graph, a weigthed graph, or a directed graph, as appropriate.

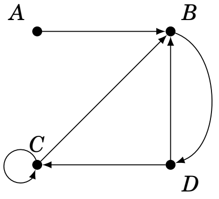
(a)



(b)



(c)



Activity 14.5 Consider the website *Facebook* as a graph where vertices are profiles and edges represent "friendship".

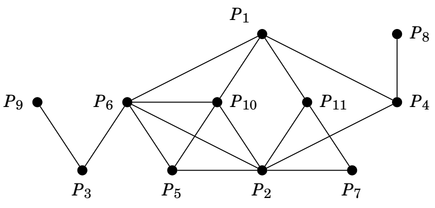
(a) Should this graph be a directed graph? Why or why not?

(b) Is this graph simple? complete? Justify your answers.

(c) What does the degree of a vertex represent?

(d) Could this graph have isolated vertices?

(e) Suppose the following graph is a subgraph of the *Facebook* graph.



(i) What is the largest party one of these people could throw where each party-goer is Facebook friends with every other party-goer? Justify your answer.

(ii) Assume all of the people in this graph live in the same geographic area. Which pair of non-friends are most likely to become friends in the future? Which pair of non-friends are least likely to become friends in the future? Justify your answers.

14.6 Exercises

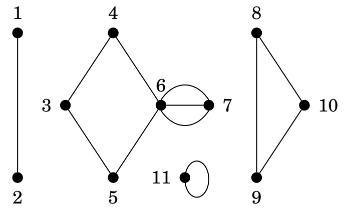


Figure 14.6.1 An example graph.

1. Consider the graph in Figure 14.6.1.

(a) Are vertices 1 and 2 incident?

(b) Are any vertices adjacent to themselves?

(c) Is vertex 3 adjacent to vertex 6?

(d) Is this a simple graph?

(e) Compute the degree of each vertex in the graph. Then verify that the sum of the degrees is equal to twice the number of edges. (See Theorem 14.2.4.)

2.

(a) How many edges does the complete graph with ten vertices have?

Hint. See Theorem 14.2.4.

(b) Generalize your result to a formula for the number of edges in the complete graph with n vertices.

3.

(a) Draw an example of a simple graph that has no vertices of odd degree.

(b) Draw an example of a simple graph that has no vertices of even degree.

4. Given a collection of sets $A_1, A_2, \ldots, A_n$, the intersection graph of the collection is the simple graph that has a vertex for each of the sets in the collection, with two vertices joined by an edge if and only if the two corresponding sets have nonempty intersection. Draw the intersection graph of the following collection of sets.

$$\begin{aligned} A_1 &= \{1,2,3,4,5\}, \\ A_2 &= \{2,4,6,8\}, \\ A_3 &= \{3,5,12\}, \\ A_4 &= \{5,8,10\}. \end{aligned}$$

Complement of a graph. Exercises 5–7 concern the following definition.

complement (of a simple graph G)
the simple graph that has the same vertex set as G, but in which two vertices are joined by an edge if and only if those same two vertices are not joined by an edge in G

5. Draw the complement of the simple graph in Figure 14.6.2.

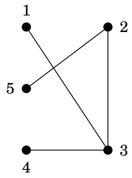


Figure 14.6.2 A simple graph.

6. What is the complement of a complete graph?

7. Suppose G is a simple graph with n vertices. Determine a relationship between the number of edges in G, the number of edges in the complement of G, and the number of edges in the complete graph K_n with n vertices.

Hint. Recall that every simple graph with n vertices is a subgraph of K_n.

Chapter 15

Paths and connectedness

15.1 Motivation

Example 15.1.1 Driving routes on a graph. Looking at a map of Alberta, you might decide that there are three reasonable driving routes from Camrose to Red Deer and three reasonable driving routes from Red Deer to Drumheller. Define a graph $G = (V, E)$ with cities as vertices and routes as edges. If we travel between Camrose, Red Deer, and Drumheller on these routes, we find that any multi-city trip is a finite sequence from $V \cup E$ which starts and ends at a vertex and alternates between vertices and edges.

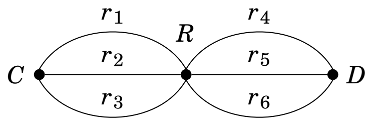


Figure 15.1.2 Driving routes between Camrose, Red Deer, and Drumheller.

For example,

$$C, r_1, R, r_4, D, r_5, R, r_2, C, r_3, R, r_6, D$$

is a trip that travels back and forth between Camrose and Drumheller, via Red Deer each time, and that never uses the same route twice. Notice that we cannot extend this trip without repeating a route. □

15.2 Walks, trails, and paths

Suppose $G = (V, E)$ is a graph.

walk a finite sequence $v_0, e_1, v_1, e_2, \ldots, v_{n-1}, e_n, v_n$ of elements from $V \cup E$, with each $v_i \in V$ and each $e_i \in E$, such that edge e_i connects vertices v_{i-1} and v_i

closed walk a walk that ends at the same vertex at which it started (that is, $v_n = v_0$)

open walk a walk that isn't closed (that is, $v_n \neq v_0$)

trail a walk that traverses no edge more than once

path a walk that passes through no vertex more than once, *except* possibly the endpoints v_0, v_n

Note 15.2.1 We may also apply the adjectives **open** and **closed** to trails and paths.

Example 15.2.2 Consider the graph in Example 15.1.1.

1. The "trip" we found in the example is a trail of maximal length starting at vertex C.
2. The walks C,r_1,R,r_4,D and C,r_1,R,r_2,C are both paths, the first open and the second closed.
3. The walk C,r_1,R,r_4,D,r_4,R is neither a path nor a trail.

□

Example 15.2.3 Paths and trails. Consider the following graph.

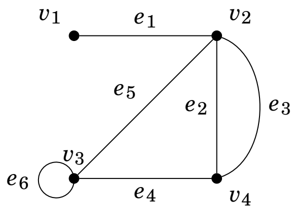


Figure 15.2.4 A example graph to illustrate paths and trails.

This graph has the following properties.

1. Every path or trail passing through v_1 must start or end there but cannot be closed, *except* for the closed paths:
   - v_1;
   - v_1,e_1,v_2,e_1,v_1;
   - v_2,e_1,v_1,e_1,v_2;
2. Walk $v_1,e_1,v_2,e_5,v_3,e_4,v_4$, is both a trail and a path.
3. Walk $v_1,e_1,v_2,e_5,v_3,e_6,v_3,e_4,v_4$, is a trail but not a path.

□

Worked Example 15.2.5 Consider again the graph in Figure 15.2.4 from Example 15.2.3. How many trails from v_3 to v_4 exist? How many of those trails are paths? Are there any paths from v_3 to v_4 that are *not* trails?

Solution. We can solve this using a graph! The graph in Figure 15.2.6 was created by mapping out all possible trails starting at v_3 and ending at v_4, moving across one edge at a time. Each node in this new (directed) graph is labelled with a partial walk that is a continuation of the walk assigned to the node above it. Each leg in the graph stops when the associated walk being followed reaches v_4 and cannot be continued without repeating another edge. To save space in the node labels, we have used "..." to mean the walk from the previous node.

Counting all the nodes in the graph of Figure 15.2.6 that are labelled with a walk that ends in v_4, we see that there are ten trails from v_3 to v_4. Also, we can easily see that only three of the trails are paths.

We can use the same technique to map out all paths from v_3 to v_4, but this time we terminate a leg when we cannot move off a vertex without repeating a vertex that is already visited in that walk. (Note that the walk v_3,e_6,v_3 is a path, but if we extend this walk in any way it will no longer be a path.)

From the result in Figure 15.2.7, we see that there are only three paths from v_3 to v_4, and each of them is a trail.

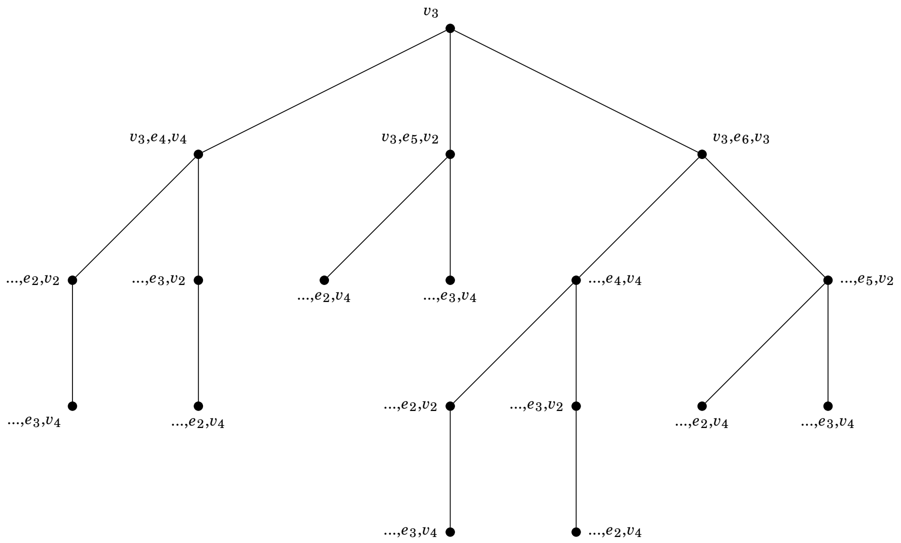


Figure 15.2.6 Mapping the trails from v_3 to v_4 in the graph of Figure 15.2.4.

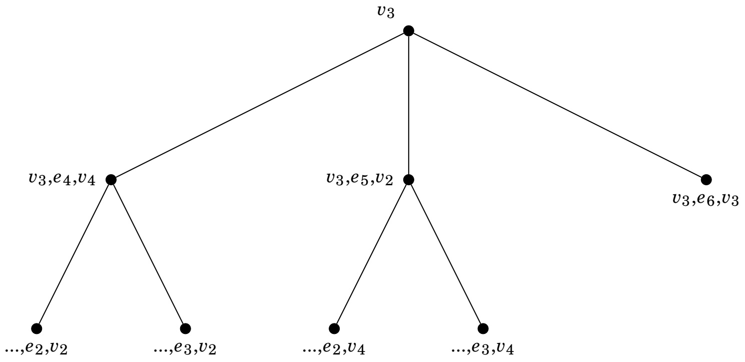


Figure 15.2.7 Mapping the paths from v_3 to v_4 in the graph of Figure 15.2.4.

□

Proposition 15.2.8 *Every open path is a trail.*

Proof. We will prove the contrapositive: a walk that is not a trail cannot be an open path. So suppose W is a walk in a graph, and that W traverses edge e twice.

Case e is a loop. Then W passes through the vertex incident to e at least three times, hence is not a path.

Case e is not a loop. Write $e = \{v, v'\}$. Initially, there are two possibilities to consider. If each of the two assumed traversals of e moves from v to v', then W passes through each of v, v' at least twice, and hence is not a path. If the two assumed traversals of e move v to v' and v' to v respectively, then W passes through v at least twice. If W traverses v twice because it both starts and ends there, then W is not open. If W is open and traverses v twice, then W is not a path. So in any case, W is not an open path. ■

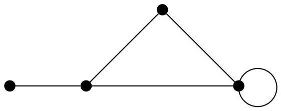
Note 15.2.9 In Activity 15.1, you are asked to create counterexamples of some statements related to the above proposition.

15.3 Connected vertices, graphs, and components

connected vertices
a pair of vertices v, v' such that there exists a walk beginning at v and ending at v'

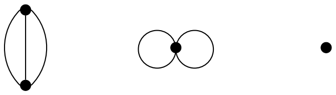
connected graph
every pair of vertices is connected

Example 15.3.1 A connected graph.



□

Example 15.3.2 A nonconnected graph.



□

Being connected is a symmetric relationship between vertices, and walks connecting vertices can be shortened to paths.

Proposition 15.3.3 Characterizations of connected vertices. *Assume v, v' are vertices in a graph. Then the following are equivalent.*

1. *Vertices v, v' are connected.*
2. *There exists a walk beginning at v and ending at v'.*
3. *There exists a path beginning at v and ending at v'.*
4. *There exists a walk beginning at v' and ending at v.*
5. *There exists a path beginning at v' and ending at v.*

Proof idea. As usual, we prove the equivalence of multiple statements using a cycle of logical implications.

Statement 1 implies Statement 2. This is the definition of **connected vertices**.

Statement 2 implies Statement 3. Suppose $v_0, e_1, v_1, \ldots, v_{n-1}, e_n, v_n$ is a walk with $v_0 = v$ and $v_n = v'$. If this walk is not a path, then there is a repeated vertex. Suppose $v_j = v_i$ with $j > i$. Then

$$v_0, e_1, v_1, \ldots, v_{i-1}, e_i, v_j, e_{j+1}, \ldots, v_{n-1}, e_n, v_n$$

is also a walk from v to v'. Keep removing repeated vertices in this way until a path is obtained.

Statement 3 implies Statement 4. Just reverse the order of the vertices and edges in the path from v to v' to obtain a walk in the other direction.

Statement 4 implies Statement 5. As before, if the walk from v' to v is not a path, each pair of repeated vertices can be eliminated by "snipping out" the portion of the walk between them.

Statement 5 implies Statement 1. Reverse the path from v' to v to obtain a walk from v to v', thereby satisfying the definition of **connected vertices**. ■

A nonconnected graph can be considered to simply be a collection of connected subgraphs.

connected component (working definition)
a connected subgraph of a graph which is not contained in any larger connected subgraph of that graph

connected component (formal definition)
a subgraph G' of a graph G that satisfies the following:

(i) G' is connected;

(ii) if G'' is a subgraph of G such that G'' is connected and $G' \preceq G''$, then $G'' = G'$

Example 15.3.4 Breaking a nonconnected graph into connected components. Considering Figure 15.3.5 below as a single graph, we have placed the connected components (of which there are three) into boxes.

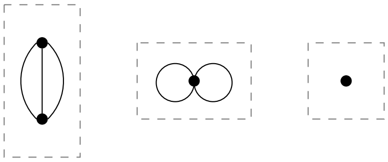
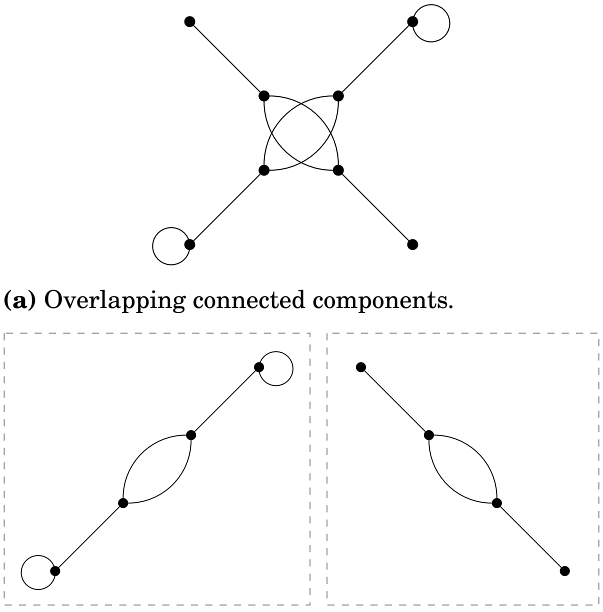


Figure 15.3.5 A nonconnected graph as a collection of connected subgraphs.

This nonconnected graph has other connected subgraphs. For example, the subgraph that contains only the left-most two vertices joined by a single edge is a connected subgraph. But that connected graph is not a connected component because it is a subgraph of a larger connected subgraph. □

Example 15.3.6 Connected components do not depend on how the graph is represented diagrammatically.



(a) Overlapping connected components.

(b) Non-overlapping connected components.

Figure 15.3.7 Two different representations of the same nonconnected graph.

The two graphs in Figure 15.3.7 are in fact the same graph, just with different diagrammatic representations. In the second version of the graph, we have again identified connected components by placing each of them in a box. □

Example 15.3.8 A connected graph has one component. If a graph is connected, then the entire graph is a largest connected subgraph possible.

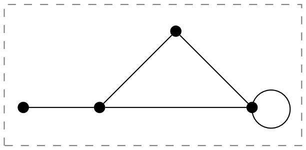


Figure 15.3.9 A connected graph with one connected component.

□

Note 15.3.10 As in our working definition, informally the **connected components** of a graph G are the "largest" subgraphs of G that are connected. The second condition in the formal definition is just a positive way of stating the working definition. We will make the general notion of "largest" more precise in a similar way in Chapter 19 (see the definition of **maximal element**, the Test for Maximal/Minimal Elements, as well as Example 19.5.7).

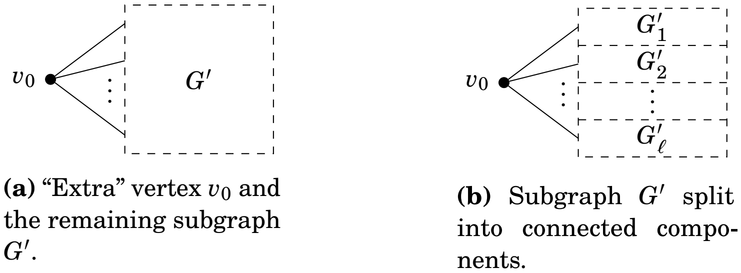
Theorem 15.3.11 A lower bound for the number of edges in a connected graph. *If $G = (V,E)$ is connected and $|V| = n$, then $|E| \geq n-1$.*

Proof. By (strong) induction.

Base case $n = 1$. Every graph with only one vertex is connected and satisfies $|E| \geq 0$.

Induction step. Assume $k \geq 1$ and the statement is true for all $n \leq k$. Let $G = (V,E)$ be a connected graph with $k+1$ vertices. We must show $|E| \geq k$.

Arbitrarily choose some vertex $v_0 \in V$, and let $G' = (V',E')$ be the graph obtained from G by removing v_0 and all edges incident to it. Unfortunately, G' might not be connected. Let $G'_1,\ldots,G'_\ell$ be its connected components. Write $G'_i = (V'_i,E'_i)$, and let $n_i = \left|V'_i\right|$. Then each G'_i is connected and has at most k vertices, so the induction hypothesis applies. Also note that $n_1 + \cdots + n_\ell = k$, since every vertex of G except v_0 is part of exactly one subgraph G'_i.



(a) "Extra" vertex v_0 and the remaining subgraph G'.

(b) Subgraph G' split into connected components.

Figure 15.3.12 Removing "extra" vertex v_0, splitting remaining subgraph into connected components.

Therefore, using our induction hypothesis we may calculate

$$\begin{aligned}|E| &= \left|E'\right| + |\{\text{edges in } G \text{ incident to } v_0\}| \\ &= \left|E'_1\right| + \cdots \left|E'_\ell\right| + |\{\text{edges in } G \text{ incident to } v_0\}| \\ &\geq (n_1 - 1) + \cdots + (n_\ell - 1) + |\{\text{edges in } G \text{ incident to } v_0\}| \\ &= k - \ell + |\{\text{edges in } G \text{ incident to } v_0\}|.\end{aligned}$$

However, since G is connected, v_0 must be the glue keeping the subgraphs $\{G'_i\}$ together. That is, for each i there must be at least one edge between v_0 and some vertex of G'_i. Therefore,

$$\begin{aligned}|E| &\geq k-\ell+|\{\text{edges in } G \text{ incident to } v_0\}|\\ &\geq k-\ell+\ell\\ &= k.\end{aligned}$$

■

15.4 Articulation vertices, bridges, and edge connectivity

articulation vertex
a vertex of a graph such that, if it were to be removed (along with any edges incident to it), the resulting subgraph would have more connected components than the original

bridge an edge of a graph such that, if it were to be removed, the resulting subgraph would have more connected components than the original

Example 15.4.1 An articulation vertex. In the graph of Figure 15.4.2, the central vertex that is common to both diamond-shaped subgraphs is an articulation vertex, as removing it and all edges incident to it would leave two unconnected "ears" on the outside of the two diamond shapes.

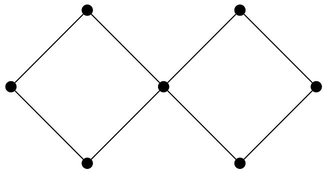


Figure 15.4.2 A graph featuring a single, central articulation vertex.

□

Example 15.4.3 A bridge between two articulation vertices. In the graph of Figure 15.4.4, edge e is a bridge, and each of v and v' are articulation vertices.

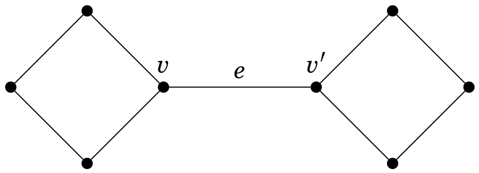


Figure 15.4.4 A graph featuring a bridge between two articulation vertices.

□

Remark 15.4.5 In the proof of Theorem 15.3.11, our conception was that "extra" vertex v_0 was an articulation vertex, where removing it would create a subgraph G' that would be split into connected components $G'_1, \ldots, G'_\ell$. (Though it is possible v_0 is *not* an articulation vertex, if subgraph G' is connected.)

edge connectivity
the minimum number of edges that must be removed from a connected graph to obtain a nonconnected subgraph

Remark 15.4.6 Edge connectivity measures redundancy in the graph, as each edge that can be removed *without* breaking the graph into nonconnected subgraphs must be incident to a pair of vertices that remain connected via some other walk through the graph.

Example 15.4.7 The edge connectivity of the graph in Figure 15.4.2 is 2. □

Example 15.4.8 Edge connectivity in a graph with a bridge. A bridge represents a "single point of failure," and every graph that contains a bridge has edge connectivity 1. For example, removing the single edge e in the graph of Figure 15.4.4 breaks the graph into two nonconnected subgraphs. □

Proposition 15.4.9 Two upper bounds for edge connectivity. *Suppose $G = (V,E)$ is a connected graph. Let $n = |V|$, $e = |E|$, and let d be the smallest degree of any of the vertices of G. Then the edge connectivity of G cannot be greater than either of the integers d or $\lfloor 2e/n \rfloor$.*

Proof. First, if v is a vertex of G with $\deg v = d$, then removing all of the edges incident to v will cause v to become isolated and G to become nonconnected. So the edge connectivity of G cannot be greater than d.

Next, recall that the sum of the degrees of the vertices of G is equal to $2e$ (Theorem 14.2.4). Using this, we have

$$2e = \deg v_1 + \deg v_2 + \cdots + \deg v_n \geq d + d + \cdots + d = nd.$$

So $d \leq 2e/n$. The number $2e/n$ is rational, but may not be an integer. However, d is definitely an integer, so we must have $d \leq \lfloor 2e/n \rfloor$. Since we have already concluded that the edge connectivity of G is no greater than d, it also can be no greater than $\lfloor 2e/n \rfloor$. ■

Remark 15.4.10 With n and e as in the statement of the theorem, $2e$ is equal to the sum of the degrees of the vertices (Theorem 14.2.4), so $2e/n$ is equal to the *average* degree of vertices in the graph.

Worked Example 15.4.11 Your tree fort rivals have set up a communication system of tin cans and strings. You have mapped out their network as in Figure 15.4.12. To minimize the risk of crab apple welts, what is the minimum number of strings you must cut to disrupt their communications?

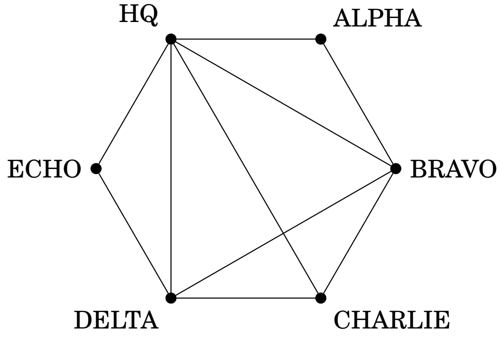


Figure 15.4.12 TreeFort CommNet.

Solution. There are 6 nodes and 10 edges, so Proposition 15.4.9 tells that the edge connectivity must be no greater than $\lfloor 20/6 \rfloor = 3$. By inspection, the edge connectivity is not 1 as there are no bridges. However, we may isolate either fort ALPHA or ECHO with two snips. □

15.5 Activities

Activity 15.1 In each of the following, devise a graph that contains the requested type of walk. (You do not need to create one graph that contains all three types of walks; you may draw three separate graphs.)

(a) A closed path that is not a trail.

(b) An open trail that is not a path.

(c) A closed trail that is not a path.

Activity 15.2 Devise a graph with exactly four vertices, each of which has degree 5, so that the graph is

(a) nonconnected.

(b) connected.

Activity 15.3 In each of the following, devise a connected graph with at least five vertices that has the requested properties. Do so *without* looking at the example graphs in this chapter. (You do not need to create one graph that contains all of the properties; you may draw a separate graph for each task below.)

(a) Contains a bridge.

(b) *Every* edge is a bridge.

(c) Contains an articulation vertex.

(d) *Every* vertex of degree at least 2 is an articulation vertex.

Activity 15.4 Does increasing the number of edges in a graph increase its edge connectivity?

Activity 15.5

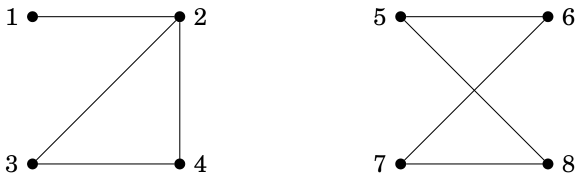


Figure 15.5.1 A nonconnected graph.

In the graph of Figure 15.5.1, explain why the subgraph formed by vertices 2, 3, and 4, along with all edges incident to these vertices, *fails* the formal definition of connected component. Identify which of the two conditions of this formal definition the subgraph fails, and explicitly describe how the subgraph fails to meet that condition.

Activity 15.6 Suppose G is a simple, nonconnected graph with n vertices that is **maximal** with respect to these properties. That is, if you tried to make a larger graph in which G is a subgraph, this larger graph will lose at least one of the properties (a) simple, (b) nonconnected, or (c) has n vertices.

What does being maximal with respect to these properties imply about G? That is, what further properties must G possess because of this assumption?

Activity 15.7 An **Euler circuit** is a closed trail in a connected graph G that traverses every edge of G. Since it must be a trail, you could say that an Euler circuit traverses each edge of G *exactly* once (as well as ending at the same node at which it begins).

Prove that if a connected graph contains an Euler circuit, then every vertex in that graph must have even degree.

15.6 Exercises

Recognizing paths and trails. In each of Exercises 1–4, you are given a walk through a graph. Determine whether the walk is a path, a trail, or neither. Also determine whether the walk is open or closed.

1. $v_1, e_1, v_2, e_2, v_3, e_3, v_4, e_{12}, v_6, e_6, v_3$.

2. $v_1, e_1, v_2, e_2, v_3, e_3, v_4, e_4, v_5, e_5, v_6, e_6, v_3, e_7, v_7, e_8, v_1$.

3. $v_1, e_8, v_7, e_7, v_3, e_7, v_7, e_8, v_1$.

4. $v_3, e_3, v_4, e_4, v_5, e_5, v_6, e_6, v_3$.

5. Consider the graph in Figure 15.6.1.

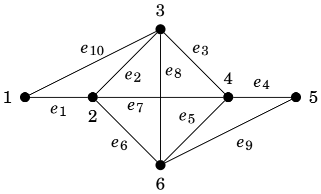


Figure 15.6.1 An example graph.

(a) Determine four different paths from vertex 1 to vertex 5.

(b) Determine four different trails from vertex 1 to vertex 5, none of which are paths.

(c) Determine four different walks from vertex 1 to vertex 5, none of which are trails.

6. Consider the graph in Figure 15.6.2.

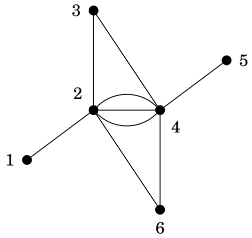


Figure 15.6.2 An example graph.

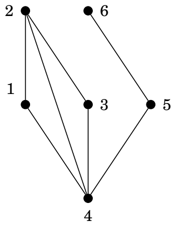
(a) How many different trails are there from vertex 1 to vertex 5?

(b) How many different paths are there from vertex 1 to vertex 5? (*Hint:* See Proposition 15.2.8.)

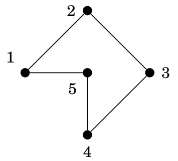
(c) How many different walks are there from vertex 1 to vertex 5?

Recognizing bridges. In each of Exercises 7–10, identify each edge that is a bridge.

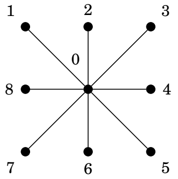
7.



8.



9.



10. The complete graph with n vertices.

11. Among all possible nonconnected graphs with n vertices, let H be one with the maximum number of edges. Prove that H has exactly two connected components.

Hint. Argue by contradiction.

12. Suppose G is a connected graph that contains a closed path that is also a trail. Prove that it is possible to remove any single edge from this path and be left with a connected subgraph of G. That is, prove that no edge in this path could be a bridge.

13. Prove that a graph in which every edge is a bridge cannot have a closed path that is also a trail.

CHAPTER 16

Trees and searches

16.1 Motivation

Example 16.1.1 Reducing redundancy. You have set up your own tree-fort communication system out of tin cans and strings. (See Worked Example 15.4.11.) However, peace has broken out and your communication system is underused. To address the crippling tin-can-and-string stilts shortage, you want to dismantle as much of your network as possible without disrupting communications.

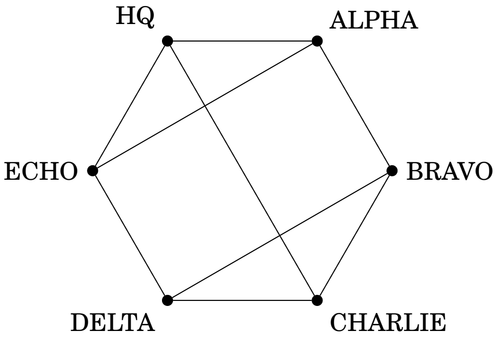
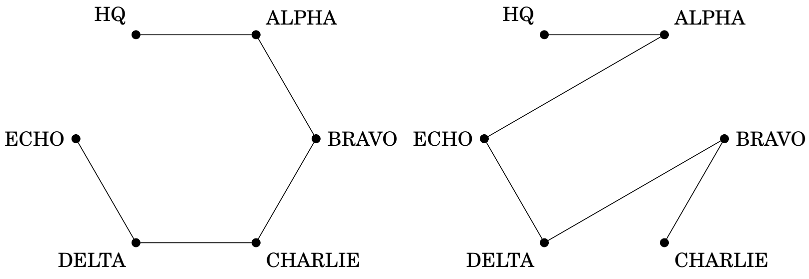


Figure 16.1.2 TreeFort CommNet.

Closed paths are redundant, as communication could be routed around such a path in two directions. So try to eliminate closed all paths; two possible solutions appear in Figure 16.1.3.



(a) One possible result of removing redundancy.

(b) Another possible result of removing redundancy.

Figure 16.1.3 TreeFort CommNet (after removing redundant communication paths).

□

Remark 16.1.4 A more difficult modification of the above example would be to include the length of string in each communication link as a weight for that edge in the graph, and then try to determine a configuration that removes the most string from the network without disrupting communications.

16.2 Basics

trivial path
a path that consists of a single vertex

cycle a closed path

proper cycle
a nontrivial cycle that is also a trail

Note 16.2.1 If G contains vertices v, v' and edge $e = \{v, v'\}$, then v, e, v', e, v is a nontrivial cycle which is not proper.

acyclic graph
contains no proper cycles

forest synonym for **acyclic graph**

tree a connected, acyclic graph

Example 16.2.2 A forest of trees. The graph in Figure 16.2.3 is acyclic. Each of its connected components is a tree.

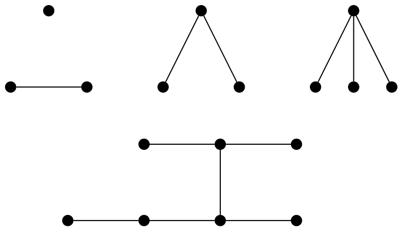


Figure 16.2.3 A nonconnected acyclic graph.

□

Example 16.2.4 Decision trees are trees. In Worked Example 15.2.5, we attempted to determine all possible trails from one node to another in a given graph. The graph in Figure 15.2.6 that we used to explore possible trails in the given graph is an example of a **decision tree** — at each node we "branched out" to new possibilities in continuing the trail. As the name suggested, the connected graph we ended up with is a tree. □

Proposition 16.2.5 Subgraphs of forests.

1. *Every subgraph of an acyclic graph is acyclic.*
2. *Every connected subgraph of an acyclic graph is a tree. In particular, each connected component of an acyclic graph is a tree.*

16.3 Identifying trees

Theorem 16.3.1 Characterizations of trees. *The following are equivalent for a graph $G = (V, E)$ with $|V| = n$ vertices.*

1. *Graph G is a tree.*
2. *Graph G is acyclic but the addition of any new edge would create a cycle.*
3. *Graph G contains no loops and contains exactly one path between each pair of distinct vertices.*
4. *Graph G is connected but every edge of G is a bridge.*
5. *Graph G is connected and has* exactly *$|E| = n-1$ edges.*
6. *Graph G is acyclic and has* exactly *$|E| = n-1$ edges.*

Proof of the equivalence of Statements 1–4.

Statement 1 implies Statement 2. Suppose G is a tree. By definition, it is acyclic. Furthermore, suppose we add an edge between vertices v, v'. Since trees are connected, there was already a path from v to v' in G. Traversing the new edge from v' back to v closes that path to a cycle.

Statement 2 implies Statement 3. Considering the contrapositive, we will assume that Statement 3 is false, and prove that this implies that Statement 2 must also be false.

For Statement 3 to be false, one of the following must be true.

(i) Loops exist in G.

(ii) Some pair of distinct vertices in G is not connected.

(iii) Some pair of distinct vertices in G is connected by more than one path.

In the first case, G would not be acyclic, as a loop is the most basic form of cycle. In the second case, adding an edge between these two vertices that were previously unconnected by a path would *not* create a cycle, as the rest of that cycle other than the new edge would have been a path between the two vertices. And in the third case, if a pair of vertices is connected by more than one path then the parts of two such paths that are different could be concatenated (one forward, one reversed) to create a cycle, so that G must be not be acyclic.

Thus, in all cases that make Statement 3 false, Statement 2 is also false.

Statement 3 implies Statement 4. Again, we will consider the contrapositive, assuming that Statement 4 is false and proving that Statement 3 is also false.

For Statement 4 to be false, one of the following is true.

(i) Graph G is not connected.

(ii) Some edge in G is not a bridge.

In the first case, G must contain at least one pair of vertices that is not connected by *any* path. For the second case, suppose edge e in G is a loop. If e is a loop, then G contains loops. If e is not a loop, then it is an edge between a pair of distinct vertices, say v and v'. But then removing e from G would leave a subgraph G' that still contains both v and v', and which is still connected. So this subgraph (and hence G) must contain a path between v and v' that does not involve e. On the other hand, v, e, v' is also a path between v and v'. So v, v' is a pair of distinct vertices in G for which there is more than one path between them.

Thus, in all cases that make Statement 4 false, Statement 3 is also false.

Statement 4 implies Statement 1. Again, we consider the contrapositive of this logical implication, assuming that Statement 1 is false and proving that Statement 4 is also false. However, since both statements contain the substatement that G is connected, we will not negate that part.

So assume that G is connected but contains a proper cycle. We aim to prove that *at least one* edge in G is not a bridge. In Activity 16.4, you are asked to prove that none of the edges in the proper cycle that G contains is a bridge, which will complete the proof. ■

Proof of the equivalence of Statement 1, Statement 5, and Statement 6.

Statement 1 implies Statement 5. Assume that G is a tree. Then it is connected. To prove that the number of edges is $|E| = n - 1$, we proceed by (strong) induction on n, the number of vertices in G.

For the base base case of $n = 1$, $|E| = 0$ is the only possibility, as loops are not allowed in a tree.

Now the induction step. Assume that every tree with $k < n$ vertices has $k - 1$ edges. Choose some edge in G. By Statement 4, removing that edge creates two connected components, G_1 and G_2. As G is acyclic, these connected components are both trees (Statement 2 of Proposition 16.2.5). Let k_1, k_2 represent the number of vertices in G_1, G_2, respectively, so that $k_1 + k_2 = n$. Since each of k_1 and k_2 must be strictly less than n, we may apply our indution hypothesis to each of G_1 and G_2, so that G_1 has exactly $k_1 - 1$ edges and G_2 has exactly $k_2 - 1$ edges.

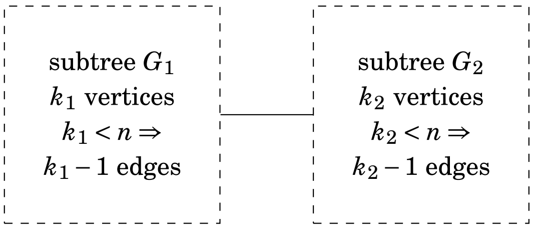


Figure 16.3.2 Tree G splits into subtrees G_1, G_2 after removal of an edge.

Adding up the number of edges in G_1 and G_2, along with the single edge in G that was removed to create these two connected components, we obtain $|E| = (k_1 - 1) + (k_2 - 1) + 1 = (k_1 + k_2) - 1 = n - 1$, as desired.

Statement 5 implies Statement 6. Consider the contrapositive of this logical implication, assuming that Statement 6 is false and proving that Statement 5 is also false. However, since both statements contain the substatement that $|E| = n - 1$, we will not negate that part.

So assume that G has $n - 1$ edges, but contains a proper cycle. We must prove that G cannot be connected in this case. Choose an edge e in the proper cycle and create a subgraph G' by removing e. Subgraph G' now has n vertices but $n - 2$ edges, and so by the contrapositive of Theorem 15.3.11 G' cannot be connected. That means that G' contains a pair of vertices v_1, v_2 between which no walk exists. If there is a walk between v_1, v_2 in G but not in G', then that walk must involve the chosen e. But then there would be another walk between v_1, v_2 in G avoiding e via the rest of the proper cycle containing e. And this other walk would be in G' since it does not involve e. Except that we assumed there was *no* walk between v, v' in G', hence there also can be no walk between them in G. Thus, G is not connected.

Statement 6 implies Statement 1. Again, consider the contrapositive of this logical implication, assuming that Statement 1 is false and proving that Statement 6 is also false. However, since both statements contain the substatement that G is acyclic (part of the definition of **tree**), we will not negate that part.

So assume that G is acyclic but not a tree, i.e. that G is not connected. We must prove that the number of edges in G is different from $n-1$, where n is the number of vertices in G. Let $G_1, G_2, \ldots, G_\ell$ be the connected components of G. Now, since G is assumed acyclic, each of these connected components is a tree (Statement 2 of Proposition 16.2.5). We have already proved above that Statement 1 implies Statement 5, so if we write k_i for the number of vertices in component G_i, then we may conclude that the number of edges in component G_i is $k_i - 1$. As the components make up the entire graph G, we may add up the vertices and edges in each component to get the totals in the full graph:

$$n = k_1 + k_2 + \cdots + k_\ell,$$

$$\begin{aligned} |E| &= (k_1 - 1) + (k_2 - 1) + \cdots + (k_\ell - 1) \\ &= (k_1 + k_2 + \cdots + k_\ell) - \ell \\ &= n - \ell. \end{aligned}$$

Since we assume G is not connected, we have $\ell \geq 2$, and so $|E| \neq n-1$ as desired. ■

16.4 Depth-first and breadth-first searches

Let G be a graph. Given vertices v, v' of G, we might wish to find a path from v to v', if one exists. We can do this by constructing a tree $T \preceq G$.

Algorithm 16.4.1 Depth-first search. *To create a tree T that is a subgraph of a graph G wherein a path (in G) from v to v' is evident, begin with T containing the single vertex v and no edges. Set $x = v$.*

1. *Look for a vertex y of G which is adjacent to x but not already in T. If such a y is found, go to Step 2. Otherwise, go to Step 3.*

2. *Adjoin y and a single edge between x and y to T. If $y = v'$, stop — a path from v to v' exists and is now contained in T. Otherwise, set $x = y$ and return to Step 1.*

3. *If you have arrived here immediately after beginning the algorithm (i.e. with x still set to be v), stop — there is no path from v to v'. Otherwise, return to the vertex z adjoined before x. Set $x = z$ and return to Step 1.*

Note 16.4.2 In Step 1 of the algorithm, there is no specification on how to choose a single y satisfying the search criteria from multiple possibilities. In other words, there is flexibility in the **implementation** of the algorithm here, and for the problem at hand there may be implementation choices that are more expedient then others.

Example 16.4.3 Carrying out a depth-first search. We perform a Depth-first search on the graph in Figure 16.4.4, attempting to find a path from vertex 1 to vertex 9. In carrying out the algorithm, if we always choose the vertex with the smallest label in Step 1, we obtain the graph in Figure 16.4.5(a). The graph in Figure 16.4.5(b) is the result of always choosing the vertex with the largest label.

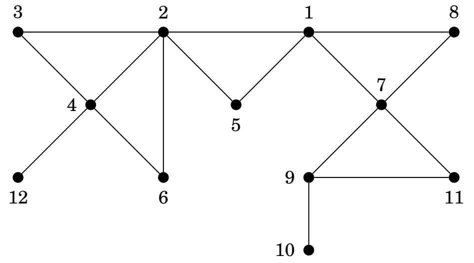
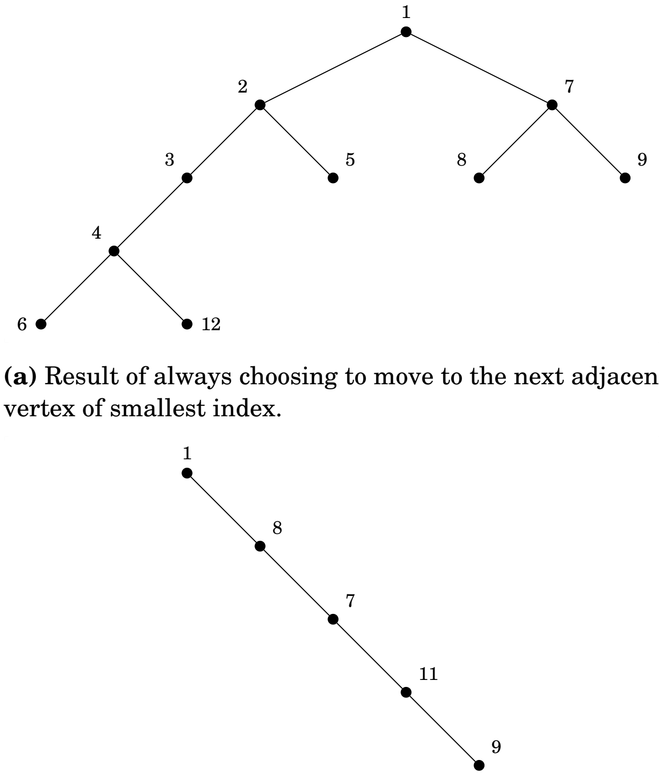


Figure 16.4.4 An example graph to illustrate depth-first search.



(a) Result of always choosing to move to the next adjacent vertex of smallest index.

(b) Result of always choosing to move to the next adjacent vertex of largest index.

Figure 16.4.5 Results of two different implementation choices in a depth-first search.

□

The depth-first search will not necessarily yield the *shortest* path from v to v'. The following algorithm will.

Algorithm 16.4.6 Breadth-first search. *To create a tree T that is a subgraph of a graph G wherein the* shortest *path in G from v to v' is evident, begin with T containing the single vertex v and no edges.*

1. *For each vertex x in T added in the last application of this step (or, in the case of the first application of this step, for $x = v$), adjoin* all *vertices of G that are adjacent to x and not already in T, along with a single edge between each such vertex and x. If at least one vertex has been adjoined to T in this step, proceed to Step 2. Otherwise, stop — there is no path from v to v' in G.*

2. *If v' was one of the vertices adjoined in Step 1, stop — a path from v to v'*

exists and is now contained in T. Otherwise, return to Step 1.

Example 16.4.7 Carrying out a breadth-first search. Below is the result of the breadth-first algorithm, carried out to find a path from 1 to 9 in the graph in Figure 16.4.4 from Example 16.4.3.

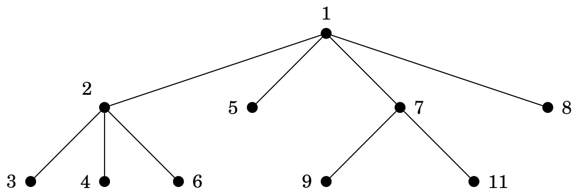


Figure 16.4.8 The result of a breadth-first search.

□

16.5 Spanning trees

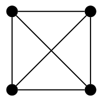
spanning subgraph

a subgraph that contains *all* the vertices of the parent graph

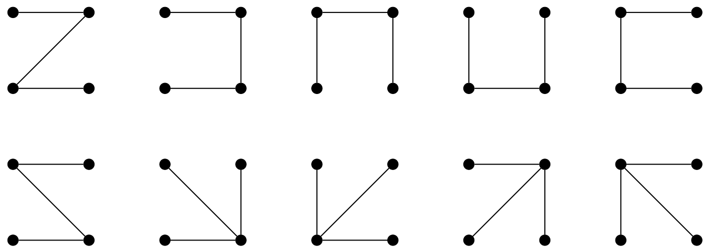
spanning tree

a spanning subgraph that is a tree

Example 16.5.1 Spanning trees for the complete graph K_4. Here is the complete graph with four vertices.



And here are ten different spanning trees for K_4.



□

If we carry out either of the depth-first or breadth-first search algorithms, but aren't looking for a path between specific vertices, the end result will be a spanning tree for the original graph.

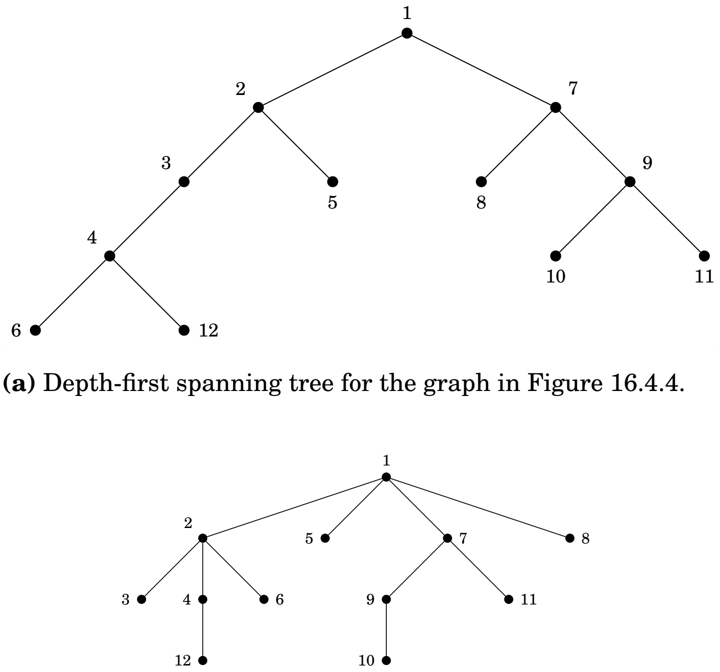
depth-first spanning tree

the result of performing the depth-first search algorithm on a graph, continuing until all vertices in the original graph appear in the search tree

breadth-first spanning tree

the result of performing the breadth-first search algorithm on a graph, continuing until all vertices in the original graph appear in the search tree

Example 16.5.2 Depth-first and breadth-first spanning trees. Figure 16.5.3 contains depth-first and breadth-first spanning trees for the graph in Figure 16.4.4, our source of examples for depth-first search (Example 16.4.3) and breadth-first search (Example 16.4.7).



(a) Depth-first spanning tree for the graph in Figure 16.4.4.

(b) Breadth-first spanning tree for the graph in Figure 16.4.4.

Figure 16.5.3 Examples of depth- and breadth-first spanning trees.

□

16.6 Binary searches

binary search tree
a tree in which every node has degree 1 or 3, except for a single node of degree 2

initial node
the unique node of degree 2 in a binary search tree

terminal node
a node of degree 1 in a binary search tree

binary search
the construction of a binary search tree through a series of "either-or" decisions

Worked Example 16.6.1 Estimate the root of $f(x) = 4x^3 + 6x^2 + 3x - 1$ that lies in $(0,1)$ to 2 decimal places.

Solution. The Intermediate Value Theorem from first-year calculus says that if f is continuous on the closed interval $[a,b]$ and $f(a), f(b)$ are nonzero and opposite signs, then f has a root in the open interval (a,b). We have $f(0) = -1 < 0$

and $f(1) = 12 > 0$, so there is indeed a root in $(0,1)$. The graph in Figure 16.6.2 was obtained by performing a binary search by splitting into subintervals.

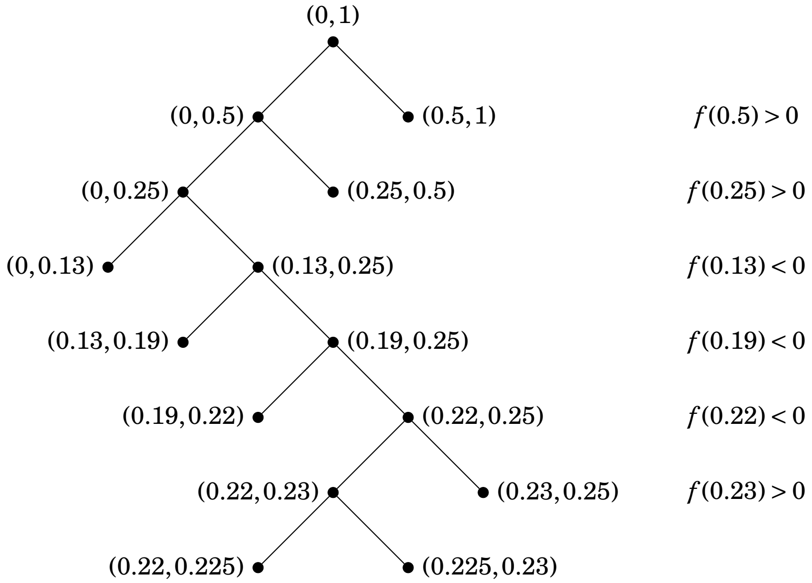


Figure 16.6.2 A binary search tree search for the root of a polynomial.

Since $f(0.225) > 0$, the root must be in the subinterval $(0.22, 0.225)$. This tells us to round down to 0.22 instead of rounding up to 0.23, so we conclude that the root is approximately 0.22. □

16.7 Activities

Activity 16.1

(a) Draw two different connected graphs with five vertices each in which every edge is a bridge.

(b) How many edges are in each of the examples that you drew in Task a?

(c) Would it be possible to add an edge to either of the examples that you drew in Task a *without* creating a cycle?

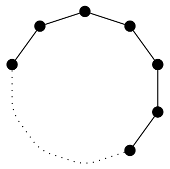
Activity 16.2

(a) Draw two different simple graphs with 5 vertices in which every pair of vertices has a single path between them.

(b) How many edges are in each of the examples that you drew in Task a?

(c) Would it be possible to add an edge to either of the examples that you drew in Task a *without* creating a cycle?

Activity 16.3 Suppose that G is a connected graph that *consists entirely of* a proper cycle. (See Figure 16.7.1.)



Graph G

Figure 16.7.1 A graph that consists entirely of a proper cycle.

Let G' represent the subgraph of G that results by removing a single edge. Argue that G' remains connected.

Activity 16.4 Suppose that H is a connected graph that *contains* a proper cycle. Let H' represent the subgraph of H that results by removing a single edge from H, where the edge removed is part of the proper cycle that Hcontains. Argue that H' remains connected.

Notes.

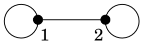
- Your argument here needs to be (slightly) different from your argument in Activity 16.3.
- Make sure you are using the technical definition of **connected graph** in your argument. What are you assuming about H, and what do you need to verify about H'?

16.8 Exercises

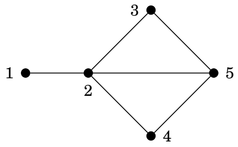
1. Prove that if a graph contains a closed trail then it also contains a proper cycle.

Spanning trees. For each of the graphs in Exercises 2–3, draw a spanning tree by inspection.

2.



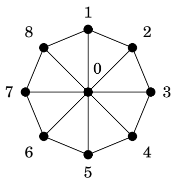
3.



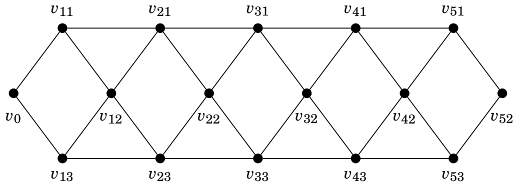
Reducing to a spanning tree. For each of the graphs in Exercises 4–5, use the following algorithm to obtain a spanning tree.

- If the graph contains a proper cycle, remove one edge of that cycle.
- If the resulting subgraph contains a proper cycle, remove one edge of that cycle.
- If the resulting subgraph contains a proper cycle, remove one edge of that cycle.
- etc..
- Continue until there are no proper cycles left.

4.

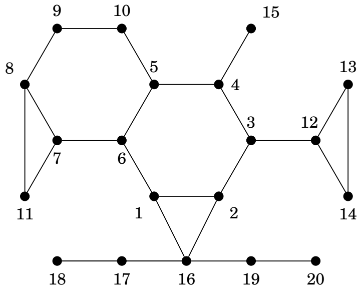


5.

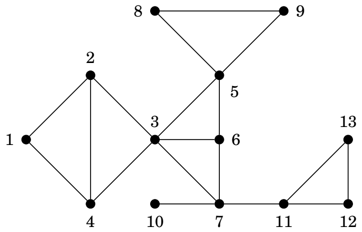


Depth-first and breadth-first spanning trees. For each of the graphs in Exercises 6–8, determine both a depth-first and breadth-first spanning tree. Use any vertex you like as the starting node.

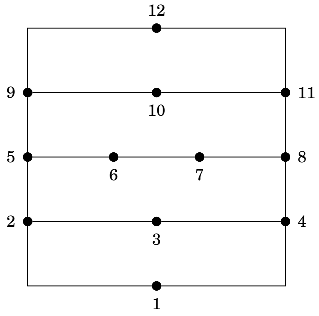
6.



7.



8.



Part V

More Set Theory

CHAPTER 17

Relations

17.1 Basics

relation (working definition)

a rule which assigns to *some* elements of a set A several elements from a set B

$a\,R\,b$ element $a \in A$ is related to element $b \in B$ by relation R

relation on a set

a relation between elements of the same set

Remark 17.1.1 Compare this working definition of **relation** with our **working definition of function** in Section 10.1.

As the name implies, a relation describes some relationship of elements of a set A to elements of a set B.

Example 17.1.2 Pet relation. Let C represent the set of living cats, and let H represent the set of living humans. Then one relationship between elements of these two sets can be expressed by writing $c\,R\,h$ to mean that cat c is the pet of human h. □

Example 17.1.3 Parent relation. Let H represent the set of living humans. Then one type of relationship between elements of this set can be expressed by writing $h_1\,R\,h_2$ to mean that human h_1 is the parent of human h_2. □

Example 17.1.4 Division relation. One type of relationship between elements of $\mathbb{N}_{>0}$ can be expressed by writing $m \mid n$ to mean that nonzero natural number m divides nonzero natural number n. □

Just as with functions, we want to avoid the use of the undefinable word "rule". Notice that a relation just pairs elements of a set A with elements of a set B; we have seen this before.

relation (formal definition)

a subset of a Cartesian product

With this formal definition, writing $R \subseteq A \times B$ becomes the same as saying "R is a relation between elements of A and B," and writing $(a,b) \in R$ becomes the same as writing $a\,R\,b$.

Note 17.1.5 With this formal definition, a relation on a set A means a subset of $A \times A$.

Remark 17.1.6 Recall that our **formal definition of function** states that a function $A \to B$ is a *special kind* of subset of $A \times B$. But *every* subset of $A \times B$ can be considered as a relation, so a function is a special kind of relation.

The difference is that a function $A \to B$ must assign *exactly one* element of B to *each* element of A, whereas a relation from A to B can assign any number of elements of B (even zero) to each element of A. That is, a relation does not have to be **well-defined**, and can be left **undefined** on some elements of A.

See. Example 10.1.16 and Example 10.1.17.

Example 17.1.7 Identity relation. Consider

$$R = \{\, (a,a) \mid a \in A \,\} \subseteq A \times A.$$

Then $a_1 \, R \, a_2$ means $a_1 = a_2$. This relation is the same as the identity function $\mathrm{id}_A : A \to A$. □

Example 17.1.8 Element relation. Consider

$$R = \{\, (a,C) \mid a \in C \,\} \subseteq A \times \mathscr{P}(A).$$

Then $a \, R \, C$ means $a \in C$. This relation is in general not a function, since it is not well-defined: an element of A can be contained in several subsets of A. □

A relation between pairs of objects, such as the ones we have considered so far, is sometimes called a **binary relation**. But we can consider relationships between collections of more than two objects.

ternary relation
a subset of $A \times B \times C$ for sets A,B,C

Example 17.1.9 A human (usually) has *two* biological parents. Let H represent the set of all living humans. Then we can define a ternary relation $R \subseteq H^3$ by taking $(h_1,h_2,h_3) \in R$ to mean that humans h_1,h_2 are the parents of human h_3. □

17.2 Operations on relations

Viewing relations as subsets of Cartesian products suggests ways to build new relations from old.

union (of relations R_1,R_2)
the relation where $a \, (R_1 \cup R_2) \, b$ means that at least one of $a \, R_1 \, b$ or $a \, R_2 \, b$ is true

intersection (of relations R_1,R_2)
the relation where $a \, (R_1 \cap R_2) \, b$ means that both $a \, R_1 \, b$ and $a \, R_2 \, b$ are true

complement (of relation R)
the relation where $a \, R^{\mathrm{c}} \, b$ means that $a \, R \, b$ is *not* true

$a \not R b$ alternative notation for $a \, R^{\mathrm{c}} \, b$

Note 17.2.1 Considering relations as subsets of Cartesian products, the above relation operations mean precisely the same thing as the corresponding set operations.

Example 17.2.2 Union of "less than" and "equal to" relations. Consider the relations $<$ and $=$ on $\mathbb{R}$, and let R be the union $< \cup =$. Then $x \, R \, y$ means that at least one of $x < y$ or $x = y$ is true. That is, R is the same as the relation $\leq$. □

Example 17.2.3 Sibling relations. Let H represent the set of all living humans. Let relations $R_F, R_M \subseteq H \times H$ be defined by

- $a\, R_F\, b$ if a, b have the same father; and
- $a\, R_M\, b$ if a, b have the same mother.

Set $R_P = R_F \cap R_M$. Then $a\, R_P\, b$ means that a, b have the same parents. □

Example 17.2.4 Complement of the subset relation. Let U be a universal set and consider the relation $\subseteq$ on $\mathscr{P}(U)$. Then $A \subseteq^c B$ means that A is *not* a subset of B, which can only happen if some elements of A are not in B. In other words, $A \subseteq^c B$ means that $A \cap B^c \neq \emptyset$.

Careful. Relation $A \subseteq^c B$ does not (necessarily) mean $A \subseteq B^c$. Draw a representative Venn diagram to see why.

□

Unlike functions, which can only be reversed if bijective, every relation can be reversed by simply stating the relationship in the reverse order.

inverse (of a relation R)
the relation where $b\, R^{-1}\, a$ means that $a\, R\, b$ is true

Note 17.2.5

- As subsets of Cartesian products, if $R \subseteq A \times B$, then $R^{-1} \subseteq B \times A$, and $(a,b) \in R$ if and only if $(b,a) \in R^{-1}$.
- A relation R and its inverse R^{-1} express the *same relationship* between elements of two sets A and B, just phrased in the opposite order. In logical terms, $b\, R^{-1}\, a \Leftrightarrow a\, R\, b$.

Example 17.2.6 Parent/child relations. Let H represent the set of all living humans, and let R represent the relation on H where $h_1\, R\, h_2$ means human h_1 is the parent of human h_2. Then $h_2\, R^{-1}\, h_1$ means human h_2 is the child of human h_1. Both relations express the same information, but in a different order. □

Example 17.2.7 Inverse of division relation. Recall that $\mid$ is a relation on $\mathbb{N}_{>0}$ where $m \mid n$ means that m divides n. Then for the inverse relation, $n|^{-1}m$ means n is a multiple of m. Both relations express the same information, but in a different order. □

Example 17.2.8 Inverse of logical equivalence. Let $\mathcal{L}$ represent the set of all possible logical statements. We have a relation $\equiv$ on $\mathcal{L}$, where $A \equiv B$ means that logical statement A involves the same statement variables and has the same truth table as logical statement B. Since $A \equiv B$ if and only if $B \equiv A$, we conclude that the logical equivalence relation on $\mathcal{L}$ is its own inverse. □

There are two more set-theoretic ideas we can reinterpret as relations.

empty relation
the relation between sets A and B corresponding to the empty subset $\emptyset \subseteq A \times B$, so that $a\, \emptyset\, b$ is always false

universal relation
the relation between sets A and B corresponding to the full subset $U = A \times B \subseteq A \times B$, so that $a\, U\, b$ is always true

17.3 Properties of relations

Here we list some important properties a relation R on a set A can have.

17.3.1 Reflexivity

reflexive $a \, R \, a$ is true for all $a \in A$

Example 17.3.1 A reflexive and a non-reflexive relation on the set of real numbers. The relation $\leq$ on $\mathbb{R}$ is reflexive, but the relation $<$ is not. □

Test 17.3.2 Reflexive relation. *To verify that relation R on set A is reflexive, prove that* $(\forall a \in A)(a \, R \, a)$.

17.3.2 Symmetry and antisymmetry

symmetric

for every pair of elements $a_1, a_2 \in A$ for which $a_1 \, R \, a_2$ is true, $a_2 \, R \, a_1$ is also true

Example 17.3.3 Sibling relation is symmetric, brother/sister relation is not. On the set of all living humans, the relation "a is the sibling of b" is symmetric, but neither the relation "a is the brother of b" nor the relation "a is the sister of b" is symmetric. □

Test 17.3.4 Symmetric relation. *To verify that relation R on set A is symmetric, prove that*

$$(\forall a_1 \in A)(\forall a_2 \in A)(a_1 \, R \, a_2 \Rightarrow a_2 \, R \, a_1).$$

antisymmetric

for every pair of *distinct* elements $a_1, a_2 \in A$, either $a_1 \not\!R \, a_2$ or $a_2 \not\!R \, a_1$ (or both)

Remark 17.3.5 The *distinct* part of the definition is important, since if $a_1, a_2 \in A$ are *not* distinct (i.e. $a_2 = a_1$), then obviously both $a_1 \, R \, a_2$ and $a_2 \, R \, a_1$ can be simultaneously true because they are the same statement.

Example 17.3.6 An antisymmetric relation on real numbers. The relation $\leq$ on $\mathbb{R}$ is antisymmetric. □

Example 17.3.7 A relation can be *neither* antisymmetric nor symmetric. On $A = \{a, b, c\}$, the relation

$$R = \{(a,b),(b,a),(a,c)\} \subseteq A \times A$$

is neither antisymmetric nor symmetric. □

Example 17.3.8 A relation can be *both* antisymmetric and symmetric. The identity relation on any set, where each element is related to itself and only to itself, is both antisymmetric and symmetric. □

Remark 17.3.9 As Example 17.3.7 and Example 17.3.8 demonstrate, antisymmetry is *not* the opposite of symmetry. However, for a relation R on set A, we may think of symmetry and antisymmetry as being at opposite ends of a spectrum, measuring how often we have *both* $a_1 \, R \, a_2$ and $a_2 \, R \, a_1$ for $a_1 \neq a_2$.

By definition, **antisymmetry** is when we *never have both*. On the other hand, **symmetry** is when we *always have both or neither*; that is, for every distinct pair $a_1, a_2 \in A$, we either have both $a_1 \mathrel{R} a_2$ and $a_2 \mathrel{R} a_1$, or we have both $a_1 \not\mathrel{R} a_2$ and $a_2 \not\mathrel{R} a_1$. However, a relation can fall *between* symmetry and antisymmetry on the spectrum, such as in Example 17.3.7, where we sometimes have both (e.g. both $a \mathrel{R} b$ and $b \mathrel{R} a$ for that example relation) and we also sometimes have only one (e.g. $a \mathrel{R} c$ but $c \not\mathrel{R} a$ for that example relation).

The equality relation on a set is a special case that is *both* symmetric and antisymmetric. In fact, equality is essentially the *only* relation that is both symmetric and antisymmetric — see Exercise 17.6.22.

In symbolic language, the definition of **antisymmetric relation** is

$$(\forall a_1 \in A)(\forall a_2 \in A)(a_1 \neq a_2 \Rightarrow a_1 \not\mathrel{R} a_2 \vee a_2 \not\mathrel{R} a_1).$$

However, in practise we usually prove antisymmetry using one of two logically equivalent formulations.

Test 17.3.10 Antiymmetric relation. *To verify that relation R on set A is antisymmetric, prove either* ***one*** *of the following logical statements.*

- $(\forall a_1 \in A)(\forall a_2 \in A)(a_1 \neq a_2 \wedge a_1 \mathrel{R} a_2 \Rightarrow a_2 \not\mathrel{R} a_1)$
- $(\forall a_1 \in A)(\forall a_2 \in A)(a_1 \mathrel{R} a_2 \wedge a_2 \mathrel{R} a_1 \Rightarrow a_2 = a_1)$

Remark 17.3.11 The first formulation for proving antisymmetry provided above can be thought of as just a different way to say that it is not possible to have both $a_1 \mathrel{R} a_2$ and $a_2 \mathrel{R} a_1$ for distinct elements a_1, a_2. The second formulation essentially says that the only possible way to have both $a_1 \mathrel{R} a_2$ and $a_2 \mathrel{R} a_1$ is if $a_2 = a_1$.

Note 17.3.12 In Exercise 17.6.21 you are asked to prove that each of the two different ways of verifying that a relation is antisymmetric provided in the test above are equivalent.

17.3.3 Transitivity

transitive for every triple of elements $a_1, a_2, a_3 \in A$ for which both $a_1 \mathrel{R} a_2$ and $a_2 \mathrel{R} a_3$ are true, $a_1 \mathrel{R} a_3$ must also be true

Example 17.3.13 Ancestry is transitive. The relation on the set of all humans who ever lived defined by "a is the ancestor of b" is transitive. □

Test 17.3.14 Transitive relation. *To verify that relation R on set A is transitive, prove that*

$$(\forall a_1 \in A)(\forall a_2 \in A)(\forall a_3 \in A)(a_1 \mathrel{R} a_2 \wedge a_2 \mathrel{R} a_3 \Rightarrow a_1 \mathrel{R} a_3).$$

17.4 Graphing relations

Recall that if R is a relation on a set A, then formally R is a subset $A \times A$. In other words, R is a collection of *ordered pairs* of elements from A.

Also recall that in a *directed* graph, the edge collection is formally defined to be a collection of ordered pairs of vertices. So when the set A is finite, we may regard A as a set of vertices and R as a collection of (directed) edges in a graph!

To summarize, we may represent a relation $R \subseteq A \times A$ by the directed graph (A, R). The vertices of the graph are the elements of A, and for elements $a_1, a_2 \in A$, we draw an arrow from a_1 to a_2 if $a_1 \mathrel{R} a_2$ is true.

Example 17.4.1 Graph of the division relation on a finite set of natural numbers. Recall that for natural numbers m and n, $m \mid n$ means "m divides n". Consider this relation on the finite set $A = \{2,3,4,5,6,7,8,9,10\}$. The graph of this relation appears in Figure 17.4.2.

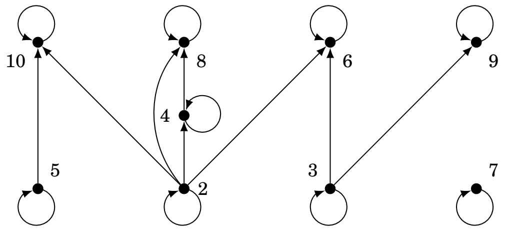


Figure 17.4.2 Graph of the division relation on a small set of natural numbers.

Note that each vertex has a loop since every number divides itself. □

Example 17.4.3 Graph of an inverse relation. Using the same division relation on the same set A as in Example 17.4.1 above, we may obtain the graph for the inverse relation by just reversing the direction of all the arrows in the graph in Figure 17.4.2. □

Question 17.4.4 How are the properties of a relation reflected in its graph? □

Reflexive relations. In this case, $a\,R\,a$ is true for every element $a \in A$, so every vertex has a loop. For example, the relation in Example 17.4.1 is reflexive, and we see this mirrored in the graph in Figure 17.4.2 by the placement of a loop at every node.

Remark 17.4.5 When a relation is understood to be reflexive, we often omit the loops from its graph to reduce visual clutter.

Symmetric relations. In this case, the conditional $a_1\,R\,a_2 \Rightarrow a_2\,R\,a_1$ is always true. Therefore, in the graph for R, whenever we have an arrow from a_1 to a_2, we must also have an arrow from a_2 to a_1.

Example 17.4.6 Graph of a symmetric relation. On the set $A = \{a,b,c,d\}$, the relation

$$R = \{(a,b),(b,a),(b,c),(c,b)\}$$

is symmetric, and we see this property reflected in the graph in Figure 17.4.7, as each pair of related (distinct) nodes has an arrow in each direction between them.

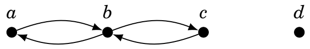


Figure 17.4.7 The graph of a basic symmetric relation.

□

Remark 17.4.8 When R is symmetric, arrows are essentially meaningless since between every pair of vertices we will have either no arrows or one arrow in each direction. So we may as well draw the graph for R as an ordinary (undirected) graph instead of a directed graph, replacing each pair of arrows with a single edge.

Example 17.4.9 Simplified graph of a symmetric relation. The relation in the previous example is more concisely depicted graphically as in Figure 17.4.10 below.

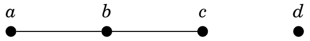


Figure 17.4.10 The simplified graph of a basic symmetric relation.

□

Antisymmetric relations. In this case, we *never* have both $a_1 \; R \; a_2$ and $a_2 \; R \; a_1$ for $a_1 \neq a_2$, so in the graph for R, *no* pair of vertices can have two oppositely-directed arrows between them.

Example 17.4.11 Graph of an antisymmetric relation. On the set $A = \{a,b,c,d\}$, the relation

$$R = \{(a,a),(a,b),(c,b)\}$$

is antisymmetric, and we see this property reflected in the graph in Figure 17.4.12, as each pair of distinct nodes has *at most* one arrow between them.

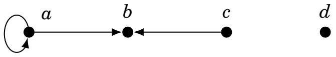


Figure 17.4.12 The graph of a basic antisymmetric relation.

□

Transitive relations. In this case, the conditional $a_1 \; R \; a_2 \wedge a_2 \; R \; a_3 \Rightarrow a_1 \; R \; a_3$ is always true. Therefore, in the graph for R, every "chain" of two arrows has a corresponding "composite" arrow.

Example 17.4.13 Graph of an transitive relation. On the set $A = \{a,b,c,d,e\}$, the relation

$$R = \{(a,b),(b,c),(a,c),(d,e),(e,d),(d,d),(e,e)\}$$

is transitive, and we see this property reflected in the graph in Figure 17.4.14, as each pair of arrows forming a "chain" between three nodes has a corresponding "composite" arrow from the first node in the chain to the third.

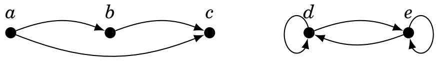


Figure 17.4.14 The graph of a basic antisymmetric relation.

□

Remark 17.4.15 In the graph of a transitive relation, we often omit the "composite" arrows to reduce visual clutter, as we can infer from "chains" of arrows where the "composite" arrows would go. For example, we did this in both the power set graph in Example 14.4.1 (see Figure 14.4.2) and in the division graph in Example 14.4.3 (see Figure 14.4.4). It should be obvious that the relations "is a subset of" and "divides" are transitive, so there was no need to clutter up the graphs of those relations with extra "composite" arrows — we could trace the fact that one set was a subset of another or the fact that one number divides another by following a chain of arrows through intermediate nodes, as necessary.

17.5 Activities

Activity 17.1 In each of the following, describe the requested combination of relations in words (i.e. in the form "a is related to b if . . . "). Try to "simplify" your description, if possible.

In Task h and Task i, the symbol $\equiv_k$ represents a relation on $\mathbb{Z}$, where $m \equiv_k n$ means that m and n have the same remainder when divided by k. (It may help to know that this is equivalent to k dividing the difference $m-n$.)

(a) $<\cup>$ on $\mathbb{R}$.

(b) Union of "longer than" and "shorter than" on Σ^* for some alphabet Σ.

(c) Union of "longer than", "shorter than", and "same length as" on Σ^* for some alphabet Σ.

(d) Intersection of "longer than" and "shorter than" on Σ^* for some alphabet Σ.

(e) The complement of $\leq$ on $\mathbb{R}$.

(f) The inverse of $\leq$ on $\mathbb{R}$.

(g) The inverse of "$x\,R\,y$ if $2x+3y=0$" on $\mathbb{R}$.

(h) The intersection of $\equiv_5$ and $\equiv_7$ on $\mathbb{Z}$.

(i) The intersection of $\equiv_2$ and $\equiv_4$ on $\mathbb{Z}$.

Activity 17.2 In each of the following, you are given a set A and a relation R on A. Determine which of the properties **reflexive**, **symmetric**, **antisymmetric**, and **transitive** R possesses.

(a) $A=\mathbb{R}$, R is $<$.

(b) A is the set of all straight lines in the plane, R means "is parallel to."

(c) A is the set of all straight lines in the plane, R means "is perpendicular to."

(d) $A=\Sigma^*$ for some alphabet Σ, R means "is the same length as."

(e) $A=\Sigma^*$ for some alphabet Σ, R means "is shorter than."

(f) $A=\Sigma^*$ for some alphabet Σ, x is some fixed choice of letter in Σ, R means "contains the same number of occurrences of x as."

(g) A is an arbitrary set, R is the **empty relation**.

(h) A is an arbitrary set, R is the **universal relation**.

Activity 17.3

(a) Suppose R is a relation on a set A. Convince yourself that $R\cup R^{-1}$ is symmetric. (See the Symmetric Relation Test.)

(b) Recall that $|$ represents the relation "divides" on sets of integers. Draw the directed graph for $|$ on the set $A=\{2,4,6,8,10,12,14,16\}$. Then describe how to obtain the graph for the symmetric relation $|\cup|^{-1}$ as an undirected graph from the graph of R using only an eraser.

Activity 17.4 For each of the properties **reflexive**, **symmetric**, **antisymmetric**, and **transitive**, carry out the following.

Assume that R and S are nonempty relations on a set A that both have the property. For each of R^c, $R\cup S$, $R\cap S$, and R^{-1}, determine whether the new relation

(i) must also have that property;

(ii) might have that property, but might not; or

(iii) cannot have that property.

Any time you answer Statement i or Statement iii, outline a proof. Any time you answer Statement ii, provide two examples: one where the new relation has the property, and one where the new relation does not. (You may use graphs to describe your examples.)

17.6 Exercises

Directed graph for a relation. In each of Exercises 1–4, you are given a relation on a specific set. Draw a directed graph that represents the relation.

1. Relation $\subsetneq$ on $\mathscr{P}(\{a,b,c\})$.

2. Relation $<$ on $\{1,2,3,4\}$.

3. Relation $\equiv_3$ on $\mathbb{N}_{<13}$.

4. Relation "has the same number of occurrences of the letter a as" on Σ_4^* for alphabet $\Sigma = \{\mathrm{a},\mathrm{z}\}$.

5. Recall that a relation on a set A is just a subset of the Cartesian product $A \times A$. Write out all relations on the set $A = \{a,b\}$ as subsets of $A \times A$. Which of these relations are reflexive? Symmetric? Antisymmetric? Transitive?

Testing reflexivity/symmetry/antisymmetry/transitivity. In each of Exercises 6–17, you are given a relation on a specific set. Determine which of the properties **reflexive**, **symmetric**, **antisymmetric**, and **transitive** the given relation possesses.

6. Relation $<$ on $\mathbb{R}$.

7. Relation $\geq$ on $\mathbb{R}$.

8. Relation $|$ on $\mathbb{Z}$.

9. Relation $\subseteq$ on $\mathscr{P}(X)$, where X is an arbitrary, unspecified set.

10. Relation "is taller than" on the set of all living humans.

11. Relation "is parallel to" on the set of all straight lines in the plane.

12. Relation "is perpendicular to" on the set of all straight lines in the plane.

13. Relation "has the same length as" on Σ^*, where Σ is an arbitrary, unspecified alphabet set.

14. Relation "is shorter than" on Σ^*, where Σ is an arbitrary, unspecified alphabet set.

15. Relation "contains the same number of occurrences of the letter x as" on Σ^*, where Σ is an arbitrary, unspecified alphabet set and x is some fixed choice of letter in Σ.

16. Relation $\Leftrightarrow$ on the set of all logical statements involving the statement variables $p_1, p_2, p_3, \ldots$.

17. Relation R defined by "$a_1 \mathrel{R} a_2$ if $f(a_1) = f(a_2)$" on a set A, where $f\colon A \to B$ is an arbitrary, unspecified function.

Properties of relations reflected in their graphs. In each of Exercises 18–19, you are given a list of properties. Draw the directed graph of a relation on the set $\{a,b,c,d\}$ that possesses the given properties.

18. Symmetric and transitive, but neither reflexive nor antisymmetric.

19. Reflexive, antisymmetric, and transitive, but not symmetric.

20. Prove that a relation is symmetric if and only if it is equivalent to its own inverse relation.

21. As described in Section 17.3, the definition of **antisymmetric relation** can be formulated in symbolic language as

$$(\forall a_1 \in A)(\forall a_2 \in A)(a_1 \neq a_2 \Rightarrow a_1 \not\mathrel{R} a_2 \vee a_2 \not\mathrel{R} a_1).$$

Prove that each of the two conditionals provided in the Antisymmetric Relation Test are equivalent to the symbolic formulation of the definition of antisymmetric given above.

22. Suppose R is a relation on a set A that is both symmetric and antisymmetric. Prove that R is a subset of the identity relation $\{(x,x) \mid x \in A\}$.

CHAPTER 18

Equivalence relations

18.1 Motivation

There are often situations where we want to group certain elements of a set together as being "the same."

Example 18.1.1 At the end of this course, your instructor will assign each student a grade. In this system, every student who receives a "B" had roughly the same performance in the course (in principle, anyways). That is, consider the set of all students in this course — from the point of view of the grading system, the students in the subset of "B" students are all **equivalent** in performance. □

18.2 Basics and examples

What properties should a relation on a set have to be useful as a notion of "equivalence"?

- Each object in the set should be equivalent to itself. So the relation should be **reflexive**.
- Equivalence should be bidirectional. That is, a pair of equivalent objects should be equivalent to *each other*. So the relation should be **symmetric**.
- We should be able to infer equivalence from chains of equivalence. So the relation should be **transitive**.

equivalence relation
a relation on a set that is **reflexive**, **symmetric**, and **transitive**

$\equiv$ symbol for an abstract equivalence relation (instead of the letter R that we've been using for abstract relations up until now)

Worked Example 18.2.1 Let $\mathcal{L}$ be the set of all possible logical statements built out of the statement variables $p_1, p_2, p_3, \ldots$. Show that logical equivalence of statements is an equivalence relation on $\mathcal{L}$.

Solution.

Reflexive. We have $A \Leftrightarrow A$ for every statement A, since A has the same truth table as itself.

Symmetric. If $A \Leftrightarrow B$, then A, B have the same truth table, so $B \Leftrightarrow A$.

Transitive. If $A \Leftrightarrow B$ and $B \Leftrightarrow C$, then A has the same truth table as B, which has the same truth table as C. So A has the same truth table as C, i.e. $A \Leftrightarrow C$. □

Here is an important equivalence relation on $\mathbb{N}$ or on $\mathbb{Z}$.

equivalence modulo n
an equivalence of integers, where two integers are equivalent if they have the same remainder when divided by n

$m_1 \equiv_n m_2$ integers m_1, m_2 are equivalent modulo n

Checkpoint 18.2.2 Verify that equivalence modulo n is an equivalence relation.

18.3 Classes, partitions, and quotients

As desired (see Section 18.1), an equivalence relation can be used to group equivalent objects together.

Example 18.3.1 Consider $\equiv_5$ on $\mathbb{N}$. Notice that the elements in each of the following sets are all equivalent to each other with respect to $\equiv_5$.

$$\{0,5,10,15,\ldots\}$$
$$\{1,6,11,16,\ldots\}$$
$$\{2,7,12,17,\ldots\}$$
$$\{3,8,13,18,\ldots\}$$
$$\{4,9,14,19,\ldots\}$$

Also notice that $\mathbb{N}$ is the **disjoint union** of the above sets.

In fact, we could do the same for every divisor n, not just for $n = 5$, as $\mathbb{N}$ is also the disjoint union of the sets

$$\begin{gathered}\{0, n, 2n, 3n, \ldots\},\\ \{1, n+1, 2n+1, 3n+1, \ldots\},\\ \{2, n+2, 2n+2, 3n+2, \ldots\},\\ \vdots\\ \{n-1, n+(n-1), 2n+(n-1), 3n+(n-1), \ldots\},\end{gathered}$$

and again elements in each of the above sets are all equivalent to each other with respect to $\equiv_n$. □

equivalence class (of an element a)
the subset of A consisting of all elements that are equivalent to the given element $a \in A$, relative to a specific equivalence relation $\equiv$ on A; i.e. the set

$$\{\, x \in A \mid x \equiv a \,\}$$

$[a]$ the equivalence class of the element $a \in A$ relative to some specific equivalence relation on A

Example 18.3.2 Equivalence classes of natural numbers modulo 5. If we divide 8 by 5, we get 1 with 3 remainder. So the equivalence class of 8 relative to $\equiv_5$ consists of all natural numbers that have remainder 3 when divided by 5:

$$[8] = \{3,8,13,18,\ldots\}.$$

Now, 3 is in this class because when we divide 3 by 5 we get 0 with 3 remainder. But if we had started with 3 instead of 8, we would have also said that the equivalence class of 3 relative to $\equiv_5$ consists of all natural numbers that have remainder 3 when divided by 5:

$$[3] = \{3, 8, 13, 18, \dots\}.$$

□

Proposition 18.3.3 Properties of equivalence classes. *Suppose $\equiv$ is an equivalence relation on a nonempty set A.*

1. *For every $a \in A$, we have $a \in [a]$.*
2. *For $a, a_1, a_2 \in A$ with $a_1, a_2 \in [a]$, we have $a_1 \equiv a_2$.*
3. *For each pair $a_1, a_2 \in A$, we have $a_1 \equiv a_2$ if and only if $[a_1] = [a_2]$.*
4. *For each pair $a_1, a_2 \in A$, we have $a_1 \not\equiv a_2$ if and only if $[a_1] \cap [a_2] = \varnothing$.*

Proof of Statement 1. This is just the reflexive property, $a \equiv a$. ■

Proof of Statement 2. If $a_1, a_2 \in [a]$, then by definition we have both $a_1 \equiv a$ and $a_2 \equiv a$. Applying symmetry to the latter equivalence, we may write $a_1 \equiv a \equiv a_2$, to which we may apply transitivity to obtain $a_1 \equiv a_2$, as desired. ■

Proof of Statement 3.

(⇒) Suppose $a_1 \equiv a_2$. To verify $[a_1] = [a_2]$, we follow the Test for Set Equality. First, assume x is an arbitrary element in $[a_1]$. Then $x \equiv a_1 \equiv a_2$, so $x \equiv a_2$ by the transitive property. Therefore, $x \in [a_2]$, as required. This shows that $[a_1] \subseteq [a_2]$; the argument to show $[a_2] \subseteq [a_1]$ is almost exactly the same, just using the symmetric property to first obtain $a_2 \equiv a_1$.

(⇐) By Statement 1 of this proposition, we have $a_1 \in [a_1]$. If we assume $[a_1] = [a_2]$, then we also have $a_1 \in [a_2]$, which means that $a_1 \equiv a_2$, as required. ■

Proof of Statement 4. Let us prove the equivalent "double contrapositive" biconditional $a_1 \equiv a_2 \Leftrightarrow [a_1] \cap [a_2] \neq \varnothing$. (See Worked Example 2.1.4.)

(⇒) Suppose $a_1 \equiv a_2$. Then $[a_1] = [a_2]$ by Statement 3, so

$$[a_1] \cap [a_2] = [a_1] \cap [a_1] = [a_1].$$

But $[a_1]$ is nonempty by Statement 1.

(⇐) Suppose $[a_1] \cap [a_2] \neq \varnothing$. Then there exists some element $x \in A$ that is in both $[a_1]$ and $[a_2]$, so that both $x \equiv a_1$ and $x \equiv a_2$. By the symmetric property, we have $a_1 \equiv x$, and combining this with $x \equiv a_2$ in the transitive property gives $a_1 \equiv a_2$. ■

Statement 3 of Proposition 18.3.3 tells us that any member of an equivalence class may be used to define the class.

equivalence class representative
an element $a \in A$ used to define the equivalence class

$$[a] = \{\, x \in A \mid x \equiv a \,\}$$

complete set of equivalence class representatives
a subset $C \subseteq A$ so that for each $x \in A$ there exists *exactly one* $a \in C$ so that $x \in [a]$

Remember that elements that are equivalent to one another relative to some equivalence relation are viewed to be “essentially the same” from the point of view of the property used to define the equivalence relation. So *different but equivalent elements* become interchangeable (see Section 18.1). When we have a complete set of representatives for the equivalence classes, we are deciding to always interchange an element for the chosen representative of the class containing that element.

Example 18.3.4 A complete set of equivalence class representatives for natural numbers modulo 5. Continuing Example 18.3.2, we *could* represent the class of numbers that have remainder 3 by the number 8:

$$[8] = \{3, 8, 13, 18, \ldots\}.$$

But it seems more “natural” to represent this class by the number 3:

$$[3] = \{3, 8, 13, 18, \ldots\}.$$

Notice that each of the numbers $0, 1, 2, 3, 4$ has a different remainder when divided by 5, so no two of them are equivalent. That also means that each is in a different class (Statement 4 of Proposition 18.3.3). But when we go past 4, the remainders when divided by 5 start repeating: each of the numbers in the list $5, 6, 7, 8, 9$ has the same remainder as the number in the corresponding position in the list $0, 1, 2, 3, 4$. And then the remainders repeat again when we go past 9. And so on. So it seems “natural” to use $0, 1, 2, 3, 4$ as a complete set of representatives of the equivalence classes for $\mathbb{N}$ modulo 5:

$$\begin{aligned}
[0] &= \{0, 5, 10, 15, \ldots\},\\
[1] &= \{1, 6, 11, 16, \ldots\},\\
[2] &= \{2, 7, 12, 17, \ldots\},\\
[3] &= \{3, 8, 13, 18, \ldots\},\\
[4] &= \{4, 9, 14, 19, \ldots\}.
\end{aligned}$$

□

The definition of **complete set of equivalence class representatives** implicitly assumes that the equivalence classes “fill up” the whole set A. But that is always precisely the case.

partition a collection of subsets of a set A that are pairwise disjoint and whose union is A

partition cell
one of the subsets that make up a partition of a set

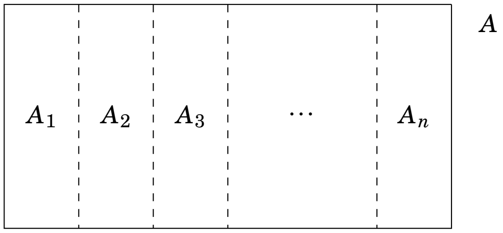


Figure 18.3.5 A diagram illustrating a **partition** of a set, so that $A = A_1 \sqcup A_2 \sqcup A_3 \sqcup \cdots \sqcup A_n$.

Remark 18.3.6 In essence, **partition** is just a synonym for **disjoint union**. So a collection of subsets form a partition when each element of the set is in *exactly* one partition cell.

Note 18.3.7 It is not necessary for a partition of a set to be made up of a *finite* (or even *countable*) number of cells; see the examples below.

Theorem 18.3.8 Equivalence classes form a partition. *If $\equiv$ is an equivalence relation on a set A, then the equivalence classes with respect to $\equiv$ are a partition of A.*

Proof idea. This theorem claims that every element of A is in *exactly one* equivalence class. But this follows from the statements of Proposition 18.3.3. ■

Example 18.3.9 Equivalence classes modulo n. Generalizing Example 18.3.4, each of the numbers $0,1,2,3,\ldots,n-1$ is its own remainder when divided by n. And then the pattern of remainders repeats, starting over at remainder 0, when we continue on to the numbers $n, n+1, \ldots$. So $0,1,2,3,\ldots,n-1$ is a complete set of equivalence class representatives, and the classes modulo n partition $\mathbb{N}$ into the disjoint subsets

$$\begin{aligned}
[0] &= \{0, n, 2n, 3n, \ldots\},\\
[1] &= \{1, n+1, 2n+1, 3n+1, \ldots\},\\
[2] &= \{2, n+2, 2n+2, 3n+2, \ldots\},\\
&\vdots\\
[n-1] &= \{n-1, 2n-1, 3n-1, 4n-1, \ldots\}.
\end{aligned}$$

□

Example 18.3.10 Let $\mathcal{L}$ be the set of all lines in the plane, and consider $\ell_1 \equiv \ell_2$ if ℓ_1, ℓ_2 are parallel. Then $\equiv$ partitions $\mathcal{L}$ into sets of parallel lines. □

Example 18.3.11 Recall that for alphabet Σ, Σ^*_n is the subset of Σ^* consisting of all words whose length is exactly n. Then

$$\Sigma^*_0, \Sigma^*_1, \Sigma^*_2, \ldots$$

is a partition of Σ^*. (See Exercise 9.9.9.)

Recall that a relation on Σ^* can be defined as a subset of $\Sigma^* \times \Sigma^*$. So consider the relation R on Σ^* defined by

$$R \;=\; (\Sigma^*_0 \times \Sigma^*_0) \;\sqcup\; (\Sigma^*_1 \times \Sigma^*_1) \;\sqcup\; (\Sigma^*_2 \times \Sigma^*_2) \;\sqcup\; \cdots$$

Then R is the equivalence relation on Σ^* where $w \mathrel{R} y$ if $|w| = |y|$, and its equivalence classes are precisely the sets Σ^*_n, $n \geq 0$. □

Theorem 18.3.12 Partitions arise from equivalence relations. *Given a partition of a set A, there exists an equivalence relation $\equiv$ on A whose equivalence classes are precisely the cells of the partition.*

Proof idea. Given a partition of A, for each $a \in A$ there exists exactly one partition cell containing a. So define $a_1 \equiv a_2$ to mean "elements a_1, a_2 are contained in the same partition cell of A." ■

Remark 18.3.13 Theorem 18.3.8 and Theorem 18.3.12 combine to provide, for each set A, a bijective correspondence

$$\{\text{equivalence relations on } A\} \;\longleftrightarrow\; \{\text{partitions of } A\}.$$

Worked Example 18.3.14 Determining an equivalence relation from a partition. Determine an explicit equivalence relation $\equiv$ on $\mathbb{Z}$ for which the equivalence classes give the following partition.

$$\mathbb{Z} \;=\; \cdots \;\sqcup\; \{-3,-2,-1\} \;\sqcup\; \{0,1,2\} \;\sqcup\; \{3,4,5\} \;\sqcup\; \cdots$$

Solution. Notice that each cell in the partition contains a multiple of 3 along with the next two consecutive integers. So one way to explicitly define the corresponding equivalence relation is: for $a,b \in \mathbb{Z}$, define $a \equiv b$ to be true if there exists $n \in \mathbb{Z}$ such that $3n \le a,b \le 3n+2$. (*Note:* Details showing that this is an equivalence relation are omitted.) □

quotient (of a set A relative to an equivalence relation $\equiv$)
the subset of $\mathscr{P}(A)$ whose elements are the equivalence classes of $\equiv$

$A/{\equiv}$ the quotient of A relative to equivalence relation $\equiv$, so that

$$(A/{\equiv}) = \{\, [a] \mid a \in A \,\}$$

Example 18.3.15 A quotient described by class representatives. Consider the partition of $\mathbb{Z}$ from Worked Example 18.3.14, and the corresponding equivalence relation $\equiv$. To describe $\mathbb{Z}/\equiv$, we just need to pick a representative of each class. The most obvious way in this case is

$$(\mathbb{Z}/{\equiv}) = \{\ldots,[-3],[0],[3],[6],\ldots\}.$$

□

Worked Example 18.3.16 Determining a quotient. Let $\equiv$ represent the equivalence relation on $\mathbb{Z}$ defined by

(i) $0 \equiv 0$, and

(ii) for non-zero $m, n \in \mathbb{Z}$,

$$m \equiv n \qquad \text{if} \qquad \frac{m}{|m|} = \frac{n}{|n|}.$$

Determine the corresponding partition and quotient of $\mathbb{Z}$.

Solution. First notice that 0 will be in an equivalence class all by itself. Next, consider the values that $m/|m|$ can possibly take.

- If $m > 0$, then $|m| = m$ so $m/|m| = 1$.
- If $m < 0$, then $|m| = -m$ so $m/|m| = -1$.

So this equivalence relation is just a fancy way of saying that m, n have the *same sign*. Therefore, all positive numbers will be in the same equivalence class, and all negative numbers will be in the same equivalence class. It now makes sense that 0 is in a class by itself, since 0 is neither positive nor negative. The partition of $\mathbb{Z}$ corresponding to $\equiv$ is then

$$\mathbb{Z} \;=\; \{\dots, -3, -2, -1\} \;\sqcup\; \{0\} \;\sqcup\; \{1, 2, 3, \dots\}.$$

To describe $\mathbb{Z}/\equiv$, we just need to pick a representative of each equivalence class. One possibility is

$$\mathbb{Z} = [-1] \sqcup [0] \sqcup [1],$$

so that

$$(\mathbb{Z}/\equiv) = \{[-1], [0], [1]\}.$$

□

natural projection (on a set A relative to an equivalence relation $\equiv$)
the function $A \to (A/\equiv)$ defined by $a \mapsto [a]$

Note 18.3.17 The natural projection $A \to (A/\equiv)$ is always surjective, but it is almost never injective.

Example 18.3.18 Natural projection modulo-5. Recall that $\equiv_5$ represents the modulo-5 equivalence relation on $\mathbb{N}$. In Example 18.3.4 we determined that there are five equivalence classes, represented by elements $0, 1, 2, 3, 4$, so that

$$(\mathbb{N}/\equiv_5) = \{[0], [1], [2], [3], [4]\}.$$

Below are some examples of images of elements under the natural projection.

$$2 \mapsto [2] \qquad 7 \mapsto [2] \qquad 104 \mapsto [4] \qquad 76 \mapsto [1] \qquad 2045 \mapsto [0].$$

□

18.4 Important examples

Example 18.4.1 Equality is the strongest form of equivalence. The "strongest" equivalence relation on a set A is the identity relation, where $a \equiv b$ if and only if $a = b$. In this case, each equivalence class is a singleton: $[a] = \{a\}$ for each $a \in A$. This equivalence relation yields the "finest" or most "granular" partition of A, into the union of all the singleton sets in $\mathscr{P}(A)$. Here, the quotient $A/\equiv$ is essentially the same as A: the natural projection $A \to (A/\equiv)$ is a bijection. □

Example 18.4.2 Even and odd. We can partition $\mathbb{N}$ into the subsets of even and odd numbers. This is the same partition obtained from the modulo-2 equivalance relation $\equiv_2$, and we have quotient

$$(\mathbb{N}/\equiv_2) = \{[0], [1]\}.$$

This quotient is how we construct *boolean algebra* (see Chapter 3). The convention $1+1=0$ in boolean algebra comes from defining addition in the quotient so that

$$[1]+[1]=[1+1]=[2]=[0].$$

□

Example 18.4.3 Modulo-n arithmetic. Similarly to Example 18.4.2, if we consider the modulo-n equivalence relation $\equiv_n$ on $\mathbb{N}$, we have

$$(\mathbb{N}/\equiv_n) = \{[0], [1], [2], \ldots, [n-1]\}.$$

We can transfer the arithmetic of $\mathbb{N}$ to $\mathbb{N}/\equiv_n$ by defining

$$[m]+[n]=[m+n], \qquad [m]\cdot[n]=[mn].$$

For example, in modulo-5 arithmetic,

$$[2]+[4]=[6]=[1]$$

and

$$[2]\cdot[4]=[8]=[3].$$

□

Checkpoint 18.4.4 (Bonus content) Properties of modulo-n arithmetic. There are a few things to check about this new modulo-n arithmetic.

1. Check that modulo-n addition and multiplication are **well-defined**; that is, make sure the result of each of these operations never depends on the choices of representatives of the equivalence classes involved.

2. Check that modulo-n addition and multiplication satisfy all the usual rules of arithmetic. That is, check that modulo-n addition and multiplication are both **associative** and **commutative**, and that multiplication distributes over addition.

3. The natural numbers 0 and 1 play special roles in $\mathbb{N}$ with respect to ordinary addition and multiplication, respectively. Do their equivalence classes [0] and [1] play the same special roles in $\mathbb{N}/\equiv_n$ with respect to modulo-n addition and multiplication, respectively?

Example 18.4.5 Same image under a function. For a function $f: A \to B$, we may consider elements of the domain equivalent if they produce the same output under f. That is, the relation $\equiv_f$ on A defined by "$a_1 \equiv_f a_2$ means $f(a_1) = f(a_2)$" is an equivalence relation. □

Checkpoint 18.4.6 Classes of the "same image" relation for an injective function. Suppose $f: A \to B$ is a function, and consider the equivalence relation $\equiv_f$ on A described in Example 18.4.5. How could one tell whether or not f is injective by looking at the equivalence classes in A under $\equiv_f$?

Example 18.4.7 Inverting a non-injective function. Equivalence relations allow us to take another point of view of the concept of inverse image of an element from Section 10.5.

Suppose $f : A \to B$, and consider the equivalence relation $\equiv$ on A described in Example 18.4.5. Then we may create a new, "induced" function

$$\begin{aligned} \tilde{f} : (A/\equiv) &\to B, \\ [a] &\mapsto f(a). \end{aligned}$$

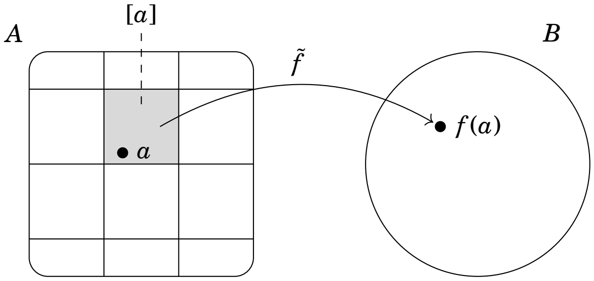


Figure 18.4.8 Diagram illustrating the induced map $\tilde{f}$.

In this function definition, an entire equivalence class is being mapped to the output image of one of the elements of that class under the original function f. But under this equivalence relation, each element in a specific equivalence class shares the same output image in the codomain as all the other elements in that class. For this reason, allowing our input-output rule definition $\tilde{f}\big([a]\big) = f(a)$ to depend on the choice of class representative a is **well-defined**, and hence *is* a function.

See. Example 10.1.17.

Moreover, the induced function $\tilde{f}$ is always injective, even if f is not. If we assume that f is surjective (or, if f is not surjective we could replace our codomain B with the image set $f(A)$ so that f *is* surjective — see restricting the domain), then $\tilde{f}$ will also be surjective, hence *bijective*. This means that $\tilde{f}$ is invertible, with inverse

$$\begin{aligned} \tilde{f}^{-1} : B &\to (A/\equiv_f), \\ b &\mapsto \{\, a \in A \mid f(a) = b \,\} = f^{-1}\big(\{b\}\big). \end{aligned}$$

In some sense $\tilde{f}^{-1}$ is an inverse of f, except that it is a function $B \to (A/\equiv_f)$ instead of $B \to A$. □

18.5 Graph for an equivalence relation

Given an equivalence relation on a finite set A, what will we observe if we draw the relation's graph?

- Since an equivalence relation is reflexive, we might as well omit the loops at each node.

- Since an equivalence relation is symmetric, we might as well replace the pairs of arrows between each related pair of nodes with a single edge, turning the directed graph into an ordinary graph.

- Since an equivalence relation partitions a set into a disjoint union of equivalence classes (Theorem 18.3.8), the graph of an equivalence relation will be disconnected, with each connected component representing a specific equivalence class.

- Since each element in an equivalence class is equivalent to every other element in the class (Statement 2 of Proposition 18.3.3), each connected component in the graph will be complete.

Example 18.5.1 Graph of the "same cardinality" equivalence relation. Let $A = \{a, b, c, d\}$, and let $\equiv$ be the equivalence relation on $\mathscr{P}(A)$ defined by $B \equiv B'$ if $|B| = |B'|$. That is, two subsets of A will be considered equivalent if they contain the same number of elements. Figure 18.5.2 contains the graph for $\equiv$, with reflexive loops and symmetric bidirectional arrows omitted.

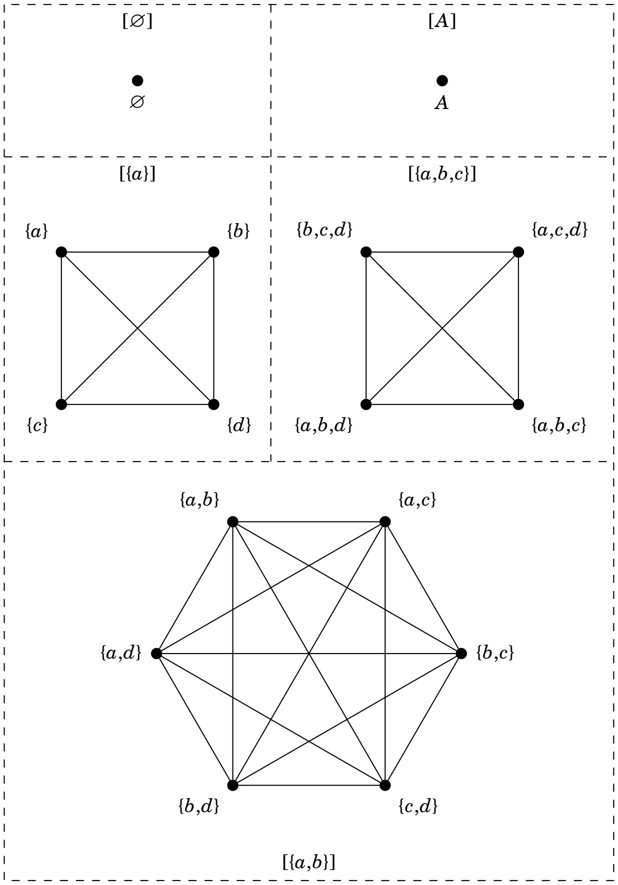


Figure 18.5.2 Graph for equivalence of cardinality on a power set.

□

18.6 Activities

Activity 18.1 For each of the relations provided, carry out the following steps.

(i) Verify that the relation is an equivalence relation on the set A.

(ii) Consider a few example equivalence classes, for the specific example representative elements provided (if applicable). What other elements are in that class?

(iii) Devise a general way to describe *every* equivalence class, using your experience from the example classes already considered (if applicable). Make your class descriptions more meaningful than just "all elements equivalent to a specific representative element."

(iv) List/describe all elements in the quotient $A/\equiv$.

(a) Relation $\equiv$ on $A = \mathbb{Z}$, where $m \equiv n$ means $m^2 = n^2$. Example equivalence classes for $1, 10, -2, 0$.

(b) Relation $\equiv$ on $A = \mathbb{R} \times \mathbb{R}$, where $(x_1, y_1) \equiv (x_2, y_2)$ means $x_1^2 + y_1^2 = x_2^2 + y_2^2$. Example equivalence classes for $(1,1), (3,4), (\sqrt{2}/2, -\sqrt{2}/2), (0,0)$.

(c) Relation $\equiv$ on $A = \mathbb{R} \times \mathbb{R}$, where $(x_1, y_1) \equiv (x_2, y_2)$ means $y_1^2 - x_1 = y_2^2 - x_2$. Example equivalence classes for $(0,0), (0,1), (1,-1)$.

(d) Relation $\equiv$ on $A = \mathscr{P}(\{a,b,c,d\})$, where $X \equiv Y$ means $|X^c| = |Y^c|$. Example equivalence classes for $\varnothing, \{a\}, \{a,b\}, \{a,b,c\}, \{a,b,c,d\}$.

(e) Relation $\equiv$ on the vertex set $A = V$ of a graph G, where $v \equiv v'$ means there exists a path in G from v to v'.

(f) Given function $f\colon A \to B$, the relation $\equiv$ on the domain A, where $a_1 \equiv a_2$ means $f(a_1) = f(a_2)$.

Activity 18.2 A sequence from a set A could also be called an **ordered list**. For example, given distinct $a_1, a_2 \in A$, the finite sequences a_1, a_1, a_2 and a_1, a_2, a_1 are different sequences, because order matters in a sequence. However, as an **unordered list**, a_1, a_1, a_2 is the *same* as a_1, a_2, a_1.

Write $\mathcal{S}_A$ for the set of all finite sequences from A. Devise an equivalence relation $\equiv$ on $\mathcal{S}_A$ such that the quotient set $\mathcal{S}_A/\equiv$ represents the set of all finite *un*ordered lists from A.

Hint. When should two different finite sequences be considered *equivalent* as unordered lists?

Activity 18.3 Suppose $\equiv$ and $\equiv'$ are equivalence relations on a set A. Determine which of the following are also equivalence relations.

(a) $\equiv^c$

(b) $\equiv \cup \equiv'$

(c) $\equiv \cap \equiv'$

See Activity 17.4.

18.7 Exercises

1. Let $\equiv$ represent the relation on $\mathbb{R} \times \mathbb{R}$ where $(x_1, y_1) \equiv (x_2, y_2)$ means $y_1 - x_1^2 = y_2 - x_2^2$.

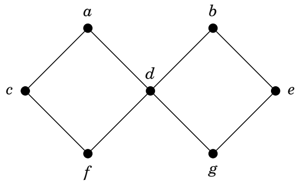
(a) Verify that $\equiv$ is an equivalence relation.

(b) Describe the equivalence classes $[(0,0)]$, $[(0,1)]$, and $[(1,0)]$ geometrically as sets of points in the plane.

2. Given a connected (undirected) graph G, we can define a relation on the set V of vertices in G as follows: let $v_1 R v_2$ mean that there exists a trail within G beginning at vertex v_1 and ending at vertex v_2 that traverses an even number of edges.

(a) Prove that R is an equivalence relation on V.

(b) Determine the equivalence classes for this relation when G is the graph below.



Equivalence relations and classes. In each of Exercises 3–12, you are given a set A and a relation R on A. Determine whether R is an equivalence relation, and, if it is, describe its equivalence classes. Try to be more descriptive than just "$[a]$ is the set of all elements that are equivalent to a."

3. $A = \{a,b,c\}$; $R = \{(a,a),(b,b),(c,c),(a,b),(b,a)\}$.

4. $A = \{-1,0,1\}$; $R = \{(x,y)|x^2 = y^2\}$.

5. A is the power set of some set; R is the subset relation.

6. $A = \mathbb{R}$; $x_1 \, R \, x_2$ means $f(x_1) = f(x_2)$, where $f: \mathbb{R} \to \mathbb{R}$ is the function $f(x) = x^2$.

7. A is some abstract set; $a_1 \, R \, a_2$ means $f(a_1) = f(a_2)$, where $f: A \to B$ is an arbitrary function with domain A.

8. A is the set of all "formal" expressions a/b, where a,b are integers and b is nonzero; $(a/b) \, R \, (c/d)$ means $ad = bc$.

Note: Do not think of a/b as a fraction in the usual way; instead think of it as a collection of symbols consisting of two integers in a specific order with a forward slash between them.

9. A is the power set of some finite set; $X \, R \, Y$ means $|X| = |Y|$.

10. A is the set of all straight lines in the plane; $L_1 \, R \, L_2$ means L_1 is parallel to L_2.

11. A is the set of all straight lines in the plane; $L_1 \, R \, L_2$ means L_1 is perpendicular to L_2.

12. $A = \mathbb{R} \times \mathbb{R}$; $(x_1,y_1) \, R \, (x_2,y_2)$ means $x_1^2 + y_1^2 = x_2^2 + y_2^2$.

Hint. Does the expression $x^2 + y^2$ remind you of anything from geometry?

CHAPTER 19

Partially ordered sets

19.1 Motivation

In many of the sets we encounter, there is some notion of elements being "less than or equal to" other elements in the set.

Example 19.1.1 Comparing numbers. In $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, we use the usual $\le$ to describe when one number is (literally) less than or equal to another. □

Example 19.1.2 Subset relationship as a measure of relative size. If A,B are subsets of a universal set U such that A is a subset of B, we might think of A as being "less than or equal to" B. The relation $\subseteq$ on $\mathscr{P}(U)$ acts very similarly to how $\le$ acts on a set of numbers. □

Warning 19.1.3 The idea of $A \subseteq B$ expressing a "less than or equal to"-like relationship between A and B is very different from cardinality-based ideas of **smaller/larger** for sets. See also Example 19.2.5.

Example 19.1.4 Subgraph relationship as a measure of relative size. Similar to Example 19.1.2, if H and H' are subgraphs of a graph G such that H' is a subgraph of H, we might think of H' as being "less than or equal to" H. That is, if we write $\mathscr{S}(G)$ to mean the set of all subgraphs of G, then we can use the subgraph relation $\preceq$ to describe when one subgraph of G is "smaller than or equal to" another. □

19.2 Definition and examples

Notice that in each of the examples in Section 19.1, the notion of "is smaller than" is defined via a relation. We will use $\le$ on $\mathbb{N}$ as our model for a relation on a set that can be thought of as expressing "is smaller than or equal in size to."

- Every element in the set should be "smaller than or equal to" itself, so the relation should be **reflexive**.
- Relative size should *never* be bidirectional for distinct elements in the set, so the relation should be **antisymmetric**.
- We should be able to infer size relationships from chains of them, so the relation should be **transitive**.

Notice that these are the same properties as for an equivalence relation, ***except*** that we have flipped **symmetric** to **antisymmetric**. Make sure to keep this straight!

partial order
a relation that is reflexive, *anti*symmetric, and transitive

partially ordered set
a set equipped with a particular partial order

$\preceq$ symbol for an abstract partial order

strictly less/smaller than
$a \preceq b$ and $a \neq b$

$a \prec b$ a is strictly less/smaller than b

Warning 19.2.1 We previously used the symbol $\preceq$ to mean exclusively "**is a subgraph of**," but that was in anticipation of the introduction of this symbol to now mean a general partial order.

Example 19.2.2 "Less than or equal to" versus "less than" on sets of numbers. The usual notion of $\leq$ is a partial order on $\mathbb{N}$ (or $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$), but $<$ is not. □

Example 19.2.3 Subset relation. For every set U, the relation $\subseteq$ is a partial order on $\mathscr{P}(U)$, but $\subsetneqq$ is not. □

Example 19.2.4 Subgraph relation. For every graph G, the subgraph relation $\preceq$ is a partial order on $\mathscr{S}(G)$, the set of subgraphs of G. □

Example 19.2.5 Comparing cardinalities. Suppose U is a universal set, and consider the collection of *finite* subsets of U. Then we have a natural way to compare sizes of these subsets: write $A \mathrel{R} B$ to mean $|A| \leq |B|$. However, this relation is not a partial order as it is not antisymmetric. This is because it is possible to have both $A \mathrel{R} B$ and $B \mathrel{R} A$ with $A \neq B$, in the case that $|A| = |B|$. Changing the relation to mean $|A| \lneqq |B|$ doesn't help, since then it wouldn't be reflexive.

Now suppose the universal set U is infinite and consider *all* (hence possibly infinite) subsets of U. In this case we have a more general idea of **smaller** and **larger**, where A is smaller than B if there exists an injection $A \hookrightarrow B$ but no bijection $A \to B$. This more general notion of size comparison via cardinality suffers the same flaws as in the finite set case, as it is not reflexive, and if we try to fix that by adding "or same size as" then it will not be antisymmetric.

However, in both finite and (possibly) infinite cases, we *can* turn cardinality comparison into a partial order using "smaller than or *equal* to", where "smaller" must mean strictly smaller in terms of cardinality, but "equal" means *equality of sets* rather than equality of cardinality. □

Example 19.2.6 English alphabetic order. Let $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$, and consider **alphabetic order** on the set of words Σ^*; e.g.

$$\mathrm{gqtiu} \preceq \mathrm{ppb}, \qquad \mathrm{aaay} \preceq \mathrm{aaaz}, \qquad \mathrm{aaa} \preceq \mathrm{aaaa}.$$

Alphabetic ordering is a partial order on Σ^*. □

Example 19.2.7 Lexicographic order. We can generalize the previous example: if Σ is a partially ordered alphabet set equipped with partial order $\preceq$, then we may inductively define a partial order $\preceq^*$ on Σ^* by:

- $\emptyset \preceq^* w$ for every $w \in \Sigma^*$, where $\emptyset$ is the empty word;
- for $a, b \in \Sigma$, considering these letters as words of length 1 in Σ^* take $a \preceq^* b$ to mean $a \preceq b$ in Σ;

- for letters $a_1, a_2 \in \Sigma$ and words $w_1, w_2 \in \Sigma^*$, take $a_1 w_1 \preceq^* a_2 w_2$ to mean that either

  (i) $a_1 \neq a_2$ and $a_1 \preceq a_2$, or

  (ii) $a_1 = a_2$ and $w_1 \preceq^* w_2$.

This is called **lexicographic** or **dictionary order** on Σ^*. □

Example 19.2.8 Ordering Cartesian products. We can employ a similar tactic for Cartesian products. If $\preceq_A, \preceq_B$ are partial orders on sets A, B, respectively, we can define a partial order $\preceq$ on $A \times B$ by allowing $(a_1, b_1) \preceq (a_2, b_2)$ to mean that either

(i) $a_1 \neq a_2$ and $a_1 \preceq_A a_2$, or

(ii) $a_1 = a_2$ and $b_1 \preceq_B b_2$.

This is also called **lexicographic order**. □

Example 19.2.9 Larger/greater than is a partial order. We can flip "smaller/less than or equal to" around to "larger/greater than or equal to." For example, for elements $m, n \in \mathbb{N}$, write $m \preceq n$ to mean $m \geq n$. Then $\preceq$ is a partial order on $\mathbb{N}$.

This is an instance of a more general pattern. Given a partial order $\preceq$ on a set A, the inverse relation $\preceq^{-1}$, where $a_1 \preceq^{-1} a_2$ means $a_2 \preceq a_1$, is also a partial order on A, called the **dual order**. □

Example 19.2.10 Transferring $\leq$ on $\mathbb{N}$ to a power set. Let $A = \{a, b, c, d\}$, and let us "encode" each element of $\mathscr{P}(A)$ by the following algorithm.

Given input element $X \in \mathscr{P}(A)$ (that is, given input X that is a subset of A):

(i) Initialize encoded value $r = 0$.

(ii) If X contains a, add 1 to r.

(iii) If X contains b, add 2 to r.

(iv) If X contains c, add 4 to r.

(v) If X contains d, add 8 to r.

(vi) Set encode(X) to be the final value of r.

For example,

$$\text{encode}\big(\{b\}\big) = 2, \qquad \text{encode}(\varnothing) = 0,$$
$$\text{encode}\big(\{a, c\}\big) = 1 + 4 = 5, \qquad \text{encode}(A) = 1 + 2 + 4 + 8 = 15.$$

This encoding process is one-to-one; that is, no two subsets of A will output the same encoded value.

Now define $\preceq$ on $\mathscr{P}(A)$ by taking $X \preceq Y$ to mean encode(X) $\leq$ encode(Y). For example,

$$\{a, b, c\} \preceq \{d\}, \qquad \{a, d\} \preceq \{b, d\},$$

and both

$$\varnothing \preceq X, \qquad X \preceq A$$

are true for every subset $X \subseteq A$.

The facts that $\leq$ is a partial order on $\mathbb{N}$ and that this encoding process is one-to-one will combine to make $\preceq$ a partial order. □

Example 19.2.11 Pulling a partial order back through an injection. Generalizing Example 19.2.10, suppose $f\colon A \hookrightarrow B$ is an injection where B is partially ordered by $\preceq_B$. Then we can "pull back" the partial order on B to create a partial order on A as follows: define $a_1 \preceq_A a_2$ to mean that $f(a_1) \preceq_B f(a_2)$ is true. Note that the assumption that f is injective is essential to guarantee that $\preceq_A$ will be antisymmetric. □

19.3 Graph for a partial order

Hasse diagram

a diagram for the graph for a partial order on a *finite* set A, omitting reflexive loops and transitive "composite" edges, and placing "smaller" elements *lower* on the diagram instead of using arrows

Example 19.3.1 Hasse diagram for division of integers. Let $A = \{2,4,6,8,10,12\}$. The Hasse diagram of the partial order $\mid$ (i.e. "divides") on A appears in Figure 19.3.2. Notice that 2 is not joined directly to either 8 or 12, since we can use transitivity and the facts that $2 \mid 4$ and $2 \mid 6$ to infer $2 \mid 8$ and $2 \mid 12$, respectively, from the diagram.

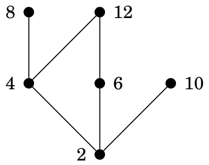


Figure 19.3.2 The Hasse diagram for the "divides" partial order on a finite set of integers.

□

Remark 19.3.3 See Example 14.4.3 for another example of a graph for the "divides" relation.

Example 19.3.4 Hasse diagram for subset order. The graph from Example 14.4.1 has been reproduced in Figure 19.3.5 as a Hasse diagram, and represents the partial order $\subseteq$ on $\mathscr{P}(\{a,b,c\})$.

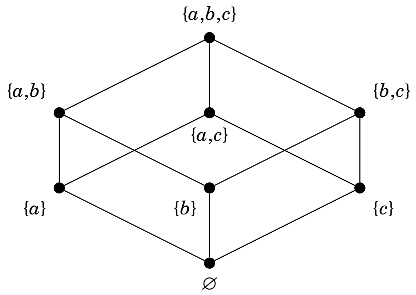


Figure 19.3.5 The Hasse diagram for the subset partial order on the power set of a finite set.

□

19.4 Total orders

comparable elements

elements a, b in a partially ordered set such that either $a \preceq b$ or $b \preceq a$

incomparable elements

elements that are not comparable

Example 19.4.1 Comparable and incomparable subsets. Let U represent some universal set containing at least two elements, and consider $\mathscr{P}(U)$ partially ordered by $\subseteq$.

- Both the empty set $\varnothing$ and the universal set U are comparable to every element of $\mathscr{P}(U)$.
- For $x, y \in U$ with $x \neq y$, then $\{x\}, \{y\}$ are incomparable.
- In fact, for every non-empty, proper subset $A \subsetneqq U$ there exists a subset $B \subseteq U$ which is incomparable to A: take $B = A^{\mathrm{c}}$.

However, do not let the second two points above lead you astray: it is not necessary for subsets to be disjoint in order to be incomparable. As long as each of a pair of subsets contains an element that the other doesn't, then the two will be incomparable by $\subseteq$. □

total order

a partial order on a set such that *every* pair of elements is comparable

totally ordered set

a set equipped with a total order

Example 19.4.2 Subset order is not total. For universal set U, order $\subseteq$ on $\mathscr{P}(U)$ is not total except when $|U| \le 1$. □

Example 19.4.3 Usual order of numbers is total. Our usual order for numbers, $\le$, is a total order on $\mathbb{N}$, on $\mathbb{Z}$, on $\mathbb{Q}$, or on $\mathbb{R}$. □

Example 19.4.4 Total order on alphabet induces total order on words. If $\preceq$ is a total order on an alphabet Σ, then the lexicographic order $\preceq^*$ described in Example 19.2.7 is a total order on the set of words Σ^*. □

Example 19.4.5 Pulling back a total order through an injection. If B is totally ordered and we use an injection $f\colon A \hookrightarrow B$ to "pull back" the order on B to an order on A (see Example 19.2.11), then the newly created order on A will also be total. □

Example 19.4.6 Countable can be totally ordered. If A is a **countably infinite** set, then there exists a bijection $f\colon \mathbb{N} \to A$. We can use the inverse $f^{-1}\colon A \to \mathbb{N}$ to "pull back" the usual total order $\le$ on $\mathbb{N}$ to a total order on A (see Example 19.4.5).

Another point of view on this is that our bijection f creates an infinite sequence

$$a_0, a_1, a_2, \ldots,$$

where each element of A appears exactly once. This sequence can be turned into a specification of the total order on A by just turning the commas into $\le$ symbols:

$$a_0 \le a_1 \le a_2 \le \cdots.$$

□

Remark 19.4.7 The pattern of Example 19.4.6 becomes even simpler when we apply it to a *finite* set: a total order on a finite set is no different than an ordering of the set elements into a list, as

$$a_0, a_1, a_2, \ldots, a_n$$

can simply be turned into

$$a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_n,$$

and vice versa.

Question 19.4.8 If A is a finite, totally ordered set, what does the corresponding Hasse diagram look like? □

The answer to this question is contained in Remark 19.4.7.

Fact 19.4.9 *A partial order on a finite set is total if and only if its Hasse diagram forms a single vertical line.*

Example 19.4.10 A totally ordered finite set. Figure 19.4.11 exhibits the Hasse diagram for the total order $|$ on the set $\{2,4,8,16,32\}$, though we have drawn the diagram on a slant from the vertical to be make it easier to see the entire diagram at a glance.

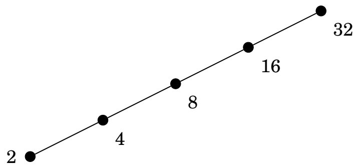


Figure 19.4.11 A Hasse diagram of a totally ordered set.

□

19.5 Maximal/minimal elements

Each of the following definitions are for a subset B of a partially ordered set A.

upper bound
an element $u \in A$ such that $b \preceq u$ for all $b \in B$

least upper bound
an upper bound for $B \subseteq A$ that is less than every other upper bound

maximum element
an upper bound for B that is contained in B

lower bound
an element $\ell \in A$ such that $\ell \preceq b$ for all $b \in B$

greatest lower bound
a lower bound for $B \subseteq A$ that is greater than every other lower bound

minimum element
a lower bound for B that is contained in B

Warning 19.5.1

1. An upper or lower bound does *not* need to belong to the subset for which it is a bound.

2. A set (or subset) does not necessarily have either a maximum or minimum element.

Fact 19.5.2 *If a subset of a partially ordered set contains a maximum element, then that maximum element is unique. And similarly for a minimum element.*

Proof. You are asked to prove this in Activity 19.5. ■

Example 19.5.3 Maximums, minimums, and bounds in $\mathbb{R}$. Consider the usual (total) order $\le$ on $\mathbb{R}$.

- The full set $\mathbb{R}$ has neither a maximum nor minimum element.
- The subset $(0,1)$ has many upper bounds (anything ≥ 1) and many lower bounds (anything ≤ 0). However, we would refer to 1 as the **least upper bound** and to 0 as the **greatest lower bound** of $(0,1)$.
- The subset $(0,1)$ has no maximum or minimum element. However, the subset $[0,1]$ has maximum 1 and minimum 0.

□

Example 19.5.4 Maximums, minimums, and bounds in a power set. Suppose U is a universal set, and consider $\mathscr{P}(U)$ partially ordered by $\subseteq$ as usual. In the full set $\mathscr{P}(U)$, the unique maximum element is U, which is just another way of saying that every element of $\mathscr{P}(U)$ is a subset of U. And the unique minimum element is $\varnothing$, which is just another way of saying that the empty set is a subset of every subset of U.

Now consider a subset $\mathcal{A} \subseteq \mathscr{P}(U)$, so that $\mathcal{A}$ is a collection of subsets of U. Because of the existence of maximum and minimum elements, these elements also serve as an upper and lower bound, respectively, for $\mathcal{A}$. However, one can also find a *least* upper bound for $\mathcal{A}$ by taking the union of all the subsets of U contained in $\mathcal{A}$, and one can find a greatest lower bound by taking the intersection of all the subsets of U contained in $\mathcal{A}$. □

The next two definitions are stated for elements in a partially ordered set, but could also be understood for elements in a subset of a partially ordered set, as every subset of a partially ordered set is also a partially ordered set.

maximal element
an element for which no other element is larger

minimal element
an element for which no other element is smaller

Remark 19.5.5 The difference between **maximum** and **maximal** is subtle. A maximum element must be larger than (and hence comparable to) *every* other element of A, while a maximal element must only be larger than every other element of A *to which it is comparable*. The distinction between *minimum* and *minimal* is similar.

Test 19.5.6 Maximal/minimal elements.

1. *To verify that $\overline{m} \in A$ is maximal, prove either the original definition*

$$(\forall x \in A)(x \ne \overline{m} \Rightarrow \overline{m} \not\preceq x),$$

or prove the equivalent contrapositive formulation

$$(\forall x \in A)(\overline{m} \preceq x \Rightarrow x = \overline{m}).$$

2. *To verify that $\underline{m} \in A$ is minimal, prove either the original definition*

$$(\forall x \in A)(x \ne \underline{m} \Rightarrow x \not\preceq \underline{m}),$$

or prove the equivalent contrapositive formulation

$$(\forall x \in A)(x \preceq \underline{m} \Rightarrow x = \underline{m}).$$

Example 19.5.7 Connected components are maximal. Consider the graph G in Figure 19.5.8.

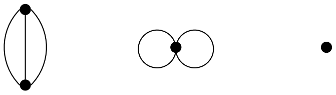


Figure 19.5.8 An example non-connected graph.

Let $\mathscr{S}(G)$ represent the collection of subgraphs of G, partially ordered by the subgraph relation. (By my count, $|\mathscr{S}(G)| = 110$.) Let $\mathscr{C}(G)$ represent the collection of *connected* subgraphs of G. (By my count, $|\mathscr{C}(G)| = 15$.) A maximal elements of $\mathscr{C}(G)$ would have to be a connected subgraph of G that is contained in no larger connected subgraph of G — but this is precisely the definition of **connected component**. Hence $\mathscr{C}(G)$ has three maximal elements, the three connected components you see in Figure 19.5.8. However, a maximum element of $\mathscr{C}(G)$ would be a connected subgraph of G which contains every other connected subgraph of G, and the existence of multiple connected components in this example non-connected graph makes such a subgraph impossible. □

Remark 19.5.9 If you compare both our informal definition and formal definition of **connected component** with our definition of **maximal element** and our Test for Maximal/Minimal Elements, you should find that the definition of **connected component** means precisely a **maximal** subgraph with respect to the property of being connected.

Worked Example 19.5.10 Let $A = \{a, b, c\}$. Given the Hasse diagram for $\subseteq$ on $P = \mathscr{P}(A) \setminus \{A\}$ in Figure 19.5.11, determine all maximal/maximum/minimal/minimum elements, if they exist.

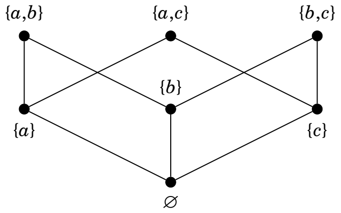


Figure 19.5.11 The Hasse diagram for $\subseteq$ on an “uncapped” power set.

Solution. The element $\{a, b\}$ is maximal, since each node in the Hasse diagram that is adjacent to $\{a, b\}$ is below it. The same reasoning confirms that $\{a, c\}$, and $\{b, c\}$ of are also maximal. However, none of them is a maximum, since none of them is larger than the other two.

The element $\varnothing$ of P is a minimal element, since each node that is adjacent to it is above it. And it is the only minimal element. Furthermore, $\varnothing$ is the minimum element, since for every other node there is a walk upwards from $\varnothing$ to that node. □

Warning 19.5.12 Just drawing a node higher or lower in a Hasse diagram does not necessarily make it larger or smaller, respectively, when compared to other elements via the partial order. For example, in the Hasse diagram of Figure 19.5.11, we could have drawn the node for $\{a, c\}$ at a higher location, but that would not make it larger than $\{a, b\}$ and $\{b, c\}$, since there still would not have been any edges or chains of edges from $\{a, c\}$ downward to those other two nodes.

Fact 19.5.13 *If the partially ordered set A has a maximum element, then that element is also the only maximal element of A. Similarly, the minimum element, if it exists, is the only minimal element of A.*

Proof idea. Assume A has a maximum element. Then every element of A is both comparable to and smaller than that maximum element, so no element is larger than it. Therefore, this maximum must be maximal. And no other element could be maximal, because to be maximal means there are no elements which are larger. But our maximum element is always larger. ■

Warning 19.5.14 A maximum element must be maximal and must be the *only* maximal. But a maximal element, even if it is the only one, need not be the maximum.

Example 19.5.15 A partially ordered set with exactly one maximal element but no maximum element. Consider

$$A = \{3\} \cup \{2, 4, 8, 16, 32, 64, \ldots, 2^n, \ldots\},$$

partially ordered by $|$, the "divides" relation. There is no element of A that is divisible by 3 (except 3 itself), so there is no element of A that is larger than 3. Therefore, 3 is maximal. Moreover, 3 is the only maximal element in A, since every power of 2 divides the next power of 2. However, there is no maximum element in A, since there is no element of A which is divisible by every other element of A. □

Example 19.5.16 A partially ordered set with infinitely many maximal/minimal elements but no maximum/minimum element. Consider $\mathcal{A} \subsetneq \mathscr{P}(\mathbb{N})$, where

$$\mathcal{A} = \mathscr{P}(\mathbb{N}) \setminus \{\varnothing, \mathbb{N}\}.$$

So $\mathcal{A}$ is the set of all nonempty, proper subsets of $\mathbb{N}$. Under the partial order $\subseteq$, $\mathcal{A}$ has neither a maximum nor a minimum element, but for every $n \in \mathbb{N}$, $\{n\}$ is a minimal element and $\mathbb{N} \setminus \{n\}$ is a maximal element of $\mathcal{A}$. □

19.6 Topological sorting

Sometimes we want to turn a partial order into a total order. What makes an order **partial** instead of **total** is the presence of pairs of incomparable elements. So to convert our partial order into a total order we just need to impose an order relation on those previously incomparable element pairs. However, for each pair of incomparable elements there is a choice to be made of which will become the smaller and which the larger in the new total order. And we cannot carry out these choices completely arbitrarily, because we risk contradicting the required properties of a partial order (see Example 19.6.1).

The following definitions apply to a partial order $\preceq$ on a set A.

compatible total order

a total order $\le$ on A such that if $a_1 \preceq a_2$ then $a_1 \le a_2$

topological sorting

a process for determining a compatible total order

Example 19.6.1 A failed attempt at topological sorting. The relation $\subseteq$ on $\mathscr{P}(\mathbb{N})$ is a partial order but not a total order. Consider what happens when we begin trying to build a total order on $\mathscr{P}(\mathbb{N})$ out of $\subseteq$ by choosing relations between previously incomparable elements arbitrarily.

- Elements $\{1\}, \{2,3\}$ are $\subseteq$-incomparable; choose $\{2,3\} \leq \{1\}$.
- Elements $\{1\}, \{2\}$ are $\subseteq$-incomparable; choose $\{1\} \leq \{2\}$.
- ...

Now, $\{2\} \subseteq \{2,3\}$ is already true, so to be compatible we must set $\{2\} \leq \{2,3\}$ in the new total order. But now $\{1\} \leq \{2\} \leq \{2,3\}$ would dictate $\{1\} \leq \{2,3\}$ to satisfy the transitive property, but this contradicts our first arbitrary choice above. □

Note 19.6.2 If A is countable, whether finite or countably infinite, then specifying a total order on A amounts to writing the elements of A in an ordered list. (See Example 19.4.6 and Remark 19.4.7.) In that case, topological sorting amounts to creating such an ordered list so that if $a \preceq b$ then a appears before b in the list.

Algorithm 19.6.3 Topological sorting. *If A is a* finite *partially ordered set with respect to $\preceq$, we can specify a compatible total order $\leq$ on A by writing the elements of A in a list*

$$a_0 \leq a_1 \leq \cdots \leq a_{n-1}$$

as follows, where $n = |A|$.

1. *Initialize $i = 0$ and $A_0 = A$.*
2. *Choose a minimal element of A_i; let a_i represent the chosen element.*
3. *Set $A_{i+1} = A_i \setminus \{a_i\}$ (i.e. create a smaller partially ordered set by discarding a_i).*
4. *Increment i by 1. If $i < n$ then go back to Step 2. Otherwise, if $i = n$ then A_i should now be empty, so stop — the desired compatible order has now been specified.*

Note 19.6.4 In Step 2 of the algorithm, if A_i contains a *minimum* element, then you must choose that element, since in that case no other minimal element can exist (see the Fact 19.5.13).

Worked Example 19.6.5 Consider

$$A = \mathscr{P}(\{0,1,2\}) = \big\{\varnothing, \{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\big\}.$$

Apply Algorithm 19.6.3 to determine a total order $\leq$ on A that is compatible with $\subseteq$.

Solution 1 (Algorithm solution). In $A_0 = A$, we must choose $a_0 = \varnothing$, since it is the minimum. Now remove a_0 so that

$$A_1 = \big\{\{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\big\}.$$

Choose a minimal element from A_1: let $a_1 = \{2\}$. Now remove a_1; set

$$A_2 = \big\{\{0\}, \{1\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\big\}.$$

Choose a minimal element from A_2: let $a_2 = \{0\}$. Now remove a_2; set

$$A_3 = \big\{\{1\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\big\}.$$

Choose a minimal element from A_3: let $a_3 = \{0,2\}$. Now remove a_3; set

$$A_4 = \big\{\{1\}, \{0,1\}, \{1,2\}, \{0,1,2\}\big\}.$$

We must choose $a_4 = \{1\}$, since it is the minimum in A_4. Now remove a_4; set

$$A_5 = \big\{\{0,1\}, \{1,2\}, \{0,1,2\}\big\}.$$

Choose a minimal element from A_5: let $a_5 = \{1,2\}$. Now remove a_5; set $A_6 = \big\{\{0,1\},\{0,1,2\}\big\}$. We must choose $a_6 = \{0,1\}$, since it is the minimum in A_6. There is only one element left; set $A_7 = \big\{\{0,1,2\}\big\}$ and choose $a_7 = \{0,1,2\}$. So we have

$$\varnothing \leq \{2\} \leq \{0\} \leq \{0,2\} \leq \{1\} \leq \{1,2\} \leq \{0,1\} \leq \{0,1,2\}.$$

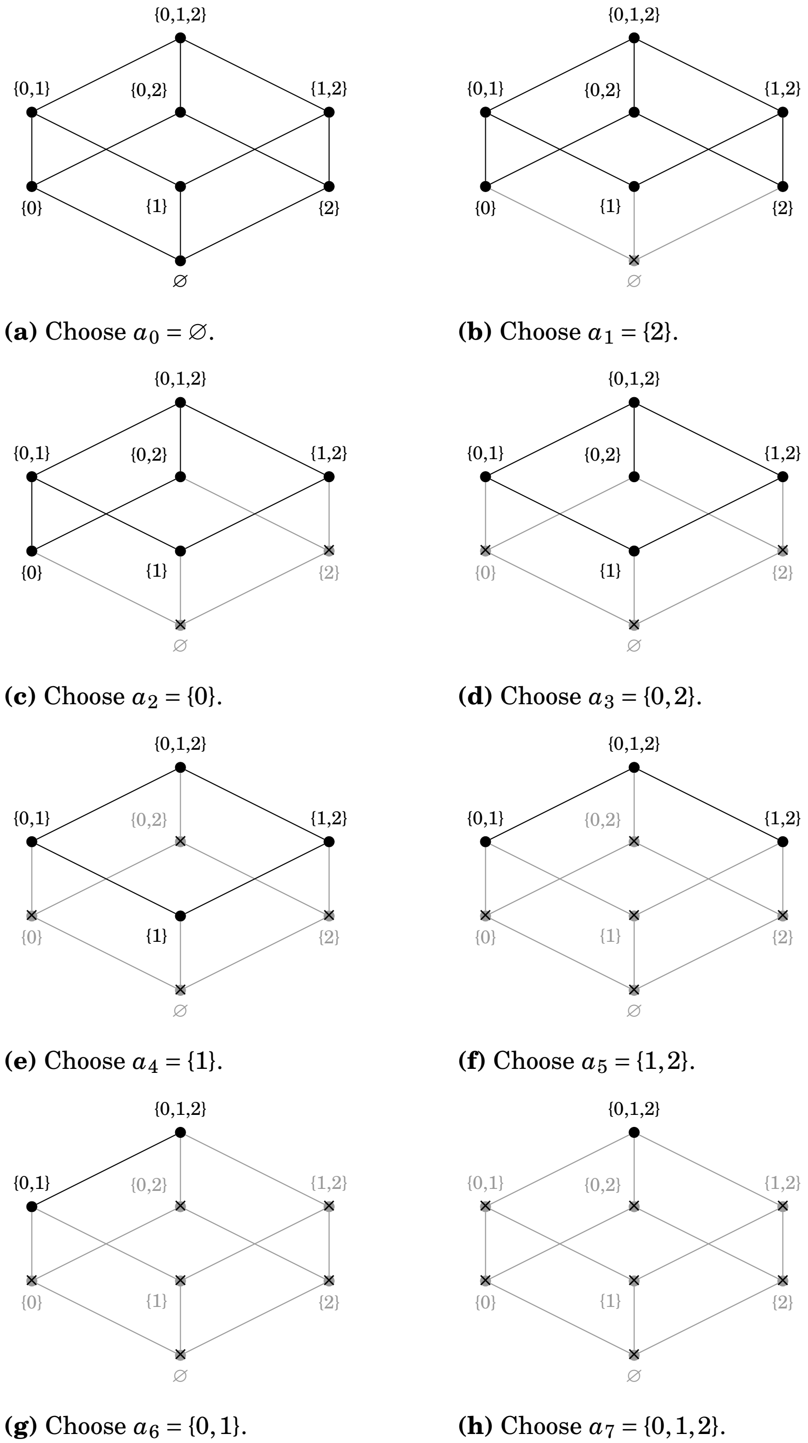
Notice that the maximum element of A ended up at the "top" of the total order and the minimum element was forced to the "bottom."

Solution 2 (Graphical solution). We can perform the algorithm of topological sorting graphically; at each step, choose a vertex that has no adjacent vertices below it in the graph, then cross that vertex and any adjacent edges out of the graph. (See Figure 19.6.6.)

Our end result is a list of our choices, in order:

$$\varnothing \leq \{2\} \leq \{0\} \leq \{0,2\} \leq \{1\} \leq \{1,2\} \leq \{0,1\} \leq \{0,1,2\}.$$

Notice that the maximum element of A ended up at the "top" of the total order and the minimum element was forced to the "bottom".



(a) Choose $a_0 = \varnothing$.

(b) Choose $a_1 = \{2\}$.

(c) Choose $a_2 = \{0\}$.

(d) Choose $a_3 = \{0,2\}$.

(e) Choose $a_4 = \{1\}$.

(f) Choose $a_5 = \{1,2\}$.

(g) Choose $a_6 = \{0,1\}$.

(h) Choose $a_7 = \{0,1,2\}$.

Figure 19.6.6 Example of a graphical approach to topological sorting.

□

Note 19.6.7 Compatible total orders are not unique: in the previous worked example, the order in which the elements of A were originally written represents another compatible total order:

$$\varnothing \leq \{0\} \leq \{1\} \leq \{2\} \leq \{0,1\} \leq \{0,2\} \leq \{1,2\} \leq \{0,1,2\}.$$

19.7 Activities

Activity 19.1 Let $F \subseteq \mathbb{N}$ represent the set of all divisors of 30. Let $A = \{a, b, c\}$.

Note: In Task c you will compare your work from Task a and Task a, so keep your work!

(a) Draw the Hasse diagram for the subset partial order $\subseteq$ on $\mathscr{P}(A)$.

(b) Draw the Hasse diagram for the "divides" partial order $\mid$ on F.

(c) Compare your two Hasse diagrams. Can you devise a function $f : F \to \mathscr{P}(A)$ that would deserve to be called an **order-preserving correspondence** between F and $\mathscr{P}(A)$?

Activity 19.2 Suppose $\preceq$ is a partial order on a set A. Verify that the inverse relation $\preceq^{-1}$ is also a partial order on A by verifying that it is **reflexive**, **anti-symmetric**, and **transitive**.

Activity 19.3 Let $A = \{a, b, c, d, e\}$. Carry out the following steps for each of the scenarios below.

(i) Draw the Hasse diagram for a partial order on A with the requested features.

(ii) In your diagram, identify all maximal/minimal elements.

(iii) Identify all pairs of incomparable elements.

(a) A has both a maximum and a minimum.

(b) A has a maximum but no minimum.

(c) A has a minimum but no maximum.

(d) A has neither a maximum nor a minimum.

Activity 19.4 Suppose $\preceq$ is a partial order on the set $A = \{0, 1, 2\}$ such that 1 is a maximal element. What are the possibilities for the Hasse diagram of $\preceq$?

Activity 19.5 Using the proper strategy for proving uniqueness (see Procedure 6.10.2), prove that if a partially ordered set A has a maximum element, then that element is the *unique* maximum element.

How can your proof be modified to show that a minimum element is also unique?

Activity 19.6 Recall that $(a, b) \subseteq \mathbb{R}$ means an **open interval** on the real number line:

$$(a, b) = \{ x \in \mathbb{R} \mid a < x < b \}.$$

Let $\leq$ be the usual "less than or equal to" total order on the set

$$A = (-2, 0) \cup (0, 2).$$

Consider the subset

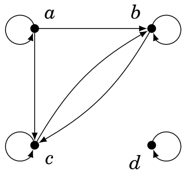
$$B = \left\{ -\frac{1}{n} \;\middle|\; n \in \mathbb{N}, n \geq 1 \right\} \subseteq A.$$

Determine an upper bound for B in A. Then formally prove that B has no *least* upper bound in A by arguing that every element of A fails the criteria in the definition of **least upper bound**.

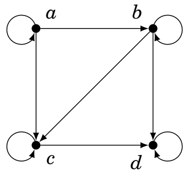
19.8 Exercises

Recognizing a partial order from its graph. In each of Exercises 1–2, you are given a directed graph for a relation on the set $A = \{a,b,c,d\}$. Determine whether the relation is a partial order. Justify your answers.

1.



2.



Testing partial orders. In each of Exercises 3–6, you are given a set A and a relation R on A. Determine whether the relation is a partial order. Justify your answers.

3. A is the set of all Augustana students; $a\,R\,b$ means that student a has a higher GPA than student b.

4. A is the power set of some finite set; $S\,R\,T$ means $|S| \le |T|$.

5. A is the set of words on some alphabet; $w\,R\,w'$ means $|w| \le \left|w'\right|$, where $|w|$ means the length of word w.

6. $A = \mathbb{R} \times \mathbb{R}$; $(x1, y1)\,R\,(x2, y2)$ means $x_1 \le x2$ and $y_1 \le y2$.

Drawing Hasse diagrams. In each of Exercises 7–8, you are given a finite, partially ordered set A. Draw the Hasse diagram.

7. $A = \mathscr{P}(\{1,2,3,4\})$ under the subset relation.

8. $A = \Sigma_4^*$, the set of words of length 4 in the alphabet $\Sigma = \{0,1\}$, under dictionary order.

9. Draw all possible valid Hasse diagrams for each of the sets $A = \{a,b\}$ and $B = \{a,b,c\}$. (Thus, you will have determined all possible partial orders on those sets.)

10. Consider the “divides” relation $|$ on $\mathbb{N}_{>0}$. Provide an example of a set $A \subseteq \mathbb{N}_{>0}$

(a) that is finite, and on which $|$ is a total order.

(b) that is infinite, and on which $|$ is a total order.

(c) on which $|$ is a partial order but not a total order.

11. Let $A = \{0,1,2\}$, and consider the partial order $\subseteq$ on the power set $\mathscr{P}(A)$. List all pairs of incomparable elements in $\mathscr{P}(A)$.

Determining maximal/maximum/minimal/minimum elements. In each of Exercises 12–16, you are given a partially ordered set A. Determine any and all maximal, maximum, minimal, and minimum elements.

12. $A = \mathbb{N}_{>0}$ under the usual $\le$.

13. $A = \mathbb{Q}_{>0}$ under the usual $\le$.

14. $A = \mathbb{N} \smallsetminus \{0,1\}$ under the “divides” relation $|$.

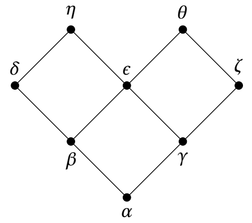
15. $A = \{2,5,11,13,22,65,110,143,496\}$ under the “divides” relation $|$.

16. A is the set of the first ten prime numbers under the “divides” relation $|$.

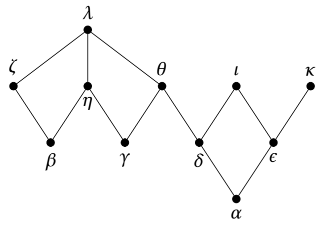
17. Suppose $\preceq$ is a partial order on the set $A = \{0,1,2\}$ such that 1 is a maximal element. What are the possibilities for the Hasse diagram of $\preceq$?

Topological sorting. In each of Exercises 18–19, you are given the Hasse diagram for a partially ordered set A. Use the Topological sorting algorithm to determine a compatible total order on A.

18.



19.



Part VI

Combinatorics

CHAPTER 20

Counting

20.1 Motivation

You probably learned to count before you even started kindergarten. But efficiently counting large collections can be difficult!

Example 20.1.1 Examples of counting large collections.

- How many different ways can you choose your winning numbers for the lottery?
- How many different possible seating charts could be made for the students in this course in the assigned classroom?
- How many different ways are there for you to choose courses to satisfy your degree requirements?
- How many bijections between the sets $\{0,1,2,3,4,5\}$ and $\{a,b,c,d,e,f\}$ exist?
- How many total orders on the set $\{0,1,2,3,4,5\}$ exist?
- How many partial orders on the set $\{0,1,2,3,4,5\}$ exist?

□

20.2 Addition and subtraction rules

As usual in mathematics, breaking a big problem into smaller parts is a useful strategy.

Theorem 20.2.1 Addition Rule. *Assume U is a finite set.*

1. *If $U = A_1 \sqcup A_2$, then $|U| = |A_1| + |A_2|$.*
2. *If $U = A_1 \cup A_2$, then $|U| = |A_1| + |A_2| - |A_1 \cap A_2|$.*

Proof idea. After recalling the definition of **disjoint union**, Statement 1 should be obvious. To prove Statement 2, apply Statement 1 to the following disjoint unions:

$$U = A_1 \sqcup (A_2 \setminus A_1), \qquad A_2 = (A_2 \setminus A_1) \sqcup (A_1 \cap A_2).$$

Then combine the resulting equalities of cardinalities. ■

Remark 20.2.2 Statement 1 of Theorem 20.2.1 can be extended to a disjoint union of any number of subsets.

Worked Example 20.2.3 Counting by breaking into cases. How many words of length 3 or less are there using alphabet $\Sigma = \{\alpha, \omega\}$?

Solution. Write $\Sigma^*_{\leq 3}$ to mean the set of words in alphabet Σ of length 3 or less. Then

$$\Sigma^*_{\leq 3} = \Sigma^*_0 \sqcup \Sigma^*_1 \sqcup \Sigma^*_2 \sqcup \Sigma^*_3,$$

so we can break into cases based on length and then apply the Addition Rule.

Count Σ^*_0. There is only one word of length 0: the empty word. So $\left|\Sigma^*_0\right| = 1$.

Count Σ^*_1. There are only two words of length 1: the single-letter words $w_\alpha = \alpha$ and $w_\omega = \omega$. So $\left|\Sigma^*_1\right| = 2$.

Count Σ^*_2. We can count be simply listing the elements:

$$\Sigma^*_2 = \{\alpha\alpha, \alpha\omega, \omega\alpha, \omega\omega\}.$$

So $\left|\Sigma^*_2\right| = 4$.

Count Σ^*_3. This time we will just use inductive reasoning. Each word in Σ^*_2 may be extended to a word in Σ^*_3 by appending either an α or an ω onto the end. So there must be twice as many words in Σ^*_3 as in Σ^*_2, i.e. $\left|\Sigma^*_3\right| = 8$.

Total count. Using the Addition Rule, we obtain the total by adding up our preliminary results:

$$\left|\Sigma^*_{\leq 3}\right| = 1 + 2 + 4 + 8 = 15.$$

□

Another common strategy in mathematics is to consider the opposite.

Theorem 20.2.4 Subtraction Rule. *Assume U is a finite set. For every subset $A \subseteq U$, we have $|A| = |U| - |A^c|$.*

Proof idea. Since $U = A \sqcup A^c$ is always true, simply apply Statement 1 of Theorem 20.2.1 to this disjoint union and rearrange to isolate $|A|$. ■

Example 20.2.5 Counting by counting the complement. For alphabet $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$, how many words in Σ^*_2 do *not* begin with the letter a? It's much easier to count the number of words in Σ^*_2 that *do* begin with a, as there are only 26 possibilities for the second letter.

Later in this chapter we will learn a rule that will allow us to easily calculate the total number of words in Σ^*_2 to be 26^2 (see Worked Example 20.3.10). Accepting this fact for the moment, we can then use the Subtraction Rule to compute

$$\begin{aligned}\#\{\text{2-letter words not beginning with a}\} &= \left|\Sigma^*_2\right| - \#\{\text{2-letter words beginning with a}\}\\ &= 26^2 - 26\\ &= 26(26-1)\\ &= 26 \cdot 25.\end{aligned}$$

□

20.3 Multiplication rule

Worked Example 20.3.1 Counting a small Cartesian product. What is $|A \times B|$ for $A = \{0,1,2,3\}$ and $B = \{-1,0,1\}$?

Solution. We can solve this by just writing out the elements of $A \times B$ and counting them.

$$\begin{aligned} A \times B = \{&(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1),\\ &(2,-1),(2,0),(2,1),(3,-1),(3,0),(3,1)\} \end{aligned}$$

So $|A \times B| = 12$. □

Worked Example 20.3.2 Counting a large Cartesian product. What is $|C \times D|$ for $C = \{a,b,c,\ldots,z\}$ and $D = \{0,1,2,\cdots,99\}$?

Solution. Writing out all the elements of $C \times D$ and then counting them all seems like a lot of work. Instead, using our experience from Worked Example 20.3.1, notice that we usually perform the task of writing the elements of a Cartesian product in a pattern to make sure we get them all. One-by-one we pick a single element of the first set C, and pair it up with *every* element of the second set D. From this pattern we see that for each $c \in C$, there are $|D|$ elements of $C \times D$ with c as the first coordinate, and there are $|C|$ such groupings of elements from $C \times D$. So we arrive at

$$|C \times D| = |C| \cdot |D| = 26 \cdot 100 = 2600.$$

□

Checkpoint 20.3.3 For sets X and Y, define an equivalence relation on $X \times Y$ whose equivalence classes partition $X \times Y$ in the manner described in the provided solution to Worked Example 20.3.2. Then describe how the number of classes and the number of objects in each class correspond to $|X|$ and $|Y|$.

Theorem 20.3.4 Multiplication Rule. *If there are m ways to perform task S and n ways to perform task T, then there are mn ways to perform task S followed by task T.*

Warning 20.3.5 The Multiplication Rule only applies to consecutive tasks S,T such that the number of ways of performing task T is *independent* of the choice made in performing task S.

Example 20.3.6 Counting Cartesian product elements by constructing an arbitrary element. To create a specific example of an element from $A \times B$, we must first choose an element of A to be the first coordinate (task S), then choose an element of B to be the second coordinate (task T). There are $m = |A|$ ways to perform task S and $n = |B|$ ways to perform task T. Therefore, the Multiplication Rule says there are mn ways to construct an element of $A \times B$, which means $|A \times B| = mn$. □

Example 20.3.7 Choosing candidates. Suppose you are a casting director and need to select both a primary actor and an understudy for the lead role in a play. If n actors audition for the role, then there are n different ways to select the primary actor. Once this choice is made, there remain $n-1$ different ways to the select the understudy. Hence there are $n(n-1)$ ways to cast the role.

Now, the actual pool of candidates for understudy will differ based on which actor is offered the lead role. However, no matter who is chosen for the lead, the *number* of remaining candidates for understudy is the same. □

Note 20.3.8 We may extend the Multiplication Rule to any (finite) number of consecutive tasks.

Example 20.3.9 Cardinality of Cartesian product of many sets. If $A_1, A_2, \ldots, A_m$ are finite sets with $\left|A_j\right| = m_j$, then

$$|A_1 \times A_2 \times \cdots \times A_\ell| = m_1 m_2 \cdots m_\ell.$$

□

Worked Example 20.3.10 Words of a given length. Recall that, given alphabet Σ and number $n \in \mathbb{N}$, Σ_n^* is the set of words of length n. If $|\Sigma| = m$, what is $\left|\Sigma_n^*\right|$?

Solution. To construct a specific example word $w \in \Sigma_n^*$, there are:

- m ways to choose the first letter,
- m ways to choose the second letter,
- $\ldots$,
- m ways to choose the n^{th} letter.

So there are

$$\underbrace{m \cdot m \cdot m \cdot \cdots \cdot m}_{n \text{ factors}} = m^n$$

ways to construct w. We conclude $\left|\Sigma_n^*\right| = m^n$. □

Worked Example 20.3.11 Words with no repeated letters. Suppose $|\Sigma| = 5$. How many words in $\left|\Sigma_5^*\right|$ have no repeated letters? (That is, in which no two letters are the same?)

Solution. To construct a specific example word $w \in \Sigma_5^*$ in which no two letters are the same, there are

- 5 ways to choose the first letter,
- 4 remaining ways to choose the second letter,
- 3 remaining ways to choose the third letter,
- 2 remaining ways to choose the fourth letter, and
- only 1 remaining way to choose the last letter.

So there are

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

ways to construct w.

Similar to Example 20.3.7, while the actual pool of candidates for the next letter at each step will differ based on which letters have been chosen already, the *number* of remaining letters is always independent of which letters have actually been chosen so far. So the Multiplication Rule can be applied to this problem exactly as we have applied it. □

Worked Example 20.3.12 Palindromes. Let $\Sigma = \{\text{a}, \text{b}, \text{c}, \ldots, \text{y}, \text{z}\}$. How many palindromes w with $3 \le |w| \le 6$ are there in Σ^*?

Solution. Break into cases based on the length of w.

Case $|w| = 3$. Once we choose the first letter, the last is chosen for us, but we are still free to choose the middle letter. So there are 26^2 palindromes of length 3.

Case $|w| = 4$. Once we choose the first two letters, the last two are chosen for us. So there are also 26^2 palindromes of length 4.

Case $|w| = 5$. Once we choose the first two letters, the last two are chosen for us, but we are still free to choose the middle letter. So there are 26^3 palindromes of length 5.

Case $|w| = 6$. Once we choose the first three letters, the last three are chosen for us. So there are also 26^3 palindromes of length 6.

Total. Applying the Addition Rule to these non-overlapping cases, we obtain

$$\begin{aligned} 26^2 + 26^2 + 26^3 + 26^3 &= 26^2(1 + 1 + 26 + 26) \\ &= 54 \cdot 26^2 \\ &= 36{,}504 \end{aligned}$$

as the number of palindromes length 3 to 6. □

Worked Example 20.3.13 Set $A = \{a, b, c\}$ and $B = \{0, 1, 2, 3, 4\}$. How many functions $A \to B$ exist? How many of these are injections? How many are surjections?

Solution.

Number of functions. A function $f : A \to B$ can be constructed in three steps: choose $f(a)$, then choose $f(b)$, then choose $f(c)$. Each of the steps can be carried out in $|B| = 5$ ways. So the number of functions is $5^3 = 125$.

Number of injections. An injection $f : A \hookrightarrow B$ can be constructed in three steps: choose $f(a)$, then choose $f(b)$ to be *different from* $f(a)$, then choose $f(c)$ to be *different from both* $f(a)$ and $f(b)$. First step has $|B| = 5$ choices. Second step has $|B \setminus \{f(a)\}| = 4$ choices. Third step has $|B \setminus \{f(a), f(b)\}| = 3$ choices. So the number of injections is $5 \cdot 4 \cdot 3 = 60$.

A look ahead. Notice that the number of injections has turned out to be

$$\frac{|B|!}{(|B| - |A|)!}.$$

We will understand better how this formula arises in Section 21.4.

Number of surjections. Suppose $f : A \to B$. Since $|A| = 3$, the *largest* that $|f(A)|$ can be is 3, which occurs when f is injective. However, even in such a largest case it is still smaller then $|B|$, so no surjections exist. That is, the number of surjections is 0. □

20.4 Division rule

Sometimes it is easier to count a related but more structured collection, where the collection we actually want to count corresponds to equivalence classes of the more structured collection.

Theorem 20.4.1 Division Rule. *Suppose $\equiv$ is an equivalence relation on a finite set A so that the equivalence classes* ***all have the same number of elements****. Then*

$$\#\{\textit{equivalence classes}\} = \frac{|A|}{\textit{common size of classes}}.$$

That is,

$$|A/{\equiv}| = \frac{|A|}{|[a]|},$$

where a is an arbitrary element of A.

Proof. Write N for the number of equivalence classes, and write C for the common cardinality of the classes. We know that the equivalence classes partition the set A, so using the Addition Rule we have

$$|A| = |[a_1]| + |[a_2]| + \cdots + |[a_N]|,$$

where $a_1, a_2, \ldots, a_N$ are a complete set of equivalence class representatives. But we have assumed that these class cardinalities are all equal to each other, with each class satisfying $\left|[a_j]\right| = C$. So

$$|A| = \underbrace{C + C + \cdots + C}_{N \text{ terms}} = NC,$$

which leads to

$$N = \frac{|A|}{C},$$

as desired. ■

Worked Example 20.4.2 Equivalent words. Let $\Sigma = \{\alpha, \beta, \gamma\}$. How many words in Σ_4^* contain exactly two αs, one β and one γ?

Solution. Write Λ for the collection of words in Σ_4^* of the type described. Instead of trying to count Λ directly, consider the following more structured collection.

Write $\Sigma' = \{\alpha_1, \alpha_2, \beta, \gamma\}$, and let Λ' be the set of words in $(\Sigma')_4^*$ that have no repeated letters. Similar to Worked Example 20.3.11, we have

$$\left|\Lambda'\right| = 4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

For each pair of these words, write $w_1 \equiv w_2$ if the following two conditions hold.

(i) At whatever position w_1 contains α_1, w_2 contains either α_1 or α_2 at that same position.

(ii) At whatever position w_1 contains α_2, w_2 contains either α_1 or α_2 at that same position.

You may check that $\equiv$ defines an equivalence relation on Λ'. Each class consists of exactly two words $\{w_1, w_2\}$, where w_2 has an α_2 where w_1 has an α_1 and an α_1 where w_1 has an α_2. For example, one class of $\Lambda'/{\equiv}$ is

$$\left[\alpha_1\beta\alpha_2\gamma\right] = \{\alpha_1\beta\alpha_2\gamma, \alpha_2\beta\alpha_1\gamma\}.$$

Effectively, the classes remove the distinction between α_1 and α_2, so that they might as well be the same letter, say, α. In other words, there is a bijective correspondence between the classes in $\Lambda'/{\equiv}$ and the words in Λ. Using the Division Rule, we have

$$|\Lambda| = \left|\Lambda'/{\equiv}\right| = \frac{\left|\Lambda'\right|}{2} = \frac{24}{2} = 12.$$

□

20.5 Pigeonhole principle

20.5.1 Regular strength version

Theorem 20.5.1 Pigeonhole Principle (formal version). *If A,B are finite sets with $|B| < |A|$, then no function $A \to B$ can be an injection.*

Proof. This principle is just the contrapositive of Statement 2 of Fact 12.2.5. ■

Corollary 20.5.2 Pigeonhole Principle. *If m objects are placed in n containers, where $m > n$, then at least one container must contain more than one object.*

Proof. Let A be the set of objects and B the set of containers, so that

$$|B| = n < m = |A|.$$

Also let $f : A \to B$ represent the function where $f(a) = b$ means that object a has been placed in container b. Then the Theorem 20.5.1 tells us that f cannot be an injection, which means that there is at least one pair of distinct objects a_1,a_2 with $f(a_1) = f(a_2)$. ■

Worked Example 20.5.3 Too few seats. Your car has five seats, but you also have five friends who need a ride home. How will everyone fit?

Solution. Using the people who need to get home (i.e. your friends *and* you) as objects and car seats as containers, Pigeonhole Principle says that someone will have to sit on someone else's lap. □

Worked Example 20.5.4 Dessert logistics. The cafeteria puts out 200 chocolate puddings and 200 tapioca puddings. If 201 students each grab a bowl of pudding, what is the *minimum* number of tapioca puddings that have been taken?

Solution. Since $201 > 200$, there is no injection

$$\{\text{students who took a pudding}\} \hookrightarrow \{\text{bowls of chocolate pudding}\}.$$

(*Or:* use students as objects, bowls of chocolate pudding as containers.)

But we can't actually have two students take the same bowl of pudding, so at least one student must eat tapioca. □

Worked Example 20.5.5 Matching pairs. Suppose $A \subseteq \{1,2,\ldots,20\}$. How big must $|A|$ be to ensure that there exist two elements of A whose sum is 21?

Solution. Collect together the (unordered) pairs of numbers that add to 21:

$$B = \big\{\{1,20\},\{2,19\},\ldots,\{10,11\}\big\} \subseteq \mathscr{P}(\{1,2,\ldots,20\}).$$

Notice that $|B| = 10$. Thinking of the elements of B as containers and elements of A as objects, place object a into container b if $a \in b$. We need one more object than container to ensure some container receives two objects, so the answer is $|A| \geq 11$. □

Worked Example 20.5.6 Matching modulo m. Suppose $m \in \mathbb{N}$ and $A \subseteq \mathbb{N}$ such that $0 < m < |A| < \infty$. Show that there exist distinct $a_1,a_2 \in A$ that have the same remainder when divided by m.

Solution. The set of possible remainders is $\mathbb{N}_{<m} = \{0,1,2,\cdots,m-1\}$. Computing remainder after division by m defines a function $A \to \mathbb{N}_{<m}$. Since $|\mathbb{N}_{<m}| = m < |A|$, this function cannot be an injection.

(*Or:* use elements of A as objects, possible remainders when dividing a number by m as containers.) □

20.5.2 Strong version

Recall that given a function $f\colon A \to B$, we can define an equivalence relation $\equiv$ on A by taking $a_1 \equiv a_2$ to mean $f(a_1) = f(a_2)$ (see Example 18.4.5). In this way, we can regard f as placing objects (elements of A) into containers (elements of B), so that object $a \in A$ is "placed" in container $b \in B$ when $f(a) = b$.

Theorem 20.5.7 Pigeonhole Principle (strong form, formal version). *Suppose $f\colon A \to B$, with A,B finite, and let $\equiv$ be the equivalence relation on A where $a_1 \equiv a_2$ means $f(a_1) = f(a_2)$.*

If $|A| > \ell \cdot |B|$ for some $\ell \in \mathbb{N}$, then at least one of the equivalence classes of A with respect to $\equiv$ has more than ℓ elements.

Proof. Consider the contrapositive:

> if *every* equivalence class of A has *no more* than ℓ elements, then $|A| \le \ell \cdot |B|$.

Since B is finite and $f(A) \subseteq B$, then also $f(A)$ is finite and we can enumerate its elements. Write $f(A) = \{b_1, b_2, \ldots, b_n\}$. Each element of $f(A)$ corresponds to an equivalence class of A.

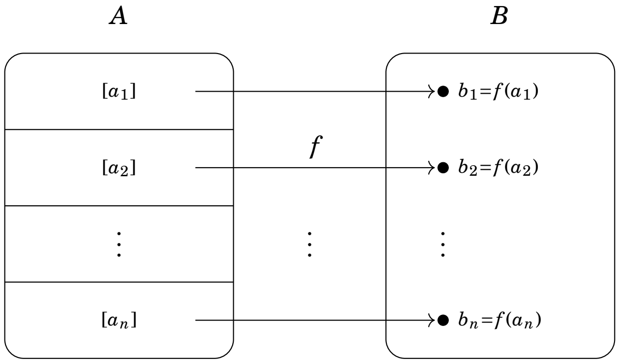


Figure 20.5.8 Diagram of equivalence classes under the "have same image" equivalence.

In this diagram, the a_i are representative elements of the class of elements of A that are mapped to b_i by f. In particular, we must have $f(a_i) = b_i$ for each index i.

We are assuming that each class $[a_i]$ contains no more than ℓ elements; i.e. $|[a_i]| \le \ell$. Since an equivalence relation always partitions a set A into the disjoint union of its equivalence classes, we have

$$\begin{aligned} |A| &= |[a_1]| + |[a_2]| + \cdots + |[a_n]| \\ &\le \underbrace{\ell + \ell + \cdots + \ell}_{r \text{ terms}} \\ &= \ell n \\ &= \ell \cdot |f(A)|. \end{aligned}$$

But $f(A)$ is a subset of the finite set B, and so $|f(A)| \le |B|$. Combining this with the calculation above gives

$$|A| \le \ell \cdot |f(A)| \le \ell \cdot |B|.$$

■

Corollary 20.5.9 Pigeonhole Principle (strong form, informal version). *If m objects are placed in n containers, with $m > \ell n$ for some $\ell \in \mathbb{N}$, then at least one container contains more than ℓ objects.*

Proof idea. Again, let A be the set of objects and B the set of containers, so that

$$|A| = m > \ell n = \ell \cdot |B|.$$

Then apply the Pigeonhole Principle (strong form, formal version). ■

Note 20.5.10 The Pigeonhole Principle (strong form, formal version) is a *generalization* of the Pigeonhole Principle (formal version). A function is an injection precisely when no two distinct elements of the domain produce the same output image, so using $\ell = 1$ in the strong form gives back the original form.

Worked Example 20.5.11 Handing out coins. Show that if thirteen coins are distributed to six children, then at least one child will receive at least three coins.

Solution. Using coins as objects and children as containers, the given statement is just the Pigeonhole Principle (strong form, formal version) with $\ell = 2$: we have 13 objects and 6 containers, and $13 > 2 \cdot 6$. (*Note:* Since coins are discrete objects, "more than two" and "at least three" are the same thing.) □

Remark 20.5.12 It is worthwhile to think about how the strong form of the Pigeonhole Principle could be proved directly. Consider the diagram in Figure 20.5.8: the "tipping point" between $|A| \le \ell \cdot |B|$ and $|A| > \ell \cdot |B|$ is when f is surjective and each of the equivalence classes has exactly ℓ elements. When f is surjective, there are $|B|$ equivalence classes in A. Since A is the disjoint union of its equivalence classes under $\equiv$, we have $|A| = \ell \cdot |B|$. If we add *one more* element to A, it will have to be included in one of the equivalence classes, and that class will now have size greater than ℓ.

Worked Example 20.5.13 Handing out pudding. The cafeteria puts out 100 chocolate, 100 tapioca, and 100 butterscotch puddings. How many students must grab a pudding before we can be certain that at least one of the flavours has at least half of the bowls taken?

Solution 1 (Using "tipping point" thinking). The "tipping point" is exactly 49 bowls of *each* flavour taken, which requires $3 \cdot 49 = 147$ students. After the 148th student, we will definitely have 50 bowls of one of the flavours taken.

Solution 2 (Direct use of the Pigeonhole Principle). Consider students as objects (m of them — this is the unknown to be determined) and flavours as containers (3 of them). To determine the appropriate value of ℓ to use, consider that "at least half" in this problem means "at least 50", which is the same as "more than 49". So choose $\ell = 49$. In that case, we need $m > 49 \cdot 3 = 147$, bringing us to the answer of $m = 147 + 1 = 148$ students. □

20.6 Activities

Activity 20.1 A standard Alberta license plate has three letters followed by three or four digits.

(a) How many different vehicles can the province license with this scheme?

(b) Do you think the province was right to expand license plates by adding another digit, or do you think it should have added another letter instead? (Or, as a third possibility, is it irrelevant in practical terms?)

Hint. The figure $26^3 = 17576$ may help you decide.

Activity 20.2

(a) You roll a six-sided die ten times. How many different sequences of rolls are possible?

(b) Describe how Task a relates to the problem of determining $\left|\Sigma_{10}^*\right|$ for a suitable alphabet Σ.

Activity 20.3 Let $\Sigma = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{y}, \mathrm{z}\}$. How many words in Σ_5^* end in the letter z? How many do not?

Activity 20.4 You and your five housemates pick names out of a hat each week to determine who is going to clean the toilet. Over a three-week period, how many different sequences of toilet bowl cleaners could be determined in this fashion

(i) if names are placed back in the hat after each draw?

(ii) if names are *not* placed back in the hat after each draw?

Activity 20.5 How many natural numbers between 1 and 1,000,000 (inclusive) contain the digit 5?

Hint. You might instead count how many numbers *don't* contain the digit 5.

Activity 20.6 How many natural numbers between 100 and 999 (inclusive) have no repeated digits? Of these, how many are odd?

Hint. There's no rule that when you "construct" an arbitrary object of this type that you have to choose the first digit first.

Activity 20.7 Use the Pigeonhole Principle to prove that in every set of three integers there exists a pair whose difference is even.

Hint. What kinds of numbers add up to an even sum?

Activity 20.8 You have a list of the names of twenty students. Ten of the students are domestic students and the other ten are out-of-province students. How many students must you select from the list to be certain to form a group that contains at least one domestic student and at least one out-of-province student?

Activity 20.9 Let n be a fixed natural number. Determine the smallest number M for which the following statement is true: every subset of

$$\mathbb{N}_{<2n+1} = \{0, 1, 2, 3, \ldots, 2n\}$$

of size M contains at least one odd number.

Activity 20.10 You're cleaning up your little nephew's toy room. There are T toys on the floor and n empty toy storage boxes. You randomly throw toys into boxes, and when you're done the box with the most toys contains N toys.

(a) What is the smallest that N could be when $T = 2n + 1$?

(b) What is the smallest that N could be when $T = kn + 1$?

(c) Now suppose that the number of toys T satisfies

$$T < \frac{n(n-1)}{2}.$$

Prove that when you are done cleaning up, there will be (at least) one pair of boxes that contain the same number of toys.

Hint. Argue the contrapositive by assuming that every box ends up a different number of toys. What is the *fewest* number of toys you could have started with?

20.7 Exercises

1. You are trying to decide how to top your ice-cream sundae. You have five choices of sprinkles, four choices of cookie crumbs, five choices of fruit, and three choices of chocolate chunks. For each category of topping, you may choose only one of the available options, or you may choose to skip that category altogether. How many different sundaes could you create out of these choices?

2. You turn eighteen and your trust fund finally starts paying out. You decide to buy a vehicle, and eventually narrow things down to a choice between five SUVs, four sports cars, and two motorcycles. How many ways are there to choose a vehicle? How many ways are there to choose one vehicle of each type?

3.

(a) Use the Multiplication Rule to demonstrate that the truth table of a logical statement with n statement variables requires 2^n rows. That is, demonstrate that there are 2^n different possible combinations of input truth values for n statement variables.

(b) How many different truth tables involving n statement variables exist?

4. Recall that if A is a finite set with $|A| = n$, then $|\mathscr{P}(A)| = 2^n$. Use the Multiplication Rule to verify this formula by considering the construction of an arbitrary subset of A as a process of making n “either-or” decisions.

5. It is the year 2030, and Alberta has succeeded in seceding from Canada and has become the landlocked Kingdom of Albertania. The King decrees that the kingdom’s citizens will all be assigned a hexadecimal ID. That is, using alphabet

$$\Sigma = \{0,1,2,3,4,5,6,7,8,9,\mathrm{a},\mathrm{b},\mathrm{c},\mathrm{d},\mathrm{e},\mathrm{f}\},$$

IDs will be words from Σ^*. However, the king is vain and doesn’t want any such ID to contain his initials, jk.

For each $n \geq 1$, let s_n represent number of allowable IDs of length n.

(a) Compute s_1, s_2, and s_3.

(b) Determine a recurrence relation for s_n which is valid (at least) for $n \geq 3$.

Hint. For each allowable word of length $n-1$ you can create a word of length n by adding a new letter onto the end. But you want your new word to also be allowable, so be careful about what you add onto the end!

CHAPTER 21

Permutations

21.1 Factorials

In counting, factorials come up a lot.

$n!$ for natural number n, notation for the computation formula

$$n(n-1)(n-2)\cdots 2\cdot 1$$

Example 21.1.1 Two factorial calculations.

$$3! = 3\cdot 2\cdot 1 = 6, \qquad 7! = 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 = 5{,}040.$$

□

Example 21.1.2 Factorial factors. A factorial contains every smaller factorial as a factor. For example,

$$\frac{7!}{3!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot \cancel{(3!)}}{\cancel{3!}} = 7\cdot 6\cdot 5\cdot 4 = 840.$$

□

Convention 21.1.3 To avoid division by zero in certain formulas, define $0! = 1$. This choice is also made to be consistent with the methods for counting permutations we will explore in this chapter.

21.2 Definition

We often want to count how many ways we can "mix up" the objects in a collection.

permutation a bijection from a finite set to itself

Remark 21.2.1 Once you have written the elements of a finite set in some order, think of a permutation as a way of *re-ordering* them.

Example 21.2.2 Permutations of three objects. Figure 21.2.3 contains tables of values for all six possible permutations of the set $A = \{a, b, c\}$. We have grouped them according to: all elements fixed; one element fixed and two mixed; all elements mixed.

x	a	b	c
$\mathrm{id}_A(x)$	a	b	c

x	a	b	c
$f_a(x)$	a	c	b
$f_b(x)$	c	b	a
$f_c(x)$	b	a	c

x	a	b	c
$s_r(x)$	c	a	b
$s_l(x)$	b	c	a

Figure 21.2.3 All possible permutations on three objects.

□

21.3 Counting permutations

Theorem 21.3.1 *For $|A| = n$, there are $n!$ permutations on A.*

Informal proof. We want to count the number of ways of constructing an ordered list of the n elements of A. There are n ways to choose the first element in the list, $n-1$ ways to choose the second, $n-2$ ways to choose the third, and so on, ending at a single way to choose the n^{th}. By the Multiplication Rule, there are

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1 = n!$$

ways to construct such a list. ■

Formal proof. By induction.

Base case $n = 1$. If $|A| = 1$, then A consists of a single element, say $A = \{a\}$. There is only one possible permutation of A, and that is the identity function $\mathrm{id}_A : A \to A$ defined by $\mathrm{id}_A(a) = a$. Thus, we have verified that there is $1! = 1$ permutation of A.

Induction step. Let $k \geq 1$ be a fixed integer. Our induction hypothesis is to assume that if B is any set with $|B| = k$ elements, then there are $k!$ permutations on B. We want to use this hypothesis to prove that if A is a set with $|A| = k+1$ elements, then there are $(k+1)!$ permutations on A.

Write $A = \{a_0, a_1, \dots, a_k\}$ and $B = \{a_1, a_2, \dots, a_k\}$. Then B is a subset of A that contains k elements, and so by our induction hypothesis there are $k!$ permutations on B. For every such permutation of B, we can construct $k+1$ permutations of A by "inserting" a_0 at different positions in the output list. For example, consider how the identity permutation on B can be turned into $k+1$ different permutations on A — see Figure 21.3.2.

x	a_1	a_2	a_3	$\dots$	a_{k-1}	a_k
$\mathrm{id}_B(x)$	a_1	a_2	a_3	$\dots$	a_{k-1}	a_k

↓

x	a_0	a_1	a_2	a_3	$\dots$	a_{k-1}	a_k
$\mathrm{id}_B^{(0)}(x) = \mathrm{id}_A(x)$	a_0	a_1	a_2	a_3	$\dots$	a_{k-1}	a_k
$\mathrm{id}_B^{(1)}(x)$	a_1	a_0	a_2	a_3	$\dots$	a_{k-1}	a_k
$\mathrm{id}_B^{(2)}(x)$	a_1	a_2	a_0	a_3	$\dots$	a_{k-1}	a_k
$\mathrm{id}_B^{(3)}(x)$	a_1	a_2	a_3	a_0	$\dots$	a_{k-1}	a_k
$\vdots$				$\vdots$			
$\mathrm{id}_B^{(k-1)}(x)$	a_1	a_2	a_3	a_4	$\dots$	a_0	a_k
$\mathrm{id}_B^{(k)}(x)$	a_1	a_2	a_3	a_4	$\dots$	a_k	a_0

Figure 21.3.2 Inserting an extra element at various positions of a permutation to create new, longer permutations.

Each of the $k!$ permutations of B can be used to construct $k+1$ permutations of A, in the same fashion as we have above for the identity permutation of B. So we have in total $(k+1) \cdot k! = (k+1)!$ permutations of A, as required. ■

Remark 21.3.3 When applying the method of mathematical induction in the formal proof, we began our base case at $n = 1$. But the formula $n!$ is still valid for the number of permutations of the empty set. In this case, $n = 0$ and so $n! = 0! = 1$ by Convention 21.1.3. And there is indeed exactly one permutation of the empty set — the empty function. (See Statement 2 of Proposition 12.1.6.)

Each of the provided proofs for Theorem 21.3.1 above contains an idea that is of practical use in counting collections.

- In the informal proof, we used the Multiplication Rule to count the number of ways to construct an ordered list, where the tasks in the construction are choosing the elements in the list one at a time. (We used this similar thinking often in Chapter 20, though we didn't explicitly connect the Multiplication Rule to ordered lists.)
- In the formal proof, we used the idea of "inserting" an object into an existing ordered list to create a new ordered list.

Worked Example 21.3.4 Distributing items. For a class of twenty students, in how many different orders can a professor hand back marked tests:

1. In total?
2. If Karishma's test must be handed back first?
3. If Elizabeth's and Ruijing's tests cannot immediately follow one another?

Solution.

1. A test distribution order is the same thing as a permutation of the students in the class, so there are 20! different handback orders (approximately 2.4 quintillion).
2. This is just the number of ways of ordering the remaining nineteen students' papers, which is 19! (approximately 122 quadrillion).
3. It is easier to count the ways that they *do* follow each other. One way to do this is as follows. Remove Elizabeth's test from the pile. There are now

19! ways to order the remaining nineteen papers. There are two ways to insert Elizabeth's test back into any such ordering — either immediately before or after Ruijing's paper. So there are $2 \cdot 19!$ orderings we *do not* want. Therefore, applying the Subtraction Rule yields answer

$$20! - 2 \cdot 19! = 20 \cdot 19! - 2 \cdot 19! = 18 \cdot 19!.$$

□

Worked Example 21.3.5 Words using the entire alphabet. For an alphabet Σ with $|\Sigma| = n$, how many words in Σ^* contain each element of the alphabet exactly once?

Solution. Here we just want to order all the elements of Σ into a word, so the answer is $n!$.

Compare. See Worked Example 20.3.11.

□

Remark 21.3.6 Worked Example 21.3.5 justifies the convention $0! = 1$, since if $|\Sigma| = 0$, then Σ^* contains exactly one word: the empty word $\emptyset$. And in this case it is vacuously true that $\emptyset$ contains each element of Σ exactly once.

Worked Example 21.3.7 Counting total orders. If $|A| = n$, how many different total orders on A exist?

Solution. Specifying a total order on A really just means ordering the elements of A:

$$a_1 \le a_2 \le a_3 \le \cdots \le a_n.$$

So there are $n!$ possible total orders. □

Worked Example 21.3.8 Counting colour patterns. How many different colour patterns can we obtain by placing three red bottles and five blue bottles on a shelf? (Assume the bottles are indistinguishable except by colour.)

Solution. Let's use the Division Rule, where first we will count a more structured collection. If the bottles of the same colour *were* distinguishable from each other, we would have 8! ways of lining them up on the shelf. Assuming indistinguishability, we now consider two orderings with the same colour pattern but mixed up red and/or blue bottles to be *equivalent*. For example, the two orderings

$$\mathrm{R}_1\,\mathrm{B}_1\,\mathrm{B}_2\,\mathrm{R}_2\,\mathrm{R}_3\,\mathrm{B}_3\,\mathrm{B}_4\,\mathrm{B}_5,$$
$$\mathrm{R}_2\,\mathrm{B}_5\,\mathrm{B}_3\,\mathrm{R}_1\,\mathrm{R}_3\,\mathrm{B}_1\,\mathrm{B}_4\,\mathrm{B}_2,$$

of distinguishable bottles create the same colour pattern, and so are equivalent. Once red and blue bottle positions are determined, we can permute the reds (3! ways) and blues (5! ways) independently, so each equivalence class inside the collection of orderings of *distinguishable* bottles contains $3! \cdot 5!$ equivalent orderings. Applying the Division Rule, we arrive at

$$\frac{8!}{3! \cdot 5!} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2} = 56$$

possible colour patterns. □

Worked Example 21.3.9 Circular orderings. How many different seating arrangements of ten people around a round table are possible, if no one is considered to be at the "head" or "foot" of the table?

Solution 1. Let's use the Division Rule, where first we will count a more structured collection. There are 10! ways to line the 10 people up. Wrapping the end of the line around to meet the beginning forms a circular seating arrangement. But “rotating” around the line (10 possible rotations) yields an equivalent circular seating arrangement. So the answer is

$$\frac{10!}{10} = 9!.$$

Solution 2. Force one particular person to *always* be the “start” of the seating arrangement, no matter what physical seat they are sitting in, and ignoring the fact that a circular arrangement really has no “start.” Then there are 9! ways to arrange the remaining 9 people around the table starting from the seat to the left of the “start” person. □

21.4 Permutations of subsets

Sometimes we want to create an ordered list of a certain length from a larger pool of candidates.

permutation of size k
an ordered list of k elements from a given set A, with $|A| \geq k$

$P(n,k)$ the number of permutations of size k taken from a set of size n

P^n_k, ${}_nP_k$ alternative notation choices for $P(n,k)$

Example 21.4.1 Visualizing $P(4,2)$. Consider $A = \{1,2,3,4\}$, so that $n = |A| = 4$. There are $4! = 24$ permutations of A.

x	1	2	3	4
$\mathrm{id}_A(x)$	1	2	3	4
$p_1(x)$	1	2	4	3
$p_2(x)$	1	3	2	4
$p_3(x)$	1	3	4	2
$p_4(x)$	1	4	2	3
$p_5(x)$	1	4	3	2
$p_6(x)$	2	1	3	4
$p_7(x)$	2	1	4	3
$p_8(x)$	2	3	1	4
$p_9(x)$	2	3	4	1
$p_{10}(x)$	2	4	1	3
$p_{11}(x)$	2	4	3	1

x	1	2	3	4
$p_{12}(x)$	3	1	2	4
$p_{13}(x)$	3	1	4	2
$p_{14}(x)$	3	2	1	4
$p_{15}(x)$	3	2	4	1
$p_{16}(x)$	3	4	1	2
$p_{17}(x)$	3	4	2	1
$p_{18}(x)$	4	1	2	3
$p_{19}(x)$	4	1	3	2
$p_{20}(x)$	4	2	1	3
$p_{21}(x)$	4	2	3	1
$p_{22}(x)$	4	3	1	2
$p_{23}(x)$	4	3	2	1

Figure 21.4.2 Permutations of a set of size 4.

Notice that the permutations above have been grouped into pairs, where the two permutations in a given pair have the same two first elements in the same order. From this, we can conclude that there are only $24/2 = 12$ permutations of size $k = 2$ from A. □

Theorem 21.4.3 Computing $P(n,k)$. *We have*

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

Proof. One way to construct an ordered list of k elements from a set A, where $|A| = n$, is as in Example 21.4.1. Form an ordered list of *all* the elements of A ($n!$ ways), and then take the first k elements from that list. But we get the same ordered list of length k no matter how the last $n-k$ elements are ordered. That is, we consider any two orderings of all n elements to be *equivalent* if the first k elements in the list are the same between the two. As there are $(n-k)!$ different ways the last $n-k$ elements could be ordered while keeping the first k elements the same, each equivalence class has size $(n-k)!$. Applying the Division Rule, we obtain the desired formula

$$P(n,k) = \frac{\#\{\text{orderings of all } n \text{ elements}\}}{\#\{\text{reorderings of the last } n-k \text{ elements}\}} = \frac{n!}{(n-k)!}.$$

■

Remark 21.4.4 The number $P(n,n)$ represents the number of ways to construct an ordered list of n elements chosen from a set of n elements, so $P(n,n) = n!$. The convention $0! = 1$ ensures that our formula for $P(n,k)$ expressed in Theorem 21.4.3 remains valid in the case $k = n$.

Worked Example 21.4.5 Elections. A class of twenty discrete mathematics students decides to elect a class president and vice-president. How many possible outcomes to the election process are there?

Solution. An arbitrary way to elect students to these offices would be to line all the students up and choose the first two students in line to be the president and vice-president, respectively. Therefore, there are

$$P(20,2) = \frac{20!}{(20-2)!} = 20 \cdot 19 = 380$$

possible outcomes to the election. □

Worked Example 21.4.6 Ranking choices. You go to the horsetrack to bet on a race. From a field of nine horses, how many ways are there to make a "Trifecta" bet (i.e. specify the first three finishers in order)?

Solution. There are

$$P(9,3) = \frac{9!}{(9-3)!} = 9 \cdot 8 \cdot 7 = 504$$

possible such bets. □

Worked Example 21.4.7 Words with no repeated letters. For alphabet $\Sigma = \{\text{a, b, c}, \ldots, \text{y, z}\}$, how many four-letter words made up of distinct letters are there in Σ^*?

Compare. See Worked Example 20.3.11.

Solution. A four-letter word with no repeated letters is the same as a permutation of size 4, so the number of such words is

$$P(26,4) = \frac{26!}{(26-4)!} = 26 \cdot 25 \cdot 24 \cdot 23 = 358,800.$$

□

Worked Example 21.4.8 If $|A| = k$ and $|B| = n$, with $k \le n$, how many injective functions $f : A \to B$ exist?

Compare. See Worked Example 20.3.13.

Solution. Fix an ordering $a_1, a_2, \ldots, a_k$ of the elements of A. Then from any ordering $b_1, b_2, \ldots, b_k$ of size k from B, we get an injective function $f : A \hookrightarrow B$ by the following table of values.

x	a_1	a_2	$\cdots$	a_k
$f(x)$	b_1	b_2	$\cdots$	b_k

That is, every permutation of B of size k corresponds to an injection $f : A \hookrightarrow B$, and so the number of such injections is $P(n,k)$. □

21.5 Activities

If you know what the **choose function** is, for this activity set pretend that you don't.

Activity 21.1 Write down all permutations of the set $A = \{c, a, t\}$. Express your permutations as *functions* from A to itself.

Activity 21.2 Write down some example permutations of size 3 from the set $A = \{t, r, u, c, k\}$.

Activity 21.3 Verify the equality

$$\frac{(n+1)!}{(k+1)!(n-k)!} = \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!}.$$

Activity 21.4 A child has the following set of refrigerator magnets: $\{A, B, C, D, E, F, G, H, I, J\}$.

(a) How many four-letter words can the child form? (Nonsense words allowed.)

(b) How many five-letter words can the child form if the middle letter must always be a vowel?

(c) If the child were able to form one word per second, and never stopped to eat or sleep, how many days would it take to form every possible word that uses all of the magnets?

Activity 21.5

(a) How many ways could student groups have been formed today if both group membership and group station location matter? (But assume that each station always has the number of students it has now.)

(b) How many ways could student groups have been formed today if only group membership matters? (Again assume that each group station always has the number of students it has now.)

Activity 21.6

(a) How many binary words of length 10 contain at least two zeros?

(b) How many binary words of length 10 contain at least at least three ones?

Activity 21.7 Consider the letters in the word $PEANUT$.

(a) How many six-letter words can be formed using these letters? (Each letter can only appear once.)

(b) How about if the vowels must be at the beginning?

(c) How about if no consonant may be isolated between two vowels?

Activity 21.8 You're cleaning up your shop and it's time to hang all your screwdrivers in a row on your pegboard. You have two slot-head screwdrivers, three Phillips-head screwdrivers, and four Robertson-head screwdrivers. (Assume that screwdrivers of the same type are of different sizes.)

- **(a)** In how many different orders can you arrange your screwdrivers?
- **(b)** How about if all the slot-heads are arranged on the left, all the Phillips-heads in the middle, and all the Robertson-heads are arranged on the right?
- **(c)** How about if all screwdrivers of a particular type are arranged together, but the types are arranged in no particular order?

Activity 21.9 You're cleaning up your shop and it's time to hang all your screwdrivers in a row on your pegboard. You have five screwdrivers of each type: slot-head, Phillips-head, and Robertson-head. (Assume that screwdrivers of the same type are all of different sizes.)

- **(a)** In how many different orders can you arrange your screwdrivers if the types must alternate: first slot-head, then Phillips-head, then Robertson-head, then slot-head, then Phillips-head, then . . . .
- **(b)** How about if the types must alternate, but with no restriction on the order of the types?

Activity 21.10

- **(a)** How many ways are there to arrange six people in a circle?
- **(b)** How about if there are two people who cannot sit beside each other?
- **(c)** How about if there is one person who cannot sit directly to the right of some other person?

Activity 21.11

- **(a)** How many ways are there to arrange three professors and three students in a circle so that professors and students alternate?
- **(b)** Answer the same question for n professors and n students.

Activity 21.12 How many ways could you choose numbers a, b, c from the set $\mathbb{N}_{<11}$, allowing repetition, so that the sum a+b+c is at least 5?

Chapter 22

Combinations

22.1 Motivation

Worked Example 22.1.1 How many different $four$-member study groups could be formed from a class of twenty students?

Solution. We will use the Division Rule, first imposing additional structure on each possible study group. Within a study group, create positions of President, Vice-President, Secretary, and Janitor (cards, anyone?). Then there are $P(20,4)$ such structured groups.

But a real study group doesn't have this structure, so we'll consider two structured groups to be **equivalent** when they have the same membership, regardless of positions. How many equivalent structured groups with a given membership are there? Within a group of four, the additional structure is just an ordering, and the number of orderings of a given group is 4! So

$$\begin{aligned}\#\{\text{study groups}\} &= \frac{\#\{\text{structured groups}\}}{\#\{\text{equivalent groups with a given membership}\}}\\ &= \frac{P(20,4)}{4!}\\ &= \frac{20!}{4!(20-4)!}.\end{aligned}$$

□

Note 22.1.2 What we have counted in Worked Example 22.1.1 is the number of subsets of size 4 in the set of students enrolled in the hypothetical class.

22.2 Basics

combination
a finite subset of a given set

$C(n,k)$ the number of combinations of size k taken from a set of size n

C^n_k, ${}_nC_k$ alternative notation choices for $C(n,k)$

choose function
the function $(n,k) \mapsto C(n,k)$

Checkpoint 22.2.1 What are the domain and codomain of the choose function?

Warning 22.2.2 ***Permutations and combinations are different.*** A permutation is a bijection from a set to itself. Given a fixed chosen ordering of the set elements in a list (considered as inputs), a permutation is essentially an re-ordering of the set elements into a second list, to line up outputs with inputs. So ***order matters in a permutation***. On the other hand, a combination is just a set, and order does not matter in a set, only membership matters. That is, two listings of some of the elements from a set are the same combination if all the same elements are listed, regardless of the order of the elements in the two lists. So ***order does not matter in a combination***.

Theorem 22.2.3 Computing $C(n,k)$. *We have*

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

Proof. Suppose $|A| = n$. Using the solution to Worked Example 22.1.1 as a model for our proof, we note that each ordered list of k elements taken from A defines a combination from A, but different orderings of the same k elements yield the same combination. Define two permutations to be "equivalent" if they are orderings of the same elements, so that equivalent permutations are associated to the same combination. Since there are $k!$ elements in each equivalence class of permutations, we may apply the Division Rule to obtain

$$\begin{aligned} C(n,k) &= \#\{\text{combinations}\} \\ &= \frac{\#\{\text{permutations}\}}{\#\{\text{equivalent permutations in each class}\}} \\ &= \frac{P(n,k)}{k!}. \end{aligned}$$

Finally, to obtain the rightmost formula in the statement of the theorem, we just need to combine the above formula relating $C(n,k)$ and $P(n,k)$ with the formula for $P(n,k)$ from Theorem 21.4.3. ■

Corollary 22.2.4 *We also have $C(n,n-k) = C(n,k)$.*

Proof. Calculate

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

■

Remark 22.2.5 Choosing is equivalent to rejecting. Interpret this last corollary as follows: from a set of n objects, choosing to include k elements in a combination is equivalent to choosing $n-k$ objects to reject.

Worked Example 22.2.6 Choosing ice cream. How many double-scoop ice cream combinations are possible if the local ice cream shop features thirty-one different flavours? (*Note:* Only flavour combinations are relevant, not which flavour goes on the cone first.)

Solution. From thirty-one flavours, there are

$$C(31,2) = \frac{31!}{2!29!} = \frac{31\cdot 30}{2} = 31\cdot 15 = 465$$

possibilities for double-scoop cones with *two distinct* flavours. However, there are an additional 31 possibilities for double-scoop cones with two scoops the same flavour. So the answer is

$$465 + 31 = 496.$$

□

Worked Example 22.2.7 Counting colour patterns (revisited). How many different colour patterns can we achieve by placing three red bottles and five blue bottles on a shelf? (Assume the bottles are indistinguishable except by colour.)

Solution. There are 8 possible positions in which to place a bottle. To create an arbitrary colour pattern, we can choose 3 of the positions to be filled by red bottles, then place blue in the remaining positions. So the answer is

$$C(8,3) = \frac{8!}{3!\,5!} = 56.$$

□

Remark 22.2.8 Compare the above solution for Worked Example 22.2.7 with the solution for the identical problem in Worked Example 21.3.8.

Worked Example 22.2.9 Choosing with constraints. How many ways are there to choose a team of five people from a pool of six first-year students and four senior students if the team must have three first-years and two seniors?

Solution. Choose the seniors for team, then the first-years (or vice-versa). Applying the Multiplication Rule to these independent, consecutive tasks yields answer

$$C(4,2)\cdot C(6,3) = \left(\frac{4!}{2!\,2!}\right)\left(\frac{6!}{3!\,3!}\right) = 120.$$

□

Worked Example 22.2.10 Creating several non-overlapping combinations. How many ways are there to choose three teams of four members each from a pool of twenty people, where no person can be on more than one team?

Solution 1. Choose the first team ($C(20,4)$ ways), then the second team ($C(16,4)$ ways), then the third team ($C(12,4)$ ways). Applying the Multiplication Rule yields a total of

$$C(20,4)\cdot C(16,4)\cdot C(12,4) = \left(\frac{20!}{4!\,\cancel{16!}}\right)\left(\frac{\cancel{16!}}{4!\,\cancel{12!}}\right)\left(\frac{\cancel{12!}}{4!\,8!}\right) = \frac{20!}{8!\,(4!)^3}.$$

possible teams. However, the way in which we have constructed our teams has imposed an order on the collection of teams (first, second, and third team), when there is no reason to assume such structure. Given a collection of teams, re-ordering the teams themselves (not the people within each team) produces an equivalent collection of teams by membership. As there are 3! ways to reorder the three teams, applying the Division Rule gives us a final answer of

$$\frac{20!}{3!\,8!\,(4!)^3}.$$

Solution 2. Initially choose the twelve people who will make up the three teams, but without yet assigning anyone to a particular team ($C(20,12)$ ways). Then, from this reduced pool of candidates, choose the first team ($C(12,4)$ ways) and the second team ($C(8,4)$ ways). The third team will now consist of the remaining four people from the twelve initially chosen. The Multiplication Rule gives a preliminary total of

$$C(20,12)\cdot C(12,4)\cdot C(8,4) = \left(\frac{20!}{\cancel{12!}\,8!}\right)\left(\frac{\cancel{12!}}{4!\,\cancel{8!}}\right)\left(\frac{\cancel{8!}}{4!\,4!}\right) = \frac{20!}{8!\,(4!)^3}.$$

But as in the first solution above, we need to account for the fact that we have artificially ranked teams as first, second, and third. Applying the Division Rule gives us a final answer of

$$\frac{20!}{3!\,8!\,(4!)^3}.$$

□

22.3 Applications

22.3.1 Distributing/choosing indistinguishable objects

Worked Example 22.3.1 How many ways are there to distribute seven coins amongst three children? (Assume the coins are indistinguishable. But children are obviously distinguishable.)

Solution. Here is one scheme by which we can decide how many coins each child will get. Line the children up in some order. (There is no need to count the number of ways to do this — see the end of the solution.) Also lay out the coins in a line:

$$\circ\circ\circ\circ\circ\circ\circ.$$

Now grab two Hickory Sticks™ from the snack table to act as dividers to split the coins up into three groups. For example,

$$\circ|\circ\circ\circ\circ|\circ\circ$$

means that the first child will receive one coin, the second will receive four, and the third child will receive two, whereas

$$\circ\circ\circ\circ\circ\circ\circ||$$

means that the first child gets all seven coins.

We are now back to the red bottle, blue bottle problem (see Worked Example 21.3.8 and Worked Example 22.2.7): how many different symbol patterns can we obtain by arranging two indistinguishable $|$ symbols and seven indistinguishable $\circ$ symbols? Just choose two of the nine available positions in the pattern to place the $|$ symbols. And so we have arrived at the answer $C(9,2) = 36$.

Now, why do we not have to take into account the ordering of the children at the beginning? Let c_1, c_2, c_3 represent the three children. Relative to that ordering of children, the symbol pattern

$$\circ\circ\circ\circ\circ\circ\circ||$$

means that the child c_1 gets all seven coins, as above. But relative to the ordering c_3, c_2, c_1, the different symbol pattern

$$||\circ\circ\circ\circ\circ\circ\circ$$

also means that child c_1 gets all seven coins, which is the same result. So if we allow both symbol patterns and orderings of children to vary, we will end up over-counting. □

Theorem 22.3.2 *There are $C(n+k-1, k-1)$ ways to distribute n indistinguishable objects amongst k distinguishable containers.*

Proof. Just as in the last example, use n $\circ$ symbols to represent the indistinguishable objects and $k-1$ indistinguishable $|$ symbols to represent the division into k containers. So each word from the alphabet $\Sigma = \{\circ, |\}$ that contains exactly n $\circ$ symbols and $k-1$ $|$ symbols represents a unique way to divide the objects into the containers. The length of such a word is $n+k-1$, and every such word can be constructed by choosing $k-1$ positions for the $|$ symbols from the $n+k-1$ available letter positions. ■

Worked Example 22.3.3 Your professor throws a discrete math party, but only nine students show up (sad face). The professor sends one of the students to the corner store to get cans of soda pop for everyone. The student decides to get a mix of four different varieties. How many possible mixes of soda varieties can the student come back with? (Assume that the cans are indistinguishable except by variety, and that the store has more than ten cans of each variety available.)

Solution. Here is one scheme by which the student can decide how to choose ten cans in some combination of soda varieties. Make four boxes labelled by soda variety. Have the student choose the soda cans while blindfolded, but has the store clerk place each can in the appropriate box as the cans are chosen (a permissible assumption, since it has no bearing on the outcome). In this way, we may assume that the cans are initially indistinguishable and remain so *until* they are placed in the appropriate box, at which time they magically become the variety specified by the box's label. The previous theorem now tells us that there are

$$C(10+4-1,4-1) = C(13,3) = 286$$

ways to do this. □

Corollary 22.3.4 *There are $C(n+k-1,k-1)$ ways to choose n objects from amongst k types of object, where objects are indistinguishable except by type, and there are at least n objects of each type available.*

Proof idea. Appeal to Theorem 22.3.2 exactly as in Worked Example 22.3.3. ■

22.3.2 Counting edges in connected graphs

Proposition 22.3.5 Edges in a complete graph. *For $n \geq 2$, the* ***complete graph*** *with n vertices has $C(n,2)$ edges.*

Proof. A complete graph has no loops and exactly one edge between each pair of vertices. So to count the edges we can just count the number of *pairs* of vertices, which is $C(n,2)$ for $n \geq 2$. ■

Let's summarize what we know about the number of edges in an arbitrary *connected* graph.

1. A connected graph with n vertices has *at least* $n-1$ edges (Theorem 15.3.11).

2. A connected graph with n vertices is a tree if and only if it has *exactly* $n-1$ edges (Theorem 16.3.1).

3. A simple graph with $n \geq 2$ vertices is complete if and only if it has exactly $C(n,2)$ edges (Proposition 22.3.5 in the forward direction, Statement 3 of Proposition 14.2.11 in the reverse direction).

The first fact tells us the *minimum* number of edges a connected graph must have, but it does not guarantee that a graph with that many edges must be connected, even if the graph is simple. The following is something of a converse to this fact, as it does provide such a guarantee: it tells use how many edges a (simple) graph must have before we can be *certain* that it is connected.

Theorem 22.3.6 *If $G = (V,E)$ is a simple graph such that $|V| = n$ and $|E| > C(n-1,2)$, then G is connected.*

Proof idea. Considering the contrapositive, assume that G is simple but *not* connected. In Activity 15.6, we discovered that such a G will be maximal when it has exactly two connected components, each of which is a complete graph. Among graphs with those two characteristics (and still n vertices), the largest possible value for $|E|$ occurs when the connected components of G are an isolated vertex v

and the complete graph K_{n-1}, in which case the number of edges is $C(n-1,2)$ (Proposition 22.3.5 for K_{n-1}). All other nonconnected, simple graphs will then have $|E| \leq C(n-1,2)$, as required to complete the proof by contrapositive. ■

22.4 Properties

Note: In this section we will use the alternative notation C_k^n in place of $C(n,k)$.

The combination values C_k^n as we vary n and k exhibit some patterns — see Figure 22.4.1 below.

$$\begin{array}{c}
C_0^0 \\
C_0^1 \quad C_1^1 \\
C_0^2 \quad C_1^2 \quad C_2^2 \\
C_0^3 \quad C_1^3 \quad C_2^3 \quad C_3^3 \\
C_0^4 \quad C_1^4 \quad C_2^4 \quad C_3^4 \quad C_4^4 \\
C_0^5 \quad C_1^5 \quad C_2^5 \quad C_3^5 \quad C_4^5 \quad C_5^5 \\
C_0^6 \quad C_1^6 \quad C_2^6 \quad C_3^6 \quad C_4^6 \quad C_5^6 \quad C_6^6 \\
\vdots
\end{array}
\longrightarrow
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
\vdots
\end{array}$$

Figure 22.4.1 Pascal's triangle.

Studying the version of Pascal's triangle involving the actual combination values, here are some of the patterns we observe.

- The values are symmetric about a vertical line through the centre of the triangle.
- It appears that every entry is the sum of the two entries immediately above.
- It appears that each row sums to a power of 2.

We have already observed the first pattern as arising from the equivalence of choosing versus rejecting elements to form a subset (see Corollary 22.2.4 and Remark 22.2.5).

The next two propositions confirm that the other two observed patterns continue throughout the triangle.

Proposition 22.4.2 *For $n \geq 2$ and $1 \leq k \leq n-1$, we have $C_k^n = C_{k-1}^{n-1} + C_k^{n-1}$.*

Proof. We could prove this equality just by comparing the factorial formulas involved on the left-hand and right-hand sides. But instead we will consider each of these combination values as representing the number of subsets of a certain size.

Write

$$A = \mathbb{N}_{<n} = \{0,1,2,\ldots,n-1\}$$

so that $|A| = n$. Then the left-hand side of the equality in the statement of the proposition represents the number of subsets of A of size k. Let's break that collection of subsets into two subcollections.

Subsets of A of size k that contain 0. Each of these subsets will consist of 0 along with $k-1$ nonzero elements. As A contains $n-1$ nonzero elements from which to choose, there are C_{k-1}^{n-1} ways to select those additional subset elements from A.

Subsets of A of size k that do not contain 0. Each of these subsets must consist of k nonzero elements. As A contains $n-1$ nonzero elements from which to choose, there are C^{n-1}_k ways to select those subset elements from A. Adding these two disjoint cases together using the Addition Rule yields the right-hand side of the equality. ■

Proposition 22.4.3 *For $n \geq 0$, we have* $\sum\limits_{k=0}^{n} C^n_k = 2^n$.

Proof. First, recall that the notation on the left-hand side is **summation notation**:

$$\sum_{k=0}^{n} C^n_k = C^n_0 + C^n_1 + \cdots + C^n_n.$$

Let $A = \mathbb{N}_{<n}$, so that $|A| = n$. Then $|\mathscr{P}(A)| = 2^n$ (Theorem 12.2.9). So the right-hand side of the equality represents the number of possible subsets of A.

On the other hand, for each index k in the sum on the left-hand side, the term C^n_k is the number of subsets of A of size k. Using the Addition Rule, the sum of these terms must also be the total number of possible subsets of A. ■

Here is one further property of the choose function.

Proposition 22.4.4 *For $0 \leq m \leq n$, we have*

$$C^{n+1}_{m+1} = \sum_{k=m}^{n} C^k_m.$$

Proof. Suppose A is a finite set with $|A| = n+1$. (For example, we could use $A = \mathbb{N}_{<n+1}$, but that might get confusing between numbers as elements of A and numbers as cardinalities of subsets of A.) Then the left-hand side of the equality is the number of subsets of A of size $m+1$. Here is a systematic way we could create that those subsets.

Choose an ordering of the elements of A, so that

$$A = \{a_1, a_2, \ldots, a_{n+1}\},$$

though we will not count this choice of ordering. Then, proceed as follows.

1. Write
$$B_1 = \{a_1, a_2, \ldots, a_{m+1}\},$$
so that $B_1 \subseteq A$ with $|B_1| = m+1$. Then there is exactly $1 = C^{m+1}_{m+1}$ subset of B_1 of size $m+1$, which is B_1 itself. And this subset of B_1 is also a subset of A.

2. Write
$$B_2 = \{a_1, a_2, \ldots, a_{m+1}, a_{m+2}\},$$
so that $B_1 \subseteq B_2 \subseteq A$ with $|B_2| = m+2$. Using only the elements of B_2, to create a new subset $X \subseteq A$ of size $m+1$ that we have not already counted we must include the new element a_{m+2}, with the remaining m elements to make up X chosen from B_1. So we get C^{m+1}_m new subsets of A of size $m+1$ from B_2.

3. Write
$$B_3 = \{a_1, a_2, \ldots, a_{m+1}, a_{m+2}, a_{m+3}\},$$
so that $B_2 \subseteq B_3 \subseteq A$ with $|B_3| = m+3$. Using only the elements of B_3, to create a new subset of $X \subseteq A$ of size $m+1$ that we have not already counted we must include the new element a_{m+3}, with the remaining m elements to make up X chosen from B_2. So we get C^{m+2}_m new subsets of A of size $m+1$ from B_3.

And so on. The last step in this process is when we create new subsets of size $m+1$ by first choosing to include a_{n+1}, and then choosing the remaining m elements from $A \smallsetminus \{a_{n+1}\}$, giving us C^n_m new subsets.

Every subset $X \subseteq A$ of size $m+1$ is accounted for in the above process, since every such subset must contain at least one element with index $m+1$ or larger. If a_N is the element in X with the largest index, then X is one of the subsets considered in step ℓ, where $N = m+\ell$.

So adding up each of these disjoint cases using the Addition Rule must yield the total number of subsets of A of size $m+1$. Replacing the C^{m+1}_{m+1} total from the first step with C^m_m (since both are equal to 1) to match the pattern of the subsequent steps, we obtain

$$C^{n+1}_{m+1} = C^m_m + C^{m+1}_m + \cdots + C^n_m,$$

as desired. ■

22.5 Activities

Activity 22.1 From a pool of eleven students (five first-year, six senior), how many ways are there to form:

(a) A committee of three students?

(b) A committee consisting of three first-year students and four senior students?

(c) A committee of six students if two of the senior students refuse to be together on the committee?

(d) A committee consisting of four first-year students and three senior students if two of the first-year students refuse to be together on the team?

Activity 22.2 From the alphabet $\Sigma = \{0,1\}$:

(a) How many words of length 10 contain exactly six 0s?

(b) How many contain at least three 1s?

Activity 22.3 From the alphabet $\Sigma = \{0,1,2\}$:

(a) How many words of length 10 contain exactly four 2s?

(b) How many contain at most seven 0s?

Activity 22.4 Figure 22.5.1 contains a diagram in a pyramid shape. The unfilled circles represent "positions" in the pyramid, and the smaller dots represent "dividers" between positions. Consider "paths" through this pyramid that begin at the peak position and end on the lowest level. The filled circles joined by line segments represent *one* such path.

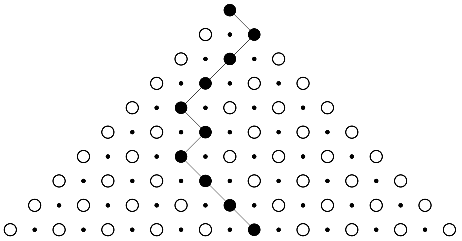


Figure 22.5.1 A Plinko™-style pyramid.

(a) How many such paths are there?

(b) How many paths are there that change direction exactly once? Exactly twice? At every step?

(For each case described in this task, you should be able to arrive at an answer without explicitly determining all such paths.)

Activity 22.5 You get to the final exam of one your courses and are faced with twelve questions. In how many ways can you fulfill the requirements exam if the instructions ask you to:

(a) Answer any ten of the questions?

(b) Answer any seven of the first eight questions and any three of the last four questions?

(c) Answer ten of the questions, at least five of which must be from the first eight questions and at least three of which must be from the last four questions?

Activity 22.6 A course instructor for a class of twenty is feeling particularly lazy and doesn't bother to mark the final exams. Instead, she decides that for each of the letter grades A, B, C, she will randomly assign that grade to exactly six students, and the last two unlucky students will be assigned a grade of D. How many different course outcomes are there?

Activity 22.7 How many ways are there to split mn people into m groups of equal size?

Activity 22.8 Suppose you have $2n$ teddy bears that are identical except for a number stitched into the paw of the right foot. Of these bears, n have the number 0 on their foot, and the remaining n bears have a unique number from $1,2,3,\ldots,n$. How many ways can you choose n of the bears, with the understanding that any of the bears labelled 0 are interchangeable?

Hint. Break into cases based on how many bears labelled 0 will be in your collection.

Activity 22.9 Consider the set $\{1,2,3,\ldots,2n\}$. How many subsets of size 2 are there such that the two elements therein have an even sum?

Activity 22.10 Consider the set $\{1,2,3,\ldots,n\}$. How many subsets of size 3 are there such that no two of the three elements therein are consecutive?

Hint. It might be easier to count the subsets of size 3 that *do* contain (at least) two consecutive numbers.

22.6 Exercises

Evaluating the combination formula. In each of Exercises 1–6, compute the value of the combination or formula of combinations. To obtain exact answers, you should simplify the factorial expressions before computing.

1. $C(4,4)$

2. $C(13,5)$

3. $C(1000000,999998)$

4. $C(7,0)$

5. $C(10,6) \cdot C(6,3)$

6. $C(10,9)/C(5,2)$

Combination formula identities. In each of Exercises 7–10, verify the equality of combination formulas. Remember to consider the left-hand and right-hand sides of each equality ***separately***, manipulating/simplifying one or the other or both sides until they are the same expression.

7. $C(n,k) = \frac{n}{k} \cdot C(n-1,k-1)$

8. $C(n,k) = \frac{n}{n-k} \cdot C(n-1,k)$

9. $C(n,k) = \frac{n-k+1}{k} \cdot C(n,k-1)$

10. $C(n+k,n) = C(n+k,k)$

11. Choose a value for m so that the equality in Proposition 22.4.4 becomes a formula for the sum $1+2+3+\cdots+n$.

CHAPTER 23

Binomial and multinomial coefficients

23.1 Binomial coefficients

binomial an expression of the form $(x+y)^n$, where $n \in \mathbb{N}$ and x, y are real numbers (or elements of any commutative ring with identity)

Example 23.1.1 Expanding binomials. Expanding binomials gets more complicated as n increases.

$$\begin{aligned}
(x+y)^2 &= x^2+2xy+y^2\\
(x+y)^3 &= (x+y)(x^2+2xy+y^2) = x^3+3x^2y+3xy^2+y^3\\
(x+y)^4 &= (x+y)(x^3+3x^2y+3xy^2+y^3) = x^4+4x^3y+6x^2y^2+4xy^3+y^4\\
(x+y)^5 &= (x+y)(x^4+4x^3y+6x^2y^2+4xy^3+y^4)\\
&= x^5+5x^4y+10x^3y^2+10x^2y^3+5xy^4+y^5\\
&\vdots
\end{aligned}$$

The symmetry in each of these expansions should be expected: we would get the same expression in the summation opposite order if we swapped x and y, since $(x+y)^n = (y+x)^n$. □

binomial coefficient
a number appearing as a coefficient in the expansion of $(x+y)^n$

$\dbinom{n}{k}$ the k^{th} coefficient in the expansion of $(x+y)^n$ $(0 \le k \le n)$

To better understand the complexity of binomial expansions, we should look for and exploit patterns. We have already expanded some binomial expressions for small exponents in Example 23.1.1 — let's extract the binomial coefficients from those expressions.

$$\begin{array}{ccccccccccccc}
&&&&&&\binom{0}{0}&&&&&& \\
&&&&&\binom{1}{0}&&\binom{1}{1}&&&&& \\
&&&&\binom{2}{0}&&\binom{2}{1}&&\binom{2}{2}&&&& \\
&&&\binom{3}{0}&&\binom{3}{1}&&\binom{3}{2}&&\binom{3}{3}&&& \\
&&\binom{4}{0}&&\binom{4}{1}&&\binom{4}{2}&&\binom{4}{3}&&\binom{4}{4}&& \\
&\binom{5}{0}&&\binom{5}{1}&&\binom{5}{2}&&\binom{5}{3}&&\binom{5}{4}&&\binom{5}{5}& \\
\binom{6}{0}&&\binom{6}{1}&&\binom{6}{2}&&\binom{6}{3}&&\binom{6}{4}&&\binom{6}{5}&&\binom{6}{6} \\
&&&&&&\vdots&&&&&&
\end{array}
\longrightarrow
\begin{array}{ccccccccccccc}
&&&&&&1&&&&&& \\
&&&&&1&&1&&&&& \\
&&&&1&&2&&1&&&& \\
&&&1&&3&&3&&1&&& \\
&&1&&4&&6&&4&&1&& \\
&1&&5&&10&&10&&5&&1& \\
1&&6&&15&&20&&15&&6&&1 \\
&&&&&&\vdots&&&&&&
\end{array}$$

Figure 23.1.2 Pascal's triangle.

Remark 23.1.3 Figure 23.1.2 above sure looks a lot like Figure 22.4.1.

Theorem 23.1.4 Binomial Theorem. *For every $x, y \in \mathbb{R}$ and every $n \in \mathbb{N}$, we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n,$$

where

$$\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!}.$$

Informal direct proof outline. Write $(x+y)^n = (x+y)(x+y)\cdots(x+y)$, with n factors. To expand this out, we generalize the FOIL method: from each factor, choose either x or y, then multiply all your choices together. Then add the results of all possible such products. For example,

$$\begin{aligned}
(x+y)^2 &= xx + xy + yx + yy = x^2 + 2xy + y^2, \\
(x+y)^3 &= xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy = x^3 + 3x^2y + 3xy^2 + y^3.
\end{aligned}$$

When forming a specific product, if you chose y for k out of n choices, you must have chosen x for the remaining $n-k$ of the n choices. The result will be $x^{n-k}y^k$. So to figure out the coefficient on $x^{n-k}y^k$, just count how many ways there are to choose y for k of the n choices. This is just C_k^n, where we choose k factors of $(x+y)$ to give us a y, and the rest to give us an x. ■

Induction proof outline.

Base case. The cases of $n = 0, 1$ are trivially true.

Induction step. Use the binomial formula for $(x+y)^{n-1}$ to obtain the binomial formula for $(x+y)^n$, by manipulating

$$\begin{aligned}
(x+y)^n &= (x+y)(x+y)^{n-1} \\
&= (x+y)\bigl(C_0^{n-1}x^{n-1} + C_1^{n-1}x^{n-2}y + \cdots + C_{n-1}^{n-1}y^{n-1}\bigr).
\end{aligned}$$

■

Worked Example 23.1.5 Expanding a binomial. Expand $(x-2)^5$.

Solution. We saw that the $n = 5$ row of Pascal's triangle is $1, 5, 10, 10, 5, 1$.

$$(x-2)^5$$

$$\begin{aligned}&\big(x+(-2)\big)^5\\&=\binom{5}{0}x^5+\binom{5}{1}x^4(-2)+\binom{5}{2}x^3(-2)^2+\binom{5}{3}x^2(-2)^3+\binom{5}{4}x(-2)^4+\binom{5}{5}(-2)^5\\&=x^5-10x^4+40x^3-80x^2+80x-32.\end{aligned}$$

□

Worked Example 23.1.6 Determining a specific coefficient in a binomial expansion. What is the coefficient on the x^4y^9 term in the expansion of $(3x+y)^{13}$?

Solution. Considering

$$(3x+y)^{13}=\big((3x)+y\big)^{13},$$

the x^4y^9 term is

$$\begin{aligned}\binom{13}{9}(3x)^4y^9&=\frac{13!}{9!4!}\cdot 3^4x^4y^9\\&=\frac{13\cdot 12\cdot 11\cdot 10\cdot 27\cdot 3}{4\cdot 3\cdot 2}x^4y^9\\&=(13\cdot 3\cdot 11\cdot 5\cdot 27)x^4y^9.\end{aligned}$$

So the desired coefficient is 57,915. □

23.2 Multinomial Coefficients

Theorem 23.2.1 Trinomial Theorem. *The expansion of the trinomial $(x+y+z)^n$ is the sum of all possible products*

$$\frac{n!}{i!j!k!}x^iy^jz^k,$$

where $0\le i,j,k\le n$ such that $i+j+k=n$.

Proof idea. Similarly to the proof of the Binomial Theorem, write

$$(x+y+z)^n=(x+y+z)(x+y+z)\cdots(x+y+z),\qquad (*)$$

with n factors. To expand this out, we generalize the FOIL method: from each factor, choose either x, y, or z, then multiply all your choices together. For any such product, the powers on x, y, and z must sum to n. To get the final expansion, add the results of all possible such products.

But we can collect terms that have the same exponent on each of x, y, and z. How many ways can we form a specific term $x^iy^jz^k$, for $0\le i,j,k\le n$ such that $i+j+k=n$? We have C_i^n ways to choose i factors from the right-hand side of $(*)$ from which to take x, then C_j^{n-i} ways to choose j factors from which to take y. But now from all remaining factors we must choose z, and there is only 1 way to do this. So the coefficient on $x^iy^jz^k$ is

$$\binom{n}{i}\binom{n-i}{j}=\left(\frac{n!}{i!(n-i)!}\right)\left(\frac{(n-i)!}{j!(n-i-j)!}\right)=\frac{n!}{i!j!k!}.$$

■

Alternative proof idea. Use the Binomial Theorem on $\big(x+(y+z)\big)^n$, then again on $(y+z)^k$ for each term $C^n_k x^{n-k}(y+z)^k$. (This would be very tedious!) ■

Worked Example 23.2.2 Expanding a trinomial. Determine the terms in the expansion of $(2x+y-3z)^3$.

Solution. First, rewrite

$$(2x+y-3z)^3 = \big((2x)+y+(-3z)\big)^3.$$

So the terms in the expansion involve products

$$(2x)^i y^j (-3z)^k.$$

We need to account for all triples of exponents i,j,k that sum to 3.

i	j	k	$\frac{n!}{i!j!k!}$	term	simplified
3	0	0	1	$(2x)^3$	$8x^3$
0	3	0	1	y^3	y^3
0	0	3	1	$(-3x)^3$	$-27z^3$
2	1	0	3	$3(2x)^2y$	$12x^2y$
2	0	1	3	$3(2x)^2(-3z)$	$-36x^2z$
1	2	0	3	$3(2x)y^2$	$6xy^2$
0	2	1	3	$3y^2(-3z)$	$-9y^2z$
1	0	2	3	$3(2x)(-3z)^2$	$-54xz^2$
0	1	2	3	$3y(-3z)^2$	$-9yz^2$
1	1	1	3!	$6(2x)y(-3z)$	$-36xyz$

Collecting this together, we have

$$\begin{aligned}&(2x+y-3z)^3\\ &= 8x^3+y^3-27z^3+12x^2y-36x^2z\\ &\quad +6xy^2-9y^2z-54xz^2-36xyz.\end{aligned}$$

□

Worked Example 23.2.3 Determining a specific coefficient in a trinomial expansion. Determine the coefficient on $x^5y^2z^7$ in the expansion of $(x+y+z)^{14}$.

Solution. Here we don't have any extra contributions to the coefficient from constants inside the trinomial, so using $n=14$, $i=5$, $j=2$, $k=7$, the coefficient is simply

$$\frac{14!}{5!2!7!} = \frac{14\cdot 13\cdot 12\cdot 11\cdot 10\cdot 9\cdot 8}{5\cdot 4\cdot 3\cdot 2\cdot 2} = 14\cdot 13\cdot 11\cdot 9\cdot 4 = 72{,}072.$$

□

The pattern of the Binomial Theorem and Trinomial Theorem continues.

Theorem 23.2.4 Multinomial Theorem. *The expansion of* $(x_1+x_2+\cdots+x_m)^n$ *is the sum of all possible products*

$$\frac{n!}{i_1!i_2!\cdots i_m!}x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m},$$

where the exponents $i_1,i_2,\ldots,i_n$ *sum to* n.

Proof idea. Use the same generalized FOIL method argument as in the Binomial and Trinomial Theorem proofs, and simplify the product of combination formulas obtained. ■

Worked Example 23.2.5 Determining a specific coefficient in a multinomial expansion. Determine the coefficient on x^2yz^6 in the expansion of $(3x+2y+z^2+6)^8$.

Solution. Rewriting

$$(3x+2y+z^2+6)^8 = \big((3x)+(2y)+(z^2)+6\big)^8,$$

we see that the four terms in this multinomial are

$$3x, \quad 2y, \quad z^2, \quad 6.$$

So what we really want to know is the total coefficient on the term involving

$$(3x)^2(2y)^1(z^2)^3 6^2.$$

The Multinomial Theorem tells us that there will be

$$\frac{8!}{2!\,1!\,3!\,2!} = 1{,}680$$

such terms in the expansion of the multinomial. Therefore, we obtain the term

$$(1{,}680)(3x)^2(2y)^1(z^2)^3 6^2 = (1{,}088{,}640)x^2yz^6$$

with a total coefficient of $1{,}088{,}640$. □

multinomial coefficient
a number appearing as a coefficient in the expansion of $(x_1+x_2+\cdots+x_m)^n$

$\binom{n}{i_1,i_2,\dots,i_m}$ the coefficient on the term $x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m}$ in the expansion of $(x_1+x_2+\cdots+x_m)^n$, where the exponents $i_1,i_2,\dots,i_m$ must sum to n

Note 23.2.6

- The Multinomial Theorem tells us $\displaystyle\binom{n}{i_1,i_2,\dots,i_m} = \frac{n!}{i_1!\,i_2!\cdots i_m!}$.
- In the case of a binomial expansion $(x_1+x_2)^n$, the term $x_1^{i_1}x_2^{i_2}$ must have $i_1+i_2=n$, or $i_2=n-i_1$. The Multinomial Theorem tells us that the coefficient on this term is

  $$\binom{n}{i_1,i_2} = \frac{n!}{i_1!\,i_2!} = \frac{n!}{i_1!\,(n-i_1)!} = \binom{n}{i_1}.$$

  Therefore, in the case $m=2$, the Multinomial Theorem reduces to the Binomial Theorem.

23.3 Applications

Proposition 23.3.1 Counting partitions of a finite set. *If $|A|=n$, then the number of ways to partition A into m disjoint subsets $A_1,A_2,\dots,A_m$, with each* *subset of predetermined size $|A_j| = i_j$, is*

$$\binom{n}{i_1, i_2, \ldots, i_m}.$$

Proof idea. There are $C^n_{i_1}$ possibilities for A_1. After choosing A_1, there are $C^{n-i_1}_{i_2}$ possibilities for A_2. After choosing A_2, there are $C^{n-i_1-i_2}_{i_3}$ possibilities for A_2. Continue in this fashion, all the way to A_m, then multiply all the combination formula expressions together. ■

Alternative proof idea. Going back to basic counting principles, we can approach this in the same way that we came up with the factorial formula for the choose function. Choosing a permutation of A ($n!$ ways) gives us an instance of the desired partition of A by setting A_1 to be the subset consisting of the first i_1 objects in the permutation, then setting A_2 to be the subset consisting of the next i_2 objects in the permutation, and so on. However, the ordering of the elements inside any such subset A_j does not matter, and we would get the same partition if we took our permutation of A and again permuted the "clusters" corresponding to each subset A_j. Since there are $i_j!$ ways to permute subset A_j, we should divide $n!$ by each of the factorials $i_j!$. ■

Warning 23.3.2 In the above theorem, the order $A_1, A_2, \ldots, A_m$ matters!

Proposition 23.3.3 Counting words with a fixed composition of letters. *Suppose $x_1, x_2, \ldots, x_m$ are distinct letters in the alphabet Σ. For $i_1 + i_2 + \cdots + i_m = n$, the number of words in Σ^* of length n which consist of exactly i_1 x_1's, i_2 x_2's, ..., and i_m x_m's is the multinomial coefficient*

$$\binom{n}{i_1, i_2, \ldots, i_m}.$$

Proof idea. If we view each letter x_i as a variable and each word made up of the letters $x_1, \ldots, x_m$ as a product of these variables, then each of the words we want to count gives us one way to achieve a term of $x_1^{i_1} \cdots x_m^{i_m}$ in the expansion of $(x_1 + \cdots + x_m)^n$. The number of such ways is the multinomial coefficient. ■

Worked Example 23.3.4 How many different 9-digit integers can we form from three 3s, four 6s and two 9s?

Solution. The number of integers of the desired digit composition is the multinomial coefficient

$$\binom{9}{3,4,2} = \frac{9!}{3!4!2!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 2} = 9 \cdot 4 \cdot 7 \cdot 5 = 1{,}260.$$

□

23.4 Exercises

1. Choose numbers x, y so that the equality in the Binomial Theorem becomes

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n.$$

2.

(a) Choose numbers x, y so that the equality in the Binomial Theorem becomes

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n \binom{n}{n} = 0.$$

(b) The equality from Task a can be rearranged to yield

$$\begin{aligned} &\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots + \binom{n}{m_1} \\ &= \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots + \binom{n}{m_2}, \end{aligned}$$

where

$$m_1 = \begin{cases} n, & n \text{ even}, \\ n-1, & n \text{ odd}, \end{cases} \qquad m_2 = \begin{cases} n-1, & n \text{ even}, \\ n, & n \text{ odd}. \end{cases}$$

What does this rearranged formula tell you about the subsets of a set of size n?

Hint. What is the sum on the left counting? What is the sum on the right counting?

Appendices

APPENDIX A

GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <http://www.fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE. The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS. This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING. You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY. If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS. You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties — for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS. You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS. You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS. A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION. Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION. You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE. The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See `http://www.gnu.org/copyleft/`.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING. "Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents. To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with… Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

APPENDIX B

Index of Notation

Symbol	Description	Page
$\neg A$	logical negation of statement A	5
$A \wedge B$	logical conjunction of statements A and B	5
$A \vee B$	logical disjunction of statements A and B	5
$A \rightarrow B$	logical conditional where statement A implies statement B	5
$A \leftrightarrow B$	logical biconditional where each of statements A and B implies the other	5
$A \Rightarrow B$	statement A logically implies statement B, so that conditional $A \rightarrow B$ is a tautology	10
$A \Leftrightarrow B$	statements A and B are equivalent	15
x'	Boolean negation	24
$A(x)$	a predicate statement A whose truth value depends on the free variable x	29
$A(x,y)$	a predicate statement A whose truth value depends on the free variables x and y	29
$\forall x$	the universal quantifier applied to the free variable x	30
$\exists x$	the existential quantifier applied to the free variable x	30
$A_1, A_2, \ldots, A_m \therefore C$	an argument with premises $A_1, A_2, \ldots, A_m$ and conclusion C	41
$\begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_m \\ \hline C \end{array}$	an argument with premises $A_1, A_2, \ldots, A_m$ and conclusion C	41
$x \in S$	object x is an element of set S	76
$\{a, b, c, \ldots\}$	a set defined by listing its elements, enclosed in braces	76
$\mathbb{N}$	the set of natural numbers	76
$\mathbb{Z}$	the set of integers	76
$\mathbb{Q}$	the set of rational numbers	76
$\mathbb{R}$	the set of real numbers	76
$\varnothing$	the empty set	78

(Continued on next page)

Symbol	**Description**	**Page**
$A \subseteq B$	set A is a subset of set B	78
$A \underset{\neq}{\subseteq} B$	set A is a proper subset of set B	80
A^{c}	the complement of A relative to some universal set	81
$B \smallsetminus A$	the complement of A relative to some universal set	81
$\mathbb{I}$	the set of irrational real numbers	81
$A \cup B$	the union of sets A and B	82
$A \cap B$	the intersection of A and B	82
$A \sqcup B$	the disjoint union of sets A and B	83
$A \times B$	the Cartesian product of A and B	84
A^n	the Cartesian product $A \times A \times \cdots \times A$ involving n copies of A	85
Σ^*	the set of words using alphabet set Σ	86
$\lvert w \rvert$	length of the word $w \in \Sigma^*$	87
Σ_n^*	for $n \in \mathbb{N}$, the subset of Σ^* consisting of all words of length n	87
$\emptyset$	the empty word	87
$\mathscr{P}(A)$	the power set of the set A	88
$f : A \to B$	f is a function with domain A and codomain B	93
$f(a) = b$	function $f : A \to B$ associates the codomain element $b \in B$ to the domain element $a \in A$	93
$a \mapsto b$	alternative notation for $f(a) = b$	93
$\Delta(f)$	graph of function f	96
$f(A)$	the image of function $f : A \to B$	98
$f(A')$	the image of function $f : A \to B$ on a subset $A' \subseteq A$	99
$f : A \twoheadrightarrow B$	function f is surjective	100
$f : A \hookrightarrow B$	function f is injective	100
$\mathrm{id}_A : A \to A$	the identity function on on set A	102
$\iota_A^X : A \to X$	the inclusion function on subset $A \subseteq X$	102
$\rho_i : A_1 \times A_2 \times \cdots \times A_n \to A_i$	the projection function onto the i^{th} factor A_i in the Cartesian product $$A_1 \times A_2 \times \cdots \times A_n$$	103
$\mathrm{proj}_i : A_1 \times A_2 \times \cdots \times A_n \to A_i$	alternative notation for ρ_i	103
$f\vert_A$	restriction of function $f : X \to Y$ to subset $A \subseteq X$	103
$f \vert A$	alternative domain restriction notation	103
$\mathrm{res}_A^X f$	alternative domain restriction notation	103
$g \circ f$	the composition of functions f and g	105
$f^{-1}(C)$	the inverse image of the subset $C \subseteq B$ under the function $f : A \to B$	107
$f^{-1} : B \to A$	the inverse function associate to bijective function $f : A \to B$	108
$\mathbb{N}_{<m}$	the set of natural numbers that are less than m	115
a_k	k^{th} term in a sequence	115
$\{a_k\}$	the collection of terms in a sequence	115

(Continued on next page)

Symbol	**Description**	**Page**
$\{a_k\}_0^m$	the collection of terms in a finite sequence	115
$\{a_k\}_0^\infty$	the collection of terms in an infinite sequence	115
$\lvert A\rvert$	cardinality of the set A	123
$\operatorname{card} A$	alternative notation for the cardinality of the set A	123
$\#\{\dots\}$	alternative notation for the cardinality of the set defined by $\{\dots\}$	123
$\lvert A\rvert = \infty$	set A is infinite	127
$\lvert A\rvert < \infty$	set A is finite	127
$\deg v$	degree of vertex v	149
$\lvert E\rvert$	the number of edges in the graph $G = (V,E)$	150
$G' \preceq G$	graph G' is a subgraph of graph G	150
K_n	the unique complete graph with n vertices	151
$a\,R\,b$	element $a \in A$ is related to element $b \in B$ by relation R	185
$R_1 \cup R_2$	union of relations R_1, R_2	186
$R_1 \cap R_2$	intersection of relations R_1, R_2	186
R^{c}	complement of relation R	186
$a \ \not R\ b$	alternative notation for $a\,R^{\mathrm{c}}\,b$	186
R^{-1}	inverse of the relation R	187
$a \varnothing b$	the empty relation between elements a and b (always false)	187
$a\,U\,b$	the universal relation between elements a and b (always true)	187
$a \equiv b$	a is related to b by the equivalence relation $\equiv$; in other words, a is somehow equivalent to b	195
$m_1 \equiv_n m_2$	integers m_1, m_2 are equivalent modulo n	196
$[a]$	the equivalence class of the element $a \in A$ relative to some specific equivalence relation on A	196
$A/{\equiv}$	the quotient of A relative to equivalence relation $\equiv$	200
$a{\preceq}b$	a is related to b by the partial order $\preceq$; in other words, a is somehow "smaller than or same size as" b	208
$a \prec b$	$a \preceq b$ but $a \neq b$	208
$n!$	factorial $n! = n(n-1)(n-2)\cdots 2\cdot 1$	237
$P(n,k)$	the number of permutations of size k taken from a set of size n	241
P_k^n	alternative notation for $P(n,k)$	241
${}_nP_k$	alternative notation for $P(n,k)$	241
$C(n,k)$	the number of combination of size k taken from a set of size n	245
C_k^n	alternative notation for $C(n,k)$	245
${}_nC_k$	alternative notation for $C(n,k)$	245
$\binom{n}{k}$	the k^{th} coefficient in the expansion of $(x+y)^n$	255
$\binom{n}{i_1,i_2,\dots,i_m}$	the coefficient on the term $x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m}$ in the expansion of $(x_1+x_2+\cdots+x_m)^n$	259

Index

Colophon

This book was authored in PreTeXt XML.